Investigating AGN variability and Tidal Disruption Events with Wide Field Optical Surveys.

Lydia Makrygianni

Department of Physics & Astronomy
The University of Sheffield

A dissertation submitted in candidature for the degree of
Doctor of Philosophy at the University of Sheffield

September 2020

Dedicated to the memory of Ioannis Bakas

Contents

1 Introduction **1**
1.1 Active Galactic Nuclei and Tidal Disruption Events in the era of wide-field surveys . . . 1
1.2 AGN variability . . . 4
1.3 Tidal Disruption Events . . . 14
1.4 Gravitational-wave Optical Transient Observer (GOTO) . . . 19
1.5 Processing pipelines of GOTO data . . . 22
1.6 This thesis . . . 23

2 Data Processing **24**
2.1 Chapter Introduction . . . 24
2.2 The LSST software stack . . . 24
2.3 The `obs_goto` package . . . 26
2.4 The processing stages . . . 30
  2.4.1 Reference catalogue production . . . 32
  2.4.2 Forced Photometry . . . 43
2.5 Accessing data through `butler` . . . 46
2.6 The LSST stack-GOTO database . . . 47
2.7 Chapter Conclusions . . . 50

3 Quality assessment of the data **51**
3.1 Chapter Introduction . . . 51
3.2 Data . . . 52
3.3 Reference catalogue . . . 53
  3.3.1 Astrometry assessment . . . 53
  3.3.2 Photometry Assessment . . . 56
  3.3.3 Survey depth . . . 56
  3.3.4 Deblending . . . 59
3.4 Forced Photometry . . . 59
  3.4.1 Internal photometric assessment . . . 60
  3.4.2 Photometric assessment with external catalogues . . . 62
  3.4.3 Observational photometric uncertainties . . . 65
  3.4.4 Lightcurves analysis . . . 70
3.5 Chapter Conclusions . . . 74

4 Studying AGN variability with GOTO **76**
4.1 Chapter Introduction 76
4.2 AGN variability methods 77
4.2.1 P_{var} and σ_{rms} 78
4.2.2 Structure Function 79
4.2.3 The Damped Random Walk 83
4.3 Data 84
4.4 AGN Lightcurves 86
4.5 Variable AGN from GOTO and ZTF lightcurves 92
4.6 A vs γ parameter space 95
4.6.1 Tests with focusing on the common sources between the GOTO and ZTF samples 103
4.6.2 Tests with downgraded ZTF data 110
4.6.3 Testing the effect of cadence and time length of GOTO lightcurves on A and γ using simulated lightcurves 112
4.7 Investigation of σ_{rms} 115
4.8 Chapter Conclusions 120

5 The variability of moderate luminosity AGN with ZTF lightcurves **121**
5.1 Chapter Introduction 121
5.2 AGN variability parameters 122
5.3 AGN physical parameters 125
5.4 Correlation between variability and physical parameters 130
5.5 r-band lightcurves 134
5.6 Chapter Discussion and Conclusions 139

6 Tidal Disruption Event discovery with GOTO **143**
6.1 Chapter Introduction 143
6.2 Detection of known TDEs 144
6.3 TDE candidates identification through GOTO lightcurves 155
6.3.1 Mock lightcurves 158
6.4 Chapter conclusions 165

7 Conclusions **166**
7.1 Future Work 170

List of Figures

1.1 The unification of Type 1 and Type 2 AGN and their typical optical spectra. *Credits:* Claudio Ricci; `http://www.isdc.unige.ch/\protect\unhbox\voidb@x\penalty\@M\{}ricci/Website/Active_Galactic_Nuclei.html` 5
1.2 Optical SF of 6 active galaxies as presented in Collier & Peterson (2001) 9
1.3 Representation of an AGN lightcurve generated with DRW and related to the SF and PSD concepts as found in Kozłowski (2016). AGN variability is usually studied using one or more of these measurements to fit, describe and analyse the AGN photometric data 10

2.1 A schematic representation of the `obs_goto` and the pipeline flow that can be used to process survey data. 29
2.2 Master bias (left), master dark (centre) and, master flat (right) from GOTO data, as generated by the LSST stack. 32
2.3 Example of the background model at the centre of an image as generated by the LSST stack. 33
2.4 The figure from Mullaney et al. (2021) shows sources that the LSST stack used to construct a model of the spatially-varying PSF within an individual GOTO frame, where the light blue points show the pixel positions of all sources detected at $> 100\sigma$ within the frame, whereas the black circles represent those that have been selected as candidates for PSF modelling based on their shape and size. Top panel shows the selection using the LSST stack default parameters whereas the bottom panel shows the PSF modelling candidates after we relaxed the selection criteria. This shows that with the relaxed criteria, we are able to have candidates across the whole frame which allows better PSF modelling on the central parts of the image as well as on the outer parts. 35
2.5 The three extensions of an calexp image (zoomed in). The first extension is the calibrated exposure (left), the second the mask (centre) and the third the variance (right). We have selected a part of the image to show the an example of masked pixels. The red colour represents pixels masked as detected whereas the green pixels mask value indicates that these pixels are affected by cosmic rays and interpolation. 36

2.6 Polynomial fitting from synthetic photometry using the Pickles library (Pickles 1998) and the respective fit from the empirical data. Using the Pickles library we pass synthetic stellar spectral models through the PanSTARRS and GOTO filter passbands to calculate the plotted differences. 39
2.7 Healpix skymap with resolution parameter equal to four, which has as results 192 tracts across the whole sky. 41
2.8 The plot shows the patches in tract 161. 41
2.9 Example of coadd image (left) and its variance (right) which ensure errors are propagated correctly. 42
2.10 Part of a individual calibrated image (left) and the respective part of the coadd image generated from images takes in a specific range of dates (see Chapter 3). The effective exposure time of the nightly images is 180 s while the respective for the coadd images is at least 180 s. 43
2.11 The LSST stack-GOTO database schema 49

3.1 The map shows the sky coverage of the GOTO data, which were processed through the LSST stack. 52
3.2 Astrometric quality against PanStarrs PS1 catalog. The plots are showing the angular distance (top), RA difference (middle) and DEC difference (bottom). 55
3.3 Mean (top) and standard deviation (bottom) for the difference between the magnitudes from aperture photometry in the reference catalog extracted by the LSST software stack and the colour corrected PanStarrs magnitudes. Both of these plots are presented in Mullaney et al. (2021). . 57
3.4 The survey depth for each patch in the reference catalogue. The data points are colour coded for the median aperture magnitude of sources with SN between 4.5 and 5.5. 58
3.5 Detection completeness as a function of magnitude in tract 94, patch 5,8 using artificial sources. 59
3.6 The photometric calibration zeropoint for the frames processed through forced photometry, which for most of them peaks around ∼22.8 with a spread and few significant peaks in lower values resulting the mean value 22.52 mag. 60
3.7 Photometric repeatability for aperture photometry from forced photometry catalogues for a field covered by the four CCDs. We see that sources brighter than 15.5 mag are precise within 0.02 RMS, which is indicated by the red line on the plot. 61
3.8 Photometric repeatability comparison for aperture photometry and PSF fit photometry. 62
3.9 The difference between GOTO aperture (top)/ PSF(bottom) photometry from LSST software stack for a full FOV field and PanSTARRS magnitudes as a function of the magnitude. The PanSTARRS magnitudes are from PSF photometry. 64

3.10 The difference between GOTO aperture (top)/ PSF(bottom) photometry from LSST softwre stack for a full FOV field and GOTOphoto magnitudes as a function of the magnitude. 66
3.11 The top panel shows the the standard deviation of the quantity of eq.(3), as a function of the median magnitude for the GOTO light curves from forced photometry for the SDSS standard stars. The middle shows the intrinsic scatter of each light curve measured as the robust standard deviation of the magnitude, as a function of the magnitude and also the theoretical noise as described in the text. The bottom panel shows the median error of each light curve as a function of the magnitude. The method evaluates the photometric uncertainties according to the Suberlak et al. (2017). In each plot, the red dots show the median values per half-magnitude bin. 69
3.12 Lightcurves of standard SDSS star from aperture and PSF photometry. Aperture photometry gives have reduced-χ^2 closer to unity which is what expected for non-variable stars. Aperture photometry is especially better on the brighter end (<15mag) of the GOTO detections. 71
3.13 Lightcurves of variable star from aperture and PSF photometry. The general shape of the PSF light curves follow the trend of the aperture photometry light curve. For the PSF light curves, however, the plots show an addtional intra-night scatter. 72
3.14 Examples of GOTO light curves from aperture photometry measured with the LSST stack, presented as phase plots, of periodic variable stars as compared with the ZTF g-band light curves. We only compare the shapes of the lightcurves as the observed difference in photometry arise from the fact that ZTF measurements are in the ZTF g-band whereas GOTO measurements are in GOTO L-band. These variable stars have period from 5 to 48 days and belong to different classes. GOTO data cover which is shorter than the total time ZTF is running and also the cadence of ZTF is higher for this period since GOTO was operating with the four UTs setup during the period that these data were collected. However the shapes of the GOTO and ZTF phase plots are very similar which shows that GOTO will be a valuable resource for time-domain studies of variable sources. 73

4.1 The figure shows a mock lightcurve with assigned photometric errors (top) spanning ∼3 years and its respective SF plot versus the time lag (bottom) using equation 4.6. 81

4.2 The number of epochs per AGN versus the rest-frame time length of the each lightcurve from the generated from LSST stack measurements (top). The red lines indicate the two criteria for well-sampled lightcurves, i.e. 200 days length in rest-frame and 20 epochs per lightcurve. There are very few data points for the low epochs-$t_{\rm rest}$ for the GOTO data processed with the LSST which is the result partially from the sample that was matched with LSST stack database but also because they are processed using forced photometry. This means that the LSST stack will perform forced photometry on the position of the AGN even if this was not detected on the nightly frame. The bottom panel shows the absolute L-band magnitude against redshift. The absolute magnitudes are not K-corrected or corrected for dust extinction. The black triangle points show the parent sample after it was matched to the LSST stack database (20,005 sources), and the red dots show these that have well-sampled lightcurves i.e. more than 20 epochs and span more than 200 days. A standard Λ cold dark matter cosmology with $H_0 = 70$ km s^{-1} Mpc^{-1},$\Omega_M = 0.27$, and $\Omega_\Lambda = 0.73$. 89

4.3 The number of epochs per AGN versus the rest-frame time length of the each lightcurve from the generated from GOTOphoto measurements (top). The red lines indicate the two criteria for well-sampled lightcurves, i.e. 200 days length in rest-frame and 20 epochs per lightcurve. The bottom panel shows the absolute L-band magnitude against redshift. The absolute magnitudes are not K-corrected or corrected for dust extinction. The black triangle points show the parent sample after it was matched to the GOTOphoto database (25,509 sources) and the red dots show these that have well-sampled lightcurves i.e. more than 20 epochs and span more than 200 days. We find that in the case of GOTOphoto the well sampled lightcurves tend to be these for the brighter sources across the redshift range, which is expected as the GOTOphoto photometry is using at the moment "blind" detection and photometry measurements on these sources. So, fainter sources are more difficult to detect when, for example, we have bad weather conditions that affect the magnitude limit. A standard Λ cold dark matter cosmology with $H_0 = 70$ km s^{-1} Mpc^{-1},$\Omega_M = 0.27$, and $\Omega_\Lambda = 0.73$. 90

4.4 The top panel shows the number of epochs per AGN versus the rest-frame time length of the each lightcurve generated from the ZTF g-band public data. The red lines indicate the two criteria for well-sampled lightcurves, i.e. 200 days length in rest-frame and 20 epochs per lightcurve. The bottom panel shows the g-band absolute magnitude against redshift. The absolute magnitudes are not K-corrected or corrected for dust extinction. The black triangle points show the parent sample after being matched to the ZTF public database for g-band photometry (8,014 sources) and the red dots show these that have well-sampled lightcurves i.e. more than 20 epochs and span more than 200 days. A standard Λ cold dark matter cosmology with $H_0 = 70$ km s^{-1} Mpc^{-1},$\Omega_M = 0.27$, and $\Omega_\Lambda = 0.73$. 91

4.5 Redshift versus magnitude distributions for all three samples i.e. GOTO LSST stack (top panel), GOTO GOTOphoto (middle panel) and ZTF (bottom panel). The gap on the top plot for the LSST stack processed GOTO data is because we choose aperture photometry for sources brighter than 17.5 and PSF photometry for sources fainter than 17.5 mag. We see, however, that for all the samples most of the sources are fainter than 17.5 mag. 94
4.6 *Top*: A vs γ parameter space for GOTO LSST stack processed data. For a better representation of the results in this plot, we have eliminated data points with errors for A or γ larger than 0.2, which indicate a failure in fitting. This reduces the sample from a total of 5,802 to 5,217. We find that the data points for the different AGN types cover essentially the same part of the plot and their distributions, as indicated by the histograms, do not show significant difference as discussed in this section.*Bottom*: Same parameter space but using contours for Type 1 and Type 2 AGN. The plot shows that the distributions cover the same region of the parameter space. 97
4.7 *Top*: A vs γ parameter space for GOTO GOTOphoto processed data. In this case of this I have also removed the data points with errors larger than 0.2 for either A or γ. This reduces the sample from a total of 3,434 objects to 3,245. Again for the GOTO data, we find that we are unable to distinguish the Type 1 from the Type 2 objects. *Bottom*: Same parameter space but using contours for Type 1 and Type 2 AGN. The plot shows that the distributions cover the same region of the parameter space. . . . 98
4.8 *Top*: A vs γ parameter space from the ZTF g-band lightcurves and the respective histograms for each parameter. For the ZTF data, we find that removing the points with larger than 0.2 error bars for either A or γ reduces the sample by just 20 objects i.e. from 5,646 to 5,626. Also, there are, in general, smaller uncertainties in the fitted parameters for ZTF than those for the respective GOTO plots (Figure 4.6, Figure 4.7) which indicate better A and γ estimates from the ZTF data. For the ZTF data, there is a clearer difference between the Type 1 and Type 2 values especially in the case of the γ parameter. *Bottom*: Same parameter space but using contours for Type 1 and Type 2 AGN. Unlike with the GOTO data, using the ZTF data we see a distinction between Type 1 and Type 2 AGN on the A vs. γ parameter space. 100
4.9 Dependence of A and γ from the rest frame time length and number of epochs for the values extracted from GOTOphoto lightcurves. There are no significant correlations of the A and γ parameters with the lightcurve properties. 105

4.10 Dependence of A and γ from the rest frame time length and number of epochs for the values extracted from GOTO LSST stack lightcurves. In this case there is an anti-correlation between the amplitude A and the rest-frame time-length of the lightcurves for the whole sample, which is also true for both types of AGN (i.e. Type 1 and Type 2 AGN). 106

4.11 Top panels: A vs. A and γ vs. γ for matched objects from ZTF and GOTOphoto. We find that there is a large scatter for both amplitude and γ values when we compare the GOTO data, processed with GOTOphoto, and the ZTF data. It is especially obvious in the case of γ values how these values for ZTF cover a larger range than these from GOTO. Bottom panels: A vs. A and γ vs. γ for matched objects from ZTF and GOTO data processed with LSST stack. From these plots we draw similar conclusions as in the case of GOTOphoto processed data 107

4.12 Top panels: γ vs. A plots for matched objects from ZTF and GOTOphoto. In the left plot, we show the data points from the ZTF lightcurves and at the right these from GOTOphoto. Right panels: γ vs. A plots for matched objects from ZTF and GOTO/ LSST stack. In the left plot we show the data points from the ZTF lightcurves and at the right these from GOTO/ LSST stack. 108

4.13 SF comparison from ZTF and GOTOphoto lightcurves for four different Type 1 AGN, where blue is the SF from the ZTF llightcurves and the orange data points represent the SF of the GOTOphoto lightcurves. The results are similar for GOTO LSST stack lightcurves. 109

4.14 A vs γ plot using the parameters extracted from the SF of the downgraded ZTF lightcurves. We see that the difference between Type 1 and Type 2 objects found in Figure 4.8 cannot be identified after introducing the GOTO RMS into the ZTF lightcurves. 112

4.15 The A versus γ for simulated lightcurves with DRW for $\tau = 300$ days and $\mathrm{SF}_\infty = 0.2$ mag for different time lengths and cadences. The short lightcurves time length is equal to ~400 days, while long are lightcurves with time length ~700 days. Low and high cadence are typical current GOTO and typical ZTF respectively. The plot shows that the time length of the lightcurves is more important for the SF fit than the cadence. . . . 114

4.16 The σ_{rms} of the GOTO lightucurves generated by the GOTOphoto (top) and LSST stack (bottom) versus the σ_{rms} calculated from the ZTF lightcurves. There is slight evidence that the ZTF results for σ_{rms} are closer to the ones the GOTO data processed with LSST stack rather than those from GOTOphoto, especially for sources ~18 mag. This is likely the result of the photometry method i.e. ZTF and LSST stack use PSF photometry for these objects, whereas GOTOphoto uses aperture photometry. 116

4.17 The σ_{rms} versus the mean magnitude for the GOTO (GOTOphoto and LSST stack) and ZTF results. We find that the across the whole range of magnitudes the σ_{rms} values of GOTO processed data (for both pipelines) are larger than these of ZTF but especially for sources fainter than 18th mag. 117
4.18 The σ_{rms} of the GOTO lightucurves generate the by the GOTOphoto (top) and LSST stack (bottom) using the RMS of non-variable sources to estimate the photometric errors versus the σ_{rms} calculated from the ZTF lightcurves. Again, we identify the same issues we explained based on Figure 4.16. 119

5.1 This corner plot shows the correlations between the different AGN variability parameters. Additionally, it shows their correlation with the rest frame time length of the lightcurves and the number of epochs. 124
5.2 The distributions of redshift and Eddington ratio for the different types of AGN in the ZTF g-band sample. We find that Type 2 AGN in this sample have typically lower redshifts than the Type 1 and NLSy1. 127
5.3 The distributions of log10($L_{5100}\,\mathrm{erg\,s^{-1}}$) for the different types of AGN in the ZTF g-band sample. We see that the sources have typically L_{5100} luminosities lower than $10^{45}\,\mathrm{erg\,s^{-1}}$. 128
5.4 Correlations between the physical parameters of the AGN in the ZTF sample g-band lightcurves (top) and the sub-sample (bottom) that has black hole masses calculated in the Shen et al. (2011). 129
5.5 Correlations between the variability parameters from the ZTF g-band lightcurves with the spectral properties of AGN. The title of each plot includes the Spearman correlation coefficients for the Type 1 AGN and the respective p-value. 132
5.6 The relationship between the Eddington ratio values in Mullaney et al. (2013), calculated using the equation $\lambda_{edd} \approx 0.43 f_{10} (\frac{L_{H\alpha}}{10^{42}\,erg\,s^{-1}})^{0.31} (\frac{FWHM_{H\alpha}}{10^{3}\mathrm{kms}^{-1}})^{-2}$ and $f_{10} = 1$ (Netzer & Trakhtenbrot 2007) and the Eddington ratio values from Shen et al. (2011) calculated using the fiducial virial mass after fitting the Hβ , C IV and Mg II lines.As expected, using the fitting to covert the values from Mullaney et al. (2013) does not affect the correlations in Figure 5.5 133
5.7 The number of epochs per AGN versus the rest-frame time length of the each lightcurve generated by the ZTF r-band public data. 135
5.8 A vs γ parameter space from the ZTF r-band lightcurves. 136
5.9 Correlations of the variability parameters from the r-band lightcurves with the spectral properties of AGN. The Spearman correlation coefficients and their respective p-values are given as reference at the top of each sub-plot. 138

6.1 The 10 TDE discovered in 2019 by the ZTF survey and reported in van Velzen et al. (2020). The objects highlighted with green are objects are TDEs with peak r-band magnitude brighter than 19, which suggests that they should have been detected by GOTO. 146
6.2 The GOTOphoto (left) and GOTO/LSST stack (right) lightcurves of 4 of the TDEs discovered by van Velzen et al. (2020). The dotted line represents the time of peak reported in the van Velzen et al. (2020) paper. 147
6.3 Realbogus values for TNS detected transients. The plot also shows the respective histograms for SN Ia, AGN and TDE. 150
6.4 Distance from the centre of the host galaxy of TNS reported transients measured as the distance of the transient from the centre of the host on the template image (top panel) and as the distance of the transient from the centre of the host reported in the PanSTARRS catalogue. 152
6.5 The g-r of host galaxies for TNS detected transients. The plot also shows the respective histograms for SN Ia, AGN and TDE. 153
6.6 Fitting to the L-band lightcurve (from difference imaging) using the TDE model in MOSFiT. 158
6.7 Examples of mock L-band TDE lightcurves with MOSFiT at redshift 0.05 (left) and 0.1 (right) for different black hole masses disrupting an 1 solar mass star. The shape of the lightcurve depends on the black hole mass, if the star is fully or partially disrupted, the mass of the disrupted star and the structure of the star, the accretion process and the time of the disruption. In our case, although I have used a complete disruption of 1 solar mass stars in all cases, I allowed the black hole mass and the parameters associated with the structure of the star to vary which results different shaped TDE lightcurves. 160
6.8 Low-sampling (10 days cadence) lightcurve and its fitting using the `tde` model (top left), the `default` model (top right), the `csm` model (bottom left) and the `slsn` model (bottom right). The median scores of the model are found at the top of each plot. 163
6.9 Well-sampled (1 day cadence) lightcurve and its fitting using the `tde` model (top left), the `default` model (top right), the `csm` model (bottom left) and the `slsn` model (bottom right). The median scores of the model are found at the top of each plot. 164

List of Tables

2.1 Table of gain and readout noise. 28
2.2 Table of LSST stack packages we need to set up in order to use the `obs_goto` to process GOTO data with the stack. 29
2.3 Table of the pixel mask types generated for each frame (if such pixel exist) and their descriptions. These masks are also translated to flags in the measurement catalogues generated by the LSST stack for each frame. 37

4.1 Table of the objects found in each survey and pipeline (for GOTO) databases after applying the various criteria described in the text. Well-sampled and variable numbers are defined after the data have been cleaned based on the flags each pipeline use to determine bad data points. 86
4.2 Table summarising the statistical values for the amplitude (A) and the logarithmic slope (γ) generated from the fitting of the SF for the different samples i.e. ZTF g-band lightcurves, GOTO data processed with GOTOphoto and GOTO data processed with LSST stack. We note that although the GOTO data processed with the LSST stack shows differences between the Type 1 and Type 2 objects especially for the γ values, the larger errors (Figure 4.6) forbid us from drawing safe conclusions. It is, however, an indication that PSF photometry is a better option for sources with fainter apparent magnitude closer to the magnitude limit. 102
4.3 The table summarises the statistical values for the amplitude A and the logarithmic slope γ, generated by fitting the SF, for the two samples of matched data; i.e., the matched data between ZTF sample with g-band lightcurves and GOTOphoto processed data and the matched data between ZTF and the LSST stack processed data. 104
4.4 The table summarises the statistical values for the amplitude A and the logarithmic slope γ generated from by fitting the SF for short-low cadence, long-low cadence and long-high cadence lightucrves. 113

5.1 This table summarises the statistical values for the amplitude A and the logarithmic slope γ generated from fitting the SF for the ZTF r-band lightcurves sample. 135

6.1 Priors and marginalised posteriors for the MOSFiT TDE model. Priors are flat within the stated ranges, except for M_*, which uses a Kroupa initial mass function. The quoted results are the median of each distribution, and error bars are the 16th and 84th percentiles. 159

Declaration

I declare that, unless otherwise stated, the work presented in this thesis is my own. No part of this thesis has been accepted or is currently being submitted for any other qualification at the University of Sheffield or elsewhere.

Much of the work presented in Chapter 2 and Chapter 3 has been published and submitted for publication on PASA in the papers:

- Makrygianni et al. - Processing GOTO survey data with the LSST software stack II: Forced Photometry and light curves
- Mullaney J. R., Makrygianni L., Dhillon V., Littlefair S., Ackley K., Dyer M., et al., 2021, PASA, 38, e004

Acknowledgements

Firstly, I would like to thank James for giving me the opportunity to do this PhD and for being very patient and supportive throughout these years. He was a great advisor and teacher during all these four years and I couldn't have asked for a better supervisor. Thank you James! I would also like to thank the rest of the AGN group, Clive, Manu, Liam, Jonny, for always learning from them at AGN coffee. A special thanks to Liam for being always helpful whenever I needed it. I would like to also thank current and past PhD students with whom I shared the office and they were all very nice people: Katie, Heloise, Becky, Martin, Gemma, James B., Emma, Christina, James W., Umar, Alex and Rebecca. A special thank you to Martin for his work on GOTO! Saying this, I want to thank the members of GOTO collaboration: Danny, Duncan, Kendall, Krzysztof and everyone else from the ever-increasing list of members.

I need to thank Paul K. who always fixed the mess I created on elara and athena.

This thesis would not be possible without the support grant from the University of Sheffield.

My family and friends know that I am not good at and hate these things. I need, however, to thank my mum who has always supported and encouraged me to achieve my goals. I would like to also mention my grandfather Filippo and my sisters Danai and Nefeli.

Finally, I want to thank Him for always helping me.

Summary

Active Galactic Nuclei (AGN) and Tidal Disruption Events (TDEs) are two phenomena that can lead to a deeper understanding of the supermassive black holes (SMBHs) that reside at the centres of all massive galaxies. Past optical surveys, such as the SDSS and PTF, have opened the way to studying both these phenomena using large samples of data. Current (e.g. GOTO, ZTF) and future (e.g LSST) surveys will provide long timespan, high cadence lightcurves for millions more galaxies than these earlier surveys, allowing us to study AGN variability and TDEs to an even greater level of detail. With this thesis, we aim to unveil the prospects of studying AGN variability and TDEs with GOTO - a recently-commissioned wide-field telescope based on La Palma, Spain.

We use the data from the telescope's prototype phase to explore the variability of moderate luminosity and low redshfit ($z < 0.4$) AGN. Most notably, we identify the barriers (e.g. systematic uncertainties) that currently prevent us from using this facility to study AGN variability for this sample, and suggest ways that the data could be improved to achieve this goal. Following the realisation that we are currently unable to study the variability of moderate luminosity AGN with GOTO, we turn to ZTF data to investigate the variability of our sample of AGN and how it relates to other physical parameters (i.e., luminosity, redshift, Eddington ratio, black hole mass), where we identify a slight anti-correlation between the amplitude of variability Eddington ratio.

Using the prototype phase data, we have also investigated the potential of GOTO as a TDE discovery machine. Using a sample of known TDEs, which were identified by

ZTF but are also found in GOTO archival data, we propose a number of ways, including host galaxy colour and pre-existed nuclear activity, to filter the GOTO data stream in order to recover these transient events. We also examine whether the difference imaging lightcurves generated from GOTO observations could be used to distinguish between TDEs and supernovae - a vital step if we wish to prioritise rapid follow-up observations.

Over the coming years, GOTO's field-of-view is set to increase by a factor of four, with a corresponding increase in cadence. With this in mind, we also explore how these upgrades – together with the longer lightcurves – will improve the prospects of studying AGN variability and TDEs with GOTO. For AGN variability they allow better statistics, whereas for TDEs, they increase the number of detections, which in turn allows faster and more confident identification of candidates.

Chapter 1

Introduction

1.1 Active Galactic Nuclei and Tidal Disruption Events in the era of wide-field surveys

Most of the large galaxies in the Universe are believed to host supermassive black holes (hereafter SMBHs) at the their centres. In the special case of active galactic nuclei (hereafter AGN) this SMBH accretes material from the galaxy's dense central regions. As this material falls toward the SMBH, its angular momentum causes it to form an accretion disk, which emits strongly at ultraviolet to near-infrared wavelengths. AGN were first studied in the early 20^{th} century (e.g Fath 1909), although it took until the discovery (Matthews & Sandage 1963) of the more luminous AGNs of this class (i.e., quasars) before it was determined that their luminosity varies over time. In the next section (i.e. 1.2), we will give more details on AGN and their optical variability, while in this section we will focus on the impact of the wide-field surveys on the study of AGN variability.

Before wide-field surveys, AGN variability studies focused on just one or a handful of objects. Starting with the Sloan Digital Sky Survey (SDSS; York et al. 2000), it became possible to study large samples of AGN (e.g., Vanden Berk et al. 2004, MacLeod et al. 2010) using multi-epoch photometry of these samples. The new idea of surveying the whole observable sky -or large parts of it- to study the ever changing sky was both captivating and challenging. After SDSS, other wide-area, time domain studies were commissioned, including, for example, the Palomar Transient Factory(PTF; Law et al. 2009, Rau et al. 2009), the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS; Hodapp et al. 2004,Chambers et al. 2016) and Hyper-Supricam Cam (Miyazaki et al. 2018).

AGN variability is a single example of the contribution of wide-field surveys to the study of the variable sky. In addition to the study of the variable sky, wide-field surveys have allowed the study of the transient sky. Transients refer to astronomical phenomena that last for as little as a fraction of a second, to up to weeks, or even years. Normally, however, transient phenomena are extreme, short-lived events that are associated with the total or partial destruction of an astrophysical object (e.g., supernovae; hereafter SNe). Tidal Disruption Event (hereafter; TDE) is a transient phenomenon, which occurs when a star passes sufficiently close to a SMBH for the tidal forces to tear the star apart. A part of the star remains bound and accretes onto the SMBH, resulting in a luminous accretion flare that can last from few months to longer than a year. While these phenomena were predicted since the 70s-80s (Hills 1975, Carter & Luminet 1983), it is only in the last ~10 years that all-sky surveys have allowed the discovery of significant numbers of them (e.g., van Velzen et al. 2011 Holoien et al. 2016). The surveys mentioned before (e.g., SDSS, PTF) were the first to identify TDE candidates in the optical wavelengths, while more recently the All-Sky Automated Survey for Supernovae (ASAS-SN[1]) project, which despite having a brighter magnitude limit (~ 18), has proved a very powerful machine for the discovery of transient events including TDEs. In section 1.3, we will give more details on the results from TDE studies using the aforementioned

---

[1] /www.astronomy.ohio-state.edu/~assassin

surveys.

Currently surveys such as the Zwicky Transient Facility (ZTF Bellm et al. 2019) are providing even more information on the variability of AGN and discovering increasing numbers of TDEs and with the commissioning of the Vera Rubin Observatory with the Legacy Survey of Space and Time (LSST; Ivezić et al. 2019) we can expect even greater amounts of data with which to study the variable and transient sky (e.g., Gibson et al. 2011, Bricman & Gomboc 2020). Within this category of new wide-area, high-cadence surveys sits the Gravitational-wave Transient Optical Observer (GOTO; Steeghs et al. in prep.). We leave a full description of the current and future technical specifications of GOTO in section 1.4. For now, it is worthwhile to explain that - when fully operational - GOTO will survey the entire sky roughly every three days in its full set-up, thereby providing a huge potential data resource for the study AGN variability and TDEs.

The aim of this thesis is to investigate the prospects of studying AGN variability and observing TDEs using GOTO survey data. Using the prototype phase (4UTs; see section 1.4), we explore the current status and identify the issues we should overcome to be able to conduct surveys on these two topics. We use roughly 1.5 year's worth of data taken when GOTO was in its prototype phase to conduct our investigations. In later chapters, we also attempt to predict, how GOTO's planned hardware developments – which will increase the cadence of the survey – will improve our study of AGN variability and TDEs. To start with, however, in the rest of this chapter we provide more background on the phenomena of AGN variability (section 1.2) and TDEs (section 1.3) highlighting key open questions. We then provide a more detailed description of GOTO (section 1.4), and its processing pipelines (section 1.5), before finishing this chapter with an outline of the remainder of this thesis (section 1.6).

1.2 AGN variability

In the previous section we introduced AGN and how wide-field surveys continue to provide us with data that can be used to answer key open questions in this field and, perhaps, uncover previously unknown behaviour. In this section we give further details on the AGN properties used throughout this thesis to investigate whether GOTO is and will be capable of studying the variability of AGN. We also describe our current understanding of AGN variability based on the findings of previous studies.

To help us put AGN variability in context, it is useful to first provide a description of the structure of a typical AGN. Our current understanding of AGN is that of a central supermassive black hole that is surrounded by an accretion disk, which spans a few light-days across (Salpeter 1964, Zel'dovich 1964). Surrounding the accretion disk is a system fast moving clouds - known as the Broad Line Region (e.g., Kaspi et al. 2005), which is responsible for the broad emission lines seen in some AGN spectra. Moving further out (to a few parsecs from the SMBH; e.g., Suganuma et al. 2006, Burtscher et al. 2013), there is the so-called "dusty torus" - a doughnut-shaped cloud of dusty, molecular gas. In the Unified Model of Antonucci (1983), it is the dusty torus that is responsible for the observed properties of the two main classes of AGN - Type 1s and Type 2s. In the case of Type 1 AGNs, the torus does not block our line-of-sight to the central accretion disk and BLR. By contrast, in the case of Type 2 AGNs, the dusty torus blocks these innermost regions from our line of sight (Figure 1.1). Since it is thought that only the accretion disk and BLR can vary on human timescales, it is expected that only Type 1 AGN will show significant levels of variability within our surveys, since these smaller regions are blocked from our view in the case of Type 2 AGN. In Chapter 4, we exploit this difference in the levels of variability between Type 1 and Type 2 AGNs to help to assess the feasibility of using current GOTO data to study AGN variability. Beyond the torus, at a distance of a few hundreds to a few thousands parsecs (e.g., Bennert et al. 2006), is the narrow

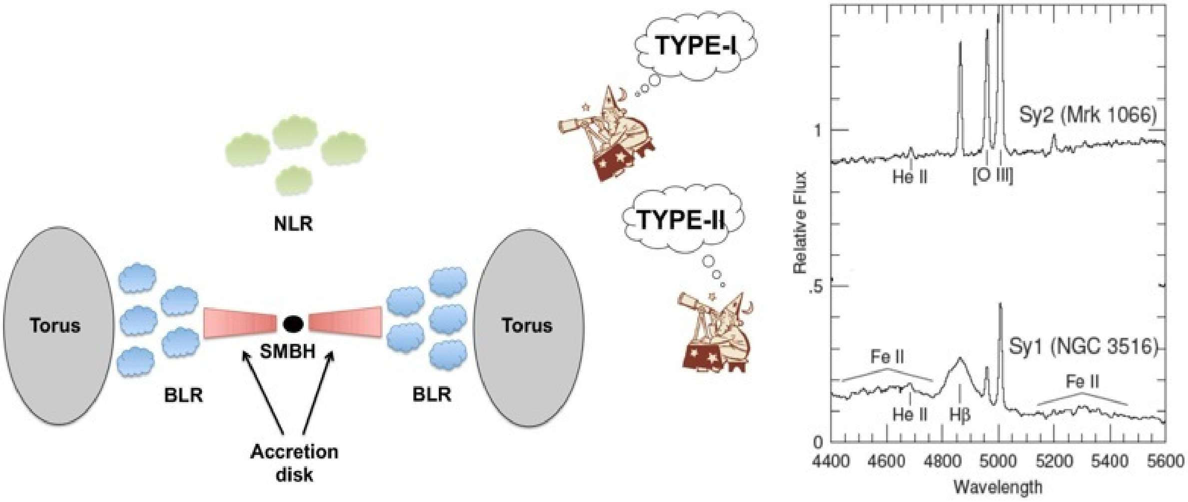


Figure 1.1: The unification of Type 1 and Type 2 AGN and their typical optical spectra. *Credits:* Claudio Ricci; `http://www.isdc.unige.ch/~ricci/Website/Active_Galactic_Nuclei.html`

line region. This is a region of low density gas that is moving at far lower velocities (~300-1000$km\,s^{-1}$) than the broad line region. It should also be mentioned that some AGN display jets (e.g., Blandford & Königl 1979) of energetic particles emanating from the nuclear regions, whereas other AGN show little or no evidence of such jets. Since these jets are detected at radio wavelengths, this has led AGN to be separated further into radio-loud and radio-quiet classes.

One of the most intriguing properties of AGN is their optical variability. Various studies have shown that the flux of Type 1 AGNs can vary on timescales ranging from hours to decades (and likely longer) across all wavelengths. Although AGN variability has been known about and studied for many years, its physical causes remain poorly understood. Different models have been proposed from theoretical predictions and observational driven studies to explain the optical/UV variability of AGN. These include accretion disk instabilities (Kawaguchi et al. 1998), microlensing (Hawkins 1993), X-ray reprocessing of disk thermal emission (Krolik et al. 1991), but also changes in the accretion rate (Pereyra et al. 2006); however, the primary origin is still under debate.

Microlensing events explain the AGN variability as an external cause rather than an intrinsic variation. In this case, quasar flux variations are explained due to cosmological lensing from stars in foreground galaxies or generally intergalactic compact objects (e.g. dark matter) in the line of sight of quasars (e.g. Hawkins 2002). Multiple imaging of quasars is needed to probe these events. In general, microlensing alone is not enough to explain the full range of quasar flux variations, but can explain the long-term variations.

Other mechanisms has been proposed for the AGN flux variations, intrinsic to central engines of these objects. One widely suggested mechanism is the so-called disc instabilities model. The accretion disc is linked to optical emission (as well to UV and X-ray emission) and so to the variations observed in optical flux. The accretion disc has been modelled through simulations (e.g. Shakura & Sunyaev 1973) and its true structure (e.g. optical and geometric depth) has been under discussion for many years. There is also much discussion on the link between the accretion disc physics and the various timescales associated with them (see next paragraph), but there are also many different reasons suggested to explain these disc instabilities. Possible explanations include magneto-rotational instabilities, accretion disc atmosphere fluctuations and thermal fluctuations. Further, other models which attempt to explain the optical variability of AGN. These involve reprocessing of X-ray emission (from the hot corona) by the accretion disc. This leads to optical/UV emission, which can explain the short-term changes in luminosity. This is supported by certain studies, which show robust time lags between X-ray, optical and UV lightcurves. All the mechanisms are currently challenged by the observations to explain the optical lightcurves of AGN.

One of the most important reasons that the models are challenged is he disagreement between the measured variability timescales and the (theoretical) physical timescales of AGN. First, there is the viscous timescale, which defines the changes of the accretion rate. This is the characteristic viscous timescale of the mass flow for a standard thin disk, which is of the order of $\sim 10^4$ years for the optical. Then, there is the dynamical timescale (months-years for the optical), which if the accretion flow is roughly Keplerian, it is given by the Keplerian frequency, which depends on the black hole mass and the radius for each part of the accretion disc. Finally, we have the thermal timescale, which corresponds to the disc cooling/heating and is related to the dynamical timescale divided by the viscosity parameter, based on the assumptions described in Shakura & Sunyaev (1973) and Czerny (2004). The thermal timescale has values from days to years. There are other timescales, defined in different models to parametrize physics, such as the free-fall timescale (from the radius r to the black hole) and the sound-crossing timescale (perpendicular to the disc). The standard accretion disc model has been challenged, since the accretion rate changes would be unable to explain the observed variability, because the observed timescales are much lower than the viscous timescale. In the era of repeated, wide field surveys, other phenomena related to AGN variability are coming to light such as changing look AGN (e.g., LaMassa et al. 2015, MacLeod et al. 2016, Gezari et al. 2017) that are challenging these models further.

In the first few years after the initial realisation that AGN vary on various timescales (see review Kawaguchi & Mineshige 1999), studies in this area focused on specific cases of AGN (e.g., Fahlman & Ulrych 1975, Peterson et al. 1994, Clavel et al. 1991). Recent wide-field, high-cadence surveys have, however, brought a major revolution in the field. Such surveys delivered larger samples of objects to be studied but also introduced the need for statistical methods to analyse the data. Another major change introduced by wide field surveys were well-sampled lightcurves for these large samples. This inspired the development of models that could be used to explain the behaviour of different AGN (e.g., Kelly et al. 2009). In terms of statistical methods, two main tools are used: the structure function (hereafter SF; e.g., Hughes et al. 1992) and power spectral density (PSD; e.g., Giveon et al. 1999). The structure function is widely used to quantify AGN variability as it can be used on sparsely sampled lightcurves, which is a major advantage as lightcurves obtained from wide-field surveys typically contain gaps due to observability, weather etc. The SF measures the root-mean-square (RMS) of the magnitude difference between two epochs of observations as a function of the time lag between the observations i.e., $SF(\Delta t) = \sqrt{\frac{1}{N}\Delta m_{ij}^2}$ (e.g., Hook et al. 1994), where N is the number of pairs in the Δt bin and Δm_{ij} is the magnitude difference between the i-th and j-th observations (i $<$ j) whose time lag $(t_j - t_i)$ falls in the Δt bin. This definition, however, does not take into account the photometric uncertainties and more recent studies use additional terms to take the photometric errors into account (see also Kozłowski 2016). In the case of AGN, the SF is well described using a power law (e.g., $SF(\Delta t) = A(\Delta t)^{\gamma}$; Schmidt et al. 2010), before it reaches a certain value, after a characteristic timescale, when it flattens. Figure 1.2 shows different examples of AGN optical SF as presented in Collier & Peterson (2001). In contrast to the SF, the PSD measures the variability in the space of frequencies using Fourier transform analysis. This measurement, however, is far more susceptible to problems arising from irregular sampling, and so it will not be further used in this work. The structure function method is considered further in Chapter 4 where we use it to analyse our various datasets.

The shape of the structure function i.e. the power law that flattens after a characteristic timescale, consists three different elements that have puzzled astronomers for years. There is the slope, the characteristic timescale at which it flattens, and the amplitude at which this happens. Kawaguchi et al. (1998) argue that the disc instabilities model best describes the observed slope of the structure function of the quasar quasar 0957+561. Hawkins (2002) tested the predictions for three models (i.e. microlensing, starburst[2] and accretion disc instability) and finds that, in terms of the logarithmic slope, the observed data for Seyfert galaxies are in better agreement with the disc instabilities model, whereas for quasars, microlensing is a better fit. For the characteristic timescales they are not able to make predictions as only one - the starburst model - makes predictions for that. The amplitude is dependent on the model's parameters, but they find that the

---

[2] The starburst model suggests that AGN are powered by multiple supernova explosions or starbursts which result in stochastic variations in their brightness.

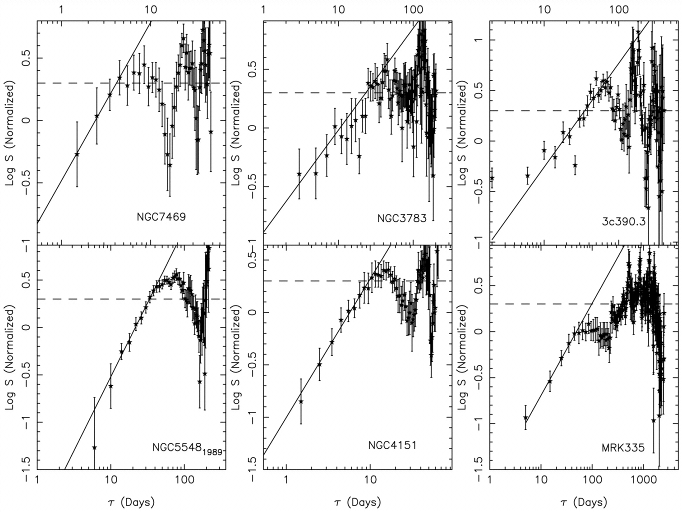


Figure 1.2: Optical SF of 6 active galaxies as presented in Collier & Peterson (2001)

microlensing model predicts the highest values and the starburst the lowest. The characteristic timescale at which the SF flattens is a great challenge for astronomers, since the standard accretion disc model predicts viscous timescales which are much longer than the observed timescales.

In terms of modelling AGN variability, various methods have been used (Emmanoulopoulos et al. 2013,Sartori et al. 2019). In 2009, however , Kelly et al. (2009) presented a way to characterize and model the stochastic AGN variability using a stochastic process called a Damped Random Walk (DRW). In the DRW model, the temporal behaviour of the AGN flux $X(t)$ can be described by the equation $dX(t) = \frac{1}{\tau}X(t)dt + \sigma_{DRW}\sqrt{dt}\epsilon(t) + bdt$ where τ is the time for the time series to become roughly uncorrelated (i.e., when after that dX does not depend on time), σ_{DRW} is the variability of the time series on time-scales short compared to τ , $b\tau$ is the mean magnitude and $\epsilon(t)$ is a white noise process with zero mean and variance equal to 1. To describe, fit or simulate the AGN variability using the DRW model two parameters are employed, a characteristic timescale τ, which is essentially the same decorrelation time as in the SF, and also the long-term standard deviation of variability σ (also related to the long timescales amplitude $SF_{\infty} = \sqrt{2}\sigma$ found in e.g. MacLeod et al. 2010). The DRW model

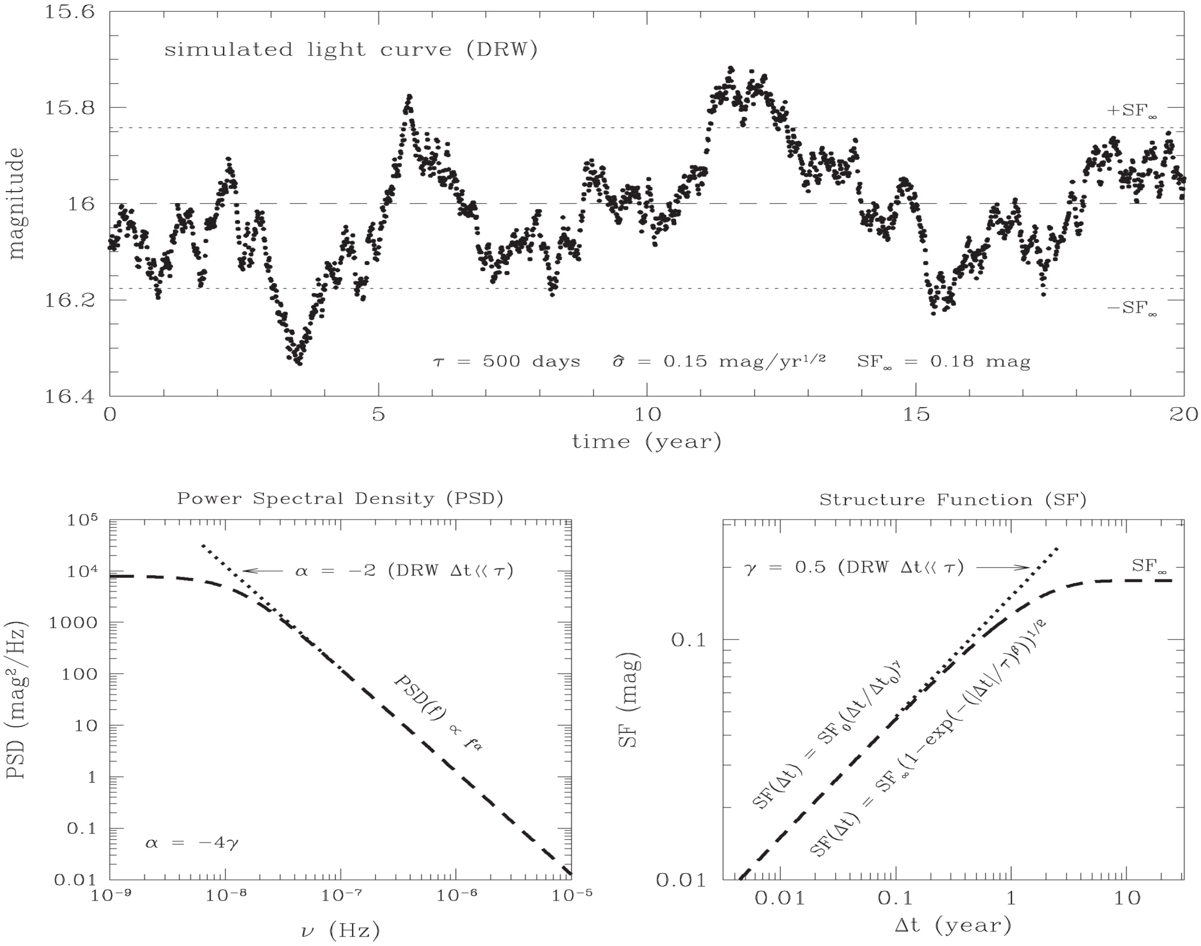


Figure 1.3: Representation of an AGN lightcurve generated with DRW and related to the SF and PSD concepts as found in Kozłowski (2016). AGN variability is usually studied using one or more of these measurements to fit, describe and analyse the AGN photometric data .

has been used to describe AGN variability, as well as to simulate and fit lightcurves in an attempt to identify AGN in repeated surveys (MacLeod et al. 2010, Kozłowski 2016, Baldassare et al. 2019). Figure 1.3 gives a schematic representation of a lightcurve generated using the DRW model and how its parameters are linked to the other two concepts (i.e., SF and PSD) of AGN variability quantification.

Equipped with more data and analysis methods, astronomers in recent years have been able to study AGN variability using thousands objects, and investigate the correlations between AGN variability and the AGN physical parameters (luminosity, black hole mass, redshfit, Eddington ratio, accretion rate). The first statistical studies of AGN variability were conducted with SDSS quasar lightcurves (e.g., Vanden Berk et al. 2004, de Vries et al. 2005,Bauer et al. 2009, MacLeod et al. 2010). In such studies the data included individual lightcurves with different numbers of data points and timespan depending on the cadence of the surveys conducted by SDSS. This means that most of the objects had lightcurves with very few epochs (e.g., of the ∼ 34,000 objects studied in MacLeod et al. 2012, 10,000 have ∼ 60 epochs, i.e., different observing nights, and the rest have 2-3 epochs). So, rather than calculating a SF for each individual object, they calculated an ensemble SF for all quasars in, for example, a given luminosity range based on the assumption that quasars with similar physical properties will vary in a similar way. This approach enabled the first statistical studies of large samples of quasars. Vanden Berk et al. (2004) studied 25,000 quasars and, from their ensemble SF, found that there is an an anti-correlation between the AGN variability amplitude and rest-frame wavelength (i.e., AGN vary more at shorter wavelengths) as well as with luminosity (i.e., more luminous AGN vary less), and a positive correlation between the amplitude of variability and redshift (i.e., higher redshift vary more). Other studies followed with the same approach. Wilhite et al. (2008), for example, by calculating the ensemble SF of 2,500 quasars, and they argued that the amplitude of variability anti-correlates with luminosity, but correlates with black hole mass. Combining this results, Wilhite et al. (2008) suggested that the main driver for quasar variability is the Eddington ratio. The positive correlation of the amplitude of variability with black hole mass was also reported in MacLeod et al. (2010) where they studied 10,000 quasars in SDSS Stripe82. Caplar et al. (2017), using the iPTF lightcurves of 28,000 luminous AGN and ensemble SF, however, find only a weak (or even no) correlation between the amplitude of variability and black hole mass. This is also true for the timescale of variability to reach a certain variability amplitude.

The aforementioned studies all explored AGN variability via the ensemble SF, after binning according to different properties. Further wide-field high cadence surveys, however, started giving data with better-sampled lightcurves and also allowed astronomers investigate the variability of less luminous AGN (e.g., Simm et al. 2016, Sánchez-Sáez et al. 2018, Baldassare et al. 2019). With these lightcurves it became possible to extract the variability information for individual objects. Simm et al. (2016) studied ∼90 X-ray selected AGN and analysed their optical light curves in the four bands (g, r, I, z) provided by the Pan-STARRS1 (PS1) Medium Deep Field 04 survey. They used the normalised excess variance (i.e., the standard deviation of the lightcurve corrected for the average photometric error and normalised for the mean magnitude) and found that it is strongly anti-correlated with the wavelength, the luminosity and Eddington ratio, but found no evidence of any dependence with black hole mass and redshift. The same conclusions were drawn using PSD analysis. Sánchez-Sáez et al. (2018) studied 1348 variable AGN lightcurves with more than 40 epochs each, and found that the amplitude of variability anti-correlates with both the rest-frame wavelength and the Eddington ratio. They also found the logarithmic gradient of the variability (from the SF power law; see more Chapter 4) does not correlate significantly with any physical parameters of the SMBHs.

In addition to studying the correlations between the AGN variability parameters and the AGN physical parameters, most of the aforementioned studies also tried to quantify the level of variability for AGN after a specific amount of time (e.g., 1 year) and/or tried to determine a characteristic timescale for AGN or groups of AGN (i.e., the timescale for the SF to flatten so the amplitude of variability will become uncorrelated with time). Studies have found that the amplitude of variability over long periods (i.e., 1-2 years) is 0.2 mag (MacLeod et al. 2010, Suberlak et al. 2017,Sánchez-Sáez et al. 2018). Using SF analysis, the value of the logarithmic gradient (i.e., γ) for the pre-flattening time lags, is found to be close to 0.5 for ∼1 year rest-frame timescale (equivalent to PSD of slope -2; e.g., MacLeod et al. 2010, Kozłowski 2016), which is also the value that would be expected for AGN lightcurves described with DRW model. Sánchez-Sáez et al. (2018) finds lower values for the excess of variance than Simm et al. (2016), although they state that the latter study used optical light curves with different coverage in time (i.e., most of their lightcurves have rest-frame time lengths 1-3 years), which introduces differences with their results.

Measurements of AGN variability can provide answers and constraints for theoretical predictions on the physical mechanism that have been proposed to describe this phenomenon. It is important, however, that any results from studies of optical variability are combined with those obtained from other wavelengths (e.g., X-rays) to be able to get a full understanding of the central engine. AGN variability can provide insights on the SMBH physics (e.g. SMBH fueling) and the central engine at spatial scales beyond the resolving power of the current and future facilities. Also, by unveiling the most important underlying correlations between AGN variability parameters other AGN properties, it will enable us to then use variability to infer physical properties for huge datasets of AGN (from, for example, LSST data).

Finally, an additional capability of AGN variability that has been explored in recent studies is whether it can be used as a tool in the search of AGN in wide-field surveys such as LSST. Schmidt et al. (2010), using SF analysis, showed that it was possible to separate quasars from non-variable or other variable sources in SDSS Stripe82 via the location of the individual sources in the A-γ plane (i.e., from $SF = A(\Delta t)^{\gamma}$). Sánchez-Sáez et al. (2019) developed a machine learning-based technique that used a trained random forest algorithm to classify objects as either AGN or non-AGN according to their variability features and optical colors. They tested this technique by performing spectroscopic follow-up observations, confirming the AGN nature of 44 among 54 AGN candidates (81.5% efficiency). Baldassare et al. (2018) and Baldassare et al. (2019) presented their results from searching for AGN-like variability by fitting the DRW model-to identify AGN in low-mass galaxies using SDSS and iPTF lightcurves respectively. Similarly, by combining multiple variability selection criteria (i.e., single-band variability amplitudes, cross-correlation of multiband light curves, and visual inspection), Kimura et al. (2020) found 491 variability-selected AGN in the COSMOS field using lightcurves obtained from the Hyper Suprime-Cam Subaru Strategic Program (HSC SSP) survey. All these studies highlight that the AGN variability will complement other known AGN selection techniques (e.g., emission line ratios, mid-IR colours), especially since it is known that variability selection identifies some AGN that are missed by all other selection methods (e.g., Baldassare et al. 2019).

GOTO, especially when it has its full complement of 16 UTs, plus its southern node in Australia (see section 1.4), will be able to conduct surveys with very high cadence (1-3 days for full set-up to cover the whole observable sky), but also can achieve a very large sky coverage. The large coverage can provide data for millions of objects and the high cadence well-sampled lightcurves for them. This raises the prospect of being able to study the variability of AGN using better statistics than ever before, while the application of machine learning techniques to these data may unveil new samples of AGN. If we also consider that the LSST will have saturation limit of r $\sim$16 with 15s exposures it become apparent that the GOTO could be a complementary facility for the study of brighter sources. With these thoughts in mind, we investigate in Chapter 4 whether we are able to utilise current GOTO data to study AGN variability and, if not, what are the stumbling blocks that are preventing us from doing so (e.g., flux limit, cadence, sample size, systematic uncertainties etc).

1.3 Tidal Disruption Events

If a star passes too close to a SMBH, it is liable — depending on the relative properties of the star and SMBH — to be ripped apart by tidal forces. Part of the stellar debris from the disruption is expelled into the galaxy, whereas a part of it becomes bound to the black hole and is expected to circularize into an accretion disc, feeding the black hole and generating a flare of radiation in the process. The study of TDEs is a rapidly developing field in astronomy, which has the potential to provide insights into the SMBHs of quiescent galaxies (e.g., Rees 1988). Komossa (2015) lists a number of additional reasons that make TDE detections interesting including: TDEs as a probe of relativistic effects and the extremes of accretion physics at high rates and near the last stable orbit, which will provide us with a new means of measuring BH spin; jetted TDEs provide new insight into the formation and early evolution of radio jets; tracing the existence of intermediate mass black holes (IMBH); stellar dynamics in galaxy cores on spatial scales, which cannot be resolved directly can be investigated using the TDE rates. Current and future wide-field surveys are expected to discover more of these events, which will allow us to gain insight for this phenomenon and enabling astronomers to address the various open questions in this field which we outline toward the end of this subsection.

The first TDEs were discovered and identified in X-rays (Bade et al. 1996). Since that first discovery, many more have been discovered in X-ray surveys or identified in archival X-ray data (e.g., Esquej et al. 2008, Maksym et al. 2010). The X-ray lightcurves of these events, however, are sparse and – as with TDEs discovered at ultraviolet wavelengths (e.g., Gezari et al. 2009) – are more difficult to disentangle from AGN activity because, while the amplitude of AGN variability at these wavelengths is much greater than in the optical, the range of this variation has not yet been well characterized (van Velzen et al. 2011). Unlike X-rays, optical ground-based surveys deliver better-sampled lightcurves, which allow us to be more confident about the candidates, while also enabling us to study these events in more detail.

Large optical surveys (e.g., SDSS, PTF, PanSTARRS) have enabled the discovery of tens of TDEs at the time of writing. The first discoveries at optical wavelengths were reported by van Velzen et al. (2011), who found two TDE candidates in SDSS Stripe82 archival data. To do so, they searched for nuclear transients excluding AGN, and found that these two events showed spectral and colour evolution which differed from that of known SN. After this first identification of TDE candidates in the optical, other studies reported more TDE candidates in both archival data and in real-time observations (e.g., Gezari et al. 2012, Arcavi et al. 2014, Holoien et al. 2016,Blagorodnova et al. 2019, van Velzen et al. 2019).

As soon as the first TDE candidates were discovered, it became clear that in order to study these events fully, they would need to be identified in real-time to enable immediate follow-up. This would allow a better analysis of these events on different levels including: better-sampled lightcurves and more detailed analysis of their colour and spectral evolution. This in turn would allow common properties of all or groups of these events (e.g., spectral lines) and to identify of correlations of the transient properties with physical properties of their host galaxies (e.g., rise/fade time with SMBH mass). Better sampled lightcurves would also allow detailed fitting of the models which provide a deeper understanding of the phenomenon. The large wide field surveys such as LSST, ZTF and GOTO will deliver thousands of different transients each year that makes the development of pipelines that will select the most interesting transient for follow-up necessary. It is estimated (van Velzen et al. 2011, Bricman & Gomboc 2020) that the LSST will detect ∼40,000-80,000 TDEs in its 10-years run. These TDEs, however, will be detected among other transients which obviously makes the efficient follow-up of these sources even more challenging. Hung et al. (2018) proposed a systematic strategy of selecting TDE candidates using iPTF data, where they used photometric criteria to down-select from a total of 493 nuclear transients to a sample of 26 sources. Using lightcurve fitting and spectroscopic follow-up they concluded that of these 26 sources 14 were Type Ia supernovae (SNe Ia), 9 were highly variable active galactic nuclei (AGNs), 2 were TDEs, and 1 was potential core-collapse supernova. This study identified the possible TDE candidates using spatial criteria (i.e., distance of the transient from the host galaxy centre) as well as photometric criteria that included host galaxy colour and transient colour.

Arcavi et al. (2014) studied 7 TDE candidates (i.e., 3 of them found in Arcavi et al. 2014 from PTF archival data and another 4 from previous studies) and found that many of the hosts showed E+A, or post-starburst, spectra. E+A galaxies are so called because the Balmer absorption features in their spectra (characteristic of A stars) appear superimposed on an old (E)arly-type galaxy population. These post-starburst spectra are characterized by a lack of strong emission lines, indicating low current star formation rates, but with strong Balmer absorption, indicating a recent burst of star formation (within the last ∼Gyr) that has now ended. Even though these galaxies are rare in the local Universe, they are over-represented among TDE hosts (see also French et al. 2016,Law-Smith et al. 2017) . van Velzen et al. (2020) studied 17 ZTF TDEs and found that host galaxies in the so-called ”green valley” (i.e., between red ellipticals and blue star-forming galaxies in the colour-magnitude diagram) are over represented in their sample by a factor of ∼5. The green valley is also known to host quiescent, Balmer-strong galaxies (including post-starburst or E+A galaxies) and so the result is in agreement with those earlier studies.

Current (e.g., GOTO, ZTF) and future (e.g., LSST) surveys aim to reveal more about the true nature of TDEs and their hosts. To achieve this, however, they need to overcome a number of different obstacles. First, as mentioned, a rapid follow-up of these events would be necessary to unveil the colour evolution and the spectra properties of large samples of these events. TDE show characteristic spectra which include hot, blue thermal continuum, and very broad (5 - 15, 000 km s-1 ; Arcavi et al. 2014; Hung et al. 2017) emission lines, which are distinct from nearly all SNe (when observed post peak) and AGN. van Velzen et al. (2019) define three different classes to classify 32 TDE candidates in their ZTF sample according to their spectra i.e.:

- TDE-H (broad Hα and Hβ emission lines).

- TDE-He (no broad Balmer emission lines, a broad emission line near He II $\lambda4686$ only).
- TDE-Bowen (broad Hα and Hβ emission lines, a broad complex of emission lines around He II $\lambda4686$ and N III $\lambda4640$ and emission at $\lambda4100$ identified as N III $\lambda4100$ instead of Hδ, and in some cases also O III $\lambda3760$).

Further, they reported different photometric behaviour among different spectroscopic classes (e.g., TDE-Bowen have longer rise times). Larger samples, however, will be needed to confirm these findings and investigate these classes as well as their hosts galaxies. To do so, however, it will be vital to be able to efficiently select candidates for follow-up. With the efficient selection and confirmation of TDE candidates, it will be possible to better estimate the TDE rates, which can be used as probes of the SMBH demographics (e.g., Stone & Metzger 2016).

Correlations between the photometric properties of TDEs (e.g., rise/fade time) and physical properties of their host galaxies (e.g., stellar mass, black-hole mass) may be proved very useful for surveys like the LSST. With an r-band magnitude limit ∼27, the LSST will detect very faint transients for which spectroscopic follow-up will be difficult or even impossible. This means that the classification of many objects will rely solely on their photometric properties (e.g., rise time, colour evolution) and the photometric properties of their host galaxies (should it be found that the correlations between host galaxy properties and TDEs prove to robust). Current wide-field surveys can contribute significantly to demonstrate such correlations (keeping in mind that these results could be biased due to the flux limit).

As previously highlighted (and described fully in the next section), GOTO has all the characteristics to be a powerful transient detection and identification facility. While its cadence during the prototype phase (i.e., 4 UTs; see next section) is considered low for transient detection, once it has its full complement of UTs and a cadence of 1-3 days, it has the potential to be an efficient transient detection machine. In this thesis, we aim to investigate the the prospects of using GOTO to quickly detect and identify TDEs from photometry alone, and explore how its planned future cadence will enhance this ability.

1.4 Gravitational-wave Optical Transient Observer (GOTO)

Since the very first detection of gravitational waves in 2015, it was clear that it would be necessary to develop instruments that would be in place to follow-up possible on-sky optical counterpart of these detections. Even though the existing telescopes around the world would try to follow-up alerts in the years after the first detection, there was an obvious gap in terms of instruments that would follow the alerts rapidly as soon as the alert was released. To fill the void, a collaboration was formed between the University of Warwick and Monash University as well as Armagh Observatory, the University of Sheffield, the University of Leicester, the National Astronomical Institute of Thailand, the Instituto de Astrofisica de Canarias (IAC), The University of Turku and Rene Breton (The University of Manchester) with the goal to establish a new observatory dedicated to the rapid follow-up of gravitational wave counterparts. The result was the Gravitational wave Optical Transient Observer (GOTO), which we describe in this section.

The rapid follow-up of gravitational wave events is very important to localise events and thus identify viable counterparts. Therefore, the primary goal of the GOTO project is to localise and identify electromagnetic counterparts to gravitational wave events. Two key aspects arise from this goal: the need for rapid follow-up of an alert and the need for recent reference images to be able to identify those sources that have recently varied or appeared, which may flag them as a potential counterpart. With these goals in mind, the GOTO collaboration decided upon a set of specifications involving flux limit, field of view, cadence and cost, which ultimately led to the final design of the observatory.

In the first phase, a single dome was installed on La Palma (GOTO north node; GOTO-N) to house an array of 40 cm-diameter astrographs (f/2.5) attached to the same mount, with each astrograph equipped with a 50M pixel detector with a field-of-view of roughly 5 sq. degrees (and a corresponding pixel scale of 1.25 arcsec). At the time of writing, GOTO consists of eight astrographs (hereafter, unit telescopes, or UTs), although the data described in this work was obtained prior to the second set of four UTs being added. The four UTs used to obtain the data analysed in this study were aligned such that they deliver a contiguous field-of-view of roughly 18 sq. degrees per mount pointing. In terms of filters, each UT is equipped with a filter wheel consisting of standard Baader R, G, and B filters, plus a broad L-band filter which covers the entire optical passband. The latter is used as the primary filter for the survey as it maximises the amount of light reaching the detectors in a given exposure. Eventually, the full GOTO-N node will consist of two domes housing a total of 16 UTs, which will deliver a total field-of-view of ∼75 sq. degrees. A southern Australian nodeis planned and part-funded. Together, both nodes would deliver a full sky coverage with a cadence of around 2-3 days.

During the four UT phase (i.e., prototype phase), the mean cadence is estimated to be 10 nights in order to cover the full GOTO visible sky. With the 8 UTs setup the cadence will drop to ∼5 days while for the full GOTO-N of 16 UTs, which can instantaneously capture ∼75 square degrees of sky, the full visible sky will be covered every ∼2.5 nights. This cadence is vital for the primary science of GOTO i.e. to detect the optical counterparts of gravitational wave events. The high cadence, however, is also important for GOTO's secondary science driver of contributing to time-domain astrophysics via the discovery of variable and transient sources and the study of their lightcurves (e.g., cataclysmic variables, supernovae, tidal disruption events, etc.). GOTO's high whole-sky cadence will allow us to obtain lightcurves for large samples (numbering tens, or possibly even hundreds of thousands) of AGN, ultimately enabling us to study the correlations between AGN variability and other properties to a level of detail unavailable before. With this high cadence and the survey volume, we also expect that we will be able to catch "changing-look" AGN (CL-AGN); i.e., AGN that appear to transition from having Type 1 to Type 2 spectra, with an associated dramatic dimming or brightening of their photometry (e.g., MacLeod et al. 2016). The rates of CL-AGN are not yet well constrained and so GOTO with its volume and cadence could catch many of these events, thereby revealing more about how frequently they occur, which may help to unveil why they undergo this transformation. Finally, the volume and cadence of the GOTO survey is such that we should obtain a statistically significant sample of TDEs (estimate ∼10-20 per year using van Velzen et al. 2011 for the full sky coverage), which will contribute to constrain the TDE rates with regards to the black hole mass and the host galaxy type.

The theoretical capabilities of GOTO come, however, with many challenges. First, we need to test whether the quality of the data allows us to study the variable sky, as in the case of AGN, e.g., free of systematics that could hide any intrinsic variability. It is also important to to identify the best method (e.g., forced photometry, difference imaging) for each scientific task (e.g., measuring variability, measuring transient flux; see section 1.5). In order to be able to identify AGN variability on different timescales (weeks, months, years), we should make sure that the lightcurves are free of systematics that could be catastrophic for this task (such as long-term systematic trends that could, for example, be introduced by changes to hardware or software). In the case of TDE candidates, there are also other things to consider, such as the capability of GOTO processing pipelines to reliably measure the location of the transient in relation to its host nucleus, which (i.e., distance from the nucleus of the host) could be especially challenging for GOTO with its large pixel scale (1.25"/pixel).

As in the case of all wide-field, high-cadence surveys, GOTO needs to address the challenge of of processing large volumes of data on short timescales. To overcome these challenges, GOTO has developed its own in-house pipeline, which we describe in the next section of this chapter. We also, however, explain the reasons we wished to develop a secondary pipeline based on the LSST data processing pipeline in order to study AGN variability, in particular.

1.5 Processing pipelines of GOTO data

In order to address the challenges presented by the data volume and rate delivered by wide-field, high cadence surveys, significant resources have been allocated to the development of efficient data processing pipelines (e.g., SDSS;Lupton et al. 2001,Pan-STARRS; Magnier et al. 2016). GOTO's primary science goal (i.e., detection of optical counterparts to GW events) requires a bespoke data processing pipeline which has been developed in-house (`gotophoto`). Since it is focused on the detection of transient point sources within galaxies (as almost all gravitational wave events are expected to be extragalactic), the main focus of the pipeline is difference imaging i.e., subtraction of the nightly science images and reference images selected for the whole observable sky. Prior to difference imaging, however, `gotophoto` (hereafter GOTOphoto) performs the usual calibration steps including bias, dark, flat-field corrections, and astrometric (e.g., ATLAS-REFCAT2 ; Tonry et al. 2018) and photometric calibration (following source detection with `SExtractor`; Bertin & Arnouts 1996). For the difference imaging, GOTOphoto uses `hotpants` (Becker 2015), the output of which is passed through `SExtractor` to identify those sources that have changed in flux in the intervening period between when the reference and science images were obtained. For this task, it is very important to have good quality reference images (e.g., good seeing) and to make sure that science and reference image are astrometrically and PSF matched.

Difference imaging is powerful means of detecting transient sources such as TDEs, and we will use the GOTOphoto outputs to investigate the prospects of observing TDEs with GOTO. In the case of AGN variability, however, it is possible that small changes could be lost by difference imaging. Further, the fact that we are dealing with variable sources embedded within extended sources may also introduce additional problems if we relied solely on difference imaging. In order to be able to to study AGN variability, we decided that the best method to generate photometry for this purpose was forced photometry, in which the flux of a source at a known position is measured in each incoming science frame. This is in contrast to the more familiar "blind photometry", in which a detection algorithm is first used to identify and locate sources prior to their flux measurement. Since GOTOphoto does not currently perform forced photometry, we decided to adapt the LSST data processing pipeline (hereafter, the LSST stack) – which does include a forced photometry module – to process GOTO data. We describe how we achieved this in Chapter 2.

1.6 This thesis

This thesis aims to investigate all the issues mentioned in the previous sections with regards to the prospects of studying AGN variability and identifying TDEs using GOTO. In the first part of the thesis (Chapters 2 and 3), however, we focus on our efforts on adapting the LSST software stack in order to process GOTO data. After that we focus on our exploration into the feasibility of using GOTO data (whether processed by GOTOphoto or the LSST stack) to study AGN variability, using ZTF data as a quality check (Chapter 4). Unfortunately, our investigations show that GOTO data is not currently of sufficient high quality to investigate AGN variability, so we instead turn to ZTF data alone to investigate the variability of moderate luminosity AGN and how it relates to other AGN properties (Chapter 5). Finally, we present an assessment of the prospect of identifying and studying TDEs using GOTO survey data (Chapter 6), before consolidating all our findings in Chapter 7.

Chapter 2

Data Processing

2.1 Chapter Introduction

This chapter gives a more detailed description of the LSST software stack (i.e., the pipeline that will process the data obtained by the Vera C. Rubin Observatory). Despite being written with the LSST survey in mind, the LSST stack has been written in such a way that it can be adapted to other surveys obtained by different observatories. To do this, it is necessary to develop a software package (known as an "obs_package") that acts as an interface between the LSST stack and the survey data. To process GOTO data with the LSST stack, we have developed our own "obs_package", the `obs_goto`. This chapter outlines the structure of `obs_goto` and the configurations, which were needed for the GOTO data. Finally, the chapter outlines the elements of the LSST stack-based pipeline used to process the data in order to achieve the goals of this thesis.

2.2 The LSST software stack

The LSST will be the largest and deepest wide-field astronomical survey ever undertaken. As such, it presents a number of major challenges in terms of the sheer quantities and rate of data that must be processed in order for the project to achieve its science goals. Over the ten years that the LSST is expected to operate, it is expected that it will result in over 500PB of cumulative processed imaging data and more than 50PB of catalogue data from the 11 data releases. The nightly amount of raw imaging data is expected to be around 15TB (Jurić et al. 2017) . In response to the need of a pipeline that will handle and process this amount of data in a fast and efficient way, the LSST management team has put considerable effort into developing a pipeline to address these challenges. The pipeline will first deliver the outputs of the nightly processing, which includes sources that have changed in brightness or position, as well as catalogues from difference imaging. In addition to measurements from nightly processing, the LSST stack will deliver the annual data releases, which will include all the data processed to date, including fluxes, shapes, variability and lightcurves for all the detected sources.

The LSST stack is an open source code [1] and its algorithm is written in two programming languages, Python and C++. Python is used as the primary language for the high level, "calling" scripts, whereas C++ is used mostly for performance reasons (Jurić et al. 2017). The LSST stack, although developed for the LSST, it is designed to be a standard processing pipeline for other wide-field surveys. The high-cadence data from GOTO and the fact that the secondary science goals of GOTO bear many similarities to those of the LSST, inspired us to investigate whether the LSST stack could be used as a secondary data processing pipeline for GOTO. To adapt the LSST stack to process GOTO data, I needed to develop the necessary interface, the `obs_goto` package. This gives the LSST stack information specific to the telescope and camera (e.g., ccd dimensions, pixel scale, etc) and configures the LSST stack tasks to process GOTO survey data

---

[1] https://github.com/lsst

2.3 The `obs_goto` package

The `obs_goto` package is a collection of five directories, each of which handles a different aspect of the processing.

- The first of these directories is the `policy` directory, which includes a file that provides information regarding the location and format of the input and output data files. Standard input and output files that are necessary for the data processing are included in the `obs_base` package of the LSST stack; these are used as a fall-back option if specific records are not contained within the policy file. The policy file itself describes data formats of the inputs/outputs together with the path of the file location, and we can also use our package's `policy` file to write intermediate products, which are not included in the `obs_base` (e.g., PSF model).

- The second directory of the `obs_goto` package is the `camera` directory, which includes files that inform the stack of the various properties of the GOTO CCDs (i.e., CCD is the same as the UT). A number of FITS tables in this directory contain information regarding the physical properties of each CCD. Each UT has its own table in which the two rows represent the left and the right amplifiers of the GOTO UTs and the columns contain a number of different properties. In their current form, the tables include the gain, the readout noise and the overscan pixels for each UT and the saturation level. Table 2.1 lists the numbers for the gain and readout noise[2] for each amplifier for each UT, which were used when processing the data. For the overscan, we have used ten pixel columns on each side of the CCD and the saturation level has been set at 65535 [ADU][3]. The `camera` directory has also a configuration file for the camera that lists some important parameters.

---

[2] Numbers for the the gain and readout noise come either from the manufacturer or report from the test performed at the University of Sheffield by Martin Dyer.

[3] This is the manufacturer's number however tests showed that this vary between ~64,100 and ~65,040 depending on the UT and future developments should take this into account.

These parameters currently are the plate scale (206.67 ”/mm), the X (6131) and Y (8175) size in pixels for the UTs and the pixel size (0.006 mm).

- The `config` directory includes configurations files for the different LSST stack tasks we run on the GOTO data. With the `config` files, the user can change parameter values or methods of measurement (e.g., magnitude limit for reference for astrometric and photometric calibration, PSF modelling method). It is also possible, via the configuration files, to override a default process with a bespoke, user-defined one (e.g., Instrument Signature Removal), a procedure known as “re-targetting”. In the next section, we will describe these tasks and we shall see the configurations we have implemented. It is important, however, to note here that these configurations are covering various steps of the processing (e.g., instrument signature removal, background subtraction, photometric calibration) and the `config` file is the primary way by which these steps are controlled.

- The fourth directory of `obs_goto` is the `python` directory which includes python scripts that will override the default processes (as mentioned above). This directory also includes one of the most important scripts within the `obs_package` - the camera `mapper`. This overrides the basic `CameraMapper` class of the LSST stack and controls the interaction between the data and the pipeline. Sub-classes of this class allow the mapper to “know” where the `policy` file resides in the file system. Via the specific (e.g., GOTO) `mapper` the code also knows where the physical properties of the detector are stored and how to read that data. As we mentioned, in this directory the LSST stack user can add their own scripts that modify a given task in order to meet their specific requirements. For GOTO, we have modified the task that ingests the data into the database, which the stack uses to store key information regarding each frame. As described later in this chapter, we have also developed our own scripts to (a) calculate the zeropoint of each frame and (b) adapt the LSST stack's forced photometry modules to our own needs (see section 2.4.2).

- Finally, the `ups` directory contains a table of the LSST stack packages that must be set up in order to process GOTO data. Table 2.2 lists the different packages together with a brief description of what each package does. We note that the main packages needed to be set up for the `obs_goto` to work is the `pipe_tasks` and the `pipe_drivers` packages which include all the necessary tasks performed in our pipeline and scripts to coordinate these tasks. We also set up the package that contains the reference catalogue used for astrometric calibration (which currently differs from the one used for photometric calibration).

	Gain(e^{-1}/ADU)		Readout noise (ADU)	
UT number	L	R	L	R
1	0.65	0.65	20.9	20.6
2	0.52	0.52	23.0	22.1
3	0.59	0.57	22.1	20.7
4	0.59	0.59	23.0	23.4

Table 2.1: Table of gain and readout noise.

In Figure 2.1 we give a schematic representation of the `obs_goto` with the basic tasks that the LSST stack can perform on scientific data. Our aim was to perform forced photometry (which was briefly presented in section 1.6 and and expanded upon in section 2.4.2) on GOTO data. While the LSST stack contains image subtraction modules, since we use GOTOphoto's image subtraction data, that particular processing step is beyond the scope of this thesis. In the next section (2.4), I give more details on the tasks of our configuration of the LSST stack pipeline and its outputs.

Package	Description
`pipe_tasks`	This includes the scripts of the astronomical data processing tasks.
`pipe_drivers`	This includes high level task coordination scripts. These tasks allow a single command line entry point to tasks which otherwise would be run serially.
`meas_extensions_photometryKron`	It includes the scripts to perform Kron photometry.
`meas_extensions_astrometryNet`	Astrometry solver that uses astrometry.net.
`astrometry_net_data ucac4`	It sets up the directory that includes the UCAC4 index file to be used from the astrometry.net.

Table 2.2: Table of LSST stack packages we need to set up in order to use the `obs_goto` to process GOTO data with the stack.

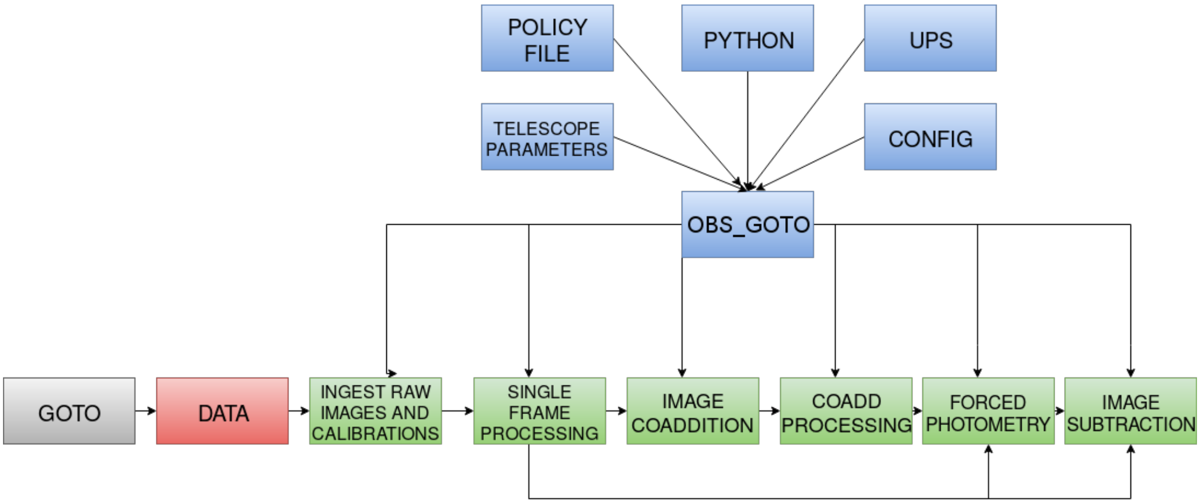


Figure 2.1: A schematic representation of the `obs_goto` and the pipeline flow that can be used to process survey data.

2.4 The processing stages

The aim of this thesis was the production of lightcurves generated from forced photometry measurements. In order to achieve this, we have used the LSST stack's forced photometry modules. By its nature, forced photometry is a two-step process. First, I created a reference catalogue of detected sources from archival GOTO observations. Once the reference catalogue had been produced, I performed forced photometry on incoming science frames at the positions of the sources in the reference catalogue. Each of these steps (i.e., the production of the reference catalogue and the forced photometry) consists of a number of different smaller steps, which are described in this chapter.

Although the production of the reference catalogue and forced photometry show differences in the processing stages, they have common elements, especially in the earlier steps of each. If we need to use data to produce a reference catalogue or we wish to perform forced photometry on nightly data, we first need to inform the stack of the presence of new data on the filesystem. To do this, we ”ingest” the data with the `ingestImages.py` script, which involves populating a database (`registry.sqlite3`) with key data from the image headers, including:

- Image type (i.e., if the raw image is a science, focus, bias, dark or flat-field).
- Date of observation (format: YYYY-MM-DD).
- Exposure time.
- Filter used for the image.
- The field that was observed (i.e., name of tile or based on trigger).
- UT number.
- Run number.

- Visit number (i.e., it is the number that groups the back-to-back exposures together and it is unique number for this group in the observations database (Dyer et al. 2018). Different UTs however can have the same visit number, which means that a combined image of the back-to-back exposures is unique for a combination of the visit and the UT number (see section 2.4.2 for more details).

- The Modified Julian Date (MJD).

During ingesting, `ingestImages.py` also creates softlinks to the raw input files in such a way that they obey the naming convention in the policy file.

Once the raw data are ingested onto the database, they are ready to be processed. Before we process the science images, we first need to create master calibration frames which will be used to correct the raw science images. In the case of GOTO data we create master, biases, master darks, and master flats. The master frames can be generated using the bias, dark or flat frames of the night or using several frames across multiple nights. We generate these using LSST stack tasks: `constructBias.py`, `constructDark.py` and `constructFlat.py`. The bias and dark frames are mean combined but the flat frames are median combined to avoid artefacts or other effects (e.g., from bright stars) on the master flat frames. These master calibration frames are then stored in their own database (`calibregistry.sqlite3`). In the case of the calibration frame databases, the table includes the type of the master calibration frame, the date of observation and the period for which the frame is valid. Examples of master GOTO calibration frames generated by the LSST stack are shown in Figure 2.2.

After the master calibration frames have been created, the next step is the processing of the science images. As we explained, to perform forced photometry using the LSST stack, we first need to produce the reference catalogue. In the next section, I describe the processes that lead to the production of the reference images and the respective reference catalogue.

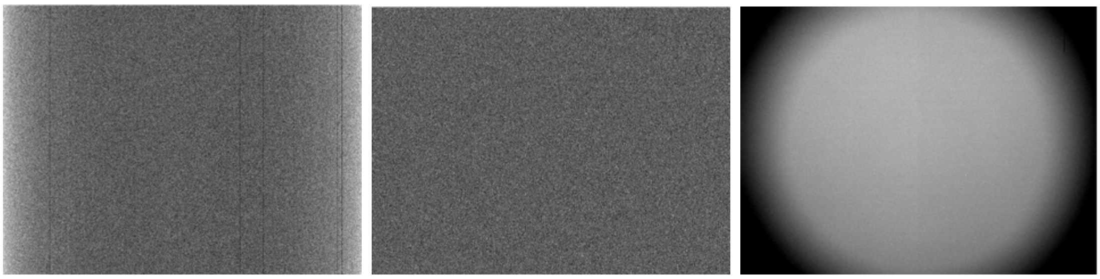


Figure 2.2: Master bias (left), master dark (centre) and, master flat (right) from GOTO data, as generated by the LSST stack.

2.4.1 Reference catalogue production

First, we describe the steps that lead to the production of the reference catalogue by performing source detection on reference images. These are produced by coadding multiple frames observed across multiple nights. To do so, we processed the science images through three LSST stack pipe drivers i.e. `singleFrameDriver.py`, `coaddDriver.py` and `multiBandDriver.py`. An LSST stack pipe driver is a collection of tasks which can be paralleled across multiple computer nodes for performance purposes (i.e., you can independently process each individual science frame since it does not depend on any others).

First, we processed the data that will be used to produce the reference images through `singleFrameDriver.py`, which includes a number of different sub-processes in order to process the raw imaging data and produce calibrated exposures (i.e., astrometrically solved and associated with a single photometric zeropoint). The first step of `singleFrameDriver.py` is instrument signature removal (ISR), which involves performing bias, dark and flat-field correction. After that, the science images are passed through the `characterizeImage` step of the `singleFrameDriver.py`. During this step, after an initial background characterization (Figure 2.3) and subtraction, the LSST stack performs a high signal-to-noise source detection. From the subsequent catalogue of detected sources, a selection of point sources is made, which the stack uses to characterise

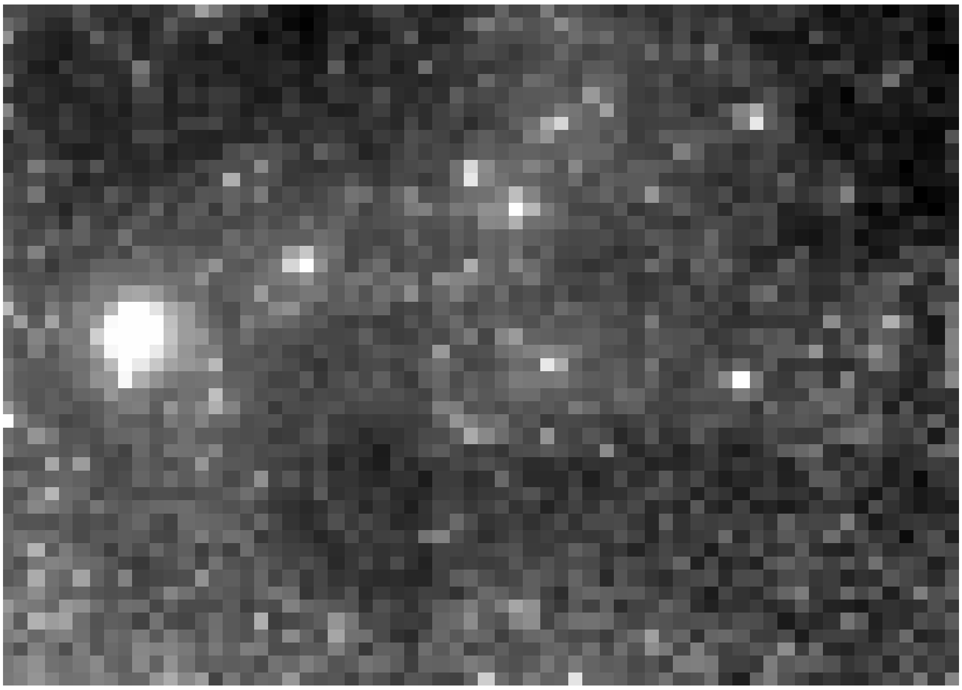


Figure 2.3: Example of the background model at the centre of an image as generated by the LSST stack.

the spatially-varying PSF across the image. Using the configuration file for this task, we have chosen a 100σ detection threshold and used the default LSST stack source selector to identify point sources. The latter uses a K-means algorithm (i.e., partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean) that looks for a cluster of small objects in a size-magnitude plot (where the size is calculated from the shape measurements of the source; Bernstein & Jarvis 2002). The large field-of-view of the GOTO UTs allows us to find many sources that meet the selection criteria, although the large FOV also means that point sources in the outskirts of the image appear extended. To avoid rejecting sources which would lead to poor PSF modelling, we have configured the task to increase the number of standard deviations in the range of source sizes –from 0.15 to 10– before they are excluded from "good" stars and we have also increased the K-means sigma clipping threshold(from 2.0 to 10.0). Figure 2.4 from Mullaney et al. (2021) shows the effect of relaxation of the point source candidates selection criteria, which allow better and fairer PSF modelling across the whole GOTO frame.

For the GOTO science images, we have used the Principal Component Analysis technique to characterise the PSF model, which is also the default method in the LSST stack. We performed tests to investigate the alternative PSFex method (Bertin 2011) but we found that PCA did a better job (for the GOTO data) than PSFEx. Due to the aforementioned issues regarding GOTO's spatially varying PSF, we have decided to relax some of the constraints for the PSF fitting. In the case of fitting, the number of eigen components for PSF kernel creation was increased from the default four to six. The reduced-χ^2 threshold above which sources are considered outliers and subsequently rejected from the next iteration was increased from 2 to 50 (i.e., similarly to the aforementioned point source selection routine) and the standard deviation threshold for rejecting sources from the the spatial fit was also increased from 3 to 50. Finally, the size of the cell to determine PSF modelling was increased from 256×256 to 512×512 considering the larger pixel scale of GOTO (1.25"/pixel).

After the `characterizeImage` step is complete `singleFrameDriver.py` performs the `calibrateImage` task which is the final task called by this driver. This step will produce the calibrated exposures (`calexp`) that will be coadded to produce the reference images. A `calexp` image is a 3-extensions FITS image, which consists the calibrated science image, a mask, which labels every pixel according to a bit-mask (see Table 2.3) and finally the per-pixel variance of the science image (Figure 2.5). The `calibrateImage` task involves the final background modelling and subtraction after source detection and also the astrometric solution and the photometric calibration of the science exposures. In this case, for source detection, we have chosen a 30σ threshold. Sources from this detection will be used by the LSST stack calibrate task to perform astrometric and photometric calibration of GOTO science images.

First, we describe how we performed astrometric calibration on the GOTO science images, which were used in the production of the reference images. The astrometric

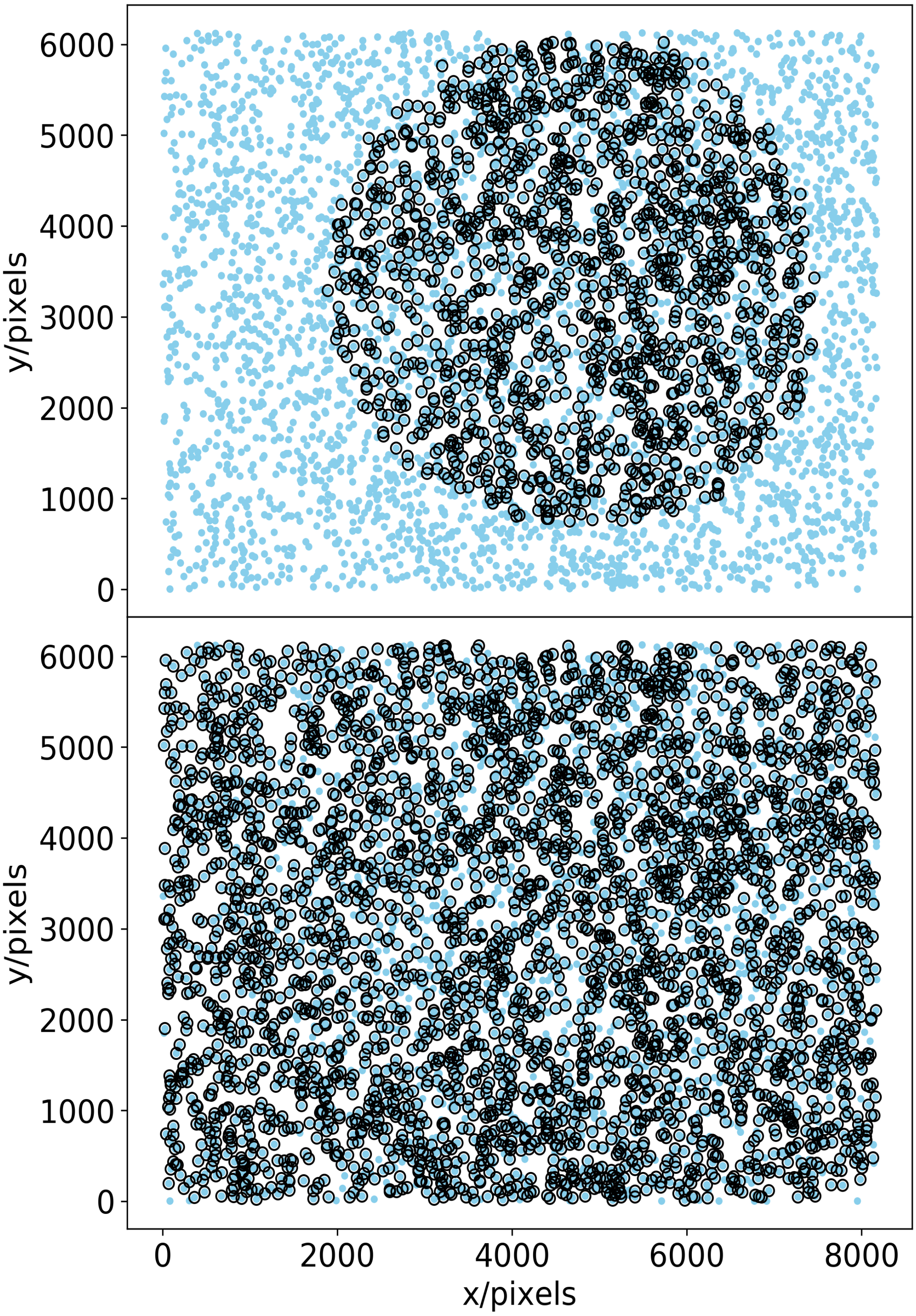


Figure 2.4: The figure from Mullaney et al. (2021) shows sources that the LSST stack used to construct a model of the spatially-varying PSF within an individual GOTO frame, where the light blue points show the pixel positions of all sources detected at $> 100\sigma$ within the frame, whereas the black circles represent those that have been selected as candidates for PSF modelling based on their shape and size. Top panel shows the selection using the LSST stack default parameters whereas the bottom panel shows the PSF modelling candidates after we relaxed the selection criteria. This shows that with the relaxed criteria, we are able to have candidates across the whole frame which allows better PSF modelling on the central parts of the image as well as on the outer parts.

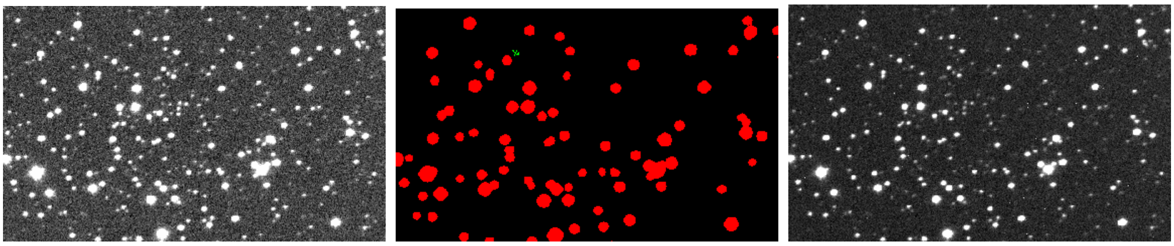


Figure 2.5: The three extensions of an calexp image (zoomed in). The first extension is the calibrated exposure (left), the second the mask (centre) and the third the variance (right). We have selected a part of the image to show the an example of masked pixels. The red colour represents pixels masked as detected whereas the green pixels mask value indicates that these pixels are affected by cosmic rays and interpolation.

solution is a very important step as the positions in the reference catalogue are what we use to perform forced photometry on GOTO nightly images. On arrival from the telescope, the GOTO raw image headers lack full WCS information, although they do contain the requested Right Ascension and Declination of the telescope mount pointing. However, since each UT points to a slightly offset position on the sky from the mount (by as much as a few degrees), the requested RA and Dec only provide a rough estimate for the boresight of each frame. To obtain an accurate astrometric solution, we run the frames through the LSST stack's astrometry task. While the stack has its own internal astrometry solver, it also offers the option to use `astrometry_dot_net` (Lang et al. 2010), which we chose for reasons of familiarity and because it is better when the true pointing is either unknown or a long way from the header astrometry. For an astrometric reference catalogue, we use the the Fourth US Naval Observatory CCD Astrograph Catalog (UCAC4; Zacharias et al. 2013), since it contains stars which are well-matched in brightness to GOTO's detection limit. For future developments, however, we would consider the implementation of other catalogues such as PanSTARRS or GAIA. The UCAC4 catalogue is an all-sky star catalogue that covers the g-band magnitude of 8 to 16 magnitude. Whilst not especially deep, it is adequate for our purposes since GOTO's wide field-of-view always contains sufficient numbers stars within this magnitude range for an astrometric solution to be found. The UCAC4 catalogue has a positional accuracy

Mask Name	Description
CR	Cosmic ray detected in the source footprint.
CR_CENTER	Cosmic ray detected in the source center.
EDGE	Could not use the full model in the fit because of proximity to the exposure border.
INTERPOLATED	Interpolated pixel (e.g., for bad columns) in the source footprint.
INTERPOLATED_CENTER	Interpolated pixel in the source center
SATURATED	Saturated pixel in the source footprint.
SATURATED_CENTER	Saturated pixel in the source center
BAD	Bad pixel in the source footprint.

Table 2.3: Table of the pixel mask types generated for each frame (if such pixel exist) and their descriptions. These masks are also translated to flags in the measurement catalogues generated by the LSST stack for each frame.

of ∼15-100 mas per coordinate (depending on magnitude). Since the only initial WCS information contained within the header is the requested mount pointing, we consider a large radius of 5 degrees around the pointing (input) RA and Dec to search for a solution. We also choose a large matching radius of 120 arcsec between GOTO detections and the reference objects to accommodate potentially large non-linearities. Despite these large search and matching radius, the solver typically only takes a few microseconds (on a Dual In- tel Xeon E5-2697v3 2.60Ghz CPU with 28 cores/56 threads with access to 256 GB of RAM) to identify a suitable astrometric solution. The astrometric solver selects a maximum number of 1000 stars, which are not masked as saturated or having pixels affected by cosmic rays in their central regions, to calibrate astrometry for a single exposure. Then, it iterates to improve the fit by rejecting those matches that have a deviation larger 3σ (we fit a third-order polynomial distortion solution in both axes). The astrometric solution is written in the calibrated exposure's FITS header.

Once an astrometric solution has been found, we use the LSST stack to perform photometric calibration on each science GOTO image. To calibrate the photometry we use point sources from the PanSTARRS DR1 catalogue (Chambers et al. 2016), which covers the whole observable sky above -30 declination, which makes it ideal for the GOTO as its coverage is similar to GOTO-N. When the southern node starts to observe, we will need to implement other catalogues (e.g., Gaia DR2). PanSTARRS uses five broadband filters i.e., g,r,i,z,y to take sky images. In order to achieve better photometric calibration, we applied some criteria for the source selection in terms of magnitude and colour. With respect to the magnitude of the selected sources, we chose the sources to be between 11 and 19 mag in the i-band. We also select the PanSTARRS sources that have $g - r > 0.0$ and $r - i < 0.5$. The first cut allows us to avoid some extreme outliers and the second cut is used to avoid stars later than K6V, which show colour variation with metallicity (Aihara et al. 2019). By default the LSST stack excludes from the sources that will be used as photometric calibrators, sources with pixels masked as saturated or interpolated (e.g., interpolated from defects, hot pixels) as well as sources that are near the edge of the image. The selected sources on GOTO images are positionally-matched to the reference sources using a matching radius of 1.25” (1 pixel).

Since GOTO’s broad L-band filter is not represented by any of PanSTARRS filters, we must calculate colour terms to account for the effect of reference stars with different-shaped SEDs on our photometric calibration. Synthetic L-band reference magnitudes (p') are calculated using the following formula:

$$p' = primary + c_0 + c_1 * (primary - secondary) + c_2 * (primary - secondary)^2, \quad (2.1)$$

where primary and secondary are the reference magnitude filters chosen for each GOTO filter e.g. to calibrate the L band magnitudes with the PS1 catalogue, we chose as primary the PS1 g-band and as secondary the PS1 r-band. The parameters c_0, c_1, and

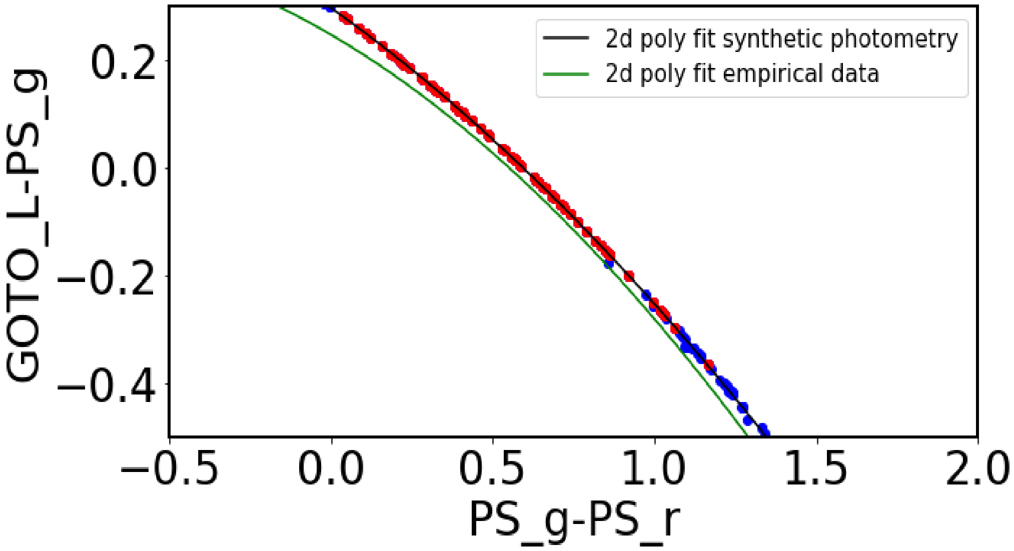


Figure 2.6: Polynomial fitting from synthetic photometry using the Pickles library (Pickles 1998) and the respective fit from the empirical data. Using the Pickles library we pass synthetic stellar spectral models through the PanSTARRS and GOTO filter passbands to calculate the plotted differences.

c_2 are found via fitting a second-order polynomial to a plot of the the colour index, GOTO$_{mag}$-PS1$_{gmag}$, versus the PS1 colour i.e., $primary - secondary$ (i.e. $PS_g - PS_r$), where GOTO$_{mag}$ is the non-colour corrected GOTO magnitudes. In this way using the PanSTARRS, we calculate synthetic calibrated GOTO L-band magnitudes.

Initially, we used empirical data (i.e. calibrated photometry using the non-colour corrected g-band magnitudes) to produce the plots (i.e. GOTO$_{mag}$-PS1$_{gmag}$ vs. $PS_g - PS_r$) that we fit to derive the colour terms. Later, however, we recalculated the colour terms using synthetic photometry and the Gunn-Stryker atlas (Gunn & Stryker 1983) and the Pickles library (Pickles 1998), obtaining a fit that is close to that obtained using empirical data (Figure 2.6). For reasons of consistency and time, we continued processing the data with the colour terms extracted from the empirical data. After matching with the reference sources and colour correcting the reference magnitudes, the LSST stack calculates the zeropoint using the reference-GOTO instrumental magnitude difference. It finds a mode and then iterates 20 times using a sigma-clipping method to remove outliers and finally calculate the zeropoint.

Deep coadded reference images are produced by combining the calibrated science exposures. Prior to co-addition, however, the LSST stack needs to define a reference grid which divides the sky into multiple (slightly overlapping) parts which are referred to as tracts. The stack then re-projects every calibrated frame onto this grid of tracts. This avoids problems associated with misalignments due to pointing errors when coadding the individual science frames. In the case of GOTO, there is an additional advantage, as future reconfigurations of the full GOTO field-of-view due to the addition/removal of UTs will not affect this grid. Tracts are then subdivided into patches which share the coordinate system of the tract. Splitting the tracts into patches is necessary for processing and storage purposes (Bosch et al. 2018). To split the sky in the reference system we have used the `makeSkymap.py` LSST stack task. For our map projection, we chose a HEALPix (Górski et al. 2005) grid with a resolution parameter of four, which corresponds to 192 tracts across the whole sky (Figure 2.7). Our patch size is 4k×4k pixels (1.78 deg^2) and each patch overlaps by 100 pixels. The result of this task is stored as `deepCoadd_skymap`. Figure 2.8 shows an example of the patches in a single tract.

Once the sky grid has been created, we process the calibrated exposures that will give the reference images and the respective catalogues through the `coaddDriver.py`. Before co-adding the calibrated exposures, they must be *warped* onto the pixel grids of patches created by `makeSkyMap.py`. These warped images are then combined (using the `assembleCoadd` task carried-out by `coadddriver.py`) to give deep coadd reference images of each tract/patch combination, with variance and mask information propagated through to the coadd. The variance images of each coadd ensures that the errors are propagated correctly as shown in Figure 2.9. The results of the coadd driver, i.e. the calibrated coadd images, are stored as `deepCoadd_calexp`, for each tract/patch combination. In Figure 2.10, we see the comparison between an individual calibrated exposure and the respective coadd image for the same part of sky. We find that more sources can

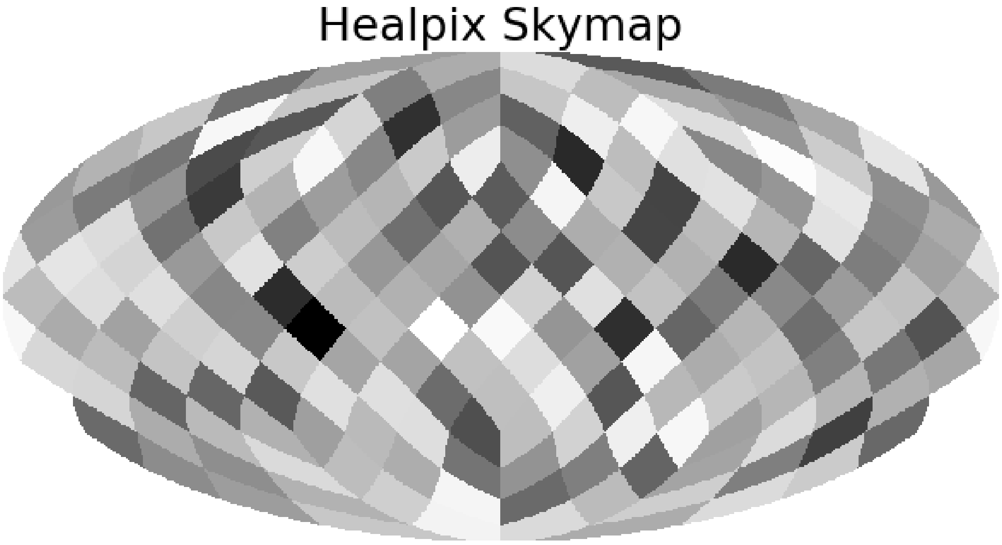


Figure 2.7: Healpix skymap with resolution parameter equal to four, which has as results 192 tracts across the whole sky.

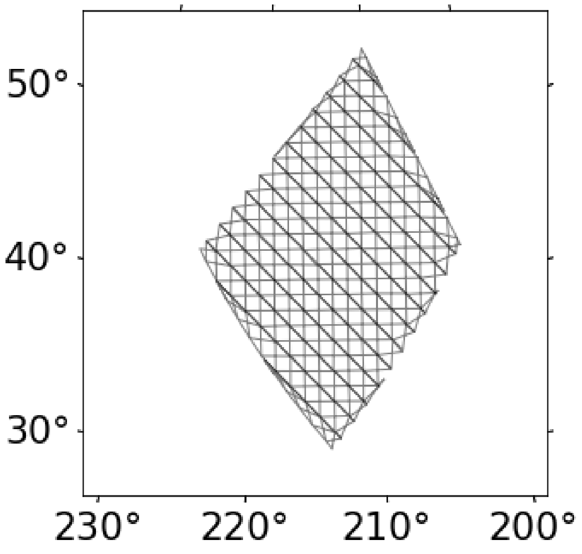


Figure 2.8: The plot shows the patches in tract 161.

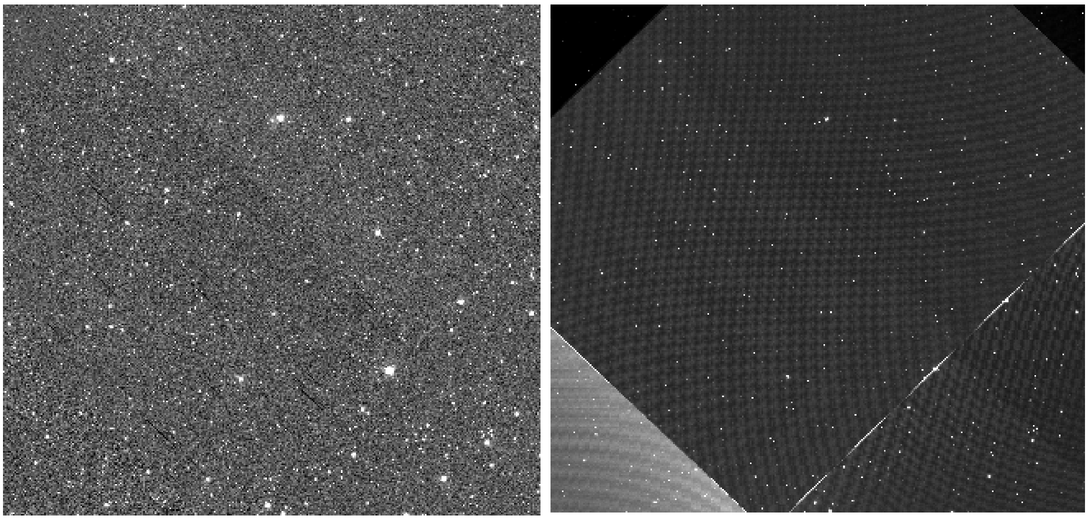


Figure 2.9: Example of coadd image (left) and its variance (right) which ensure errors are propagated correctly.

be identified even by eye on the coadd images compared to the single images.

The final step is the production of the reference catalogue generated by the `multiBandDriver.py`, which works in tract/patch space. Once the deep coadd reference images have been produced, the LSST stack runs a detection and measurement algorithm on each. The first step is to run the detection algorithm on each image (`detectCoaddSources` task), which gives a list of detected sources in each patch per filter. During the detection of the sources, we have implemented a local background subtraction in bins of 32×32 pixels, which improves the deblending. As its name suggests, `multiBandDriver.py` is designed to detect sources in multiple bands and merge the resulting source lists into a single multi-band catalogue. In our case, however, we have only processed L-band data, meaning we do not need to merge across bands. The detected footprints (area of pixels with values above the detection threshold) are then deblended (`deblendCoaddSources`) into the different sources if needed, and once this is done the LSST stack performs flux and shape measurements on the blind detections for each filter (`measureCoaddSources`). After performing direct measurements, the reference catalogue is produced (`mergeCoaddMeasurements` task). The resulting reference

Figure 2.10: Part of a individual calibrated image (left) and the respective part of the coadd image generated from images takes in a specific range of dates (see Chapter 3). The effective exposure time of the nightly images is 180 s while the respective for the coadd images is at least 180 s.

catalogue (stored as `deepCoadd_ref`) is the one that it is used as the reference for forced photometry on single CCD images. The reference catalogue is also divided to the sub-regions (patches) of sky we described in the previous section. This reference catalogue contains the centroid positions in both sky and image coordinates, shape measurements obtained using the SDSS algorithm (i.e., adaptive moments for ellipticity [4]; Bernstein & Jarvis 2002) and various photometry measurements for all the detected sources. In our case, we obtain aperture fluxes for selected apertures radii (3.72”, 5.58”, 7.44”, 11.16”, 14.88”, 29.76”, 59.52” based on integer and half-integer pixel numbers) along with PSF flux and Kron flux (Kron 1980).

2.4.2 Forced Photometry

After the reference catalogue was generated, we were able to perform forced photometry on nightly GOTO data. To perform forced photometry on GOTO images in the most efficient way in terms of time and memory, we took into account the way GOTO is

---

[4] https://www.sdss.org/dr12/algorithms/classify

conducting its survey. While in its standard survey mode, GOTO takes multiple (usually three) back-to-back exposures for each pointing, which are grouped together according to their visit and CCD numbers. Each visit is identified via a unique identification number which associates it with a given pointing (see Dyer et al. 2018 for further details), as opposed to the run number which characterises an individual exposure (4 UTs). Since the exposures in a given visit are taken back-to-back, it is unlikely that there will be much change in the flux of sources between exposures. So, rather than perform forced photometry on every incoming exposure, we decided, instead, to perform forced photometry after co-adding the back-to-back exposures to increase the depth of the forced photometry. In the standard survey mode, the number of back-to-back exposure is typically three, although the stack is able to adapt to any number of back-to-back exposures.

To perform forced photometry on our nightly GOTO science frames, we used a modified version LSST's `forcedPhotCcd.py` task. First, this modified version of `forcedPhotCcd.py` finds sources within the reference catalogue that overlap with the incoming science frames by identifying the tract in which the centre of the image is located. It then performs various (user-specified) measurements at the positions of those sources, such as flux and shape measurements. Because we use the reference catalogue to perform forced photometry, every measurement in the incoming science frame is associated with a unique object ID within the reference catalogue. This makes extracting light curves for a given object straightforward, as the user simply needs to specify the object ID of the source they are interested in. We shall see more in the last section of this chapter where we describe the LSST stack-GOTO database.

To achieve our goal of performing forced photometry on the combined back-to back exposures, we have created our own driver that performs a number of tasks before performing forced photometry (`singleVisitDriver.py`). When we process data taken on a single night, `singleVisitDriver.py` groups exposures according to their visit and

CCD (i.e UT) number. Then each of the exposures that belong in each group requires instrument signature removal (ISR) and warping to a common WCS (after being astrometrically solved with the same process as described in the previous section). We chose to warp them onto the WCS system of the first exposure. After the back-to-back images are warped, we then use the `snapCombine` task to co-add [5] the back-to-back exposures for each CCD (i.e., UT). Hereafter, when we talk about nightly images/exposures we are referring to the result of the co-addition of the back-to-back exposures.

I have made some further modifications relating to how the uncertainties associated with the photometric zeropoints are calculated for each nightly exposure. In v18 of the LSST stack, which was used for the processing, the zeropoint uncertainty is calculated as $\sqrt{\Sigma(1/\sigma^2)}$, where σ is the so-called "instrumental error" associated with each measured source arising from photon noise. While this would hold true if the *only* source of error was counting statistics, in GOTO's case there are other sources of errors (e.g., varying conditions across the CCD) that would not get captured by this method. To obtain a more appropriate estimate of the uncertainty in the zeropoint (Δzp), we instead use the standard deviation of the difference between the instrumental magnitudes (m_i) and the calibrated magnitudes (m_c) of the stars used to obtain the zeropoint, i.e.,

$$\Delta_{\mathrm{zp}} = \mathrm{std\_dev}\left(m_i - m_c\right), \tag{2.2}$$

which is essentially the standard deviation of the zeropoint. Within the LSST stack, the error on the zeropoint is then added in quadrature to the instrumental error for each measured source.

We note finally that, to speed up processing, we perform only forced photometry on the nightly frames (i.e., we do not perform blind detection on the coadded visit exposures). Due to limited resources, we were also unable to write and archive all

---

[5] We choose to coadd, rather than take a median of, the three input exposures because taking a median can affect the PSF in non-trivial ways.

the processed images. However, we need information that is stored in the metadata of the image. As a workaround, we instead separately saved only the metadata (i.e., without the associated image) for each image. In order to achieve that, we modified the `singleVisitDriver.py` but also the policy file of the `obs_goto`. In addition, we have saved the PSF model of the image, but also the calib (i.e., the file that consists the calibration information for each image), which will be further described in the last section of this chapter. For each exposure we processed, we stored the catalogue as `visitCoadd_forced_src`. containing forced aperture (radii: 5.58", 7.44", 11.16", 14.88", 29.76", 59.52") and PSF photometry at the positions of each reference source contained within the bounds of each nightly exposure.

2.5 Accessing data through `butler`

To access the data produced by the LSST stack, it is necessary to work with the LSST stack's `butler`. For example, we have used the butler to access the the forced photometry catalogues using:

```
srcs = butler.get("visitCoadd_forced_src", dataId,immediate=True),
```

which includes the aperture and PSF fluxes of the sources. We are using the

```
calib = butler.get("visitCoadd_forced_calib", dataId,immediate=True)
```

to retrieve the the file that includes photometry calibration information, so we can transform the fluxes and their uncertainties into calibrated magnitudes with their respective uncertainties. Finally, calling the metadata of the visit exposure:

```
meta = butler.get("visitCoadd_forced_meta", dataId,immediate=True)
```

and the psf model of the image

```
psf = butler.get("visitCoadd_forced_psf", dataId,immediate=True)
```

we are able to get observational and processing information. Currently we retrieve the following information:

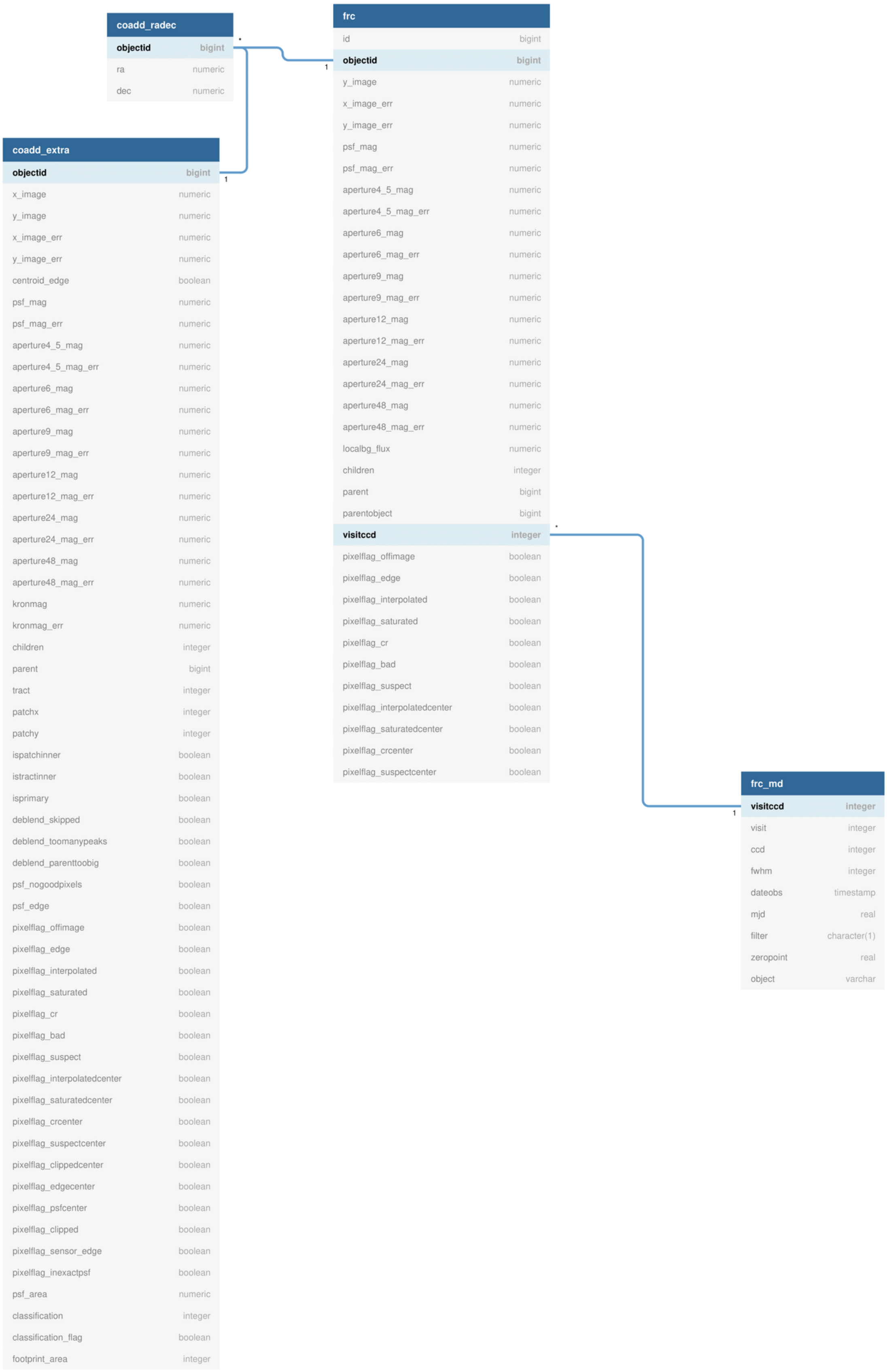
- the observation date (in YYYY-MM-DD format);
- the Modified Julian Date (MJD);
- the FWHM of the PSF at the centre of the image;
- the number of the visit;
- the number of the UT;
- the tile/target name of the pointing;
- the photometric zeropoint.

Accessing the data through the butler allowed us to generate catalogues in a convenient way to populate the LSST stack-GOTO database, by transforming both the reference catalogues and the forced photometry catalogues generated by the stack. In the next section, we describe the structure the LSST stack-GOTO database.

2.6 The LSST stack-GOTO database

For storage and data analysis purposes, we have developed a database to store and organise the data generated by `multiBandDriver.py` and `singleVisitDriver.py`. The schema of the database is shown in Figure 2.11 and consists of four tables. The two tables `coadd_radec` and `coadd_extra` are created to contain data related to the reference catalogue. The `coadd_radec` contains source ID numbers (`objectid`) together with their RAs and Decs. The `coadd_extra` includes information associated with these sources such as position on CCD and and flux measurements (aperture and PSF). We split the reference catalogue data in two tables, in order to achieve better performance for database queries, especially in the case where the user needs only the Id values of the reference sources to retrieve their lightcurves. A large number of columns only contain boolean values (i.e., True or False), corresponding to the various flags used by the LSST stack. We use the flags to identify sources with saturated pixels or other pixel failures (e.g., interpolated pixels affected by hot pixels). We also use the `isPrimary` flag, which is generated by the LSST stack, to avoid duplicates. This include cases from tract/patch overlap or deblending as both the parent and all the children (i.e., in this case `primary` source is the children source) of the source are kept in the catalogues. This means that it is possible in a positional matching to retrieve both the parent and a child or retrieve just the parent source.

The two other tables i.e. `frc` and `frc_md` host measurements and the metadata of the forced photometry processing. The `frc` table includes the forced photometry measurements (aperture and PSF fluxes) but also the flags associated with these measurements (i.e., bad pixel flags, see Figure 2.11 or flags associated with deblender e.g. if the footprint was considered too large and was not deblended). The sources in the `frc` are linked to the sources in the reference catalogue via the `objectid` number. This means that each source in `frc` has two identifiers: the first one (`id`) is the identification number for the source after forced photometry processing, and the second one is the `objectId`, which is the identification number of the source in the reference catalogue with which the source is associated. The `frc_md` includes all the metadata referred to the previous section (section 2.5). The `frc` and `frc_md` are linked via the `visitccd` number. This number is created by the combination of the visit and CCD numbers (format: `visitccd`) of a frame and is unique for each frame, so each source in the `frc` table is also associated with the `visitccd` number of the frame in which was detected. Using this `visitccd` number, we are then able to find information such as the MJD of the detection or quality information of the frame (e.g., FWHM of the PSF at the centre of the frame) to generate good quality lightcurves to use during our analysis.



coadd_radec	
objectid	**bigint**
ra	numeric
dec	numeric

frc	
id	bigint
objectid	**bigint**
y_image	numeric
x_image_err	numeric
y_image_err	numeric
psf_mag	numeric
psf_mag_err	numeric
aperture4_5_mag	numeric
aperture4_5_mag_err	numeric
aperture6_mag	numeric
aperture6_mag_err	numeric
aperture9_mag	numeric
aperture9_mag_err	numeric
aperture12_mag	numeric
aperture12_mag_err	numeric
aperture24_mag	numeric
aperture24_mag_err	numeric
aperture48_mag	numeric
aperture48_mag_err	numeric
localbg_flux	numeric
children	integer
parent	bigint
parentobject	bigint
visitccd	**integer**
pixelflag_offimage	boolean
pixelflag_edge	boolean
pixelflag_interpolated	boolean
pixelflag_saturated	boolean
pixelflag_cr	boolean
pixelflag_bad	boolean
pixelflag_suspect	boolean
pixelflag_interpolatedcenter	boolean
pixelflag_saturatedcenter	boolean
pixelflag_crcenter	boolean
pixelflag_suspectcenter	boolean

coadd_extra	
objectid	**bigint**
x_image	numeric
y_image	numeric
x_image_err	numeric
y_image_err	numeric
centroid_edge	boolean
psf_mag	numeric
psf_mag_err	numeric
aperture4_5_mag	numeric
aperture4_5_mag_err	numeric
aperture6_mag	numeric
aperture6_mag_err	numeric
aperture9_mag	numeric
aperture9_mag_err	numeric
aperture12_mag	numeric
aperture12_mag_err	numeric
aperture24_mag	numeric
aperture24_mag_err	numeric
aperture48_mag	numeric
aperture48_mag_err	numeric
kronmag	numeric
kronmag_err	numeric
children	integer
parent	bigint
tract	integer
patchx	integer
patchy	integer
ispatchinner	boolean
istractinner	boolean
isprimary	boolean
deblend_skipped	boolean
deblend_toomanypeaks	boolean
deblend_parenttoobig	boolean
psf_nogoodpixels	boolean
psf_edge	boolean
pixelflag_offimage	boolean
pixelflag_edge	boolean
pixelflag_interpolated	boolean
pixelflag_saturated	boolean
pixelflag_cr	boolean
pixelflag_bad	boolean
pixelflag_suspect	boolean
pixelflag_interpolatedcenter	boolean
pixelflag_saturatedcenter	boolean
pixelflag_crcenter	boolean
pixelflag_suspectcenter	boolean
pixelflag_clippedcenter	boolean
pixelflag_edgecenter	boolean
pixelflag_psfcenter	boolean
pixelflag_clipped	boolean
pixelflag_sensor_edge	boolean
pixelflag_inexactpsf	boolean
psf_area	numeric
classification	integer
classification_flag	boolean
footprint_area	integer

frc_md	
visitccd	**integer**
visit	integer
ccd	integer
fwhm	integer
dateobs	timestamp
mjd	real
filter	character(1)
zeropoint	real
object	varchar

Figure 2.11: The LSST stack-GOTO database schema

2.7 Chapter Conclusions

In this chapter, we described how successfully the LSST stack was adapted for the GOTO telescope by developing the `obs_goto` package from scratch. We have also presented the aforementioned package and then we explained the way we performed forced photometry on GOTO nightly images using the LSST stack. We have also reported the different modifications that were needed in order to process GOTO data. After we adapted the LSST stack for GOTO and processed the data, it was necessary to confirm the quality of the products from the two tasks described in this chapter i.e. the reference catalogue productions and the forced photometry measurements. In order to characterise the results from these tasks, we have performed a number of different tests, which are presented in the next chapter.

Chapter 3

Quality assessment of the data

3.1 Chapter Introduction

The data products of the processing stage include the reference catalogue for the forced photometry and the nightly catalogues created by forced photometry. It is important to characterise the quality of the data before they are used for scientific analysis. The reference catalogue, which will be used for the forced photometry, is tested in terms of two important aspects: deblending i.e. the ability of our pipeline to separate astronomical sources on GOTO images and the astrometric quality of the catalogue. Further, as an additional test, we assess the photometric quality of the reference catalogue, which helps us to confirm that the sources are identified correctly in the reference catalogue.

The nightly results from forced photometry were assessed on their photometry both using internal characterisation i.e. the RMS of the nightly flux measurements and using external catalogues. To further characterise the performance of the pipeline for the forced photometry task, we have also assessed the quality of the photometric uncertainties. The photometric uncertainty measurement are of high importance, as they are used in various definitions of variability detection using lightcurve analysis.

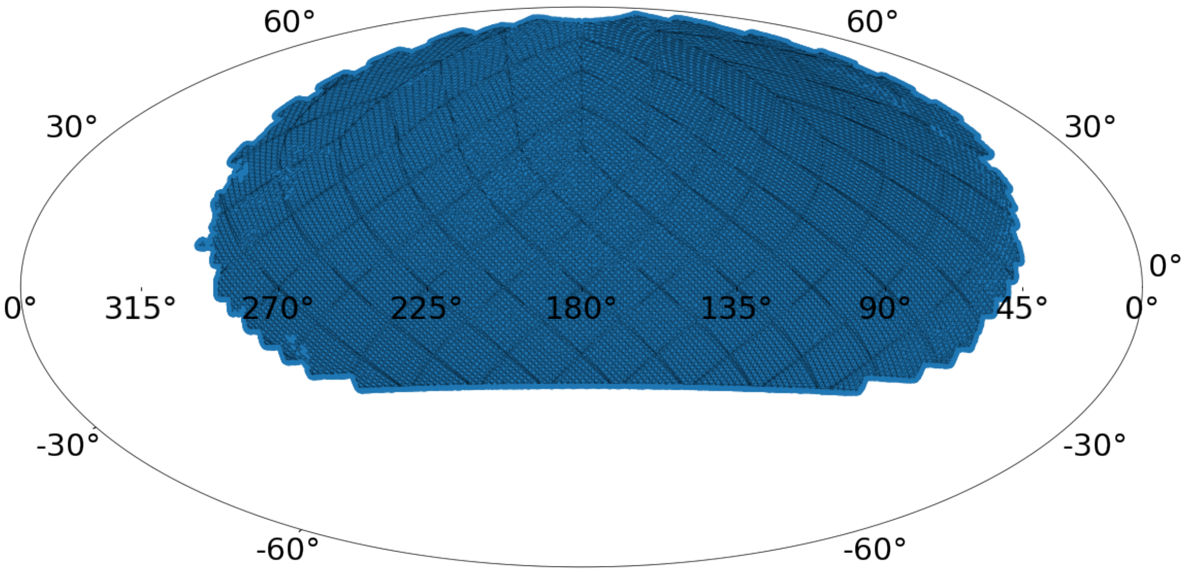


Figure 3.1: The map shows the sky coverage of the GOTO data, which were processed through the LSST stack.

3.2 Data

The reference catalogue was produced from source detection and measurements performed on coadd images constructed from individual exposures taken between the 24 February to 12 March, 2019 (see Mullaney et al. 2021). The data cover ∼ 50% of the GOTO observable sky(see Figure 3.1). The reference catalogue for this part of sky consists of ∼160 million sources.

To characterise the results from forced photometry, we have used data from the nightly processing of raw images observed between the 24 February and the 25th October, 2019. For both the reference catalogue and the forced photometry, the raw images are taken with the broad L-band and for each pointing there are (typically three) back-to-back 60 s exposures. Nightly forced photometry was performed on the mean-combined back-to-back exposures as described in section 2.4.2.

3.3 Reference catalogue

The reference catalogue that was produced from GOTO data from 24 February to 12 March, 2019 includes the reference sources on which we perform forced photometry. As we said, these data cover the ∼50% of the GOTO observable sky and the depth will vary according the number of images coadded for each part of sky. Since we use this limited time period for our reference images, it is noted that this is not the maximum depth GOTO is able to achieve by co-adding multiple images spanning multiple nights and choosing those with the best seeing conditions. Our reference catalogue, however, is deep enough for our aim i.e. performance of forced photometry on nightly GOTO data to investigate the variability of moderate luminosity AGN (see Chapter 4). It will be shown in this chapter that the depth achieved from the coadd reference images is slightly better than the depth from the nightly data. We have tested the reference catalogue in terms of astrometry, which is extremely important for the forced photometry task which relies on the positions in the reference catalogue. We have also tested the photometry of the reference catalogue in order to confirm that the source detection is accurate. It is also important to characterise the photometry before determining the survey depth. Finally, we estimated the survey depth, using the reference catalogue to ensure that sources detected on nightly images are included in the reference catalogue. We note that transient events located in hosts fainter than the depth limit of the reference catalogue will not have forced photometry measurements.

3.3.1 Astrometry assessment

We have estimated the accuracy of the astrometry by comparing the reference catalogue positions with the positions in the PanSTARRS DR1 catalogue after cross-matching the two catalogues. PS DR1 catalogue has a standard deviation of the mean residuals 2.3 mas for RA and 1.7 mas for Dec compared to Gaia DR1 (Chambers et al. 2016), which means that its quality is sufficient for comparison with the GOTO reference catalogue produced by the LSST stack. It is also convenient to use the PanSTARRS catalogue because as mentioned in the previous chapter, its coverage is similar to the GOTO-N observable sky.

For primary sources which are not flagged as having bad pixels in their footprint, we find that the mean angular distance between the sources in the reference GOTO catalogue and the matched sources in the PanSTARRS DR1 catalogue is 0.31", the median distance 0.25" and the standard deviation is 0.23". The standard error on the mean of the angular distance is 0.00016. We find that there is a systematic mean difference in declination of 0.069" (median of 0.073") and its standard deviation is 0.264". For the right ascension separations, the mean is 0.045", the median is 0.047" and the standard deviation 0.261". The standard error on both the mean of the right ascension and the declination differences is 0.00019. These results show that the typical mean angular distance between the GOTO detections and the PS1 matches is less than half GOTO's pixel scale (i.e., $1.25''$).

We visualised our astrometry results with a map that shows the angular distance between the GOTO and PanSTARRS positions, but also their difference for right ascension and declination, to show their variations across the sky(see Figure 3.2). We understand from the plot that the astrometry is slightly worse in the region close to $RA = 180°$ and $Dec = 40°$ and for the region around $RA = 290°$ and $Dec = 30°$. We have investigated these tracts and found that this is due to the quality of some (i.e. affected by mount instabilities) of the warp images that are used to create the deep coadds of these tracts. The mean angular distance, however, even for these tracts is less than half GOTO pixel.

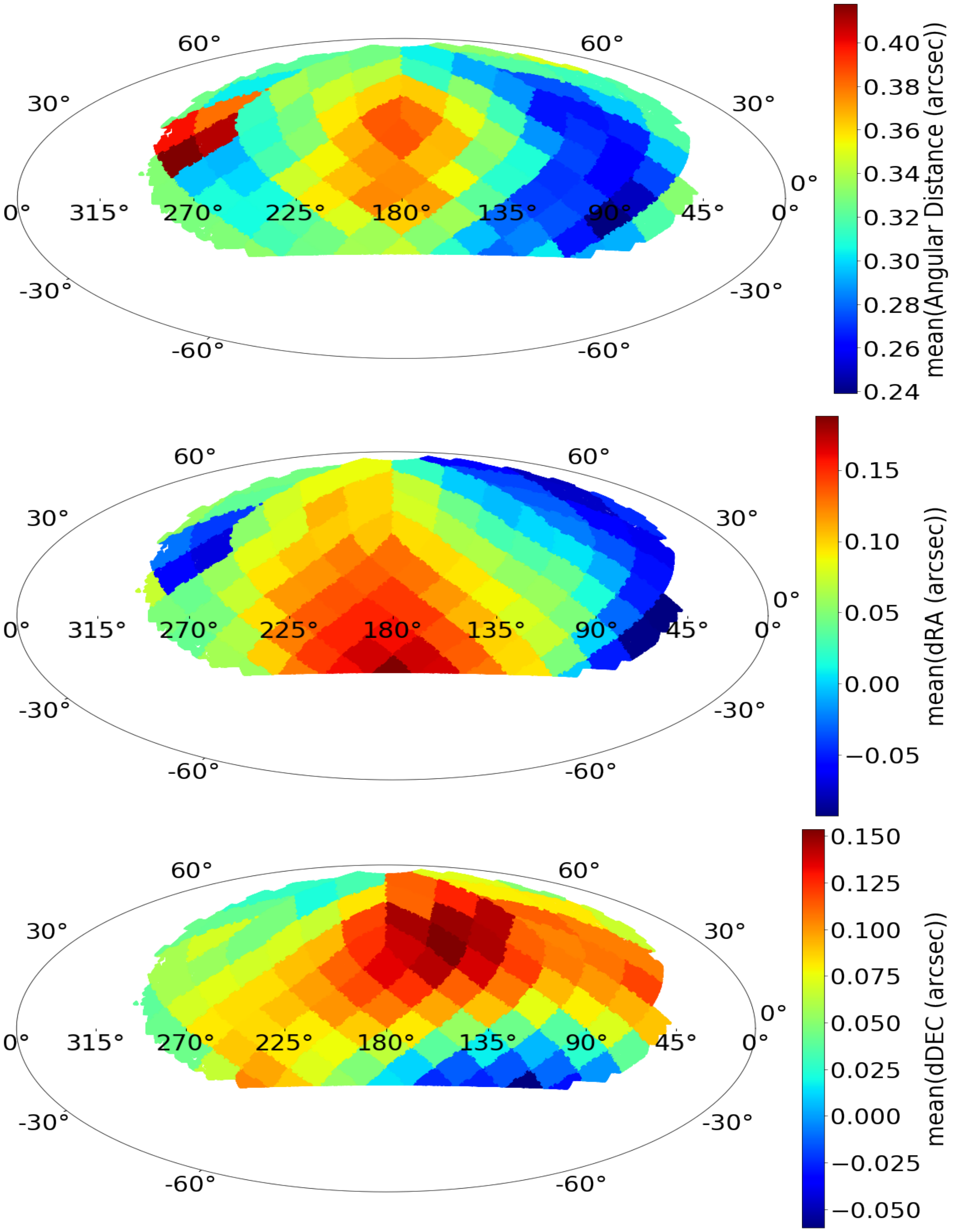


Figure 3.2: Astrometric quality against PanStarrs PS1 catalog. The plots are showing the angular distance (top), RA difference (middle) and DEC difference (bottom).

3.3.2 Photometry Assessment

We have also tested the photometric quality of the reference catalogue as an additional way to ensure the accuracy of source detection and as preparation for our estimation of the survey depth. We compare the aperture (11.16" radius [1]) photometry of the reference catalogue against the PanSTARRS g-magnitudes, which are measured from PSF photometry (we choose to use these magnitudes, as the PanSTARRS collaboration suggests that these are the most accurate measurement especially for point sources). For fair comparison, we have colour-corrected the PanSTARRS g-magnitudes with the coefficients reported in the previous chapter. We find, in Figure 3.3, that for the patches in the region close to $RA = 180°$ and $Dec = 40°$ there are larger deviations from the colour corrected g-band PanSTARRS magnitudes. This is due to the quality of a few of the images that produced the reference coadd images for these tracts. In this case the PSF and the photometry are being affected by weather and technical issue (i.e. clouds affected some of the images that went into the coadds, and more importantly some of them are affected by mount instabilities). In future implementations of the stack these issues can be avoided by selecting images to be coadded based on their quality (e.g., PSF selection is already available in the LSST stack). Excluding those tracts with affected by poor PSF, we find that the standard offset between the GOTO LSST stack photometry and the PS photometry is within 0.2. magnitude or less with standard deviation of 0.2-0.4 mag.

3.3.3 Survey depth

With astrometry and photometry evaluated, it was important to determine the survey depth of the reference catalogue. We find the magnitude limit for each tract of sky using

---

[1] The radius was chosen because this allows the aperture to be 3.4-4.5 times the FWHM of the seeing at the centre of the images. Considering the larger seeing on the outskirts of the images this large radius is preferred with the current optics. Smaller apertures present larger scatters and offsets. This is subject to future investigation.

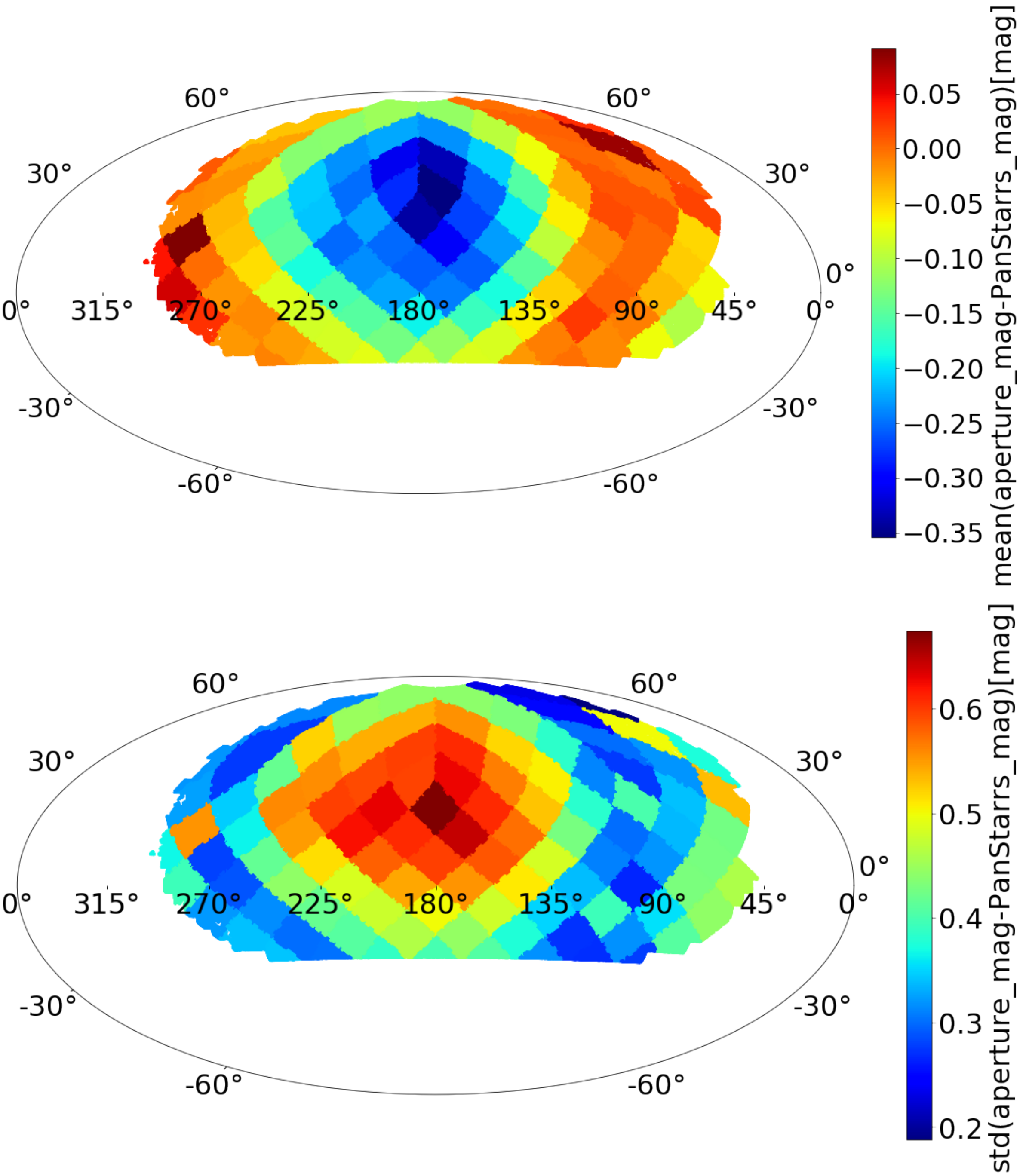


Figure 3.3: Mean (top) and standard deviation (bottom) for the difference between the magnitudes from aperture photometry in the reference catalog extracted by the LSST software stack and the colour corrected PanStarrs magnitudes. Both of these plots are presented in Mullaney et al. (2021).

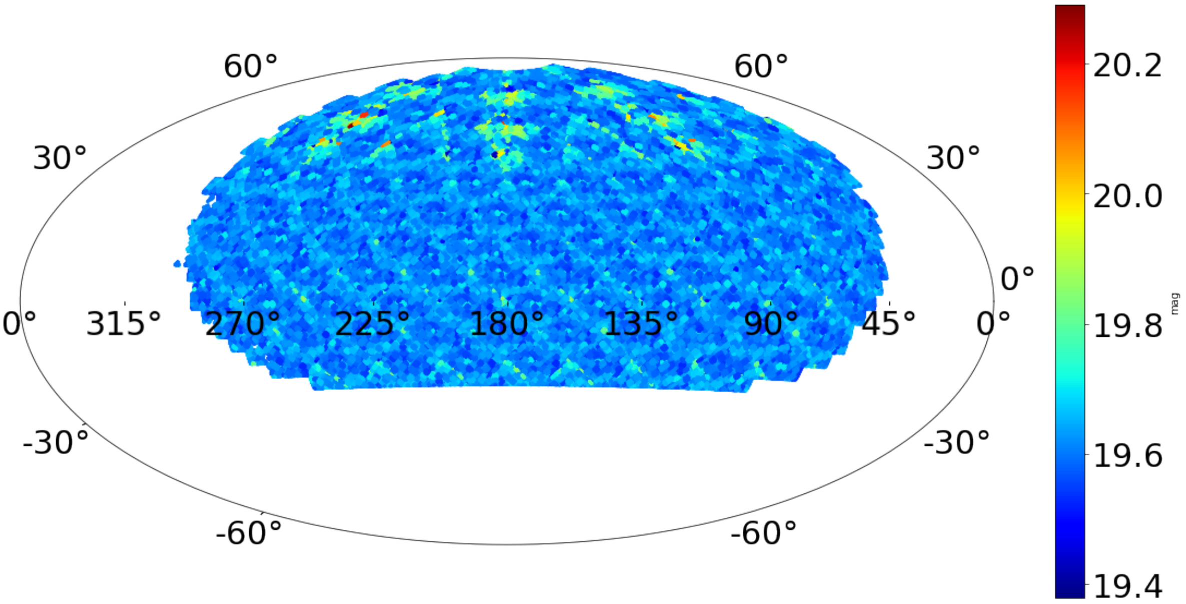


Figure 3.4: The survey depth for each patch in the reference catalogue. The data points are colour coded for the median aperture magnitude of sources with SN between 4.5 and 5.5.

the median measurement of sources with SN between 4.5 and 5.5 (i.e. to estimate for the 5σ survey depth) using aperture photometry. This way we make an estimate of the 5σ depth of the patch, assuming a flat distribution for the sources in this range (Aihara et al. 2018). We find (see Figure3.4) that typically the survey depth for the coadded images is $\sim$ 19.7 in the L-band (AB mags). We note that this is not the maximum achievable survey depth as we have coadded only limited data due to time and storage constraints. This, however, is enough for the studies we perform in chapters 4 and 6.

We have also estimated the detection completeness for our reference coadd images by inserting fake sources in random coadd images to investigate how many of them we retrieve, , and we present the results for one of them (tract:94, patch:5,8). We inserted sources with magnitudes in the range between 13 and 23 and we found that the survey is 100% complete up to $m_{\mathrm{L}} = 19.5$ (comparable to the survey depth determined earlier), 50% at $m_{\mathrm{L}} \sim 21$ mag, while we do not detect any sources fainter than $m_{\mathrm{L}} \sim 22.5$ (Figure3.5) .

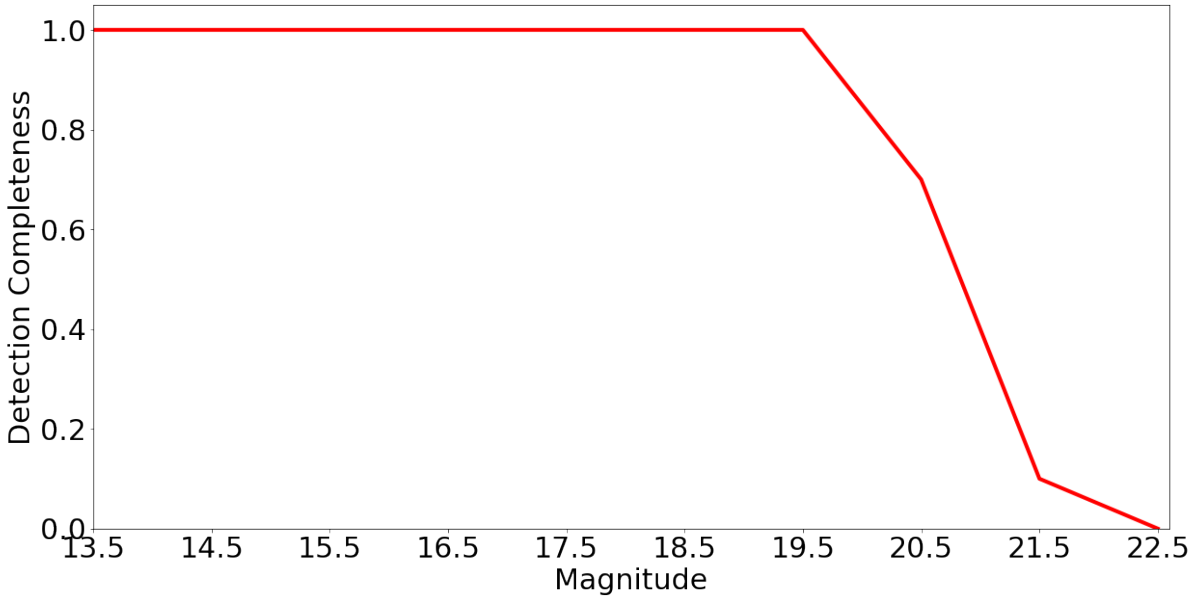


Figure 3.5: Detection completeness as a function of magnitude in tract 94, patch 5,8 using artificial sources.

3.3.4 Deblending

As part of the characterisation of the quality of the reference catalogue, we have also tested the performance of the deblending task. To do that, we have again used fake sources i.e. pairs of inserted sources with different fluxes and with different separations. The aim was to investigate the minimum distance for which blended sources will be successfully deblended. Typically, we found that a separation of 6 arcseconds is the closest that two sources can be to achieve a success rate approaching 100%, although there is some dependence on relative brightness of the blended sources.

3.4 Forced Photometry

In this section, we assess the quality of the forced photometry using both internal assessments (i.e., the rms of the nightly flux measurements) and also by comparison with external catalogues i.e. the PanSTARRS DR1 catalogue and the catalogues from GOTOphoto pipeline. We have also assessed the quality of the photometric uncertainties

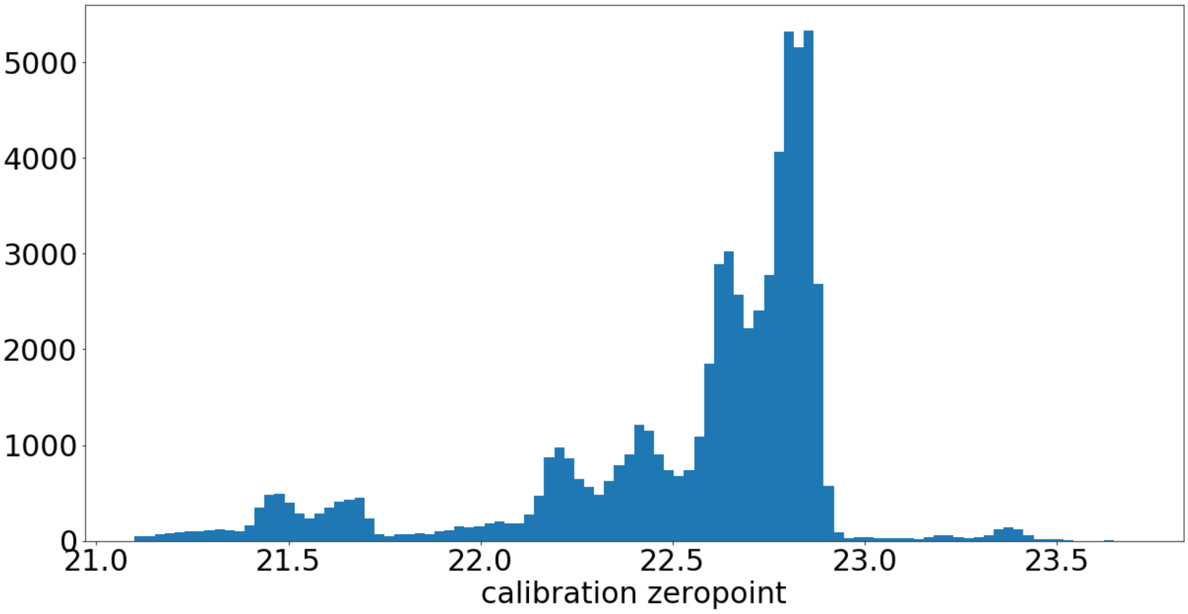


Figure 3.6: The photometric calibration zeropoint for the frames processed through forced photometry, which for most of them peaks around ∼22.8 with a spread and few significant peaks in lower values resulting the mean value 22.52 mag.

using the GOTO lightcurves generated by the LSST stack and a catalogue of standard SDSS stars.

3.4.1 Internal photometric assessment

To evaluate the quality of the photometry internally, we tested the photometric repeatability, which estimates the level of consistency between multiple photometric measurements of same source across multiple nights. We measure how the repeatability changes as a function of the magnitude. The repeatability test was performed for both types of photometry measurements obtained using our implementation of the LSST stack, i.e., 11.16" aperture photometry and PSF photometry. For this test, we have chosen a part of the sky which has been visited the most number of times by GOTO between 2019-03-01 and 2019-07-31 (inclusive) and then calculated the RMS of the magnitude of all sources in one pointing (from all four UTs). To clean the photometric data of spurious measurements, we used the pixel flags provided by the LSST stack to remove data flagged

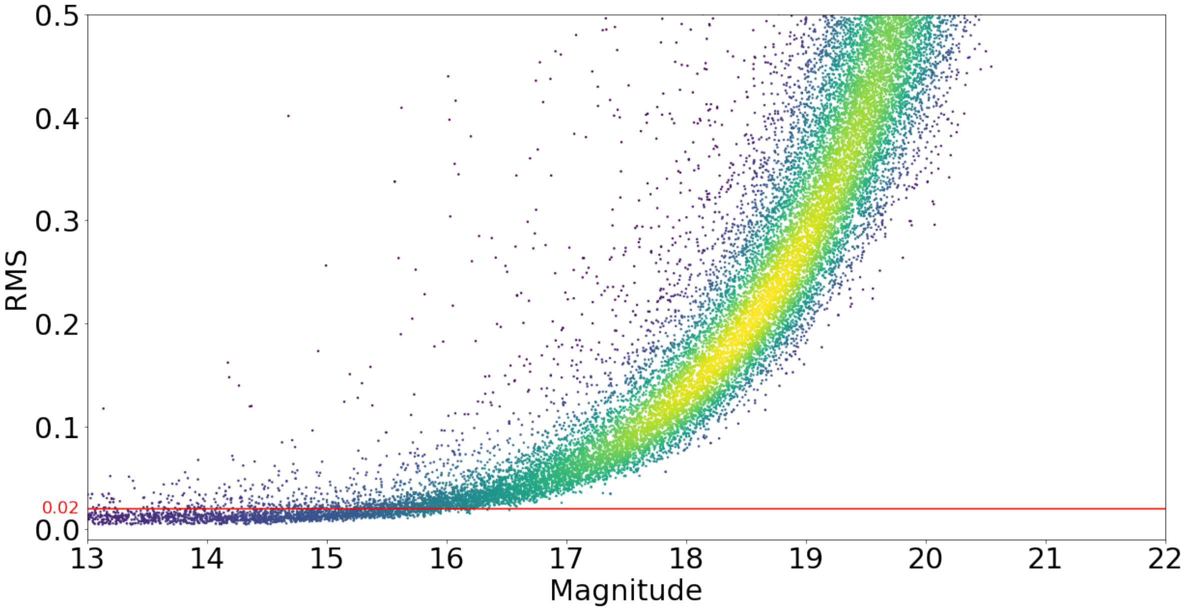


Figure 3.7: Photometric repeatability for aperture photometry from forced photometry catalogues for a field covered by the four CCDs. We see that sources brighter than 15.5 mag are precise within 0.02 RMS, which is indicated by the red line on the plot.

for pixel issues. Further, we removed measurements from any frame whose zeropoint deviates by more than 3σ (typical standard deviation: $\sigma = 0.5\,\mathrm{mag}$) from 22.52 (i.e., the mean zero-point reported in Figure 3.6). We note, however, that we did not filter for known variable sources. Although these sources will increase the measured inter-night RMS, they represent a small proportion of sources that we do not expect them to have any measurable effect on our repeatability measurements.

Figure 3.7 shows the plot of the inter-night RMS of aperture photometry measurements from the aforementioned observations. From this plot we see that, for sources brighter than 15 mag, the internal photometric precision, as measured by photometric repeatability, is typically below $\sim$0.02 mag. The RMS increases with decreasing brightness due to the increase in size of the photometric uncertainties.

We compare the repeatability of PSF and aperture photometry in Figure 3.8. This plot shows that PSF photometry is less consistent between nights than aperture photometry for sources brighter than around $m_{\mathrm{L}} = 17$. This is to be expected due to the various

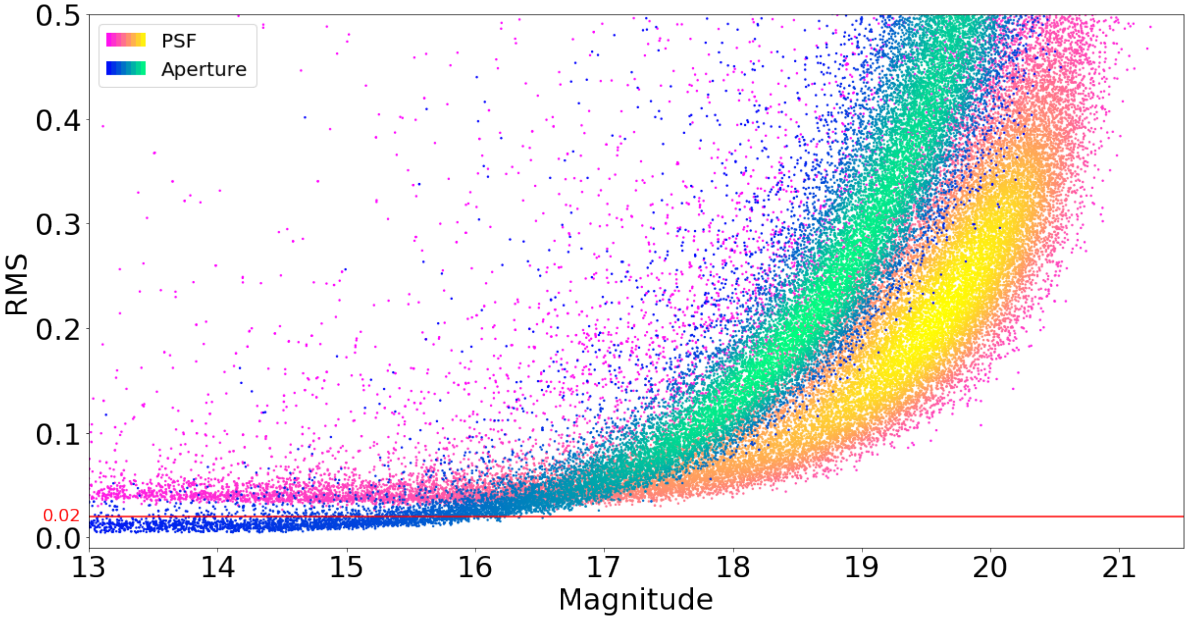


Figure 3.8: Photometric repeatability comparison for aperture photometry and PSF fit photometry.

difficulties associated with performing PSF photometry on bright sources arising from, e.g., bright spikes or saturated pixels. PSF photometry, however, is found to perform better than aperture photometry for fainter sources.

3.4.2 Photometric assessment with external catalogues

To evaluate the forced photometry measurements produced by the LSST stack using external catalogues, we first estimated their quality by comparing against the magnitudes reported in the PanSTARRS DR1 catalogue (Magnier et al. 2016). Once again the choice of PanSTARRS is motivated by the fact that it has a very similar sky coverage to the GOTO survey. Also, it is significantly deeper than the GOTO survey which means that almost all sources detected by GOTO *should* have a PanSTARRS counterpart. This means that we can obtain comparison statistics down to the detection limit of GOTO (i.e., we are not limited by the depth of PanSTARRS). In Figure 3.9 we present plots showing the magnitude difference between the GOTO magnitudes measured by the

LSST stack and the colour corrected PanSTARRS g-band magnitude versus the GOTO magnitude. We present results from both aperture (11.16” ; top) and PSF (bottom) photometry. This plot includes sources from all four CCDs for a single pointing. We, however, performed the test for various pointings and obtain similar results for them (caveat those pointings filtered-out via the method outlined in section 3.4.1). This comparison with PanSTARRS suggested that the PSF photometry is more precise for this particular epochal pointing, although as we saw in the previous section from the repeatability test, aperture photometry results are in general more accurate, especially for sources brighter than ∼17 magnitude. Both PSF and aperture photometry suggest that, for sources fainter than $18^{\rm th}$ down to the detection limit (i.e., ∼19.5), GOTO photometry as measured by the LSST stack is within 0.2 magnitude RMS of the PanSTARRS photometry. Between 16th and 18th magnitude, this is ∼0.03-0.06 mag, and for sources brighter the 16th magnitudes, the RMS is ∼0.01-0.02 mag.

Finally, for external comparison, we have compared outputs from the LSST stack against those obtained with GOTOphoto, the in-house photometric pipeline developed by the GOTO collaboration. It is important to note that we are comparing forced photometry measurements (from the LSST stack) against measurements of sources obtained via blind detection (from GOTOphoto). This should not be a concern for isolated sources, which form the vast majority, however, it could mean that some sources are deblended in the blind catalogue, but not in the forced photometry catalogue (or vice versa). GOTOphoto measurements are based on SExtractor’s `MAG_AUTO` (Bertin & Arnouts 1996) aperture photometry performed on the same (but median-combined, rather than coadded) back-to-back exposures as those we performed forced photometry on. In Figure 3.10, we present plots showing the magnitude difference between the GOTO magnitudes measured by the LSST stack and the GOTOphoto magnitudes versus the GOTO magnitude. The plots show that the vast majority are within and RMS of 0.1 mag, even in the case of the faintest sources. There is, however, a systematic offset

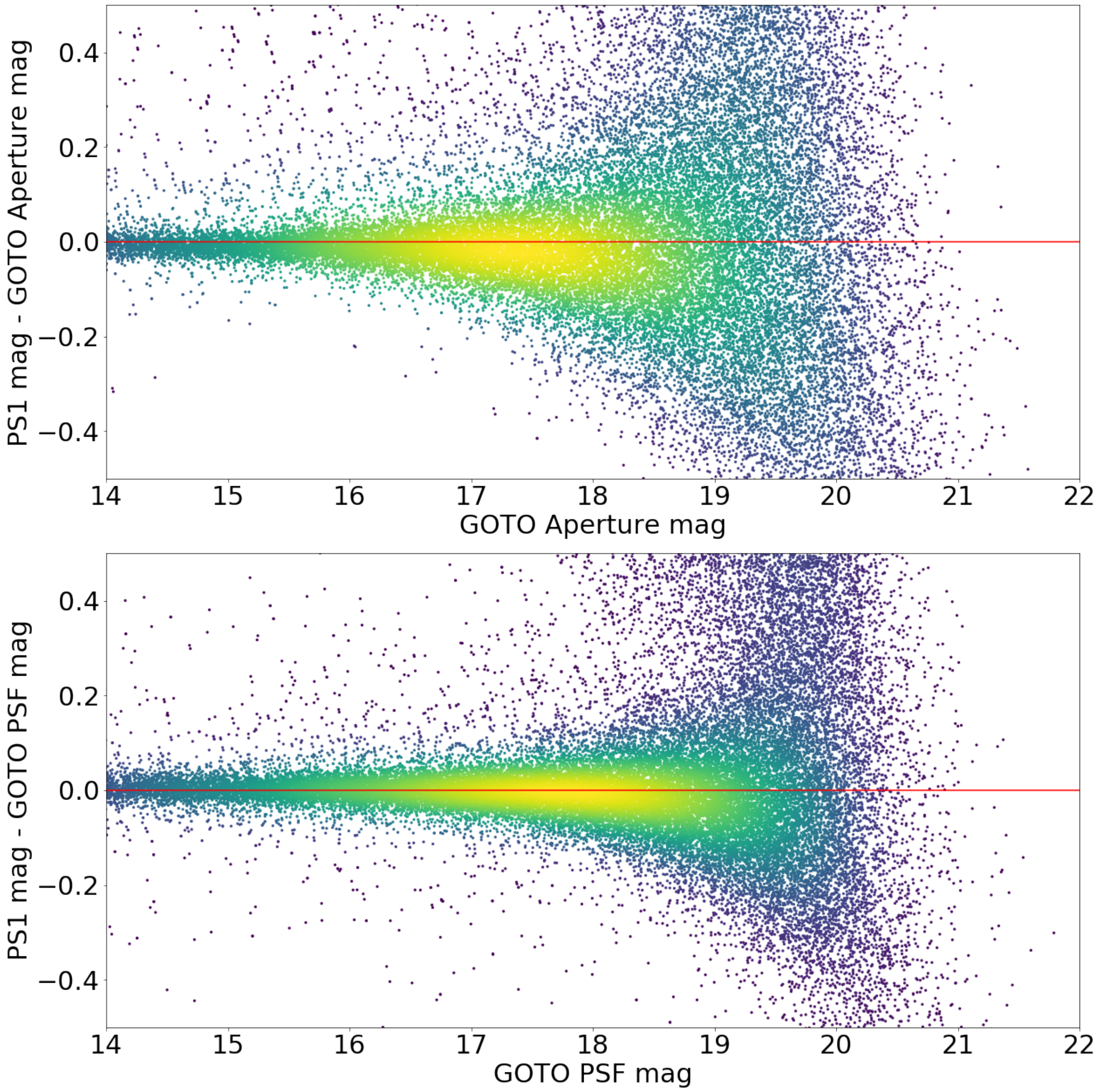


Figure 3.9: The difference between GOTO aperture (top)/ PSF(bottom) photometry from LSST software stack for a full FOV field and PanSTARRS magnitudes as a function of the magnitude. The PanSTARRS magnitudes are from PSF photometry.

between the results of the two pipelines which is likely due to the application of colour terms when we process the data with the LSST stack. The PSF photometry appears once again to result in a smaller scatter than aperture photometry for this particular pointing, although we note that other pointings produce similar overall results. In the previous sections, however, we showed the accuracy of PSF photometry vs. aperture photometry through repeatability measurements, which show that the quality of the aperture photometry is superior to that from PSF photometry especially for sources brighter than 17th magnitude.

3.4.3 Observational photometric uncertainties

In order to work with GOTO lightcurves, as we do in the following section and also the following chapters, it was also important assess their uncertainties. In the case of variability studies, this is particularly true, since we need to know whether differences in the measured photometry over multiple nights are physical in origin (i.e., genuine), or simply due to the uncertainties in our measurements. As such, we must carefully evaluate whether we are over or under-estimating our uncertainties.[2]

There are different sources of noise for CCD astronomical observations, which affect the precision of our observations (e.g. Figure3.7). Since we speak of photon measurements there is a minimum level of noise, the so-called shot noise (Poisson noise). Apart from the noise in the detected photo-electrons from the stellar source, we also have the noise in the detected photo-electrons from the sky background, the noise in the thermally excited electrons (dark current) and the readout noise. Apart from the readout noise all the other sources are Poissonian.

As described in section 2.4.2, the uncertainties reported by the LSST stack on each photometric measurement are obtained by combining, in quadrature, instrumental uncer-

---

[2] An example of the effects of poor error estimation is presented in Suberlak et al. 2017, in which they find that the quasar variability levels observed in CRTS data (Graham et al. 2014) actually arises from underestimated errors.

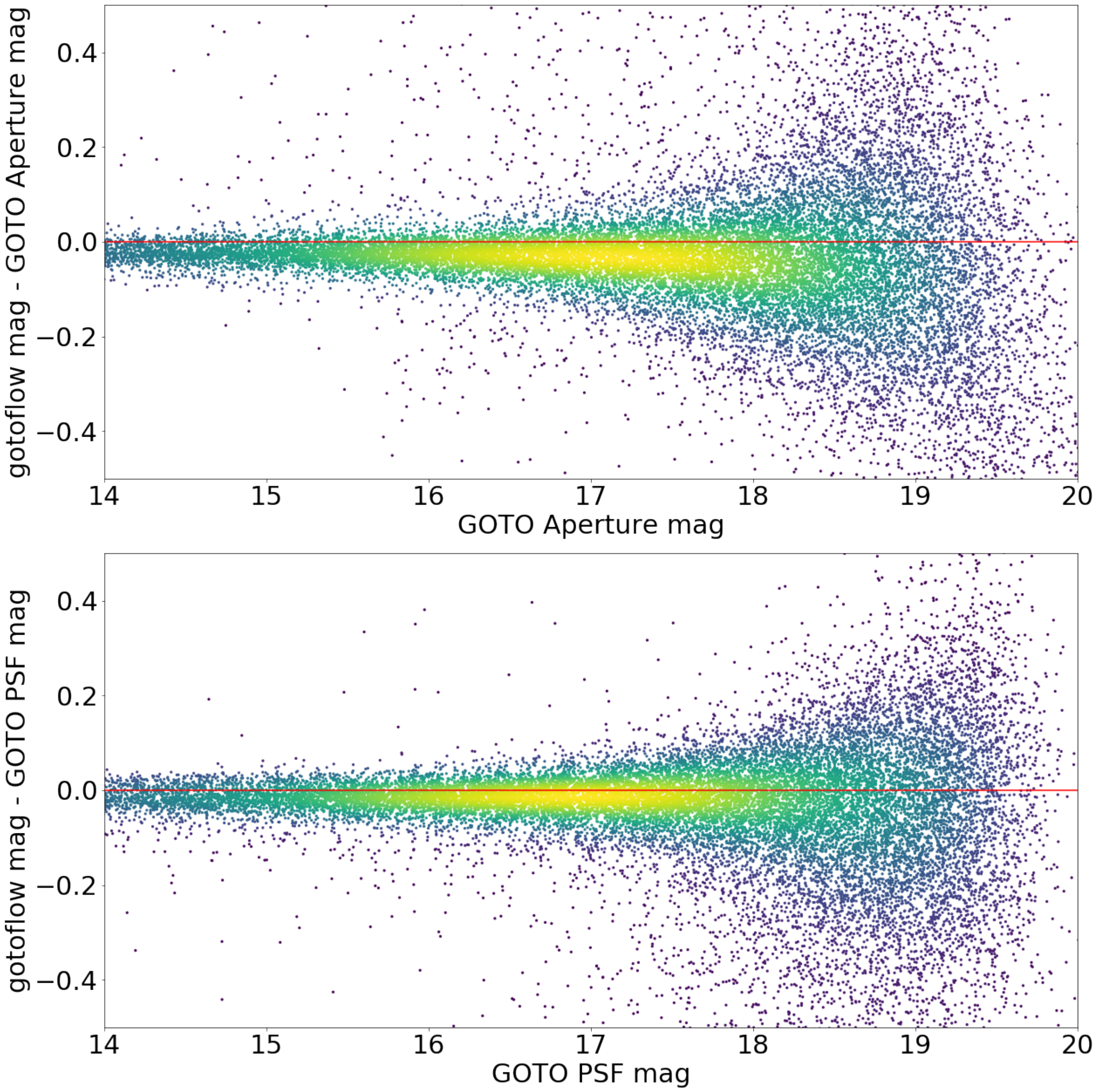


Figure 3.10: The difference between GOTO aperture (top)/ PSF(bottom) photometry from LSST softwre stack for a full FOV field and GOTOphoto magnitudes as a function of the magnitude.

tainties (i.e., those principally arising from photon noise) with calibration uncertainties (i.e., those arising from uncertainties in the zeropoint). We have assessed the quality of these uncertainties using the method outlined in Suberlak et al. (2017), which was used to assess the quality of the uncertainties reported by the Catalina Real-time Transient Survey (CRTS). In short, this involves repeatedly measuring the photometry of stars extracted from the Ivezić et al. (2007) catalogue of standard stars.[3] Next, for each standard star covered by our repeat GOTO observations we calculate z_i:

$$z_L^i = \frac{m_L^i - \bar{m}_L}{\epsilon_L^i}, \tag{3.1}$$

where m_L^i and ϵ_L^i are the measured L-band photometry and associated uncertainties from observation i, and $\bar{m}_L$ the mean of all m_L^i weighted according to inverse uncertainty. We then take the standard deviation of these z_i values for each star, using the definition of standard deviation used in Suberlak et al. (2017), which is less affected by outliers:

$$\sigma_L(z_L^i) = 0.741 \times IQR, \tag{3.2}$$

where IQR is the 25%-75% interquartile range of the z_L^i values for each standard star.

There were ∼13000 standard SDSS stars within the GOTO reference catalogue used to perform our nightly forced photometry. The r-band magnitude range of these stars is 14-20. In the top panel of figure 3.11 we plot σ_L for each of these stars. For non-variable stars, it is expected that the σ_L values would follow a distribution centred at unity and display no dependence on magnitude. However, as for CRTS, we find that this is not the case for GOTO photometry as measured by the LSST stack. Instead, we find that the photometric uncertainties are overestimated by a factor of ∼2 in the case of sources brighter than $m_L \sim 15$, and underestimated by a factor of 1.2–1.3 for sources

---

[3] To create this catalogue, Ivezić et al. (2007) used repeat measurements of ∼1 million $m_r = 14 - 22$ stars in SDSS Stripe82 to verify that they are non-variable.

fainter than $m_L \sim 17$. We also present in the middle and lower panels of Figure 3.11 the standard deviation of the magnitude difference (again calculated using Eq. 3.2, but with $m_L^i - \bar{m_L}$ in place of z_L^i) and the median error of each of our sources. From these two plots, we find that the standard SDSS stars brighter than $m_L \sim 15$ have a standard deviation in $m_L^i - \bar{m_L}$ that is less than ~ 0.015 whereas the minimum uncertainty for these magnitudes is ∼0.032. Again, this implies that the uncertainty estimate is too large by a factor of ~ 2.

In the middle plane of Figure 3.11, we also plot known noise contributions for GOTO data, as we mentioned earlier in this section. To calculate the theoretical noise, we use the equation:

$$\sigma_{th} = \sqrt{N_o + n_{pix} \times (N_{sky} + N_{dark} + R^2)}, \tag{3.3}$$

where N_o is the number of photons from the source, N_{sky} is the average number of photons per pixel from the sky background, N_{dark} is the dark current contribution in electrons per pixel, R is the readout noise contribution in electrons per pixel, and n_{pix} is the number of pixels in the photometric aperture. The exposure time is taken into account, and we get the conversion of electrons to photons using the gain values in the previous chapter. The numbers for the other contributions (e.g. theoretical sky contribution, theoretical dark contribution) are those reported in Dyer 2020 (chapters 2.2-2.4). After converting the error in magnitude terms and plot the results in Figure 3.11, we find that the theoretical noise lies below the vast majority of the observed RMS. This leads to a first indication that systematics (e.g. from haedware) affect the observations (see section 4.6).

Since the aforementioned results are based on standard deviation measurements, it is important to ensure that they are not affected by artificial factors such as the number of observations (e.g., the standard deviation will only start to approximate to the size of the uncertainty after a large number of measurements). To test for this, we explore

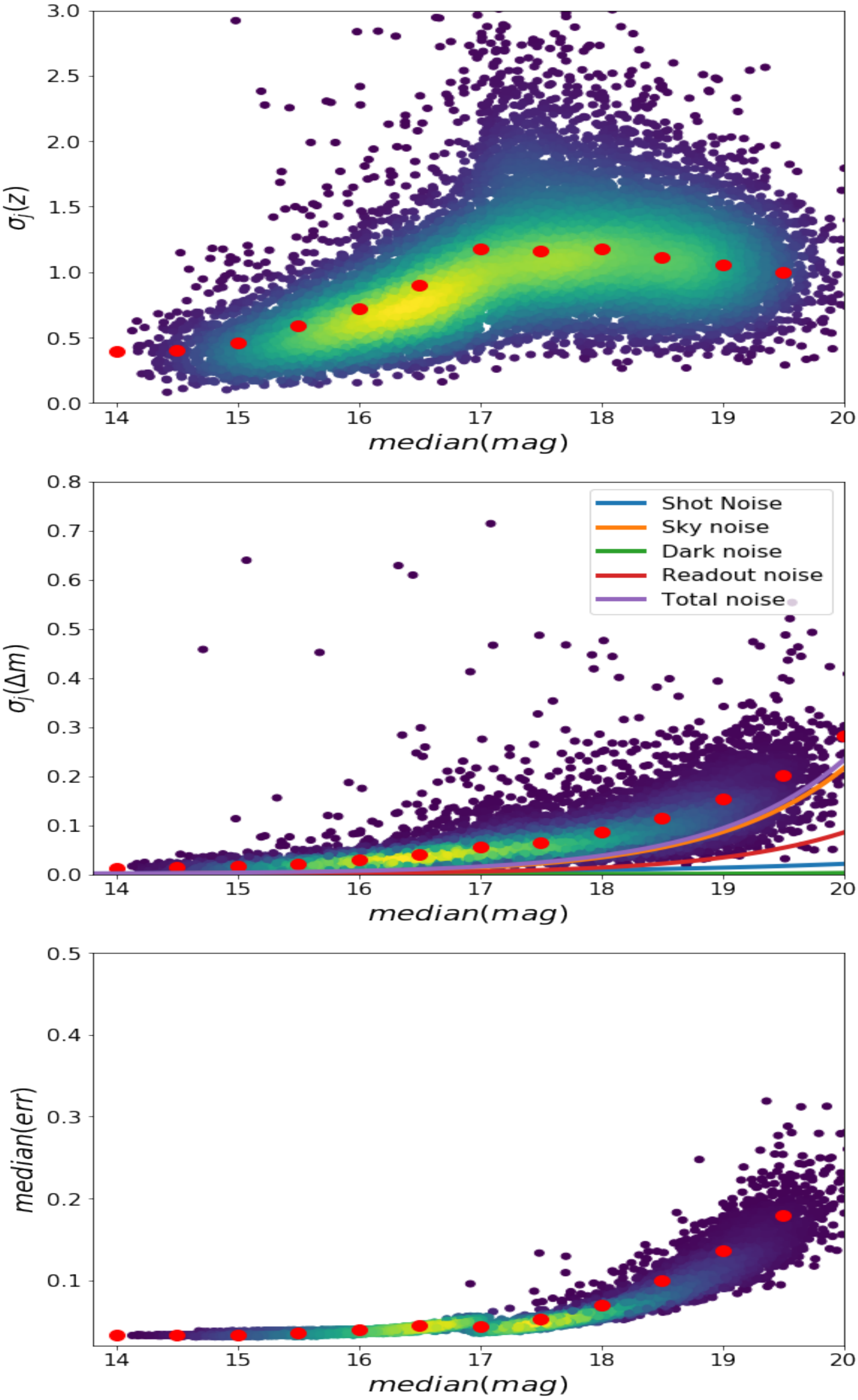


Figure 3.11: The top panel shows the the standard deviation of the quantity of eq.(3), as a function of the median magnitude for the GOTO light curves from forced photometry for the SDSS standard stars. The middle shows the intrinsic scatter of each light curve measured as the robust standard deviation of the magnitude, as a function of the magnitude and also the theoretical noise as described in the text. The bottom panel shows the median error of each light curve as a function of the magnitude. The method evaluates the photometric uncertainties according to the Suberlak et al. (2017). In each plot, the red dots show the median values per half-magnitude bin.

whether the trend seen in the top plot of Figure 3.11 changes as a function of the number of epochs, however, we do not find strong evidence that this is the case, suggesting that the observed results are not spuriously affected by the number of epochs.

To address the uncertainty overestimate/underestimate issue we have introduced correction factors to the error bars, such as those presented in Suberlak et al. (2017) for the CRTS data, prior to using the photometric uncertainties when studying source variability. Such magnitude-dependent correction factors are also included in the following section where we present some examples of lightcurves measured using forced photometry on GOTO data by the LSST stack.

3.4.4 Lightcurves analysis

Using forced photometry on the GOTO data, we have extracted, as discussed in the previous chapter, both PSF and aperture photometry. To evaluate which of these two flux measurements delivers more accurate results, we have performed reduced-χ^2 analysis on non-variable sources. The "model" in this case is constant magnitude, so the the reduced chi-squared should be 1, if the object is non-variable and the errors are OK. We have used stars included in the Ivezić et al. (2007) catalogue which was described in the previous section. For stars brighter than ~17.5 in L-band we found that the aperture photometry light curves are better in terms of reduced-χ^2 than the light curves obtained from PSF photometry, especially at the very bright end (<15 mag). Some example of lightcurves and their respective reduced-χ^2 measurements are presented in Figure 3.12. We find, however, that for stars fainter than ~17.5 in L-band both types of photometry give similar values of reduced-χ^2, with the PSF photometry perhaps giving slightly better measurements, as implied by Figure 3.8. In the case of variable stars, we can see, in Figure 3.13, that while the PSF light curve shape generally tracks that obtained via aperture photometry, the former display an additional nightly scatter for

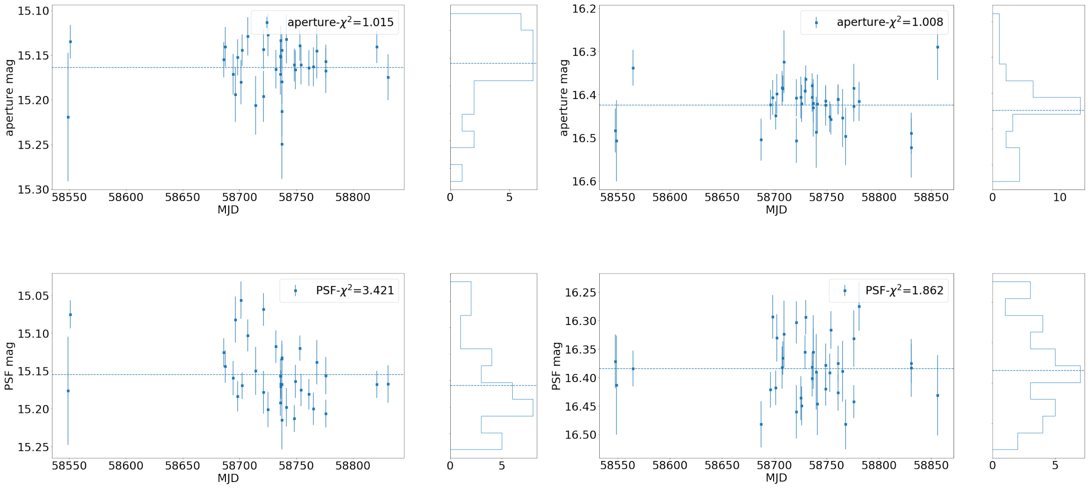


Figure 3.12: Lightcurves of standard SDSS star from aperture and PSF photometry. Aperture photometry gives have reduced-χ^2 closer to unity which is what expected for non-variable stars. Aperture photometry is especially better on the brighter end (<15mag) of the GOTO detections.

measurements from different UTs. As in the case of standard stars, for fainter sources PSF photometry performs better than aperture photometry.

Finally, in order to test our lightcurves, we have also performed a comparison of the lightcurves of variable stars as measured from GOTO data using the LSST stack against those delivered by the ZTF survey. ZTF is wide-field survey that uses the Palomar 48 inch Schmidt telescope and with a dedicated camera of 47 deg^2 field of view (Graham et al. 2019). Its large FOV allows ZTF to scan the northern sky with a 3-day cadence in g and r bands. It also scans the visible Galactic plane every night. The survey has a median limiting g-band and r-band magnitude of ∼20.8 and ∼20.4, respectively. The ZTF survey is similar to GOTO's high-cadence survey, although GOTO's cadence with 4 UTs is lower than that of ZTF. However, when GOTO is fully deployed (with 16 UTs in the Northern Hemisphere and 16 UTs in the Souther Hemisphere), its cadence will surpass that of ZTF.

Figure 3.14 shows the lightcurve plots of four selected variable stars (two cepheids, one BY Draconis type star and one beta-Lyrae) from the General Catalogue of Variable

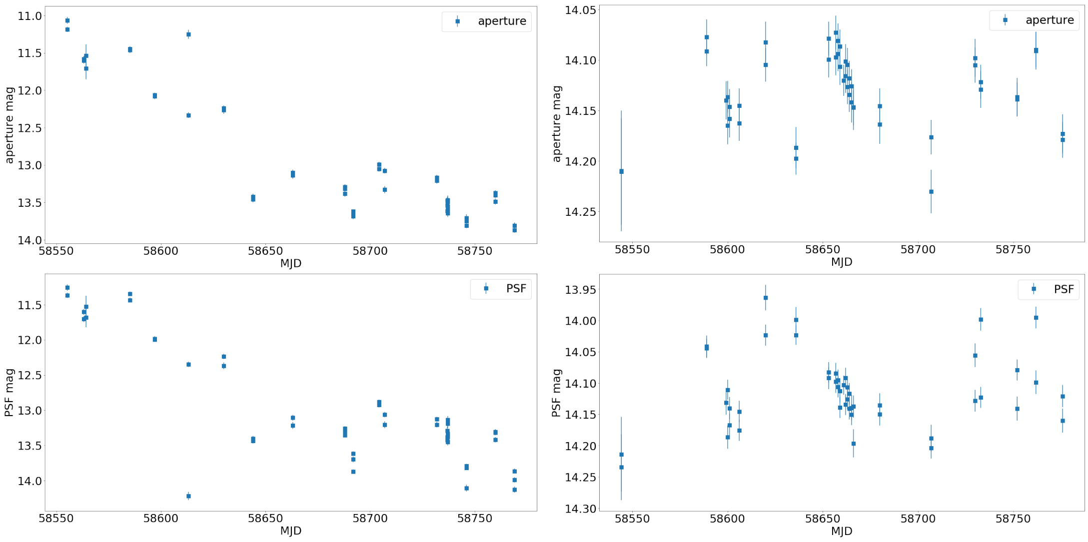


Figure 3.13: Lightcurves of variable star from aperture and PSF photometry. The general shape of the PSF light curves follow the trend of the aperture photometry light curve. For the PSF light curves, however, the plots show an addtional intra-night scatter.

Stars (GCVS 5.1; Samus' et al. 2017) with periods of 5 to 48 days (these periods are those reported by that same catalogue). The ZTF light curves are generated using data from their second data release (DR2) and only include epochs from the public survey. Since the observations from GOTO and ZTF DR span different time periods and flux measurements are not necessarily taken on the same nights, we are presenting the lightcurves as phase plots in which repeated observations of the same part of the cycle are effectively plotted “on top” of each other. To calculate the phase of a given observation , we subtract a reference starting time (the same start time is used in the case of both GOTO and ZTF), dividing by the period of the source, then take the remaining fraction of the period. We note that after applying the correction for the errors as implied by Figure 3.11, the uncertainties generated by our modified forced photometry task are larger than those on the ZTF light curves. The larger number of datapoints on the ZTF light curves are partly due to ZTF having been operating for longer than the period we are considering in this study (i.e., the latter being between February and October, 2019), and its higher cadence over this time. Having said that, it is clear the GOTO

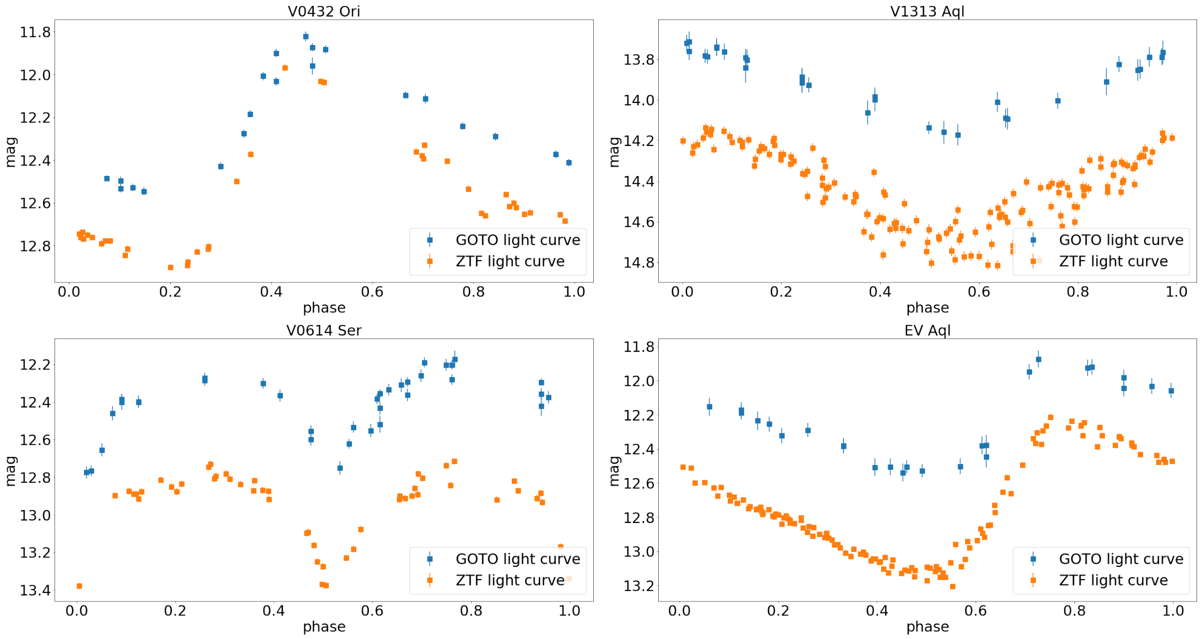


Figure 3.14: Examples of GOTO light curves from aperture photometry measured with the LSST stack, presented as phase plots, of periodic variable stars as compared with the ZTF g-band light curves. We only compare the shapes of the lightcurves as the observed difference in photometry arise from the fact that ZTF measurements are in the ZTF g-band whereas GOTO measurements are in GOTO L-band. These variable stars have period from 5 to 48 days and belong to different classes. GOTO data cover which is shorter than the total time ZTF is running and also the cadence of ZTF is higher for this period since GOTO was operating with the four UTs setup during the period that these data were collected. However the shapes of the GOTO and ZTF phase plots are very similar which shows that GOTO will be a valuable resource for time-domain studies of variable sources.

lightcurves track the shape of their respective ZTF light curves very closely. As such, this comparison makes us confident that GOTO will be valuable resource for measuring variable sources, particularly when it has its full complement of UTs.

3.5 Chapter Conclusions

In this chapter, we have quantified the quality of the reference catalogue and the outputs of the forced photometry. We have tested the reference catalogue astrometry by comparing it with the PanSTARRS catalogue and found that the two agree within 0.31" (mean) for the most of the tracts. We also found that the mean difference between the reference catalogue and the PanSTARRS catalogue is 0.2 mag for the most tracts (excluding the problematic ones described in this chapter). We have also tested deblending, where we found that 6" is the closest that two sources can be to achieve a success rate approaching 100% . We also assessed the forced photometry results. Using repeated photometry of sources detected by GOTO, we found that the precision of our photometry is typically better than 20 mmag for sources brighter than 16 mag. We also compared our photometry measurements (from forced photometry) against color-corrected PanSTARRS photometry, and find that the two agree to within 10 mmag 1σ for bright (i.e., $\sim 14^{\text{th}}$ mag) sources, and 200 mmag for faint (i.e., $\sim 18^{\text{th}}$ mag) sources. These results are similar for the comparison with the GOTOphoto catalogues. From both the reference catalogue and forced photometry we measure a 5σ survey depth of between 19 and 20 magnitudes. We also showed that PSF photometry is better for fainter sources ($>\sim$17.5). Finally, we evaluated the photometric uncertainties for the forced photometry measurements using repeated photometry of non-variable stars. The characterisation of the results was really important in order to confirm that we are able to use the outputs of the LSST stack to investigate AGN variability. In the next chapter, we investigate if AGN variability studies with GOTO are possible at present. In order to investigate that we analyse both the LSST stack and the GOTOphoto outputs, but also we use ZTF data for comparison purposes.

Chapter 4

Studying AGN variability with GOTO

4.1 Chapter Introduction

As outlined in the Introduction chapter, GOTO's wide field of view and high cadence means it has the potential to study the variability of large samples of AGN. Along these lines, GOTO may also be able to use variability as a means of identifying AGN that may have been missed by other selection techniques. This chapter investigates the feasibility of using GOTO to study the variability of moderate luminosity AGN. We start this chapter by presenting the methods that I will use to study AGN variability with GOTO data, before describing a sample of known moderate luminosity AGN whose variability we attempt to investigate using GOTO. I find, however, that there are a number of problems with the current GOTO data that prevent us from performing this analysis. As we shall see, these issues mostly originate from systematic uncertainties that afflict the GOTO data. These are in spite of using various data flags to clean the AGN lightcurves of spurious photometry. Unfortunately, these problems appear to affect all of the standard methods used to quantify AGN variability. To assess whether our findings are related to our choice of processing pipeline (i.e., the LSST stack), I also analyse data produced by GOTOphoto, but reach the same conclusions (i.e., that the current GOTO data is not currently of sufficiently high enough quality to study AGN variability).

In order to further verify the conclusions of this study, I also decided to use ZTF data as a comparison dataset. The ZTF survey is similar to GOTO's high-cadence survey and although GOTO's cadence with 4 UTs is lower than that of ZTF, when GOTO is fully deployed (with 16 UTs in both the northern and southern hemispheres), its cadence will match that of ZTF and will cover the whole sky, as opposed to the northern sky accessible by ZTF. The ZTF AGN lightcurves of the same sample are analysed in the same way to allow us to investigate whether our results arise from intrinsic properties of the studied AGN sample or indeed GOTO data drawbacks. The identification of these issues through specialised studies allow us to give feedback to the hardware and software teams in the GOTO collaboration to be able to address these complications and contribute to resolve these issues with targeted actions. In the case of the LSST stack, it also allows specific problems to be identified (e.g., photometric calibration, photometric errors) which we need to address in order to improve the prospects of studies that intend to use the forced photometry lightcurves generated by the LSST stack.

4.2 AGN variability methods

Before I present the data and the results of our analysis of AGN variability using GOTO data, it is important to explain the methods I used for this study. In order to describe how sources vary, it is necessary to first identify which sources are considered variable from their lightcurves. To do so, two parameters are used (i.e., the probability of an object to be variable P_{var} and the excess variance, σ_{rms}) which I describe in more detail in the following subsection. Objects can be classified as variable or non-variable depending on their values of P_{var} and σ_{rms}. Once a source has been classified as variable, I then use the "structure function" (see section 4.2.2) to quantify its variability. Finally, I describe the "damped random walk" model which I use to generate simulated AGN lightcurves (see section 4.2.3).

4.2.1 P_{var} and σ_{rms}

I have used two different quantities to classify objects as either variable or non-variable. The first quantity that is used is the probability of an object to be variable, P_{var}. In order to calculate the probability of an object to be variable, I first find the observed χ-squared i.e.:

$$\chi^2_{obs} = \sum_{i=1}^{N} \frac{(m_i - \bar{m})^2}{\sigma^2_{err,i}}, \tag{4.1}$$

for a light curve with N measurements of magnitude m_i and mean magnitude $\bar{m}$ and individual photometric errors of the measurements, $\sigma_{err,i}$. For an intrinsically non-varying source, the probability that a χ^2 larger than the χ^2_{obs} could just occur by chance due to Poisson noise is:

$$P(\chi^2 \geq \chi^2_{obs}) = \int_{\chi^2_{obs}}^{\infty} f(\chi^2, N-1) d\chi^2, \tag{4.2}$$

where $f(\chi^2, N-1)$ is the probability density function of the χ^2 with $N-1$ degrees of freedom. Given the probability of a source to be variable, I calculate the variability index (e.g., Lanzuisi et al. 2014) which is defined as $V = -\log_{10} P(\chi^2 \geq \chi^2_{obs})$. One of the benefits of using this approach to classify an object as being variable is that it has been used in previous studies, allowing us to compare our results with those from the literature (Cartier et al. 2015, Sánchez-Sáez et al. 2018, Simm et al. 2016). Consistent with those previous studies, we class an object as variable if V>1.3. This threshold rejects the null hypothesis of an intrinsically non-variable source with 95% confidence.

The σ_{rms} is called excess variance and in this work it will be used both as one of the quantities to identify the variable sources as well as one of the means to quantify the amplitude of AGN variability as used in e.g., Sánchez-Sáez et al. (2018) and Simm et al. (2016) . The σ_{rms} is defined as the Root Mean Squared (RMS) of the standard deviation of a time-varying quantity (x) after correcting for its average error (σ_{err}) and normalising by its mean value ($\overline{x}$):

$$\sigma^2_{rms} = \frac{1}{\bar{x}^2}\left(\sum_{i=1}^{N}\frac{(x_i-\bar{x})^2}{(N-1)} - \sum_{i=1}^{N}\frac{\sigma^2_{err,i}}{N}\right). \tag{4.3}$$

where N is the number of measurements.

The uncertainty on σ_{rms} due only to Poisson noise is calculated using the following equation (Vaughan et al. 2003):

$$err(\sigma^2_{rms}) = \sqrt{\left(\sqrt{\frac{2}{N}}\cdot\frac{\sigma^{\bar{2}}_{err}}{\bar{x}}\right)^2 + \left(\sqrt{\frac{\sigma^{\bar{2}}_{err,i}}{N}}\cdot\frac{2\sqrt{\sigma^2_{rms}}}{\bar{x}}\right)^2}. \tag{4.4}$$

In addition to the V>1.3 condition we used to classify objects as variable or non-variable, a variable source should also satisfy:

$$\sigma^2_{rms} - err(\sigma^2_{rms}) > 0. \tag{4.5}$$

Only sources that satisfy both conditions are classified as variable. For those sources which are classified as variable by the two constrains described in this subsection, I next analyse their variability via their structure function, which I describe in the next subsection.

4.2.2 Structure Function

The structure function (SF) is one of the most widely-used methods to describe AGN variability. One of the main reasons why it is so widely utilised is that the SF can be used on sparsely-sampled AGN lightcurves, which is a great advantage for ground-based surveys since their cadences depend on their FOV are also affected by “gaps” in their lightcurves where data is missing due to, for example: seasonal gaps, weather, changes in cadence because of scheduled surveys or triggers (i.e., in GOTO's case, when there is a GW or GRB alert). Many different ways of calculating the structure function have been proposed but all essentially describe the RMS of the magnitude difference as a function of the time separation between observations (hereafter observational SF). There are also different ways of modelling the observed SF.

In this work, I used three different ways to determine the structure function of the studied lightcurves in order to verify the truth of our analysis. First, I used the definition from Schmidt et al. (2010), where the RMS magnitude difference between observations separated by a time lag is corrected for the photometric uncertainties of the observations:

$$\mathrm{SF}(\Delta t) = \left\langle \sqrt{\frac{\pi}{2}} |\Delta m_{i,j}| - \sqrt{\sigma_i^2 + \sigma_j^2} \right\rangle, \tag{4.6}$$

in which $\Delta m_{i,j}$ is the magnitude difference between two observations separated by a time lag Δt and σ_i, σ_j are the photometric uncertainties of the i and j measurements respectively. I calculated the SF for each AGN lightcurve i.e. for each individual object. Figure 4.1 shows an example of a lightcurve and the calculated SF using the aforementioned definition.

In our analysis, it was necessary to determine the variability as a function of the time lag in terms of just RMS without correcting for the photometric errors to verify whether the root of the issues was the RMS or an over/under-estimate of the photometric errors. To do so, I have used the robust RMS measurement as defined in MacLeod et al. (2012)

$$\mathrm{SF}(\Delta t) = 0.741 \times IQR, \tag{4.7}$$

where IQR is the 25%-75% interquartile range of the magnitude difference (Δm distribution separated by the Δt time lag), and N is the number of Δm values in this time lag. The observed SF can be corrected for the photometric error by $SF = \sqrt{SF^2 - 2\sigma^2}$

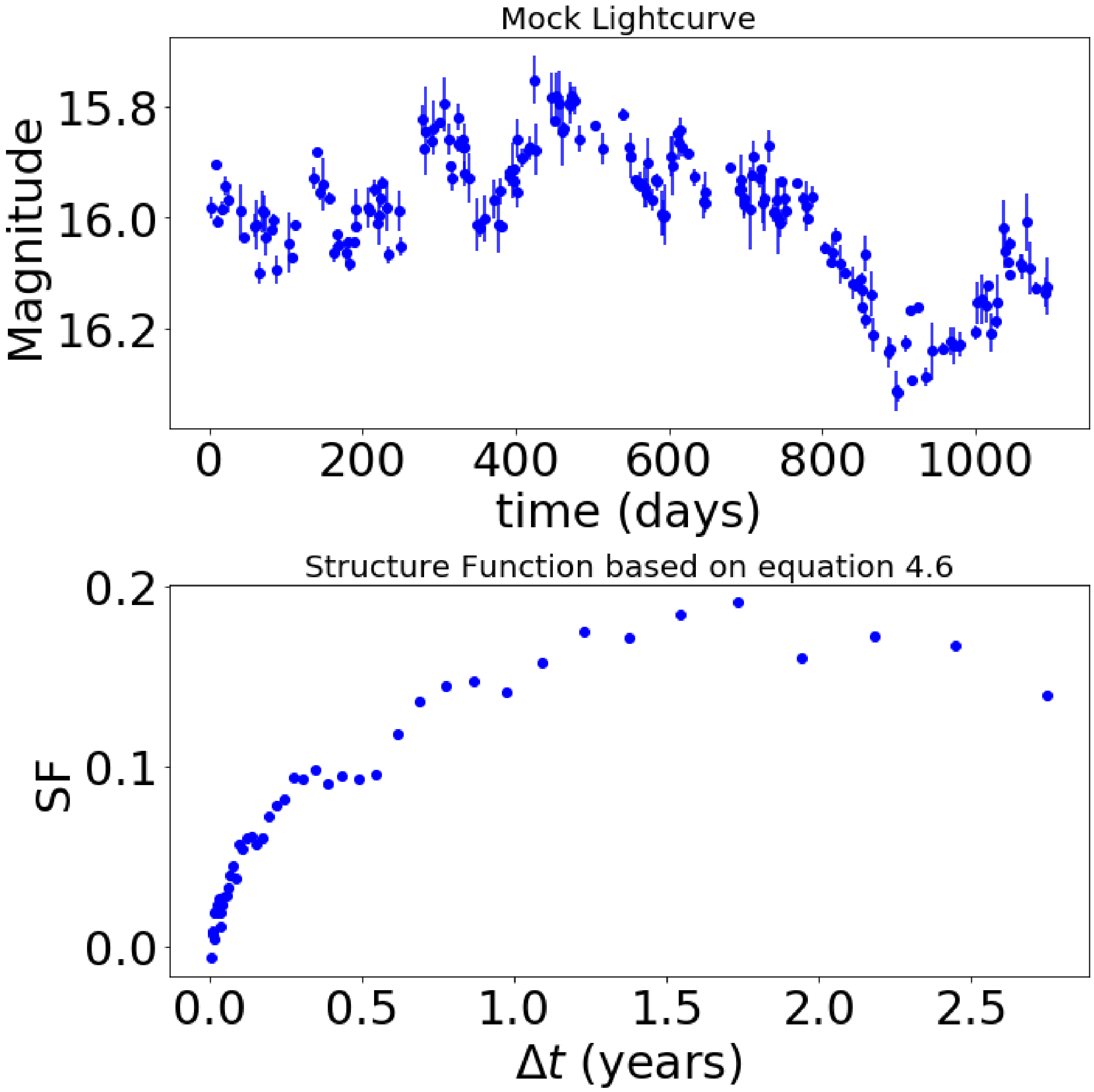


Figure 4.1: The figure shows a mock lightcurve with assigned photometric errors (top) spanning ∼3 years and its respective SF plot versus the time lag (bottom) using equation 4.6.

where σ is the photometric error value representative for the sample based on the median RMS scatter for each band (e.g., *ugriz* for SDSS; Sesar et al. 2007).

I have also tested the Kozłowski (2016) definition for the observational structure function, where they use an alternative way to measure the contribution of the photometric uncertainties on the SF:

$$\mathrm{SF}^2(\Delta t) = 0.549(IQR^2(\Delta t) - IQR^2(n)), \tag{4.8}$$

where IQR(Δt) is the interquartile range between 25% and 75% of the magnitude measurements and IQR(n) is the interquartile range of the magnitude distribution for $\Delta t < 2$ days, which is used as an estimate for the photometric uncertainty as SF$\rightarrow$ 0 when $\Delta t \rightarrow 0$.

The observational structure function is found to be well described using a broken power-law (e.g., Kawaguchi et al. 1998, MacLeod et al. 2010) i.e. that variability increases linearly with time difference up to a certain point, beyond which it is uncorrelated. In this work, I have used the model from Schmidt et al. (2010) which describes the power-law part of the structure function as:

$$\mathrm{SF}(\Delta t) = A \left(\frac{\Delta t}{1yr}\right)^{\gamma}, \tag{4.9}$$

where the amplitude A is the RMS of the magnitude difference on a one year timescale, and γ is the logarithmic gradient of the mean change in magnitude. One of the ways to analyse the variability of different types of AGN, or how variability changes with other physical parameters (e.g., SMBH mass, AGN luminosity) is to plot different sources in the A-γ parameter space.

4.2.3 The Damped Random Walk

As mentioned in the introduction of this thesis (Chapter 1), the stochastic nature of AGN variability was studied by Kelly et al. (2009) and they suggested that it could be described as a Damped Random Walk (DRW). This model statistically explains the observed lightcurves of AGN, even though the physics behind it have yet to be identified. In this process, AGN flux $X(t)$ is estimated by the equation:

$$dX(t) = \frac{1}{\tau}X(t)dt + \sigma_{DRW}\sqrt{dt}\epsilon(t) + bdt,$$

where τ is the time needed for the time series to become uncorrelated and σ_{DRW} corresponds to the amplitude of variability on timescales short compared to τ, $b\tau$ is the mean magnitude and $\epsilon(t)$ is a white noise process with zero mean and variance equal to 1. DRW has a PSD of slope -2 for time lags shorter than the characteristic timescale which is equivalent to a power (e.g., γ in Eqn. 4.9) of 0.5 for the SF (Kelly et al. 2009, MacLeod et al. 2010).

While previous studies have used the DRW model to describe, quantify or identify AGN variability (e.g., MacLeod et al. 2010, Baldassare et al. 2019), I chose not to quantify the variability in this way, because for a proper estimate of τ, the length of the lightcurves should be larger than $10 \times \tau$ (see Kozłowski 2017). Since at the time of writing, our lightcurves only span around 1 year in the rest frame, while the expected characteristic timescale of Type 1 AGN is of the order of hundreds of days, I chose to avoid this parametrisation in our study. In this work, I have used a DRW to simulate AGN lightcurves given the current and the future GOTO cadence to investigate the dependence of any results on the cadence of the observations. I have also tested the differences between short (< 400 days) and long light curves (> 2 years).

4.3 Data

As mentioned in Chapter 1, during this study, the GOTO project went through numerous developments both in terms of hardware and software. Although GOTO saw its first light on July 2017, prior to February 2019, the first four UTs suffered from being poorly collimated. While the data obtained during this early phase was useful for preparing the data processing pipelines (including the LSST stack), it was of too low a quality for scientific studies such as those outlined in this chapter. Using the data after February 2019, science working groups within GOTO collaboration focused on the quantification of the quality of the data with regard to the scientific goals of its group. This means that we need to use the current status to identify the limitations (e.g., studies of fainter sources) and any problems (e.g., systematic uncertainties) of the delivered data and also those aspects of the pipeline that require improvements (e.g., absolute photometry, error measurement, etc.). The data used in this study span the period from the 22nd of February 2019 to the 30th June 2020, which means that the lightcurves are currently limited by their baseline as they span 494 days in the observer's frame which is a limitation since our aim is to study variability over 1 year in the rest-frame. For sources at higher redshifts, this would be a significant problem as the rest-frame time length will be $\Delta t_{restframe} = \Delta t_{obs}/(1+z)$, where Δt_{obs} is the time length at the observer's frame and z is the redshift of the source. Also these data were taken with the 4 UTs setup so the mean cadence is $\sim$10 days in general. However, we shall see later how the time length and the number of epochs is distributed for the AGN sample I study and how this affects the uncertainties of our results.

In this work, I investigated whether it is possible to study AGN variability of moderate luminosity AGN with GOTO at redshift $z < 0.4$. While, there are few high luminous AGN in the sample that I used in this study ($L[OIII] > 10^{45}$ ergs s^{-1}), the majority of them are of moderate luminosity ($L[OIII] = 10^{40} - 10^{45}$ ergs s^{-1}) AGN, simply due to the fact that less luminous AGN are more common than high luminosity AGN (e.g., Aird et al. 2010). The parent sample consists of 25,670 optically selected AGN (Mullaney et al. 2013). Mullaney et al. (2013) studied 25,670 SDSS selected sources which were initially classified as AGN from emission-line ratios and line widths using the SDSS line fitting routine. In this method, using emission line ratios parameter space (e.g. [NII]6584/Hα versus [OIII]5007/Hβ) AGN are separated from star-forming galaxies by their location on the parameter space (BPT diagrams;Baldwin et al. 1981). Mullaney et al. (2013), however, then used a multicomponent fitting routine that used up to two Gaussians to fit the [NII]λ6548, and [OIII]λ5007 emission lines and up to three Gaussians to fit the Hα and Hβ (i.e., two for the narrow component and one for the broad component) to obtain a better classification for the 25,670 sources (see Mullaney et al. 2013 for more details on the fitting method). Using this method, they have identified 10,548 Type 1 AGN i.e., these are sources that needed an additional broad component for Hα fitting beyond the components required to fit the forbidden lines. From these Type 1 AGN, 9,455 are classified as broad-line Seyfert 1s (BLSy1;hereafter Type 1 AGN) and 1,093 as narrow-line Seyfert 1s (hereafter NLSy1) based on the FWHM of the Hα broadest component i.e., if it is larger or less than 2,000 km s^{-1}. To identify Type 2 AGN in the remaining sample, they used BPT diagnostics using the narrow component flux alone and find 13,716 Type 2 AGN, in their sample. The remaining 1,406 sources were excluded from their study, although I included them in this study, but label them as a separate class of AGN. According to Mullaney et al. (2013), these latter sources are mainly located in the LINER region of the BPT diagram. For all these objects, we attempt to generate lightcurves using data collected by GOTO and processed separately using the LSST stack and GOTOphoto. In the case of the LSST stack lightcurves, I used aperture photometry for sources brighter than 17.5 magnitude and PSF photometry for the fainter than that sources. GOTOphoto uses aperture photometry throughout. We also note that for the LSST stack data the uncertainties are corrected based on the polynomial fitting described in section 3.4.3.

4.4 AGN Lightcurves

This study involves analysing AGN lightcurves from GOTO and ZTF DR3 for the SDSS AGN described in the previous section. The parent catalogue contains 25,670 AGN which are classified according to their spectra into different types. I cross-matched the parent catalogue, using a 1" matching radius, with the GOTOphoto database, the GOTO/ LSST stack database and the ZTF database. For GOTOphoto the cross-matching returned the lightcurves of 25,509 objects (99,4 %), for GOTO/ LSST stack, due to the coverage of the reference catalogue, the cross-matching returned 20,005 lightcurves (77.9%) and finally for ZTF I recovered 24,037 (93.6%). For the ZTF data, we note that these are the public data which are subject to certain rules[1] (e.g., unreleased epochs). The second column of Table 4.1 lists the numbers after the 1" cross-matching with the three different databases.

Survey and Pipeline	1" match	Well-Sampled	Variable
ZTF	24,037	5,973 g-band 7,280 r-band	5,646 g-band 7,073 r-band
GOTO LSST stack	20,005	8,259	5,802
GOTO GOTOphoto	25,509	4,902	3,434

Table 4.1: Table of the objects found in each survey and pipeline (for GOTO) databases after applying the various criteria described in the text. Well-sampled and variable numbers are defined after the data have been cleaned based on the flags each pipeline use to determine bad data points.

As I described in Chapters 1 and 2, GOTO takes three back-to-back exposures for each pointing whilst in its standard survey mode. The GOTO lightcurves from both

---

[1] `https://www.ztf.caltech.edu/page/dr3`

pipelines are generated using the photometry from the combined back-to-back exposures for each pointing. Before I analyse the lightcurves generated by the two pipelines, we need to clean them of any spurious photometric data. Both pipelines use flags that can be used to identify such bad data points. Reasons for bad data points can be astrometry failures, sources containing saturated pixels and bad weather conditions that are reflected in the PSF FWHM of the images. For the LSST stack-processed data, I have used the `pixelFlag_interpolated` flag to identify data points where the sources were affected by bad pixels and remove them as this flag identifies pixels affected due to, for example, hot pixels or saturated pixels. For each lightcurve, I also removed data points whose errors are larger than two times the mean photometric error of the lightcurve's data points, and data points from images whose zeropoints deviate by more than 3σ from the mean zeropoint. Additionally, I exclude any data from images whose PSF is larger than FWHM 7.0" (i.e., $\sim 1\sigma$ larger than the mean PSF FWHM for the GOTO images as found from the LSST stack processing) at the center of the image.

For the GOTOphoto pipeline, I used the `quality` and `quality_template` flag, which is a bitmask where different bit values flag different issues (e.g., high PSF FWHM, bad astrometry, high ellipticity). Using the `quality` flag, I filter out data that have been flagged because the measurement is coming from an image that has a bad astrometric solution. Then, using the `quality_template` flag, we also filter out data, which are flagged because the largest FWHM across the grid and its standard deviation are larger than the limits (i.e. $>$ 7.5-10.5 depending on the UT for the FWHM highest values and 1-1.8 for the standard deviation). I tried different combinations of filtering using the flags but this did not affect the output. The quality flags for the GOTO/ LSST stack as well as the GOTOphoto data do not affect the number of objects, just the number of epochs.

In order to clean the ZTF data I have used the `catflags` column in the data catalogues which flags pixel failures (e.g., saturated), bad seeing, high median background level etc.[2], requiring that this flag is equal to zero. After using the flag cleaning method, there are 12,873 ZTF lightcurves in the r-band and 8,014 in the g-band. I used the g-band ZTF lightcurves for the comparison with the GOTO data part (sections 4.4-4.3), and analyse both g-band and r-band in section 5.1, where we study AGN variability using the ZTF lightcurves alone.

In order to study AGN variability using the statistical methods I described in section 4.2 I decided to analyse the objects with well-sampled lightcurves. This leads us to our next selection. In this analysis, I defined as well-sampled those lightcurves that span more than 200 days in the rest frame (i.e., $t_{rest} = t_{obs}/(1+z)$ where t_{obs} is the lightcurve length in the observer's frame). In addition, to be considered well-sampled, I required that a lightcurve has more than 20 epochs. We define an epoch as one night's worth of measurement; i.e., five separate measurements in one night constitute a single epoch, whereas five measurements on five separate nights constitute five epochs. In Figures 4.2 and 4.3, I present the number of epochs versus the rest frame time length for 20,005 AGN for the GOTO LSST stack processed data and 25,509 AGN from the GOTO/GOTOphoto processed data. Using the aforementioned cuts, we will continue the analysis for 8,259 object for LSST stack processed data and 4,902 using GOTOphoto processed data. In Figure 4.4 we show the number of epochs for the ZTF g-band lightcurves, where 5,793 satisfy the cuts for the rest-frame time length and the number of epochs.

---

[2] http://web.ipac.caltech.edu/staff/fmasci/ztf/ztf_pipelines_deliverables.pdf

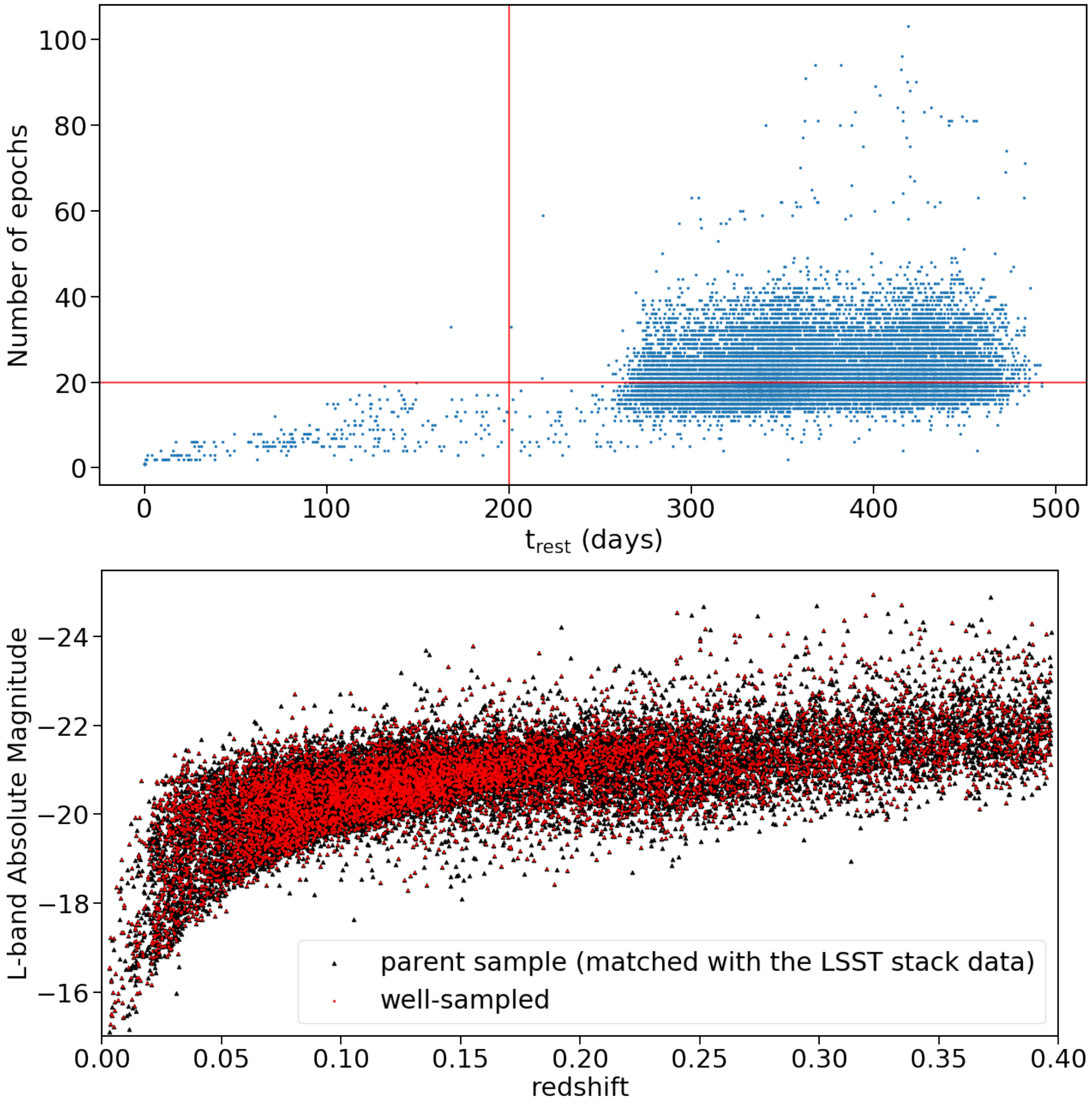


Figure 4.2: The number of epochs per AGN versus the rest-frame time length of the each lightcurve from the generated from LSST stack measurements (top). The red lines indicate the two criteria for well-sampled lightcurves, i.e. 200 days length in rest-frame and 20 epochs per lightcurve. There are very few data points for the low epochs-$t_{\rm rest}$ for the GOTO data processed with the LSST which is the result partially from the sample that was matched with LSST stack database but also because they are processed using forced photometry. This means that the LSST stack will perform forced photometry on the position of the AGN even if this was not detected on the nightly frame. The bottom panel shows the absolute L-band magnitude against redshift. The absolute magnitudes are not K-corrected or corrected for dust extinction. The black triangle points show the parent sample after it was matched to the LSST stack database (20,005 sources), and the red dots show these that have well-sampled lightcurves i.e. more than 20 epochs and span more than 200 days. A standard Λ cold dark matter cosmology with $H_0 = 70$ $\mathrm{km\,s^{-1}\ Mpc^{-1}}$,$\Omega_M = 0.27$, and $\Omega_\Lambda = 0.73$.

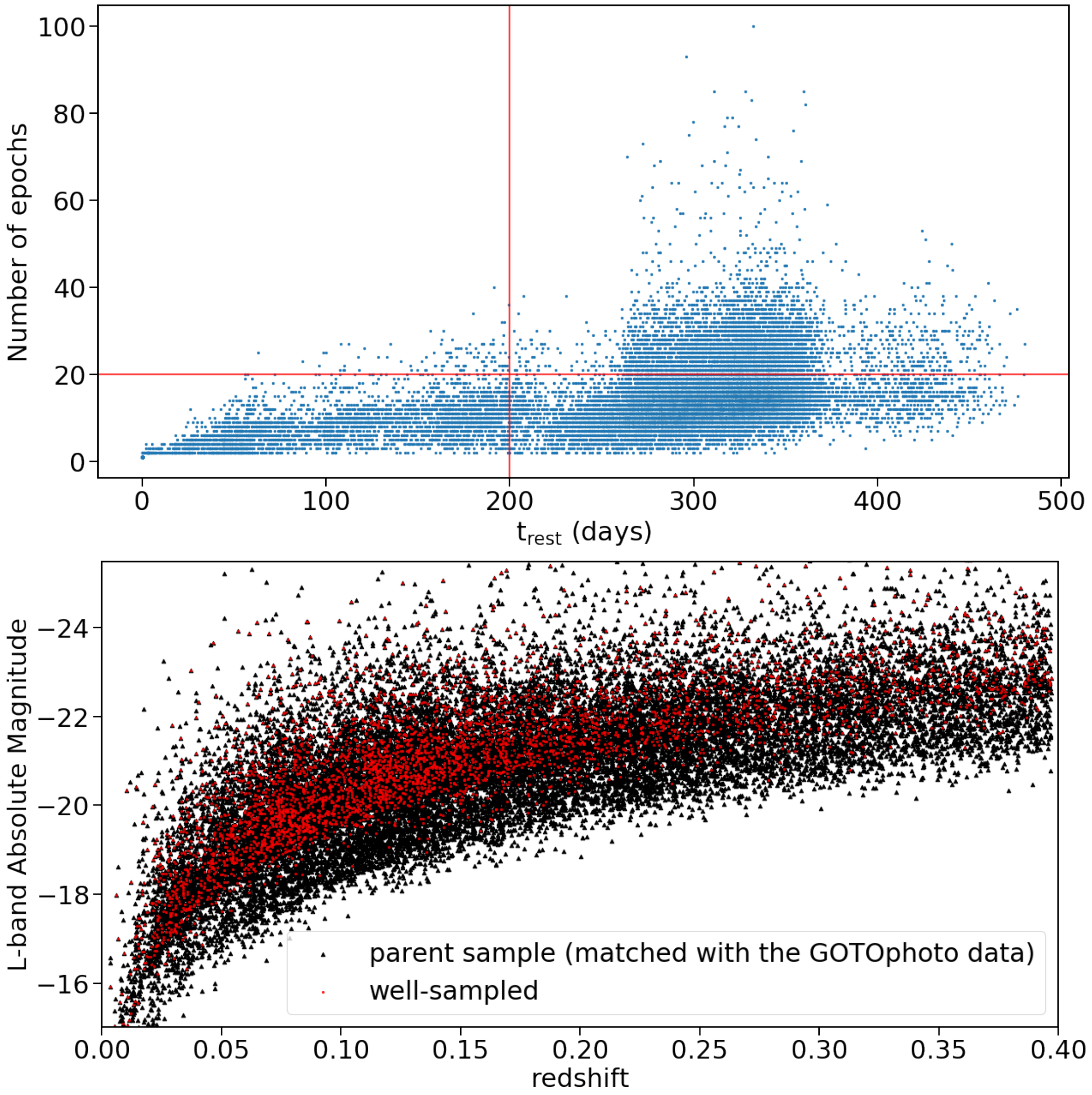


Figure 4.3: The number of epochs per AGN versus the rest-frame time length of the each lightcurve from the generated from GOTOphoto measurements (top). The red lines indicate the two criteria for well-sampled lightcurves, i.e. 200 days length in rest-frame and 20 epochs per lightcurve. The bottom panel shows the absolute L-band magnitude against redshift. The absolute magnitudes are not K-corrected or corrected for dust extinction. The black triangle points show the parent sample after it was matched to the GOTOphoto database (25,509 sources) and the red dots show these that have well-sampled lightcurves i.e. more than 20 epochs and span more than 200 days. We find that in the case of GOTOphoto the well sampled lightcurves tend to be these for the brighter sources across the redshift range, which is expected as the GOTOphoto photometry is using at the moment "blind" detection and photometry measurements on these sources. So, fainter sources are more difficult to detect when, for example, we have bad weather conditions that affect the magnitude limit. A standard Λ cold dark matter cosmology with $H_0 = 70\ \mathrm{km\,s^{-1}\ Mpc^{-1}}, \Omega_M = 0.27$, and $\Omega_\Lambda = 0.73$.

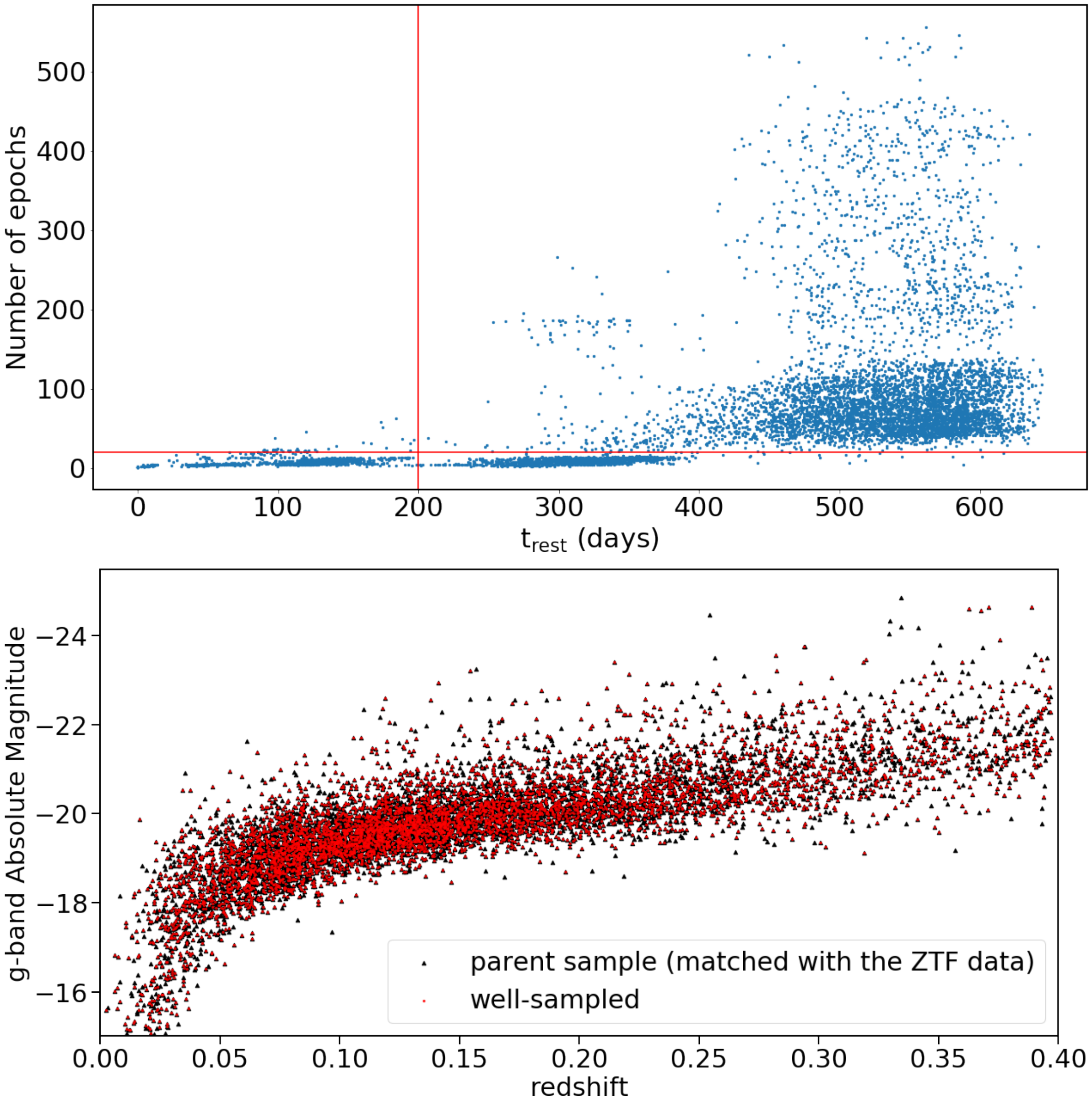


Figure 4.4: The top panel shows the number of epochs per AGN versus the rest-frame time length of the each lightcurve generated from the ZTF g-band public data. The red lines indicate the two criteria for well-sampled lightcurves, i.e. 200 days length in rest-frame and 20 epochs per lightcurve. The bottom panel shows the g-band absolute magnitude against redshift. The absolute magnitudes are not K-corrected or corrected for dust extinction. The black triangle points show the parent sample after being matched to the ZTF public database for g-band photometry (8,014 sources) and the red dots show these that have well-sampled lightcurves i.e. more than 20 epochs and span more than 200 days. A standard Λ cold dark matter cosmology with $H_0 = 70\ \mathrm{km\,s^{-1}\ Mpc^{-1}}$,$\Omega_M = 0.27$, and $\Omega_\Lambda = 0.73$.

4.5 Variable AGN from GOTO and ZTF lightcurves

After selecting a sample of AGN with well-sampled lightcurves, our next step is to identify how many of the sources are variable using the definitions in section 4.2.1. From the LSST stack analysis we have a total of 8,259 well sample lightcurves. Of these 2,863 are Type 1 AGN (i.e., BLSy), 4,658 are Type 2 AGN, 327 are NLSy1 and 411 are galaxies with a weak AGN presence (typically LINERS). From these sources, and using the constraints based on P_{var} and σ_{rms} described in section 4.2.1, we find that 2,015 of the Type 1 (70.4%), 3,271 (70.2%) of the Type 2 , 216 of the NLSy1 (66.0%) and 300 (73.0%) of those with the weak-AGN are found to be variable.

Of the 4902 objects with well-sampled lightcurves from the GOTOphoto data, 1,600 are Type 1 AGN (i.e., BLSy), 2,828 are Type 2 AGN, 196 are NLSy1 and 278 are galaxies with weak AGN presence. After applying our variability criteria, we find that, of these, 1,160 of the Type 1 (72.5%), 1,944 (68.7%) of the Type 2, 134 of the NLSy1 (68.3%) and 196 (70.5%) of the weak-AGN are classed as variable.

In the case of the ZTF g-band lightcurves, we find that of the 5,793 lightcurves, 1,863 are Type 1 AGN, 3,403 are Type 2 AGN, 223 are NLSy1 and 304 have weak AGN presence. We find that: 1,788 (95.9%) of the Type 1 sources, 3,361 (98.8%) of the Type 2, 206 (92.4%) of the NLSy1, 291 (95.7%) of the weak presence are variable based on their lightcurves and the conditions of variability index and the σ_{rms}. The large inconsistencies between the results from the different samples may be caused partially by the different well-sampled samples that consist of different objects (especially in the case of ZTF), however, it is highly probable that it is also caused by the issues explained in sections 4.6 and 4.7.

An odd result in the case of our sample is the small differences between the Type 1 and Type 2 percentages of variable sources in all three cases (GOTO-LSST stack, GOTO-GOTOphoto, ZTF). For comparison, based on a much smaller sample than ours and a combination of X-ray and optical spectroscopic classification, but using the same method (P_{var} and σ_{rms} to identify variable lightcurves), Cartier et al. (2015) reported that, of 55 and 24 respectively, 80% of broad line AGN are variable and 21% of narrow-line AGN are variable, which differs from the numbers found in this work for all the aforementioned samples. Cartier et al. (2015) used QUEST-La Silla lightcurves that span more than 600 days in the observer's frame, however they do not report the exact number of epochs per lightcurve. The length and number of epochs can be, to some degree, the reasons for the divergence. This can be investigated in the future when both GOTO and ZTF will allow us to get higher cadence and longer lightcurves for each object in our sample.

Before we move on to the next section in which I present our results from the structure function analysis, I plot the source distributions in mean magnitude versus the redshift (Figure4.5). We find that most of the sources have mean magnitudes fainter than 17.5 mag. We recall that we have corrected the uncertainties of the lightcurves generated by the LSST stack based on section 3.4.3 but we recognise that the low apparent magnitude of the sources and the results on the uncertainties described in the aforementioned section may still introduce spurious variability. The underestimation of the uncertainties however should not be the case for ZTF (i.e. we find no dependence of the uncertainties on the magnitude for the ZTF data), where we still find a high proportion of variable lightcurves even for Type 2 AGN.

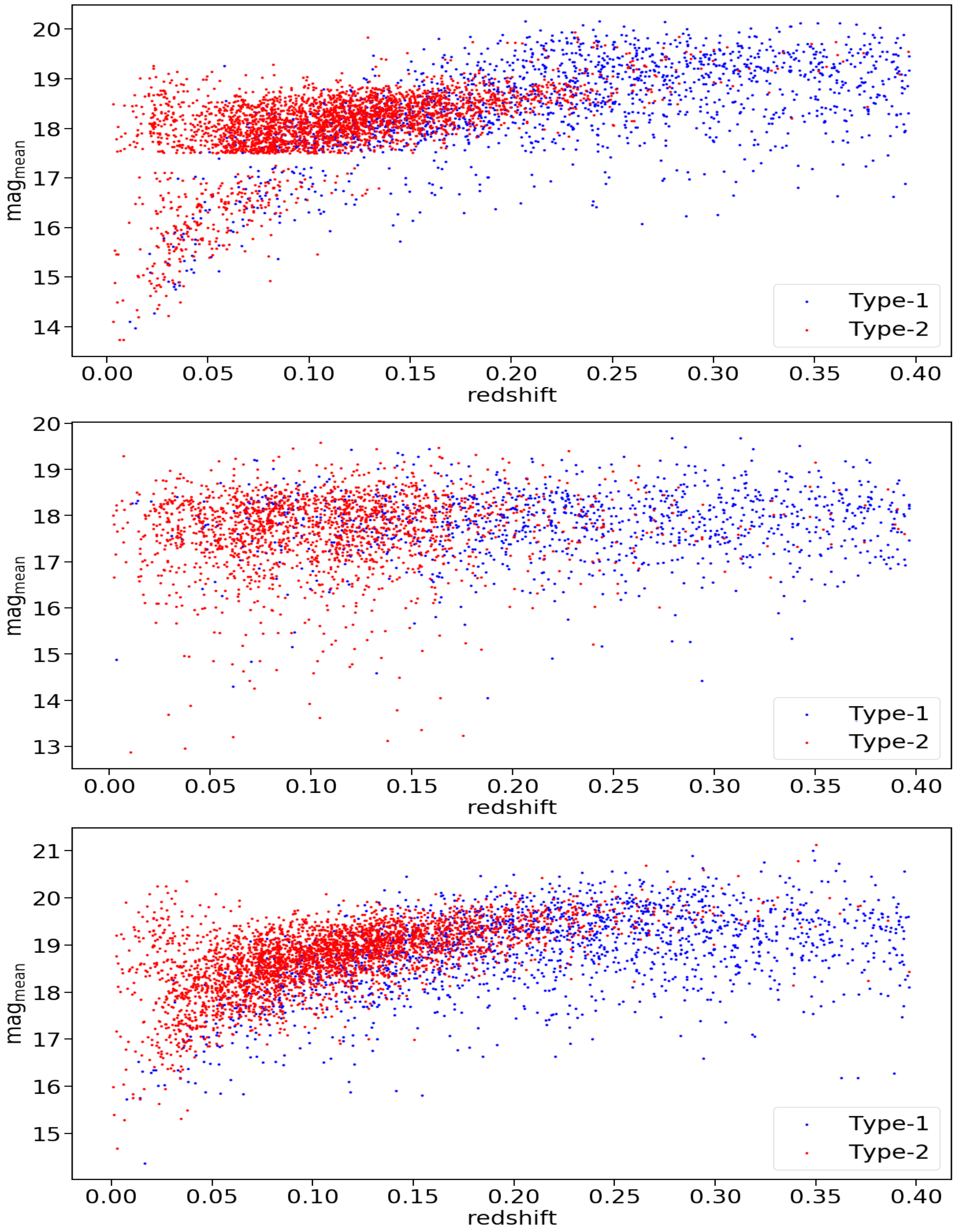


Figure 4.5: Redshift versus magnitude distributions for all three samples i.e. GOTO LSST stack (top panel), GOTO GOTOphoto (middle panel) and ZTF (bottom panel). The gap on the top plot for the LSST stack processed GOTO data is because we choose aperture photometry for sources brighter than 17.5 and PSF photometry for sources fainter than 17.5 mag. We see, however, that for all the samples most of the sources are fainter than 17.5 mag.

4.6 A vs γ parameter space

After selecting well-sampled lightcurves and identifying variable objects, our next step is to quantify the variability of the various types of AGN in our three samples (from the LSST stack-processed data, from the GOTOphoto data, and from the ZTF data) using the Structure Function (SF). As the SF measures the RMS of the magnitude difference of an AGN as a function of the time lag between observations, fitting a power-law to our measurements of the RMS of the magnitude differences per time lag (corrected for the photometric uncertainties) is expected to give the amplitude of variability at various timescales (in our case I calculate the amplitude at 1 year timescale) and the logarithmic gradient of the change. The optical variability of AGN is linked to the central engine of the AGN, which is powered by the accretion disk around the SMBH. As suggested by the unified model of AGN, in the case of the Type 2 AGN the torus, as well as obscuring the broad line region (and hence the resulting broad emission lines), also obscures the central engine and thus any variability intrinsic to the accretion disk. Cartier et al. (2015) studied a small number of AGN (see section 4.5 for more details on the sample), and by measuring A and gamma of their SFs (see section 4.2.2 for this parameters), found that Type 2 AGN cluster around (0,0) on the plot of A vs. γ. By contrast, we expect Type 1 AGN to deviate from these values as the flux measured of these objects varies. For variable AGN, as the time lag between observations increases the RMS of the magnitude difference should also increase until it reaches a certain time lag, after which the two values become uncorrelated, as found in previous observational studies (e.g., Collier & Peterson 2001, MacLeod et al. 2012). On the other hand, non-variable sources should have a flat SF with amplitudes that are close to zero (close to, rather than exactly zero due to errors and systematics which are difficult, if not impossible, to entirely eliminate).

For each sample of lightcurves (GOTO/ LSST stack, GOTOphoto, ZTF), I fitted the SF with a power law (Eqn. 4.9), optimising the values of A and γ using Markov Chain Monte Carlo (MCMC). This Bayesian method was introduced by Schmidt et al. (2010), and it has been demonstrated from previous studies (e.g., Sánchez et al. 2017) that it is the preferred approach for irregularly sampled and noisy lightcurves. Using the model described by Eqn. 4.9, we calculate the amplitude of variation at a time lag, Δt, of 1-year (rest frame) and the slope, γ, of the SF at $\Delta t < 1$ year (rest frame). I only consider time lags of between 10 days and 1 year, while fitting the median magnitude difference and using the 15.9 and 84.1 percentiles for upper and lower errors.

I fit all the viable lightcurves (i.e., after filtering on all the criteria outlined in sections 4.4 and 4.5) from each of the GOTO datasets (LSST and GOTOphoto) and the ZTF dataset, obtaining best fit values for A and γ for each lightcurve. In Figure 4.6 and Figure 4.7, I plot the distributions of A and γ derived from the GOTO data. In the scatter plots of Figures 4.6 and 4.7 we see that for both pipelines the Type 1 and the Type 2 objects cover the same region without showing any significant difference and both types appear to have a large range of amplitude and similar values for the logarithmic gradient. This similarity for the A and γ values of Type 1 and Type 2 AGN becomes clearer when considering the histograms of these values. This is contrary to the expected result whereby Type 1 AGN should show evidence of being more variable as a result of our direct line of sight to the varying central engine. Also, Type 1 γ values should be distributed around 0.5 for lightcurves that follow the DRW model, or at least different than these of Type 2 which are expected to have flat SFs. It was necessary then to investigate if this is intrinsic for the sample of moderate luminosity AGN, or whether there are current issues that hide the true variability values of these sources.

Our inability to distinguish between Type 1 and Type 2 AGN in terms of variability is a major concern, as this should be among the easiest differences to detect. It therefore hampers the prospects of more detailed studies that explore how AGN variability relates to other parameters. The fact that we get the same qualitative results for both the

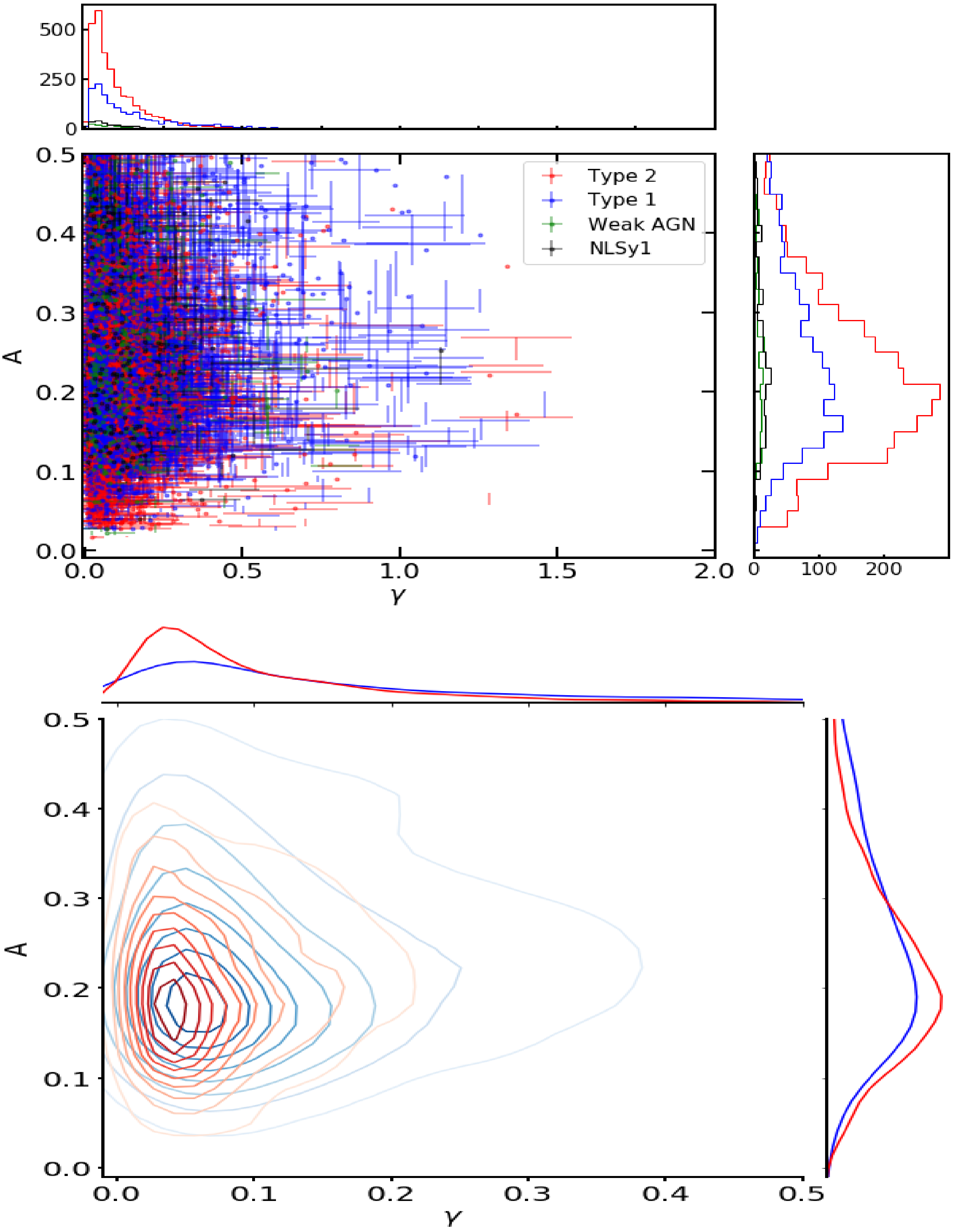


Figure 4.6: *Top*: A vs γ parameter space for GOTO LSST stack processed data. For a better representation of the results in this plot, we have eliminated data points with errors for A or γ larger than 0.2, which indicate a failure in fitting. This reduces the sample from a total of 5,802 to 5,217. We find that the data points for the different AGN types cover essentially the same part of the plot and their distributions, as indicated by the histograms, do not show significant difference as discussed in this section.*Bottom*: Same parameter space but using contours for Type 1 and Type 2 AGN. The plot shows that the distributions cover the same region of the parameter space.

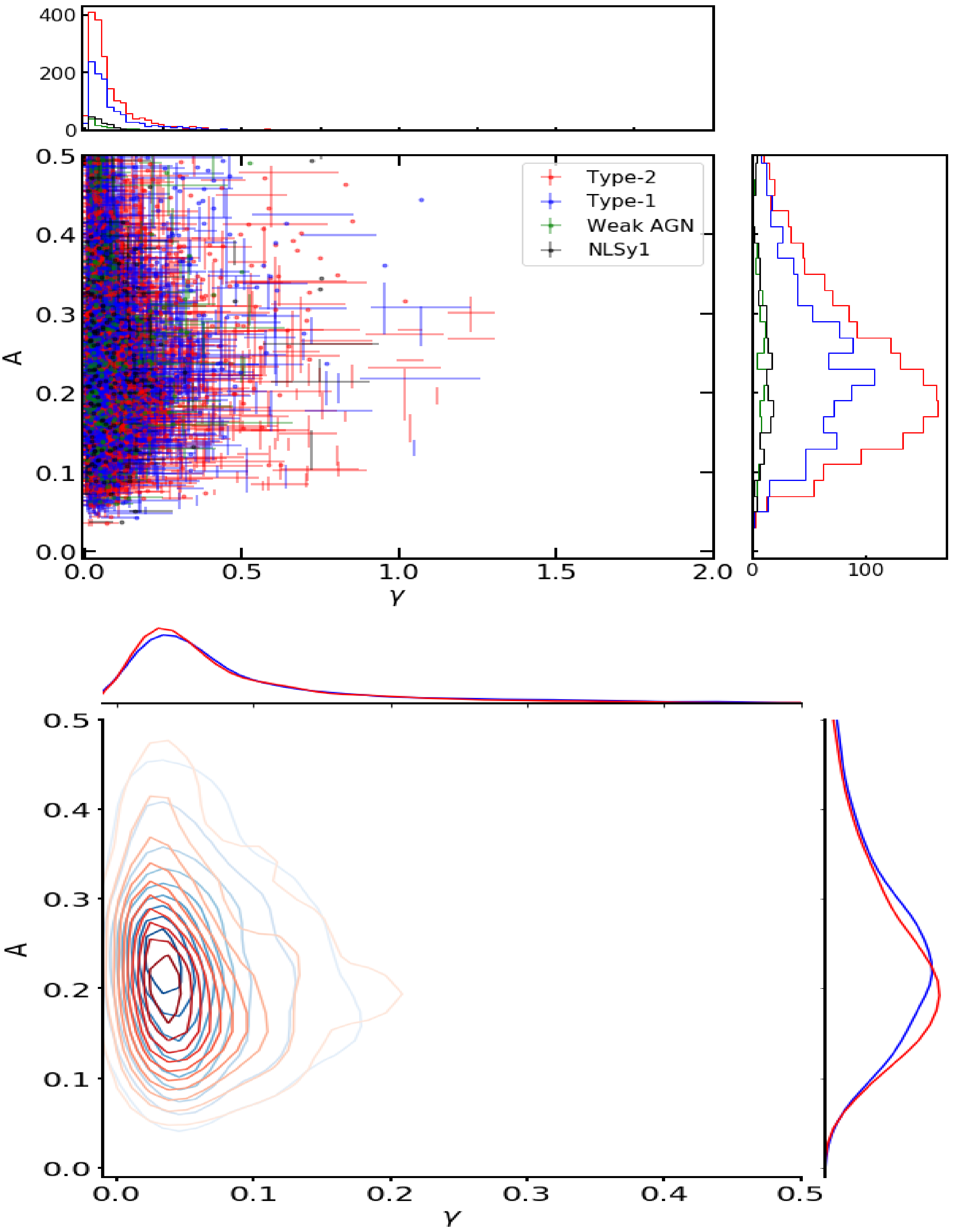


Figure 4.7: *Top*: A vs γ parameter space for GOTO GOTOphoto processed data. In this case of this I have also removed the data points with errors larger than 0.2 for either A or γ. This reduces the sample from a total of 3,434 objects to 3,245. Again for the GOTO data, we find that we are unable to distinguish the Type 1 from the Type 2 objects. *Bottom*: Same parameter space but using contours for Type 1 and Type 2 AGN. The plot shows that the distributions cover the same region of the parameter space.

LSST and GOTOphoto processed data suggests that the problem likely does not lie in one or the other pipeline, but may be intrinsic to GOTO itself, or it may perhaps be telling us that it is much harder to conduct variability studies of the predominately moderate-luminosity AGNs considered in this study.

To investigate this, I turned to the ZTF data to compare the GOTO results with those extracted using the ZTF lightcurves. The studied GOTO and the ZTF data cover different periods (i.e the GOTO data cover the period from 2019-02-22 to 2020-06-30 whereas ZTF data cover the period from 2018-03-17 to 2019-12-31), and even for the overlapped regions data are collected on different nights and with different sampling, due to the different cadences of the different telescopes at the time when their respective data were collected. While GOTO and ZTF data cover different periods and have different sampling rates, in a statistical analysis such as this the expected outcome should be the same. i.e. that Type 1 AGN should show clear signs of being more variable than Type 2 AGN. In contrast to the results derived from GOTO data (Figure 4.6, Figure 4.7), the results derived from the ZTF data show a clearer distinction in the A and γ values between Type 1 and Type 2 AGN. This difference in the γ parameters values between Type 1 and Type 2 AGNs can also be seen in the histogram plots in Figure 4.8. First, in the scatter plot, we see that the data points for the Type 1 AGN tend to cover a different region than that covered by the Type 2 AGN, whereas on the respective plot with GOTO data we showed that the data points for the two different Type 1 and Type 2 AGN overlapped. From the histograms on Figure 4.8 we can again see the differences, especially in the case of the γ distributions, of the different types which can be quantified in the parameters of the distributions.

From the ZTF data, we find that the amplitude A has a mean value of 0.23, a median of 0.20 and a standard deviation of 0.13 for the Type 1 AGN, whereas for Type 2 AGN it has a mean of 0.18, a median of 0.17 and a standard deviation of 0.07. The NLSy1 galaxies and those with a weak presence of an AGN are less well-represented in the ZTF

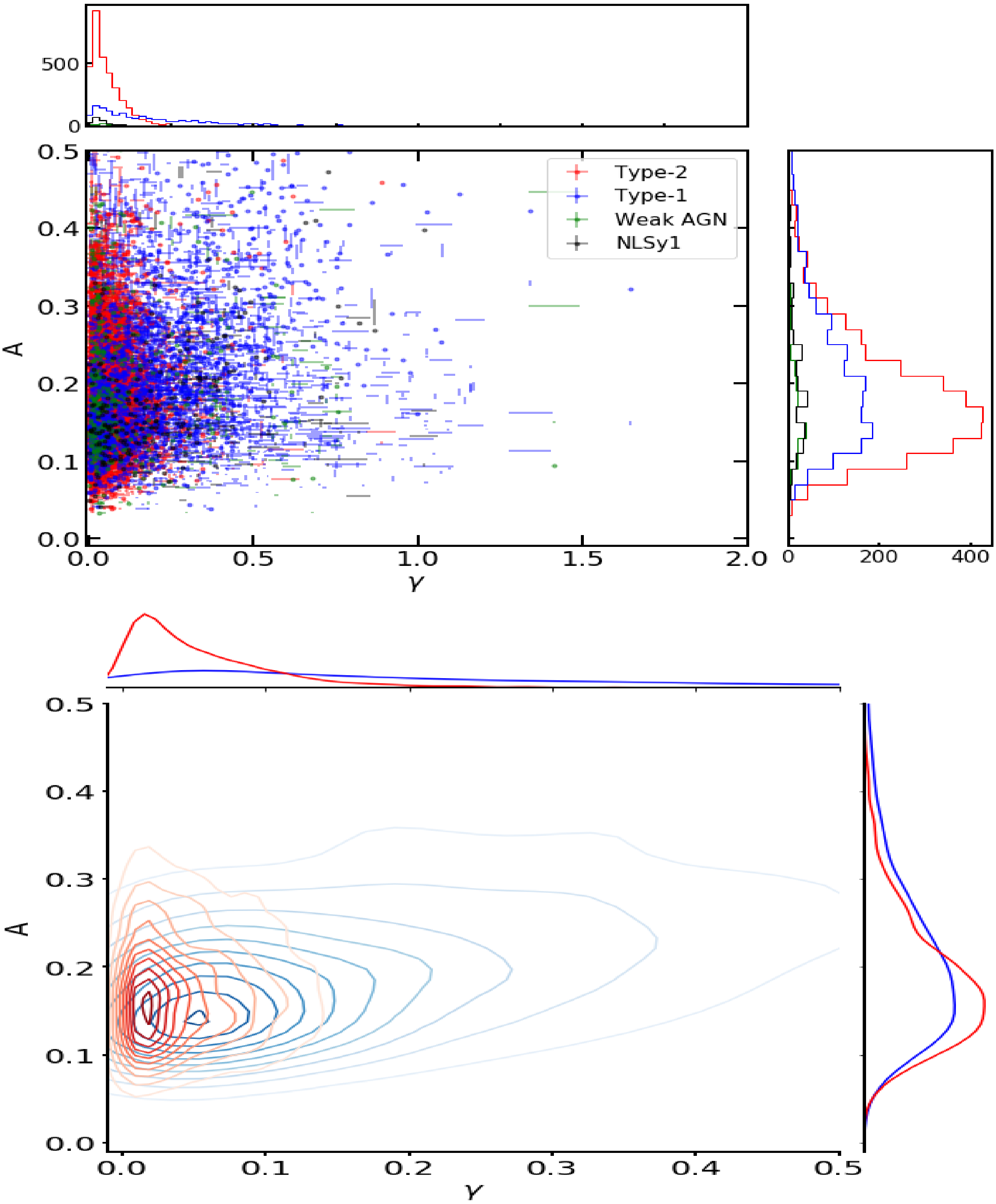


Figure 4.8: *Top*: A vs γ parameter space from the ZTF g-band lightcurves and the respective histograms for each parameter. For the ZTF data, we find that removing the points with larger than 0.2 error bars for either A or γ reduces the sample by just 20 objects i.e. from 5,646 to 5,626. Also, there are, in general, smaller uncertainties in the fitted parameters for ZTF than those for the respective GOTO plots (Figure 4.6, Figure 4.7) which indicate better A and γ estimates from the ZTF data. For the ZTF data, there is a clearer difference between the Type 1 and Type 2 values especially in the case of the γ parameter. *Bottom*: Same parameter space but using contours for Type 1 and Type 2 AGN. Unlike with the GOTO data, using the ZTF data we see a distinction between Type 1 and Type 2 AGN on the A vs. γ parameter space.

sample, but we report that the amplitude A for the former has a mean of 0.21, a median of 0.17 and a standard deviation 0.19. For the latter, it has a mean of 0.21, a median of 0.18 and a standard deviation of 0.11. I will further discuss the ZTF results in Chapter 5, but we note at this point that the expected value of A for Type 1 AGNs based on previous studies is around 0.2 mag (e.g., MacLeod et al. 2012, Sánchez-Sáez et al. 2019). In terms of γ, from the ZTF data, we find that Type 1 AGN have a mean value of 0.22, a median of 0.15 and a standard deviation of 0.22, whereas the respective numbers for Type2 AGN they are 0.06, 0.04 and 0.07, for NLSy1 are 0.19, 0.14, 0.19 and for weak-AGN they are 0.12, 0.04 and 0.17. For the γ values, we recall that, if the Type 1 AGN lightcurves indeed follow a DRW model, then the γ values should be close to 0.5. I will explore this further in section 5.1.

The outcome from the ZTF lightcurves will be discussed in the next chapter, but at this point we need to understand the reason for the discrepancy between the GOTO and the ZTF data. As mentioned previously, as both data cover a period longer than 1 year, the fits to the structure function should give similar values for the GOTO and the ZTF data. First, we note that the fitting errors for GOTO (for both pipelines) are larger than those of ZTF and also show a strong anti-correlation with the number of epochs; they have a Spearman coefficient of -0.25 for the γ and -0.51 for the A errors in the case of GOTOphoto. The numbers for GOTO LSST stack processed data are similar to the ones for the GOTOphoto ones. Although we keep in mind the difference in the errors, we first look at the numbers for A and γ for the GOTO data. The A values from GOTO AGN Type 1 lightcurves generated by GOTOphoto and LSST stack have a mean of 0.26 (0.30), a median of 0.24 (0.26) and a standard deviation of 0.13 (0.20) whereas the respective numbers for Type 2 AGN they are 0.25(0.24), 0.22 (0.21), 0.14 (0.14). The γ values for Type 1 AGN have a mean of 0.10 (0.18), a median of 0.06 (0.11) and a standard deviation of 0.12 (0.21) whereas the values for Type 2 AGN they are 0.10 (0.11), 0.05 (0.07) and 0.13 (0.12). We see that for the GOTO data, the statistics of the distributions of the Type 1 and the Type 2 AGN do not show a significant difference, as we see with the ZTF data. Despite for GOTO LSST stack we see some difference with the values approaching the numbers from ZTF which could indicate that the PSF photometry used in the case of GOTO LSST stack improves things, the larger errors do not allow us to draw any safe conclusions. I summarise the results in Table 4.2.

	Amplitude (A)		Logarithmic Gradient (γ)	
Survey-Pipeline	Type 1	Type 2	Type 1	Type 2
ZTF	mean = 0.23 median = 0.20 std = 0.13	mean = 0.18 median = 0.17 std = 0.07	mean = 0.22 median = 0.17 std = 0.22	mean = 0.06 median = 0.04 std = 0.07
GOTO GOTOphoto	mean = 0.26 median = 0.24 std = 0.13	mean = 0.25 median = 0.22 std = 0.14	mean = 0.10 median = 0.06 std = 0.12	mean = 0.10 median = 0.05 std = 0.13
GOTO LSST stack	mean = 0.30 median = 0.26 std = 0.20	mean = 0.24 median = 0.21 std = 0.14	mean = 0.18 median = 0.11 std = 0.21	mean = 0.11 median = 0.07 std = 0.12

Table 4.2: Table summarising the statistical values for the amplitude (A) and the logarithmic slope (γ) generated from the fitting of the SF for the different samples i.e. ZTF g-band lightcurves, GOTO data processed with GOTOphoto and GOTO data processed with LSST stack. We note that although the GOTO data processed with the LSST stack shows differences between the Type 1 and Type 2 objects especially for the γ values, the larger errors (Figure 4.6) forbid us from drawing safe conclusions. It is, however, an indication that PSF photometry is a better option for sources with fainter apparent magnitude closer to the magnitude limit.

For ZTF data, we saw that in terms of the amplitude values, Type 2 objects had a slightly lower mean and median and more significantly a lower standard deviation than those of Type 1 AGN. By contrast, we find that the amplitudes, A, derived from GOTO data cover a much broader range of values, irrespective of AGN type. This is a first indication of high RMS values for the magnitude change. In terms of γ values, for the ZTF data we see a clear difference in the distributions of Type 1 and Type 2 AGN which is mirrored in its mean, median and standard deviation. This means that while Type 2 AGN have lower γ values that indicate a flatter SF function. The respective values for the Type 1 objects deviated from zero with a mean of 0.22 compared to the 0.06 of the Type 2 objects, demonstrating that for Type 1 AGN, the RMS increases with time lag. This result is expected from the nature of the two types of objects, whereby, due to orientation effects, Type 2 AGN are not expected to show signs of variability, whereas Type 1 AGN are.

4.6.1 Tests with focusing on the common sources between the GOTO and ZTF samples

In this subsection, I investigate the discrepancy between the GOTO and ZTF data focusing on the common sources between GOTO (for the two different pipelines) and ZTF. The reason I choose to plot both the pipeline results against the ZTF data is that the GOTO pipelines used here are not consistent with one another. Further investigation of this discrepancy will be needed as these two pipelines develop but potential reasons could be: the different methods I have used to clean the lightcurves from the two pipelines and that GOTOphoto uses aperture photometry, whereas the LSST stack uses PSF photometry for the majority of the sources.

Before investigating this discrepancy, we look into the effects that the rest frame time length and the number of epochs have on the A and γ values measured from the GOTO data. In Figure 4.9 it is shown that A and γ from GOTOphoto do not show any significant correlation or anti-correlation with the lightcurve properties (i.e., length and number of epochs). The respective plot for the GOTO data processed with the LSST stack is presented in Figure (4.10). Using the Spearman method, we find a more significant anti-correlation between the amplitude, A, and rest frame time length (ρ_s = -0.21), which means that shorter lightcurves give higher amplitudes. This anti-correlation, however, is true for both AGN types so there should be additional reasons for the GOTO results i.e. the non-detectable difference between the A and γ parameters for Type 1 and Type 2 AGN. I investigate the effects of the cadence and the lightcurve length using simulated data in section 4.6.3.

To demonstrate this discrepancy between GOTO and ZTF data, I plotted A from GOTO data vs. A from ZTF data, and γ from GOTO vs. γ from ZTF, where, ideally, the same parameters measured from different datasets should lie along a 1:1 line. Figure 4.11, however, shows that there is a large scatter in these plots.

In Figures 4.12, we present the SF parameter space, but this time focusing on matches between ZTF objects and GOTOphoto and ZTF and GOTO LSST stack sources, which lead to similar conclusions to those from the whole ZTF, GOTOphoto and GOTO LSST stack sample. In this case, we compare the matched ZTF data with the GOTOphoto and GOTO LSST stack respectively. The problems related to the distributions of A and γ for the Type 1 and Type 2 AGN from GOTO data are clear, there being no difference for these parameters between the two types. However, there is a clear difference for these parameters between Type 1 and Type 2 AGN in the ZTF data, I summarise the results for the two matched samples in Table 4.3.

Matched Sample	Survey	Amplitude (A)		Logarithmic Gradient (γ)	
		Type 1	Type 2	Type 1	Type 2
ZTF-GOTOphoto	ZTF	mean = 0.23 median = 0.20 std = 0.11	mean = 0.18 median = 0.17 std = 0.07	mean = 0.21 median = 0.16 std = 0.19	mean = 0.05 median = 0.03 std = 0.06
	GOTO GOTOphoto	mean = 0.28 median = 0.24 std = 0.15	mean = 0.26 median = 0.23 std = 0.16	mean = 0.11 median = 0.06 std = 0.13	mean = 0.11 median = 0.05 std = 0.15
ZTF-LSST stack	ZTF	mean = 0.23 median = 0.19 std = 0.12	mean = 0.19 median = 0.18 std = 0.07	mean = 0.21 median = 0.14 std = 0.21	mean = 0.06 median = 0.04 std = 0.06
	GOTO LSST stack	mean = 0.27 median = 0.23 std = 0.16	mean = 0.23 median = 0.21 std = 0.14	mean = 0.17 median = 0.14 std = 0.20	mean = 0.11 median = 0.07 std = 0.13

Table 4.3: The table summarises the statistical values for the amplitude A and the logarithmic slope γ, generated by fitting the SF, for the two samples of matched data; i.e., the matched data between ZTF sample with g-band lightcurves and GOTOphoto processed data and the matched data between ZTF and the LSST stack processed data.

The discrepancy of A and γ values between GOTO and ZTF, which, even for Type 2 AGN objects, show a large range of values, led us to look at the SF of individual AGN in each sample. Figure 4.13 shows the SF (calculated for ZTF and GOTO data processed

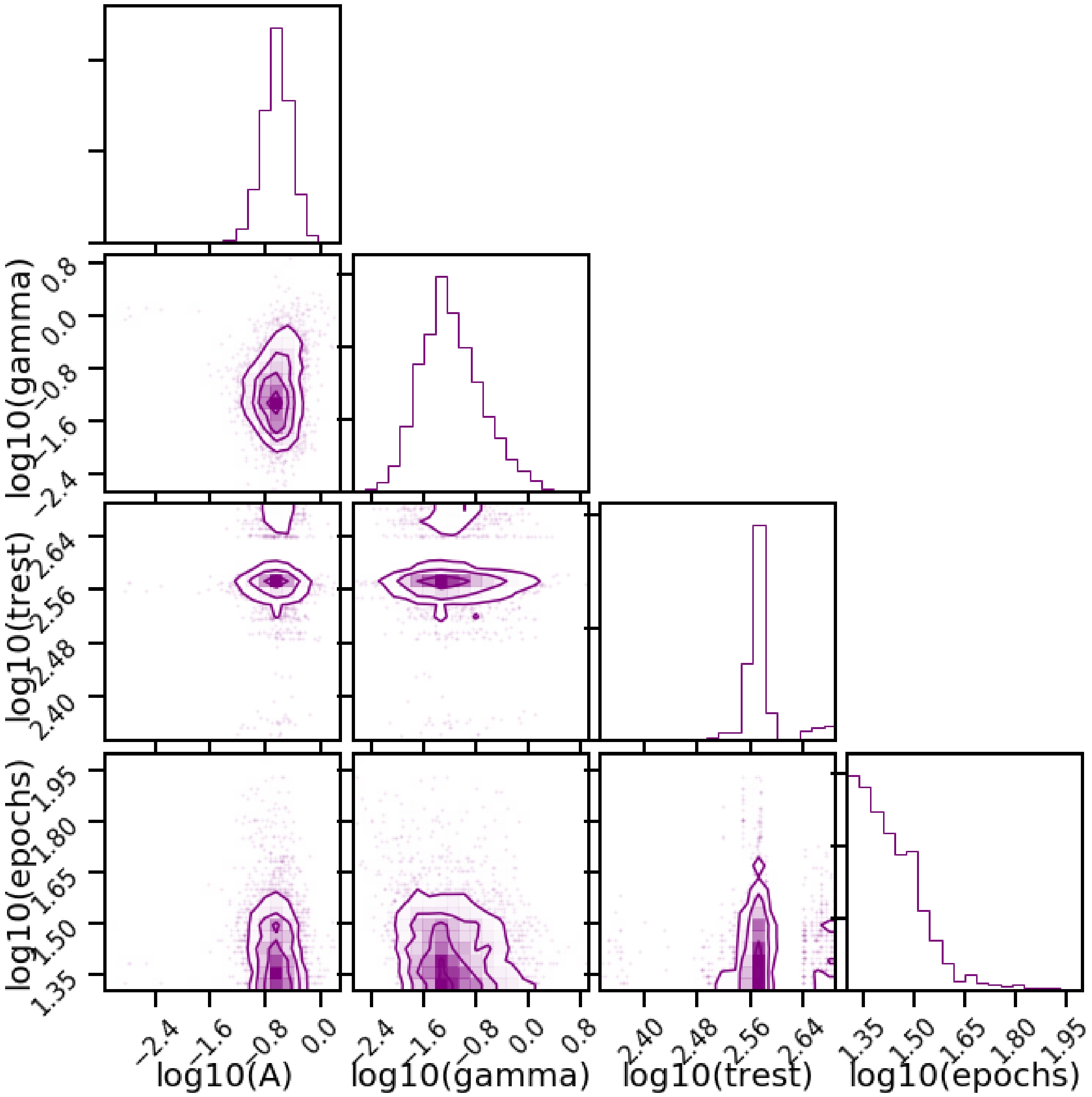


Figure 4.9: Dependence of A and γ from the rest frame time length and number of epochs for the values extracted from GOTOphoto lightcurves. There are no significant correlations of the A and γ parameters with the lightcurve properties.

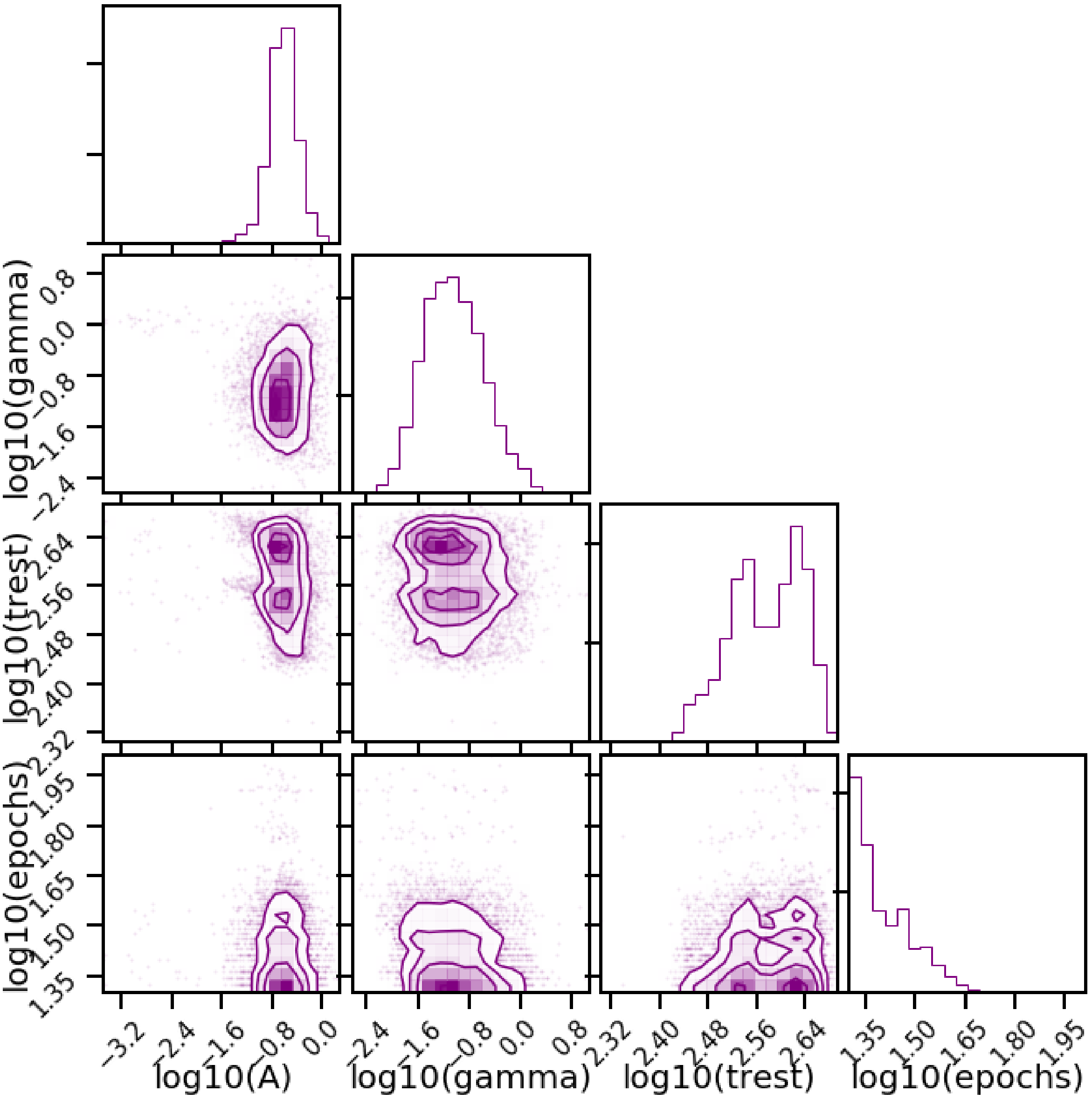


Figure 4.10: Dependence of A and γ from the rest frame time length and number of epochs for the values extracted from GOTO LSST stack lightcurves. In this case there is an anti-correlation between the amplitude A and the rest-frame time-length of the lightcurves for the whole sample, which is also true for both types of AGN (i.e. Type 1 and Type 2 AGN).

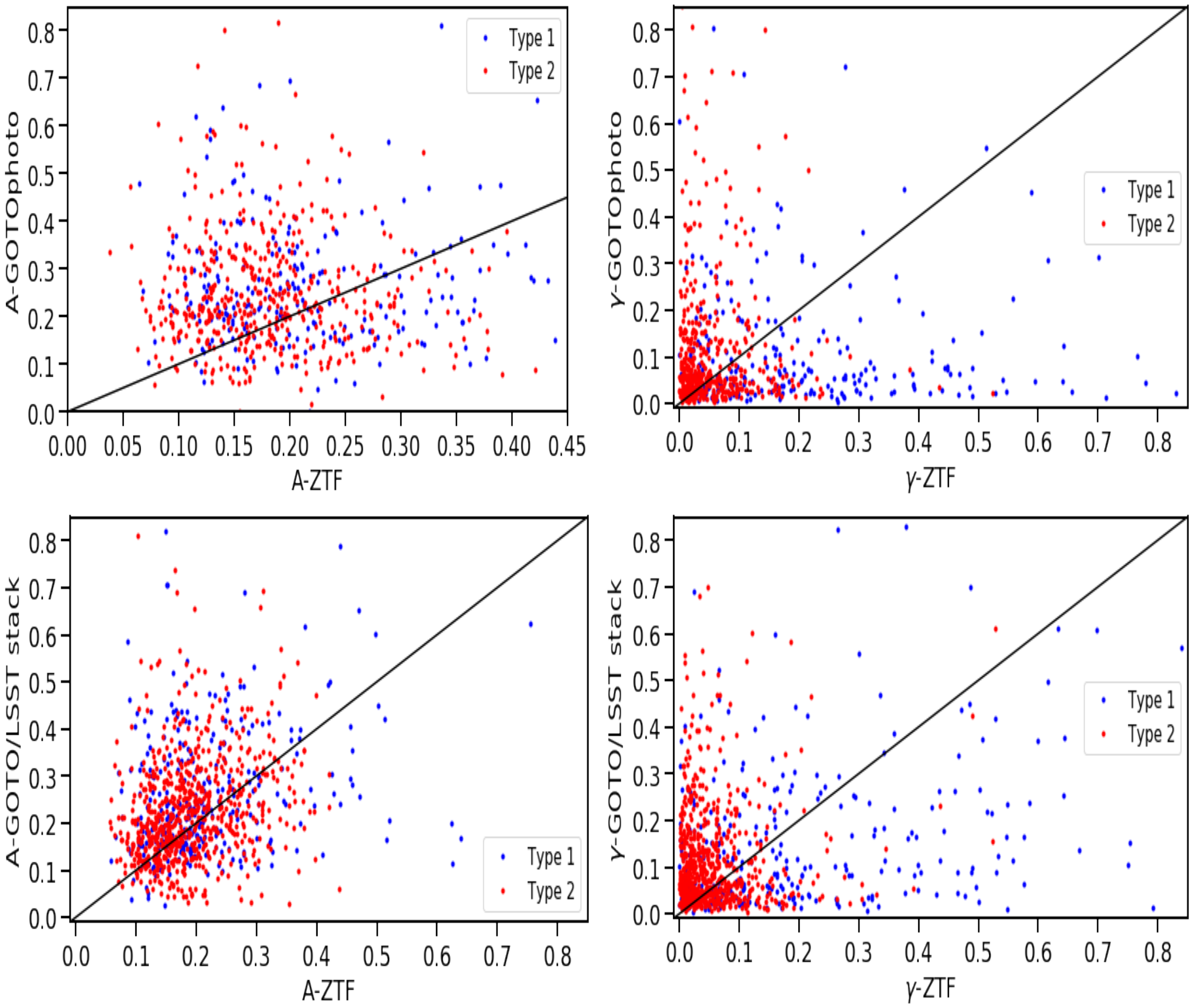


Figure 4.11: Top panels: A vs. A and γ vs. γ for matched objects from ZTF and GOTOphoto. We find that there is a large scatter for both amplitude and γ values when we compare the GOTO data, processed with GOTOphoto, and the ZTF data. It is especially obvious in the case of γ values how these values for ZTF cover a larger range than these from GOTO. Bottom panels: A vs. A and γ vs. γ for matched objects from ZTF and GOTO data processed with LSST stack. From these plots we draw similar conclusions as in the case of GOTOphoto processed data

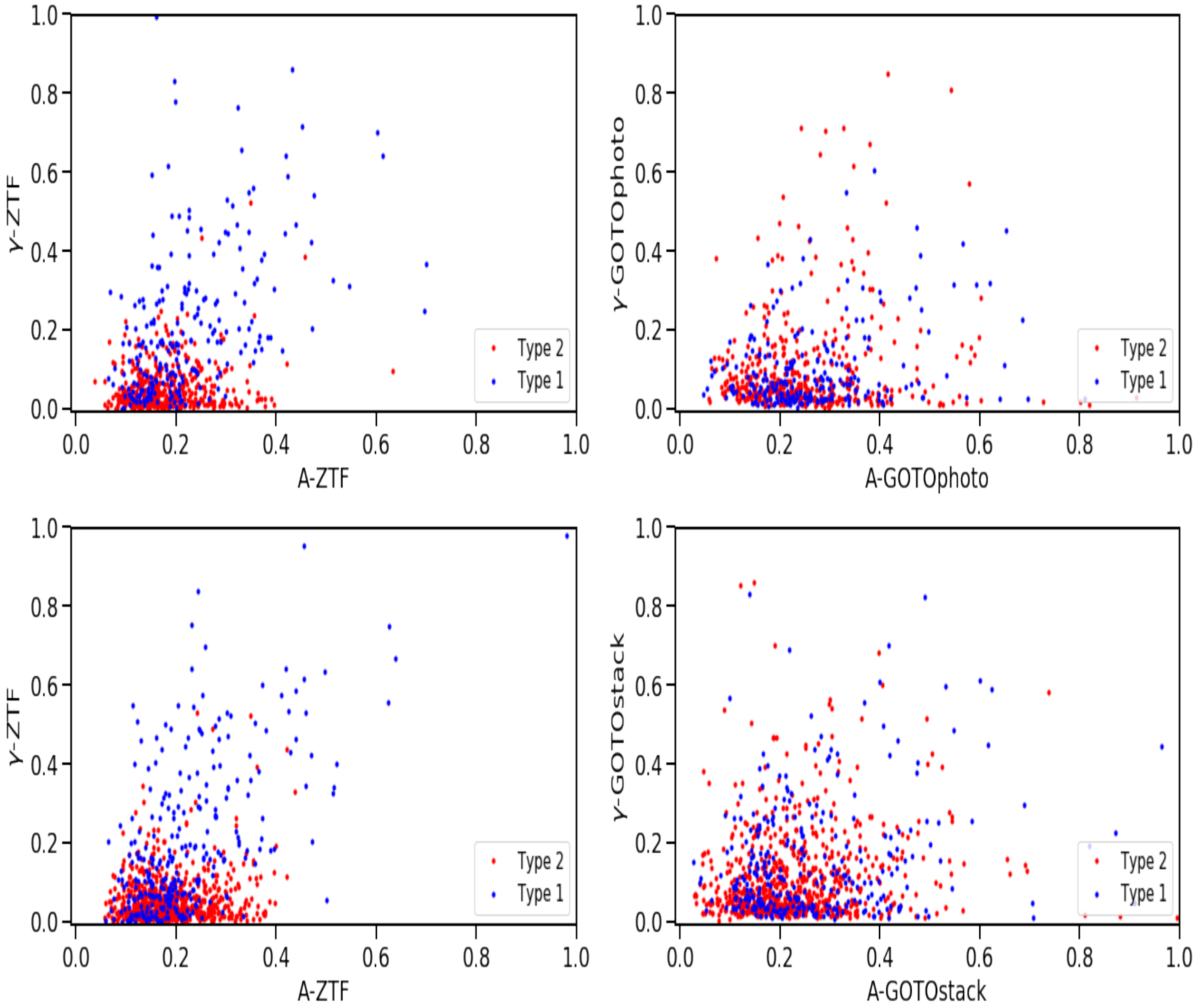


Figure 4.12: Top panels: γ vs. A plots for matched objects from ZTF and GOTOphoto. In the left plot, we show the data points from the ZTF lightcurves and at the right these from GOTOphoto. Right panels: γ vs. A plots for matched objects from ZTF and GOTO/ LSST stack. In the left plot we show the data points from the ZTF lightcurves and at the right these from GOTO/ LSST stack.

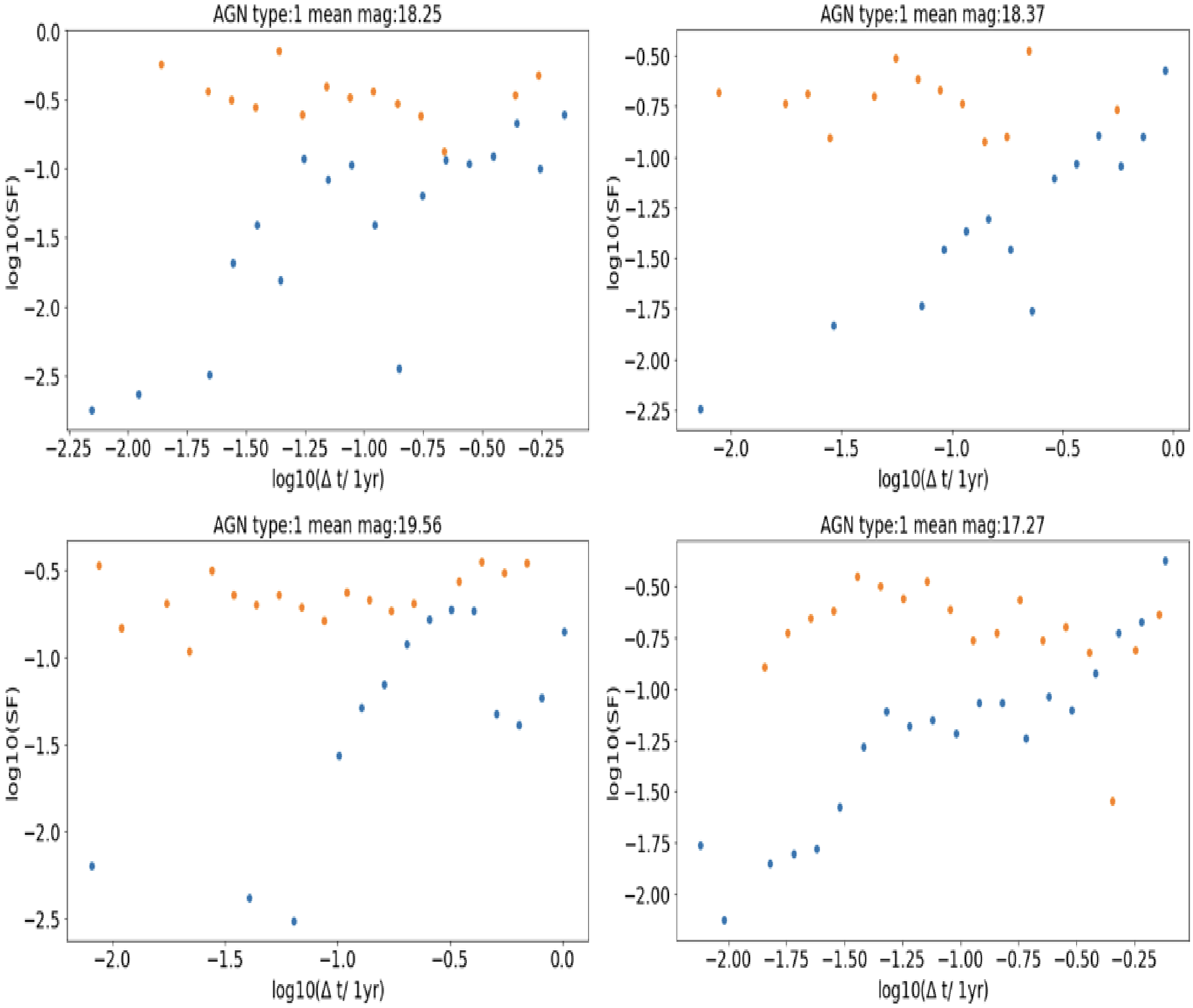


Figure 4.13: SF comparison from ZTF and GOTOphoto lightcurves for four different Type 1 AGN, where blue is the SF from the ZTF llightcurves and the orange data points represent the SF of the GOTOphoto lightcurves. The results are similar for GOTO LSST stack lightcurves.

with GOTOphoto) versus the time lag for four different Type 1 AGN. The plot shows that the shape of the respective SF from the ZTF and the GOTO data is very different in terms of the logarithmic slope i.e., GOTO is flat, whereas ZTF shows a SF increase for larger time lags. Inspection of these SFs shows that those derived from GOTO data have high RMS values at short time lags, unlike those from the ZTF data. We note that although the SF definition takes into account the photometric errors, the GOTO SF is still unable to track the variability, as a function of time lag. It is also important to note that the short-timescale data suffer from low-statistics due to the current cadence of GOTO. The flatness of the SF is mirrored in the γ results we get from the GOTO data for Type 1 AGN, which are inconsistent with the expected outcome; i.e., that they should be typically larger than the γ values of Type 2 AGN.

4.6.2 Tests with downgraded ZTF data

In order to investigate the reason(s) why the GOTO data are not able to reproduce the ZTF results but also the effect of the RMS due to systematics on the GOTO data, I attempted to downgrade the ZTF data in different ways. The ZTF data differ from the GOTO data in cadence and timespan (Figure 4.4; top panel) but also in photometric quality. In order to test whether I could reproduce the GOTO results using the ZTF data, I downgraded the ZTF data based on the aforementioned differences step-by-step. Each time, I tested whether the difference in Type 1 and Type 2 objects for the ZTF data disappeared as in the case GOTO data. First, I reduced the length of the ZTF lightcurves to resemble that from GOTO (i.e., I did not allow the lightcurves to be longer than one year in rest-frame) but this had no effect on the ZTF results i.e., the difference between Type 1 and Type 2 objects was still identifiable. Then, I reduced the sampling of the ZTF lightcurves i.e., keeping a random half of the epochs from the initial lightcurve in those cases where they had more than 50 epochs. Reducing the epochs, had no effect on the results, as well.

Then I attempted to deviate the ZTF photometry by a random value chosen from a normal distribution centered at 0 and variance equal to the flux uncertainty of a "typical" (i.e., median of non-variable sources as a function of magnitude) GOTO source of similar magnitude. The results of this was that the A vs. γ plot for the downgraded ZTF data look more like that obtained from GOTO data (Figure 4.14). We note that, when calculating the SF of each source, I have combined the ZTF photometric errors with the (added) GOTO errors in quadrature. The fact that introducing the RMS of GOTO non-variable sources in the ZTF lightcurves have indeed affected the results, in such a way that hides the difference in distributions between Type 1 and Type 2 objects is some confirmation that with the current GOTO data we are not able to study AGN variability of moderate luminosity AGN using the SF. The the current GOTO uncertainties seem to lead to large RMS in the magnitude difference even for short timescales.

The systematic errors associated with the GOTO data may result from two main sources i.e. instrumental issues and data handling (analysis software issues). Even though GOTO had a major upgrade in February 2019 there were still hardware updates throughout period. Despite this, it is still thought that there are problems related to the mirrors, in the prototype phase. The new design of the future GOTO UTs involves a change of the Optical Telescope Assembly (OTA) manufacturer (from APM to ASA) which we hope it will resolve these problems. Also, issues with the processing of the data have been identified. Since GOTO is a new survey, the in-house and LSST stack pipelines still need to fully address issues that have been reported by the various science working groups within the collaboration. Both pipelines need to better handle the photometric calibration of the data and develop a clearer and more efficient way of flagging bad data. Bad photometric calibration and general use of bad photometric data points introduce systematic errors in the lightcurves. We speculate that addressing the aforementioned issues will drastically improve the the prospects of using GOTO data for

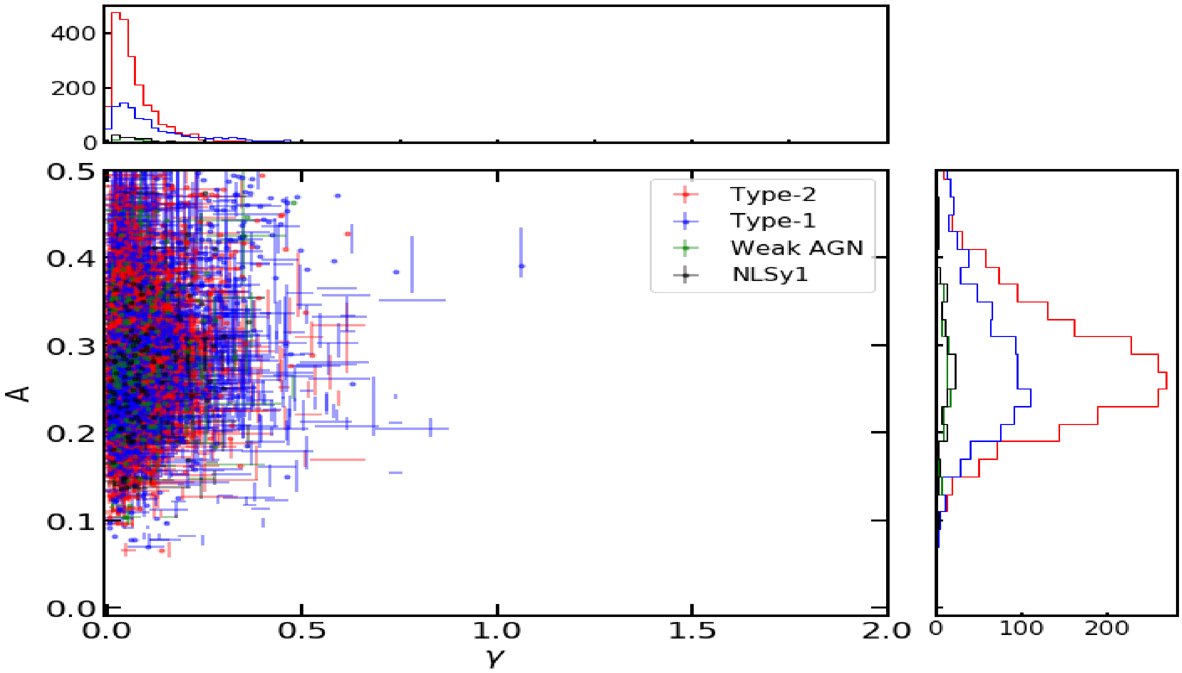


Figure 4.14: A vs γ plot using the parameters extracted from the SF of the down-graded ZTF lightcurves. We see that the difference between Type 1 and Type 2 objects found in Figure 4.8 cannot be identified after introducing the GOTO RMS into the ZTF lightcurves.

scientific analysis, but further investigation of the future data will be needed to confirm this.

4.6.3 Testing the effect of cadence and time length of GOTO lightcurves on A and γ using simulated lightcurves

In this subsection, as a final test on the A and γ parameters, I aim to investigate how higher cadence and longer lightcurves will affect the A and γ distributions. In order to do that, I have generated Type 1 AGN lightcurves using the DRW model. I simulate lightcurves for three different cases: short lightcurves (~400 days) with a cadence typical of this of GOTO, long lightcurves (~700 days) with a cadence typical of GOTO and finally long lightcurves (~700 days) with a cadence typical of ZTF. As we mentioned in section 4.2.3, to simulate the lightcurves we need the characteristic timescale (τ, the amplitude of variability for $t >> \tau$ (SF_∞) and the mean magnitude. I simulate for each case (i.e., short-low cadence, long-low cadence, long-high cadence) 1000 lightcurves with characteristic timescales of 300 days and SF_∞ equal to 0.2 mag. I also varied the magnitude measurements according to the levels of uncertainty typical of GOTO for the respective magnitude. Figure 4.15 shows the results, after calculating the SF parameters for the different lightcurves, using Eqn. 4.6 and fitting them with the model described in Eqn. 4.9. The mean, median and standard deviation of the amplitude A and the logarithmic slope γ are summarised in Table 4.4. We find that both in terms of parameter distributions (tighter distribution in the case of long lightcurves) and the fitting errors (smaller in the case of long lightcurves), the time length of the lightcurves is more important than the cadence of the lightcurves. This also indicates that for longer GOTO lightcurves, it might be possible to overcome some of the existing issues (e.g., inability to distinguish between Type 1 and Type 2 in terms of variability). For this, however, further tests will be needed as more GOTO data become available.

Sample	Amplitude (A)	Logarithmic Gradient (γ)
Short-Low cadence	mean = 0.18 median = 0.15 std = 0.11	mean = 0.42 median = 0.36 std = 0.33
Long-low cadence	mean = 0.17 median = 0.15 std = 0.08	mean = 0.42 median = 0.38 std = 0.28
Long-High cadence	mean = 0.17 median = 0.16 std = 0.06	mean = 0.43 median = 0.40 std = 0.21

Table 4.4: The table summarises the statistical values for the amplitude A and the logarithmic slope γ generated from by fitting the SF for short-low cadence, long-low cadence and long-high cadence lightucrves.

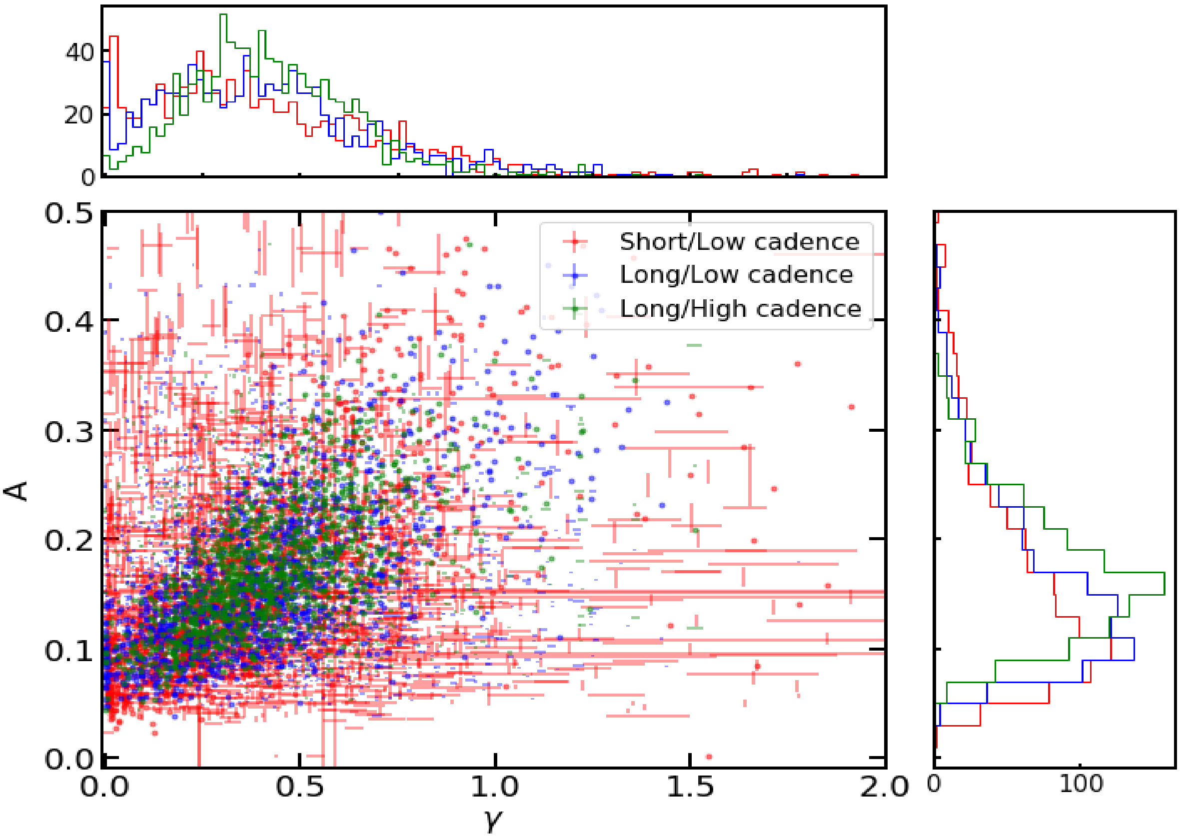


Figure 4.15: The A versus γ for simulated lightcurves with DRW for $\tau = 300$ days and $\mathrm{SF}_{\infty} = 0.2$ mag for different time lengths and cadences. The short lightcurves time length is equal to ∼400 days, while long are lightcurves with time length ∼700 days. Low and high cadence are typical current GOTO and typical ZTF respectively. The plot shows that the time length of the lightcurves is more important for the SF fit than the cadence.

4.7 Investigation of σ_{rms}

In the previous section, I explained how I used the SF to extract the A and the γ parameters after fitting the power law model of Eqn. 4.9 for the GOTO lightcurves and I also presented the issues that do not allow the detection of the variability of moderate luminosity AGN using current GOTO data. In this section, we focus on another available tool to measure the AGN variability, i.e. the excess variance (σ_{rms}).

At this stage, as I mentioned, the GOTO lightcurves have rest-frame time length of ∼1 year which means that the amplitude A parameter will be strongly correlated with the σ_{rms} value we introduced in section 4.2.1. The σ_{rms} is another way to measure the amplitude of variability of a lightcurve which essentially is a measurement of the lightcurve RMS corrected for the photometric error and normalised for the mean magnitude. As both ZTF and GOTO have lightcurves of ∼1 year rest frame length, it is safe to assume that the σ_{rms} values should be similar especially if we do not normalise by the mean magnitude (i.e., not including the $1/\bar{x}$ in Eqn. 4.3 to avoid any differences in mean magnitude due to the different filters). I compare the ZTF σ_{rms} values with those from both GOTOphoto and GOTO LSST stack and present the results in Figure 4.16.

We find that in the majority of cases, the σ_{rms} values from GOTOphoto lightcurves are larger than those meseared from ZTF lightcurves . This is to be expected, given the findings described in the previous section, where I showed that the RMS of the difference of the magnitude from its mean is large for the current GOTO data for both pipelines. Additionally, the plot shows that the result is independent of the mean magnitude of the GOTOphoto lightcurves demonstrating that this is even an issue for the brightest AGN in our sample. For GOTO/ LSST stack processed data we find again that the sources have in general higher σ_{rms} values than those extracted from the ZTF lightcurves, however, we note that in this case a few of the brightest sources ($m_L < 16$) appear to have very low σ_{rms} values compared to those from ZTF. In Figure 4.17, I plot the σ_{rms} values

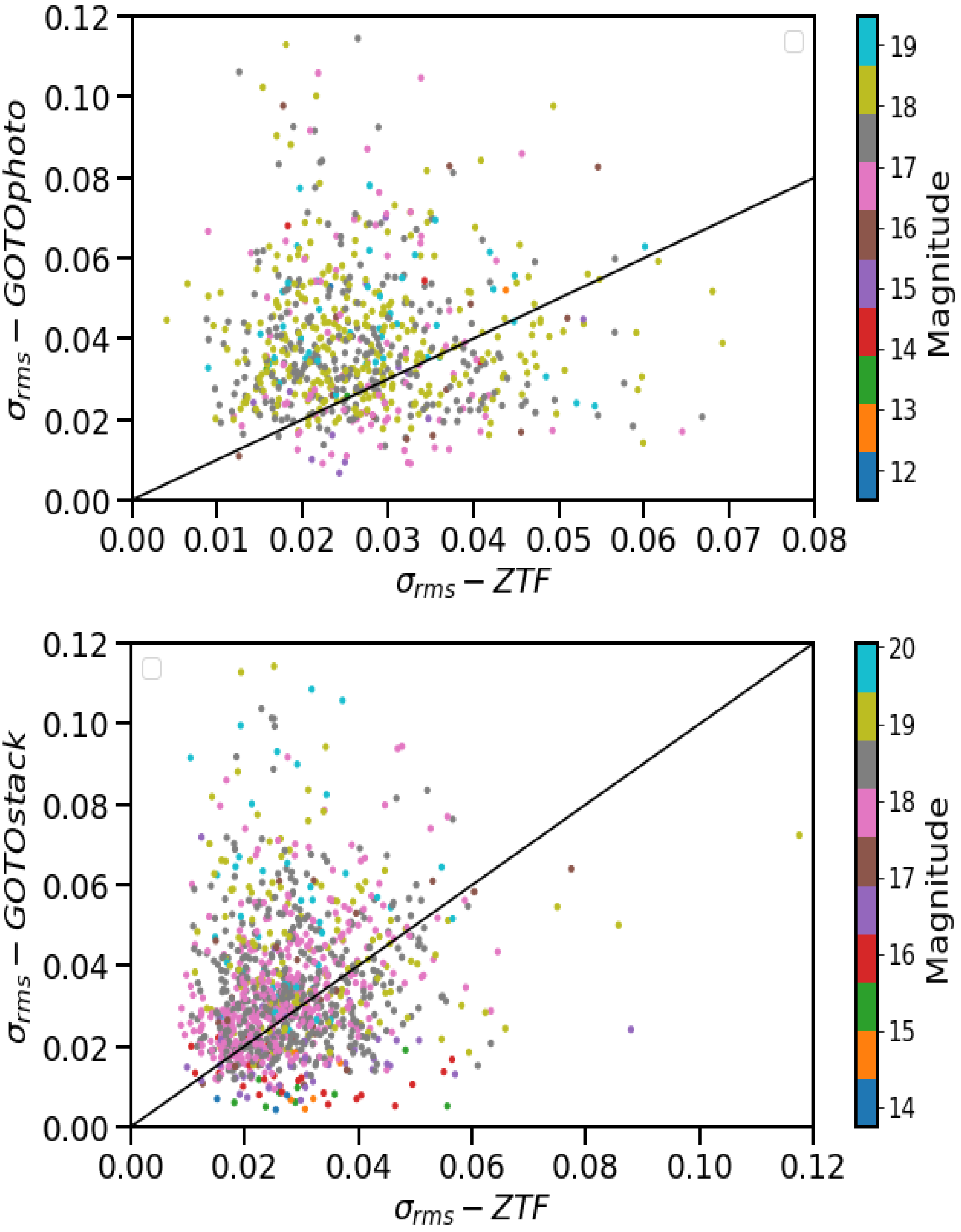


Figure 4.16: The σ_{rms} of the GOTO lightucurves generated by the GOTOphoto (top) and LSST stack (bottom) versus the σ_{rms} calculated from the ZTF lightcurves. There is slight evidence that the ZTF results for σ_{rms} are closer to the ones the GOTO data processed with LSST stack rather than those from GOTOphoto, especially for sources ∼18 mag. This is likely the result of the photometry method i.e. ZTF and LSST stack use PSF photometry for these objects, whereas GOTOphoto uses aperture photometry.

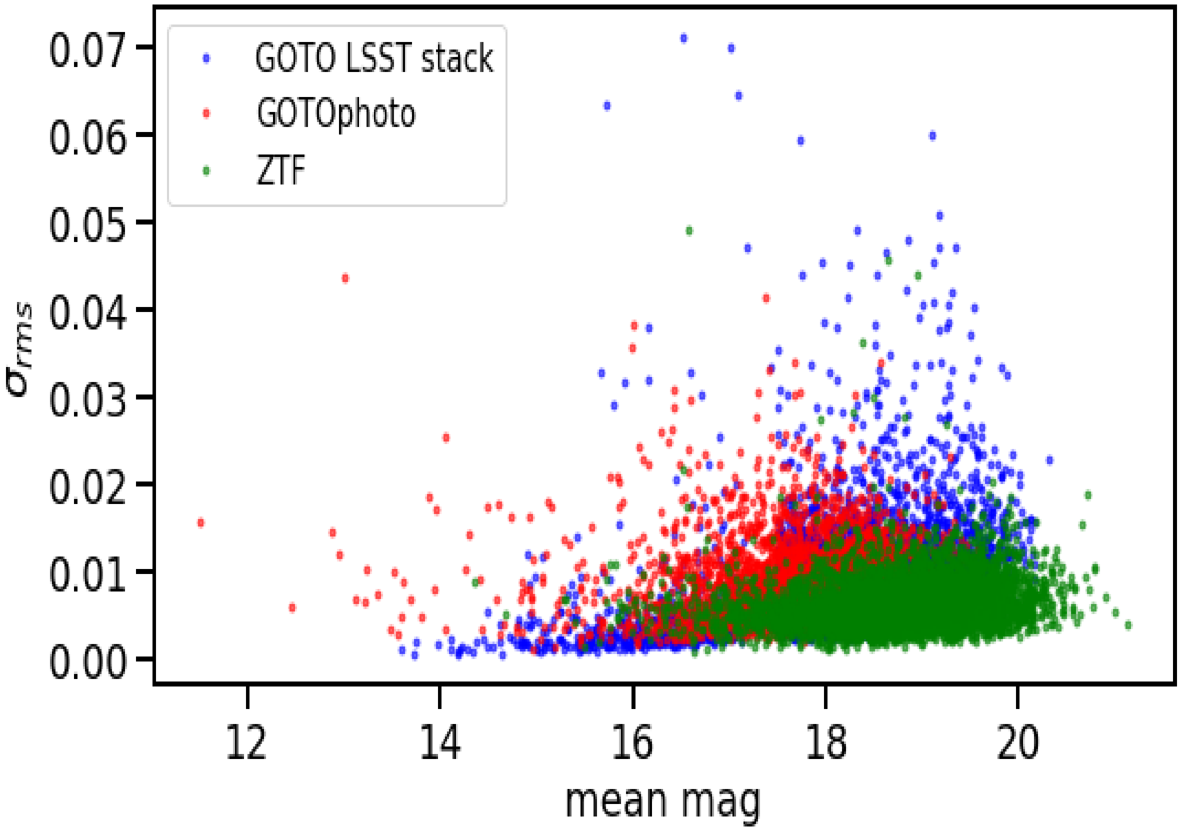


Figure 4.17: The σ_{rms} versus the mean magnitude for the GOTO (GOTOphoto and LSST stack) and ZTF results. We find that the across the whole range of magnitudes the σ_{rms} values of GOTO processed data (for both pipelines) are larger than these of ZTF but especially for sources fainter than 18th mag.

as a function of mean magnitude for the three different datasets (ZTF, GOTOphoto, GOTO LSST stack). This plot also indicates that there are large σ_{rms} values for GOTO, especially in the case of fainter sources, which makes the study of lower luminosity AGN harder. The σ_{rms} values are large even for brighter objects ($m_L < 18$), but since there are not many of such objects in our sample, it is not safe to draw any conclusions on them.

Since the σ_{rms} depends on the photometric error and because the photometric errors are not representative of the total error (i.e., as they do not fully include systematic errors), I now investigate whether we can solve the σ_{rms} discrepancy between the GOTO and ZTF data using a different approximation for the GOTO photometric errors. To do this, I re-calculate the photometric errors by calculating the standard deviation of known non-variable sources in the GOTO data. I performed the same analysis on GOTO data for both pipelines and for the non-variable sample I used the SDSS sample used in chapter 3 to evaluate the error bars. I used the robust-to-outlier standard deviation method described earlier in this chapter to calculate the RMS across the same timespan as the AGN lightcurves. I apply this technique to determine the RMS of both the GOTOphoto and GOTO LSST stack lightcurves. This change to the uncertainties reduces the number of variable sources obtained from both pipelines. Using these updated errors, we now have 4,263 variable sources from the GOTO LSST stack data (down from 5,802), of which 812 are matched with ZTF sources. From the GOTOphoto data, we now identify 2,375 sources as variable (down from 3,434) using these updated errors, of which 528 are matched in ZTF. The use, however, of the non-variable RMS as the error does not fix the discrepancy between the ZTF and GOTO data as we see in Figure 4.18. We therefore do not pursue this line of investigation any further.

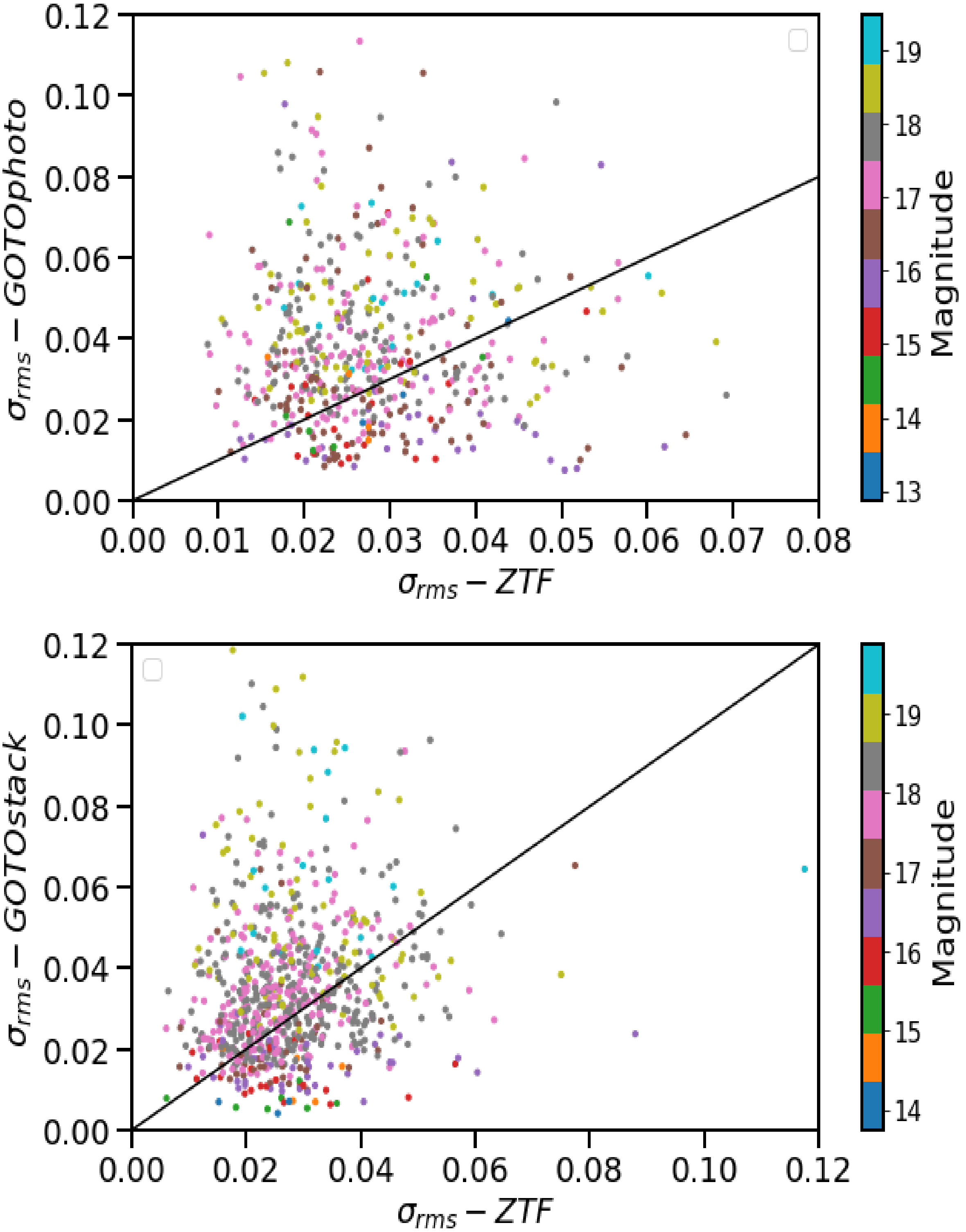


Figure 4.18: The σ_{rms} of the GOTO lightucurves generate the by the GOTOphoto (top) and LSST stack (bottom) using the RMS of non-variable sources to estimate the photometric errors versus the σ_{rms} calculated from the ZTF lightcurves. Again, we identify the same issues we explained based on Figure 4.16.

4.8 Chapter Conclusions

The two previous sections led us to the conclusion that GOTO data do not currently satisfy the criteria to be used as a means to study the variability of moderate luminosity and low-redshift AGN. I have tested the data generated by the two pipelines, the in-house developed GOTOphoto and the LSST software stack, which I have adapted for the GOTO telescope as part of this thesis. At present, the data generated by both pipelines suffer from systematic uncertainties which introduce the problems presented and analysed in the previous sections. The data processed with the LSST stack showed (when compared with ZTF data), however, that for these studies of sources that are close to the GOTO magnitude limit, PSF photometry is potentially a better choice than aperture photometry. In this chapter, I have presented the plots for the comparison between GOTO and ZTF, using the ZTF g-band lightcurves but the comparison of GOTO data with ZTF r-band lightcurves give very similar results.

It is important to note that we have not attempted the other two main ways presented in the literature to study AGN variability i.e. the DRW modelling of the AGN lightcurves and the PSD method. It is suggested from previous studies (e.g., Kozłowski 2017), however, that these methods are even more sensitive to bad systematics, sparse sampling and short lightcurves than the SF and the σ_{rms} methods. We conclude that the A, γ and σ_{rms} values measured by the GOTO lightcurves, obtained either by GOTOphoto or GOTO LSST stack, cannot be used at present to study AGN variability. In the next section, I use the ZTF data to investigate the variability of the sample presented in section 4.3, using the SF parameters and the σ_{rms} and I also examine the correlations of these parameters with physical parameters of the AGN (e.g., luminosity). I then compare the results with previous studies of higher luminosity and higher redshift AGN.

Chapter 5

The variability of moderate luminosity AGN with ZTF lightcurves

5.1 Chapter Introduction

In this section, we use the ZTF lightcurves to study the variability of moderate luminosity AGN. In section 4.6, we presented some results from fitting the SFs derived from the ZTF lightcurves. Now, I will investigate further the amplitude (A) and logarithmic slope (γ) parameters, which are generated from the SF fitting, as well as the σ_{rms} to evaluate the variability of these objects considering various caveats. I will also explore whether these parameters correlate with other key physical parameters of AGN (e.g. luminosity, Eddington ratio) and compare them to previous studies. It is important to understand the variability of these objects but also the causes of variability in moderate luminosity, low redshift AGN and compare them against results obtained for brighter AGN and/or AGN at higher redshifts. It is, however, also important to be aware of limitations with the data or analysis (e.g. short or not well sampled lightcurves) that may cause spurious correlations. In the previous chapter, where we compared the ZTF results with those from GOTO we presented the g-band lightcurves as the results (in terms of comparison of the parameters with those from GOTO) were similar for both ZTF bands. In this chapter, we also consider the ZTF r-band lightcurves and compare the AGN variability results from these longer wavelengths to those derived from the g-band measurements.

5.2 AGN variability parameters

In this section, we compare the results of the AGN variability parameters from the g-band analysis with those from previous studies. These earlier studies, which will be described further in this section and the following, focus on brighter AGN which are typically at higher redshifts. In section 4.6, we extracted the SF parameters for 1 year timescale after calculating the SF from the g-band ZTF lightcurves. Figure 4.8 shows the different distributions of A and γ for the different types of AGN in the sample i.e., Type 1 (=BLSy1), Type 2, NLSy1 and LINERS. We recall that this is the classification from Mullaney et al. (2013) and so it is possible that a few of these objects may belong to Changing-Look AGN. The plot shows the differences between the different types of objects in both the scatter and histogram plots. As discussed in section 4.6, the amplitude A has a mean value of 0.23, a median of 0.20 and a standard deviation of 0.13 for the Type 1 AGN whereas for Type 2 AGN it has mean value of 0.18, a median 0.17 of and a standard deviation of 0.07. For NLSy1 galaxies and these with a weak presence of an AGN, we find that the amplitude A for the former has a mean of 0.21, a median of 0.17 and a standard deviation of 0.19 and the for the latter, it has a mean of 0.21, a median of 0.18 and a standard deviation of 0.11. Even with generally short lightcurves (i.e., ∼1 year in the restframe), we find that the amplitude of Type 1 AGN is larger than those of Type 2 AGN and it is also close to the values reported for bright quasars (i.e., 0.2 mag; e.g., MacLeod et al. 2012 or from Sánchez-Sáez et al. 2018). For σ_{rms}, which is another way to estimate the amplitude of variability, its mean and standard deviation for Type 1 AGN are $4.9{\times}10^{-5}$ and $4.6{\times}10^{-5}$. These numbers disagree with those found by Sánchez-Sáez et al. (2018) (i.e., $2.8{\times}10^{-5}$ and $3.3{\times}10^{-5}$ respectively), however, σ_{rms} depends on the length of the lightcurves, which is most likely what introduces this discrepancy. Longer lightcurves mean, usually, larger numbers of epochs which in turn affect the σ_{rms} calculation. We note, however, that the Sánchez-Sáez et al. (2018) sample consists of higher redshift ($z > 0.7$), higher luminosity sources, as we will see later in this chapter.

For the γ distributions, the difference between the Type 1 and Type 2 AGN is even more clear as the values have a mean of 0.22, a median of 0.17 and a standard deviation of 0.22 whereas the respective numbers for Type2 AGN are 0.06, 0.04 and 0.07. This shows the slope of the Type 1 SF deviates from the zero value, which indicates that the RMS of the magnitude difference changes with time lag for this type of objects. It is also worth noting that the NLSy1 have mean, median and standard deviation values equal of 0.19, 0.14, 0.19 respectively which are closer to those from broad-line Seyfert galaxies rather than the values from Type 2 AGN. As our γ is lower than the expected 0.5 for a DRW process, our results do not support the claim that a DRW is a good description of the variability of the AGN in our sample.

Next, we consider whether any correlations exist between the various variability parameters measured from the g-band lightcurves, and also between the rest frame timespan or number of epochs of the lightcurves. We plot these values against one another in Figure 5.1, where we found that there are some significant correlations, whereas others do not show any dependence. We note. first, that there is a strong correlation (i.e the Spearman coefficient is 0.93) between A and σ_{rms} which is expected as the lightcurves have a ~1 year time length. We also find a correlation (Spearman coefficient = 0.28 for the whole sample, 0.55 for Type 1 objects) between A and γ, which, however, as Sánchez-Sáez et al. (2018) point out is not a physical relationship but is simply the

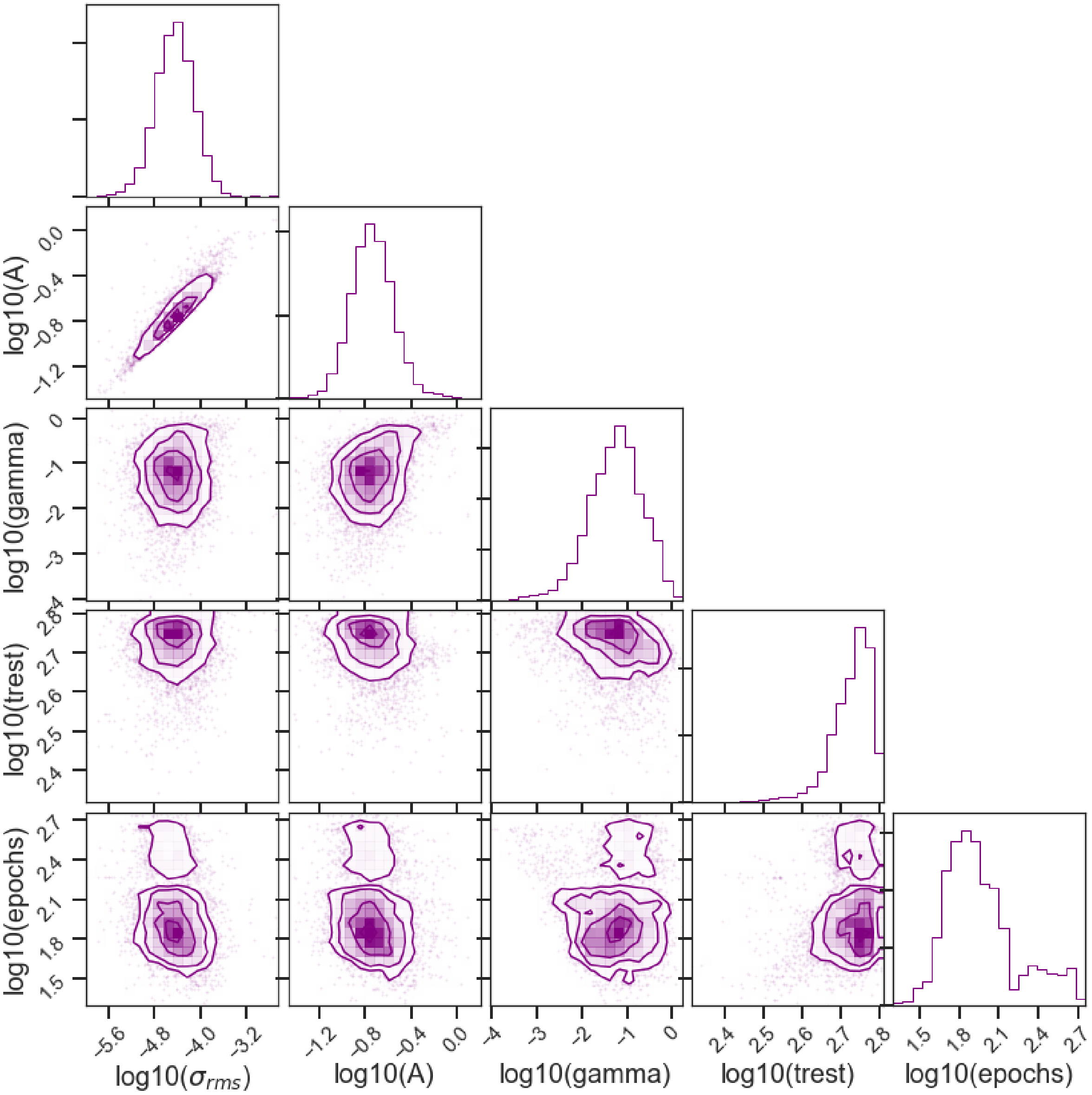


Figure 5.1: This corner plot shows the correlations between the different AGN variability parameters. Additionally, it shows their correlation with the rest frame time length of the lightcurves and the number of epochs.

result of the method by which the SF is obtained.

In terms of epochs and rest frame time length, we find, first, that σ_{rms} shows no significant correlation with the rest frame time length and only a slight anti-correlation ($\rho_s = -0.21$) with the number of epochs. The amplitude A shows only a slight anti-correlation, $\rho_s = -0.16$ and $\rho_s = -0.19$ with time length and number of epochs respectively, which shows that it is basically unaffected by both of them. The γ values show a stronger anti-correlation ($\rho_s = -0.30$ but only -0.15 for Type 1 objects) with the rest-frame time length, which we will keep in mind while we investigate its correlation with AGN physical parameters. This anti-correlation is more likely a result of the fact that a good proportion of the lightcurves have have a rest frame timespan of less than a year, which may affect the calculation of the logarithmic slope of the SF.

5.3 AGN physical parameters

We recall that the sample of 25,670 AGN from Mullaney et al. (2013) was cross-matched with the ZTF sample to give a matched sample of 8,014 AGN with g-band lightcurves. This ZTF g-band sub-sample consists of 5,646 well sampled and variable lightcurves of which 1,788 are Type 1 AGN, 3,361 are Type 2 AGN, 206 are NLSy1 and 291 display only weak evidence of AGN nuclear activity (which we universally dub LINERs). For the physical parameters, we have used the numbers reported in Mullaney et al. (2013). In this catalogue, the redshift is equal to that reported by SDSS and the Eddington ratio is calculated from the formula $\lambda_{edd} \approx 0.43 f_{10} (\frac{L_{H_\alpha}}{10^{42} \text{ ergs s}^{-1}})^{0.31} (\frac{FWHM_{H_\alpha}}{10^3 \text{kms}^{-1}})^{-2}$ where f_{10} is the is the dimensionless L_{5100}-L_{AGN} bolometric correction factor (Netzer & Trakhtenbrot 2007). Mullaney et al. (2013) assumes $f_{10} = 1$ for their calculations. This formula will give the Eddington ratios for just Type 1 AGN (as expected), which are the ones that will be used to compare with previous results. In this work, I do not calculate the

Eddington ratios for the Type 2 AGN,[1] which is subject to future work. The Eddington ratios from Mullaney et al. (2013) were, however, compared to those from Shen et al. (2011) where they calculate the Eddington ratio based on the fiducial virial black hole mass after fitting Hβ , C IV and Mg II, and we find that they produce similar results (see section 5.5). We have also attempted to convert the Mullaney et al. (2013) values based on the Shen et al. (2011) values to investigate how this affects the correlation with the variability properties (sections 5.4, 5.5). Figure 5.2 shows the distributions of redshift and Eddington ratio (from Mullaney et al. 2013) for the different types of AGN in the the ZTF g-band sub-sample, where we find that the sample contains sources with redshfit lower than 0.4. We find that the Type 1 AGN in our sample (including NLS1s) span the redshift range 0.05 to 0.4, whereas the distribution of redshifts of Type 2 AGN in our sample shows a strong peak around 0.08-0.09. For the Eddington ratio, we find that most of the sources have values less than 0.05 with few Type 1 and NLSy1 AGNs having larger values. In terms of luminosity, we used the continuum at 5100Å luminosity (L_{5100}). We calculated this luminosity using the Hα luminosity as calculated in Mullaney et al. (2013) and the fitting parameters of Table 5 in Mejía-Restrepo et al. (2016) for the L_{5100}-L_{H_α} relationship. Using Equation 1 from Greene & Ho (2005) to convert Hα to 5100Å luminosity, we found similar values for the Type 1 objects to these calculated from the Mejía-Restrepo et al. (2016) relationship but lower values (by around half order of magnitude) for Type 2 objects. We also used the [OIII]5007 luminosity from Mullaney et al. (2013) and equation 15 from Shen et al. (2011) to calculate the L_{5100} luminosity and we found values that are half an order of magnitude higher for both types compared to those obtained using the method using Mejía-Restrepo et al. (2016).

In Figure 5.3, we show the range of luminosities per AGN type. We see that they are objects with moderate luminosity where the mean luminosity is $\log(L_{5100}\ \mathrm{erg\ s^{-1}}) \sim 44.0$ and standard deviation equal to ∼1.2. It is important to note the the mean luminosity

---

[1] They have values equal to zero in the (Mullaney et al., 2013) catalogue.

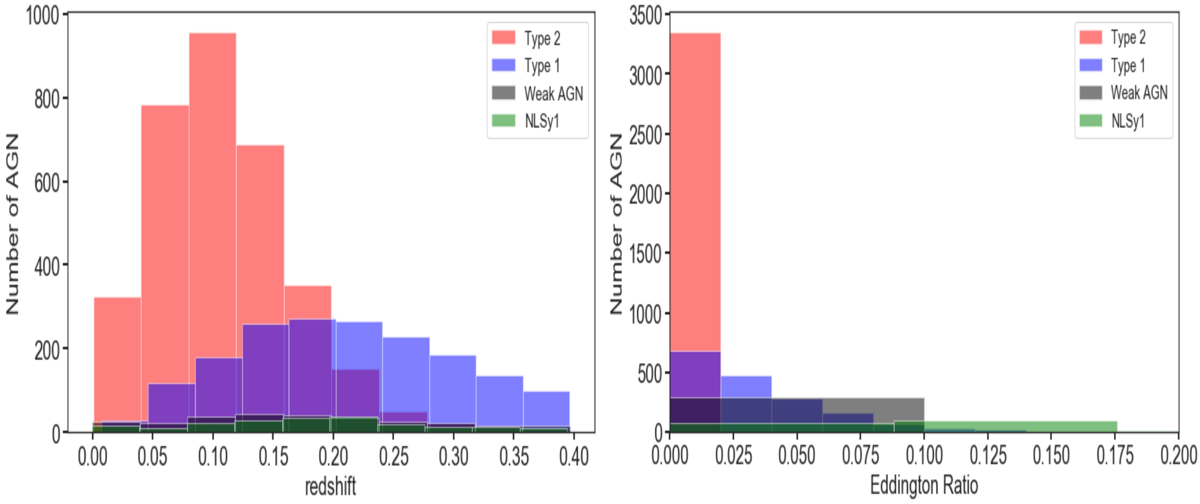


Figure 5.2: The distributions of redshift and Eddington ratio for the different types of AGN in the ZTF g-band sample. We find that Type 2 AGN in this sample have typically lower redshifts than the Type 1 and NLSy1.

for Type 1 AGN, i.e $\log(L_{5100}\ \mathrm{erg\ s^{-1}}) \sim 44.3$ with standard deviation~ 1.6, and NLSy1, i.e., $\log(L_{5100}\ \mathrm{erg\ s^{-1}}) \sim 43.7$ and standard deviation $\sim$1.3, as these are the AGN types that we will focus on in our analysis and discussion. Our sample contains mostly lower moderate and lower luminosity AGN which are also at very low redshift. Sánchez-Sáez et al. (2018) studied a sample of Type 1 AGN, with mean luminosity $\log(L_{5100}\ \mathrm{erg\ s^{-1}}) \sim 45.0$ but which contains sources with luminosities as low as $\log(L_{5100}\ \mathrm{erg\ s^{-1}}) \sim 44.0$ and redshift as low as 0.7. It is also worth noting that it is possible to sometimes overestimate our L_{5100} luminosity as we use the Hα luminosity, while Sánchez-Sáez et al. (2018) use the L_{6200}, L_{3000}, and L_{1450} to estimate the L_{5100}. For the black hole masses, we have positionally (1" match radius) cross-matched the parent (Mullaney et al. 2013) sample with the Shen et al. (2011) catalogue. From the 1788 Type 1 AGN and 206 NLSy1, there are 576 cross-matched sources. Although, we used these to investigate the relationship between AGN variability and black hole mass, we stress caution due to the low number of sources with black hole masses and the fact that Shen et al. (2011) provide virial black hole masses, which means that they have inherited dependencies on the luminosities.

We present the correlation plots for the physical parameters in Figure 5.4. Prior to

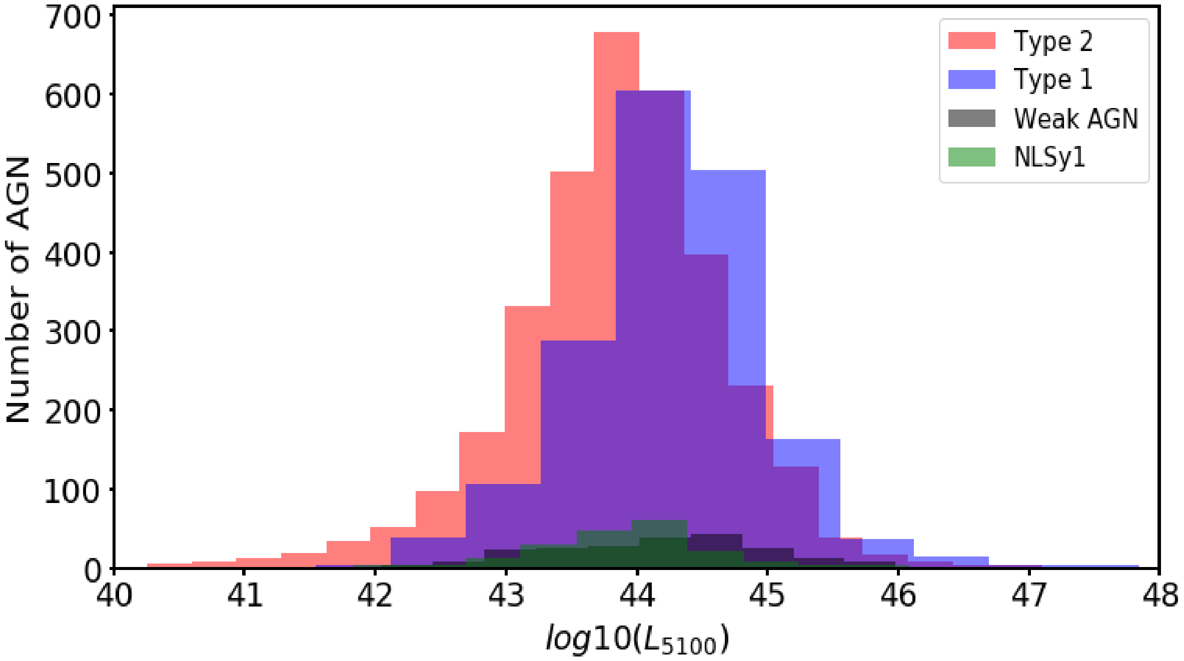


Figure 5.3: The distributions of log10($L_{5100}\,\mathrm{erg\,s^{-1}}$) for the different types of AGN in the ZTF g-band sample. We see that the sources have typically L_{5100} luminosities lower than $10^{45}\,\mathrm{erg\,s^{-1}}$.

interpreting these plots, however, it is important to consider any possible underlying trivial correlations, which may stem from, for example, straightforward luminosity-redshift correlations. The top panel of fig. 5.4 shows the correlations between the parameters (redshift, luminosity and Eddington ratio) for all the AGN in the ZTF g-band sample, while the bottom plot shows the same for the sub-sample that have black hole masses from the Shen et al. (2011) catalogue. We find that the strongest correlation for the whole ZTF g-band sub-sample is between redshift and the L_{5100} luminosity ($\rho_s = 0.50$ for the whole sample, $\rho_s = 0.50$ for Type 1 objects), which is expected due to the flux limit of the survey which means that more distant galaxies will, in general, be more luminous. Also, for the sub-sample of sources that have black hole masses, which cover a similar range as previous studies (e.g. Sánchez-Sáez et al. 2018), there is a strong anti-correlation between the black hole mass and the Eddington ratio ($\rho_s = 0.75$), which is likely to be due to another selection effect.

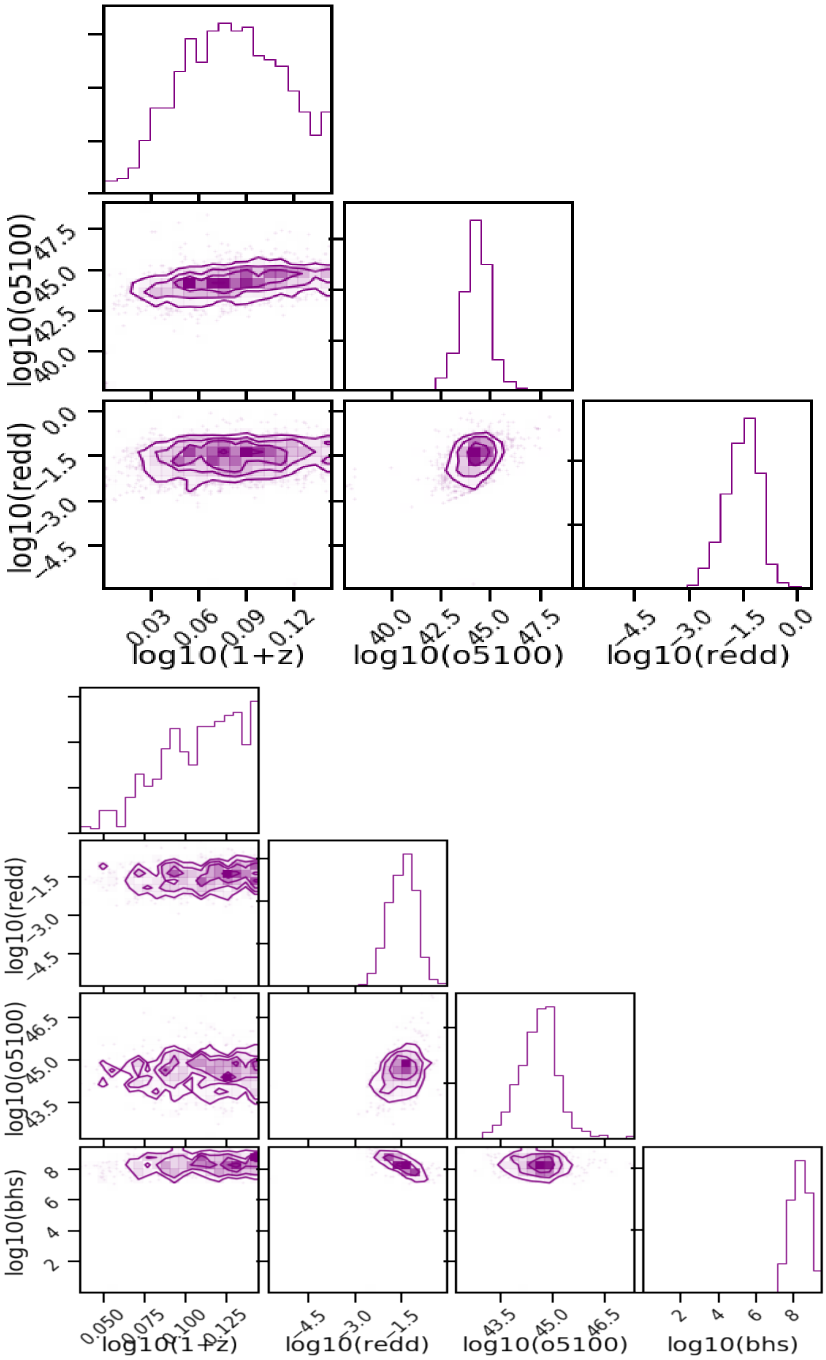


Figure 5.4: Correlations between the physical parameters of the AGN in the ZTF sample g-band lightcurves (top) and the sub-sample (bottom) that has black hole masses calculated in the Shen et al. (2011).

After exploring the correlations between the different physical parameters, the next step is to compare these values with the AGN variability parameters i.e., the SF parameters (A,γ) and the σ_{rms}. For this, we focus on the Type 1 and NLSy1 AGN as they display clearer evidence of variability (as expected), although we also include the Type 2 AGN and LINERs in our plots for comparison. In our plots, we include the Type 2 and the weak-AGN in the background for comparison between the two sub-samples. As any kind of variability in Type 2 objects should be spurious, these values should not show any kind of variability aside from the rare case of Changing-Look quasars in this sample.

5.4 Correlation between variability and physical parameters

In order to investigate if there is any dependence of the AGN variability parameters on the spectral parameters of AGN, we look at the bivariate correlation between the three AGN variability parameters (A, γ and σ_{rms}) and the four physical parameters that we consider (redshift, luminosity, Eddington ratio and black hole mass). In Figure 5.5, the different plots represent the various correlations with colours used to represent the different types of AGN in our sample. The Spearman correlation coefficients for each pair is written on each sub-plot with the respective p-value.

For σ_{rms}, we find no correlation with redshift or luminosity and a weak anti-correlation with Eddington ratio in agreement with Sánchez-Sáez et al. (2018). The σ_{rms} parameter, as well as the amplitude A, are expected to measure the amplitude of variability. We find that the SF amplitude parameter (A) shows similar behaviour to σ_{rms} with regards to the luminosity, L/L_{Edd} and black hole mass, there is, however, a weak correlation between A and redshift. As in Sánchez-Sáez et al. (2018), we find no correlation between the amplitude of variability and luminosity, which was found in previous studies with high luminosity AGN. Sánchez-Sáez et al. (2018) argue that the anti-correlation of A with luminosity in their sample may be hidden due to the positive correlation between luminosity and redshift. As they explain, their sources cover a different redshift range (∼0.7-2.1) and considering the established anti-correlation between the amplitude of variability and the rest-frame wavelength (e.g., MacLeod et al. 2010, Sánchez-Sáez et al. 2018) , this would lead to a positive correlation between the amplitude and the redshift. Since in their sample the redshift correlates positively with luminosity, the anti-correlation with luminosity could therefore be hidden. They conclude, however, using a multiple linear regression, that a dependence between the amplitude (A) and the luminosity is not needed to explain the findings from their sample. In our sample, there is a weaker correlation between luminosity and redshift for the Type 1 AGN as our study focuses on sources that cover a narrower range of redshift values than previous studies (e.g., Sánchez-Sáez et al. 2018), such that the anti-correlation between the amplitude and luminosity is absent in our sample. This is also supported by the fact that we do not see any anti-correlation between σ_{rms} and luminosity in our sample, but also no positive correlation with redshift. In the smaller sample of Type 1 AGN, which have black hole masses (574 sources), we find a weak positive correlation between the amplitude of variability, for both A and σ_{rms} values, and the black hole mass. The derived black hole masses of our sample, however, show a strong anti-correlation with Eddington ratio, which may cause the positive correlation between the amplitude measurements and the black hole masses as the anti-correlation of the amplitude with Eddington ratio is more well-established in previous works (e.g., MacLeod et al. 2010,Simm et al. 2016,Sánchez-Sáez et al. 2018). It is, however, something that we will need to investigate further in the future as the aforementioned relationship of the Eddington ratio and the black hole mass can be a selection effect and the dependence of the variability parameters with black hole mass,as mentioned in the introduction of the thesis, is still unclear (e.g., Wilhite et al. 2008, Caplar et al. 2017).

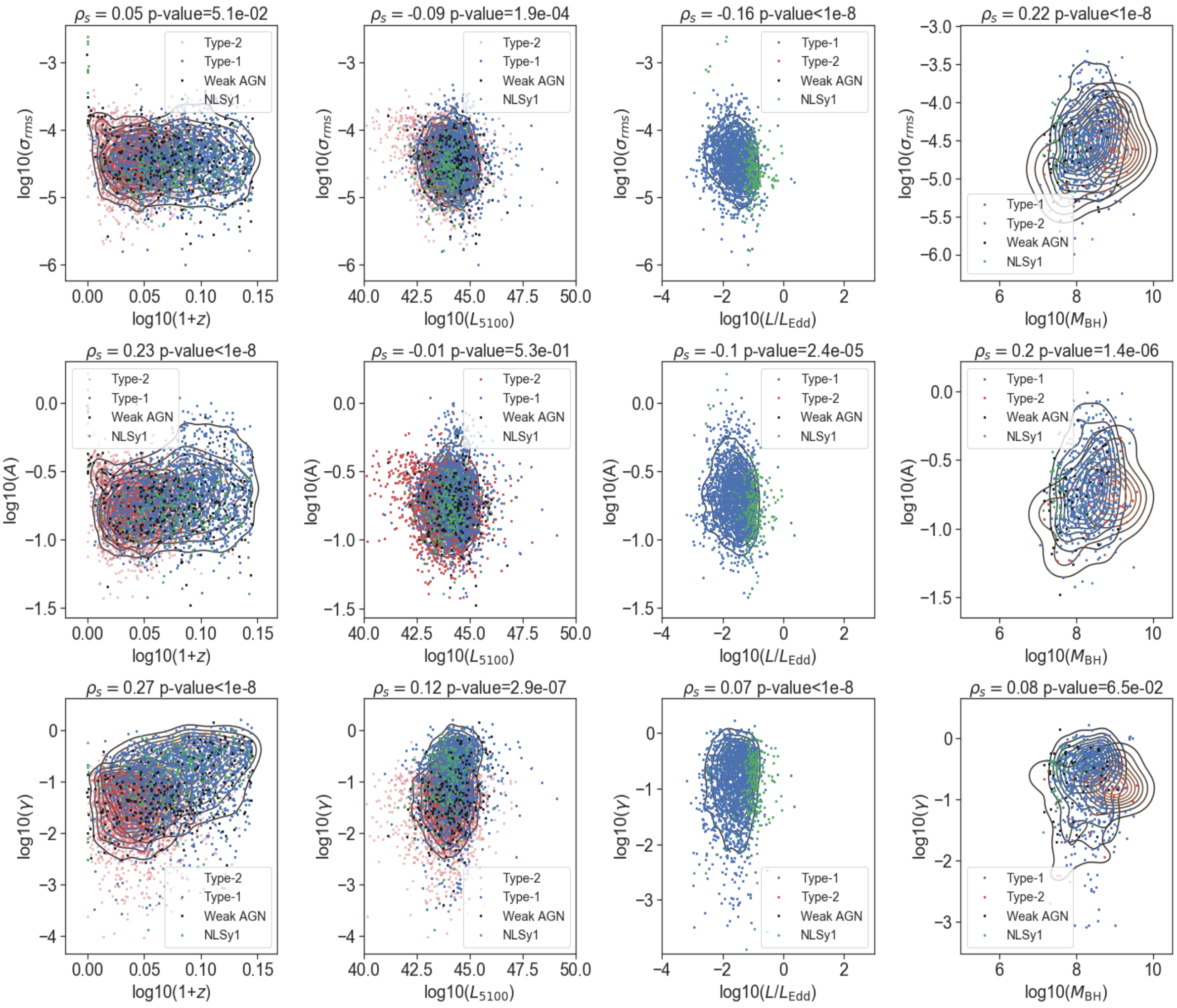


Figure 5.5: Correlations between the variability parameters from the ZTF g-band lightcurves with the spectral properties of AGN. The title of each plot includes the Spearman correlation coefficients for the Type 1 AGN and the respective p-value.

For the γ values, we find no dependence on the luminosity, Eddington ratio or black hole mass, and there is only a weak correlation with redshift. We need to be very cautious, however, with these results because as suggested previously, the γ values show a stronger anti-correlation with the rest frame timespan of the lightcurves, which may introduce spurious correlations.

We mentioned, previously (section5.3), that we have used the Eddington ratio values from the Mullaney et al. (2013) catalogue. Plotting these values against the values from Shen et al. (2011) after positionally cross-matching the two catalogues and finding 4,547

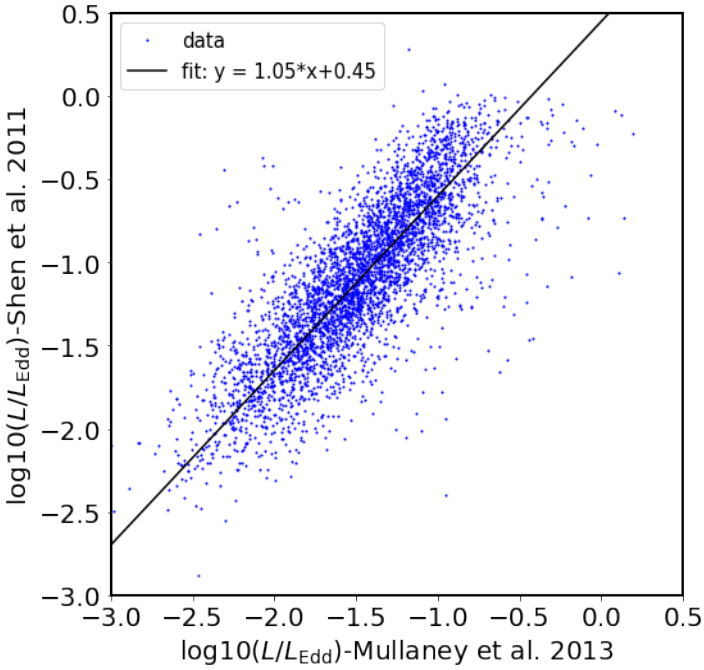


Figure 5.6: The relationship between the Eddington ratio values in Mullaney et al. (2013), calculated using the equation $\lambda_{edd} \approx 0.43 f_{10}(\frac{L_{H_\alpha}}{10^{42}\, erg\, s^{-1}})^{0.31}(\frac{FWHM_{H_\alpha}}{10^{3}{\rm kms}^{-1}})^{-2}$ and $f_{10} = 1$ (Netzer & Trakhtenbrot 2007) and the Eddington ratio values from Shen et al. (2011) calculated using the fiducial virial mass after fitting the Hβ, C IV and Mg II lines.As expected, using the fitting to covert the values from Mullaney et al. (2013) does not affect the correlations in Figure 5.5

matches, we show in Figure 5.6 the relationship between the Eddington ratio values in the aforementioned catalogues and the respective fitting. We find that converting the values in Mullaney et al. (2013) according to this fit has no effect on the correlations shown in Figure 5.5. A more systematic calculation (using different spectral lines) of the black hole masses and the respective Eddington ratios is intended for future work. Although Hα and Hβ are included in the Mullaney et al. (2013) catalogue, other lines used in previous studies (e.g., Mg II) are not measured. The systematic measurement of all previously used lines will allow a better estimate of the correlation between the AGN variability parameters and black hole mass, a relationship which remains a riddle across the whole range of AGN luminosities. It will also allow a fairer comparison to previous studies (e.g., Sánchez-Sáez et al. 2018). The Eddington ratio values from Mullaney et al. (2013), which was also evaluated using the Shen et al. (2011) catalogue, however, allow us to look at relative changes of the AGN variability parameters with the Eddington ratio.

5.5 r-band lightcurves

In this section, I present the results using the same analysis as the three previous sections but for, well sampled, variable, r-band lightcurves from the ZTF survey for the same parent sample (Mullaney et al. 2013). We also examine how the different wavelength affects the the AGN variability parameters especially in the case of lower redshift galaxies to verify results identified in previous studies (e.g., Kimura et al. 2020). From our sample, there are 12,873 object with r-band photometry in the ZTF DR3 data release. Of these, 12,873 AGN, 7,073 have well-sampled lightcurves (i.e., have more than 20 epochs covering a rest-frame timespan longer than 200 days, fig.5.7) and are found to be variable (i.e., V>1.3 and $\sigma^2_{rms} - err(\sigma^2_{rms}) > 0$). Of these 7,073 sources, 2,682 are classified as Type1 (BLSy1), 3,700 as Type 2 AGN, 283 as NLSy1 and 408 as weak-AGN (LINERs).

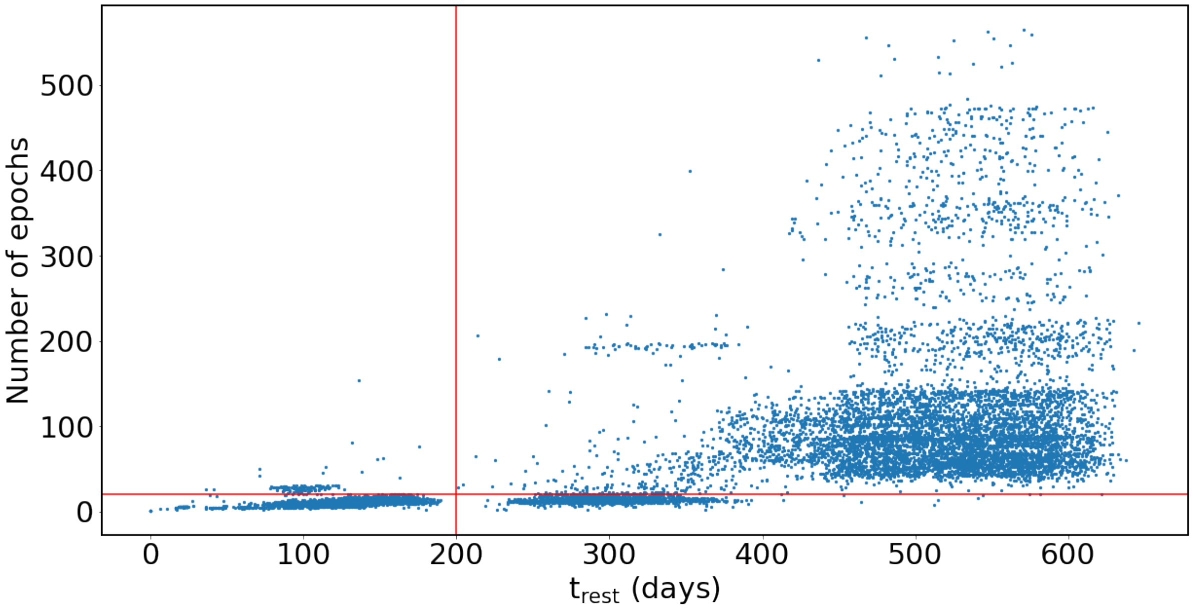


Figure 5.7: The number of epochs per AGN versus the rest-frame time length of the each lightcurve generated by the ZTF r-band public data.

We find that the plot of A versus γ for the r-band lightcurves (fig. 5.8) is quite similar to the one for g-band lightcurves. We see that the γ values for the Type 1 AGN start to deviate from the close to zero values where the Type 2 AGN mostly lie. We summarise the statistical values for A and γ of the different AGN types in Table 5.1.

	Amplitude (A)		Logarithmic Gradient (γ)	
Survey	Type 1	Type 2	Type 1	Type 2
ZTF r-band	mean = 0.15 median = 0.13 std = 0.09	mean = 0.15 median = 0.13 std = 0.07	mean = 0.23 median = 0.15 std = 0.23	mean = 0.05 median = 0.03 std = 0.06

Table 5.1: This table summarises the statistical values for the amplitude A and the logarithmic slope γ generated from fitting the SF for the ZTF r-band lightcurves sample.

For the amplitude, we conclude that we do not find a significant difference in the amplitude distributions across the different types of AGN in our sample. We point out, however, that the values from the r-band for Type 1 and NLSy1 AGN are lower than those obtained from the g-band lightcurves. This supports the argument that the host galaxy contamination (e.g., Kozłowski 2017) is expected to be higher at longer wavelengths

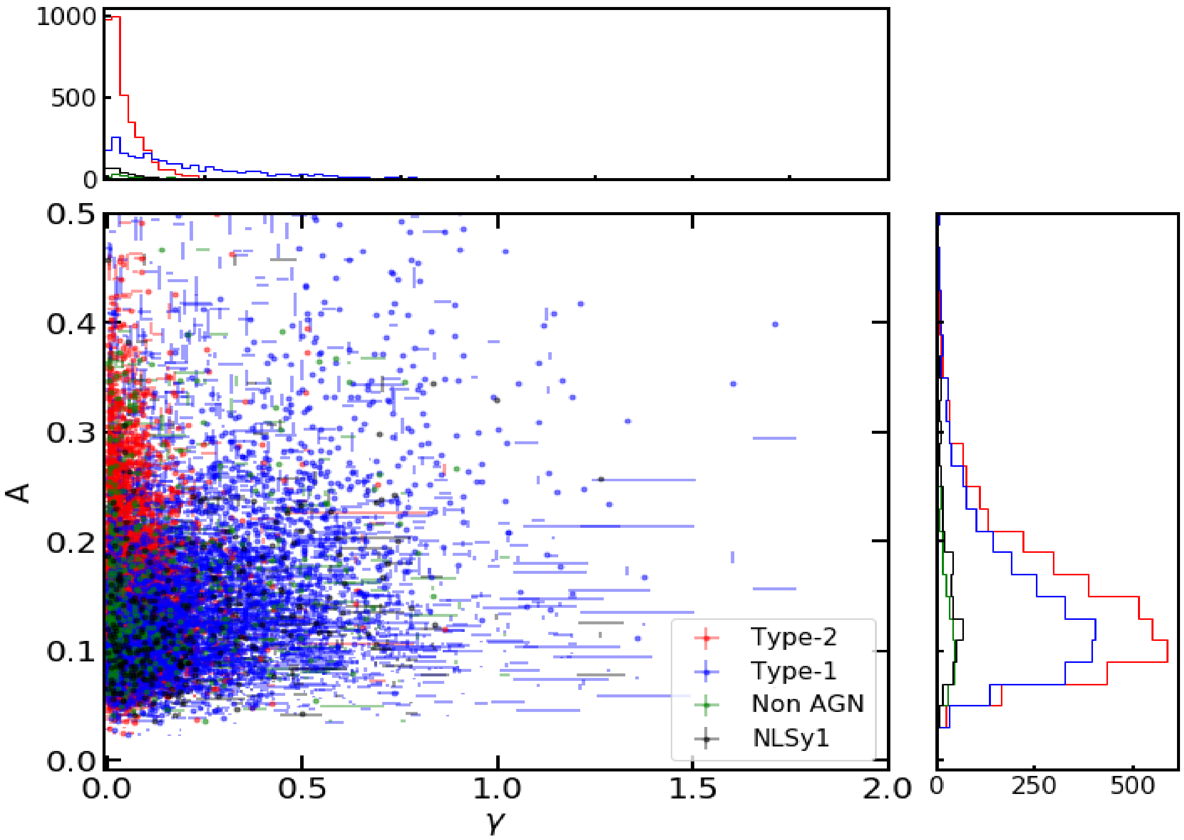


Figure 5.8: A vs γ parameter space from the ZTF r-band lightcurves.

(e.g., Kimura et al. 2020). Future work will include a more detailed consideration of the level of host contamination. At this point, there were not enough AGN with well-sampled lightcures in both the g-band and r-band lightcurves to explore the effects of host contamination. Analysis of the variability amplitude of large samples of low-redshift AGN, with regards to wavelength, will allow a better estimate of host contribution. The σ_{rms} from the r-band has a mean and a standard deviation of 2.4×10^{-5} and 2.4×10^{-5} respectively for the Type 1 AGN r-band lightcurves. These are lower than those found for g-band (i.e., 4.9×10^{-5} and 4.5×10^{-5}) and closer to those found by Sánchez-Sáez et al. (2018) (i.e., 2.8×10^{-5} and 3.3×10^{-5}) which indicates that this measurement is also affected by the wavelength in which we measure it.

We find the distribution quantities for the SF γ values for the r-band lightcurves are very similar those derived from the g-band lightcurves. This supports the idea that the γ values are not affected by the host contamination. The γ values from the r-band lightcurves again show a much lower mean and median than the 0.5 which is expected from DRW, although, they show a large range of values.

Finally, we plot the bivariate correlations of the AGN variability parameters and the spectral properties of AGN, using the values from the r-band lightcurves. In Figure 5.9, we present the same correlation plots as in Figure 5.5. The results for the correlation of σ_{rms} calculated from the r-band lightcurves with the physical parameters of AGN are similar, in terms of their correlation coefficients, to those from the g-band lightcurves by showing a weak anti-correlation with the Eddington ratio and a weak correlation with the black hole mass. We find that the amplitude A has a weaker correlation with redshift than those from the g-band lightcurves, a similar weak anti-correlation with the Eddington ratio and a weaker correlation with the black hole mass. For the γ values from the r-band ZTF lightcurves, there is a slightly stronger correlation with redshift and similar correlation with the luminosity than those from the g-band lightcurves. However, these correlations are not very strong and, as with those derived from the g-band lightcurves, show no correlation with Eddington ratio or the mass of the black hole.

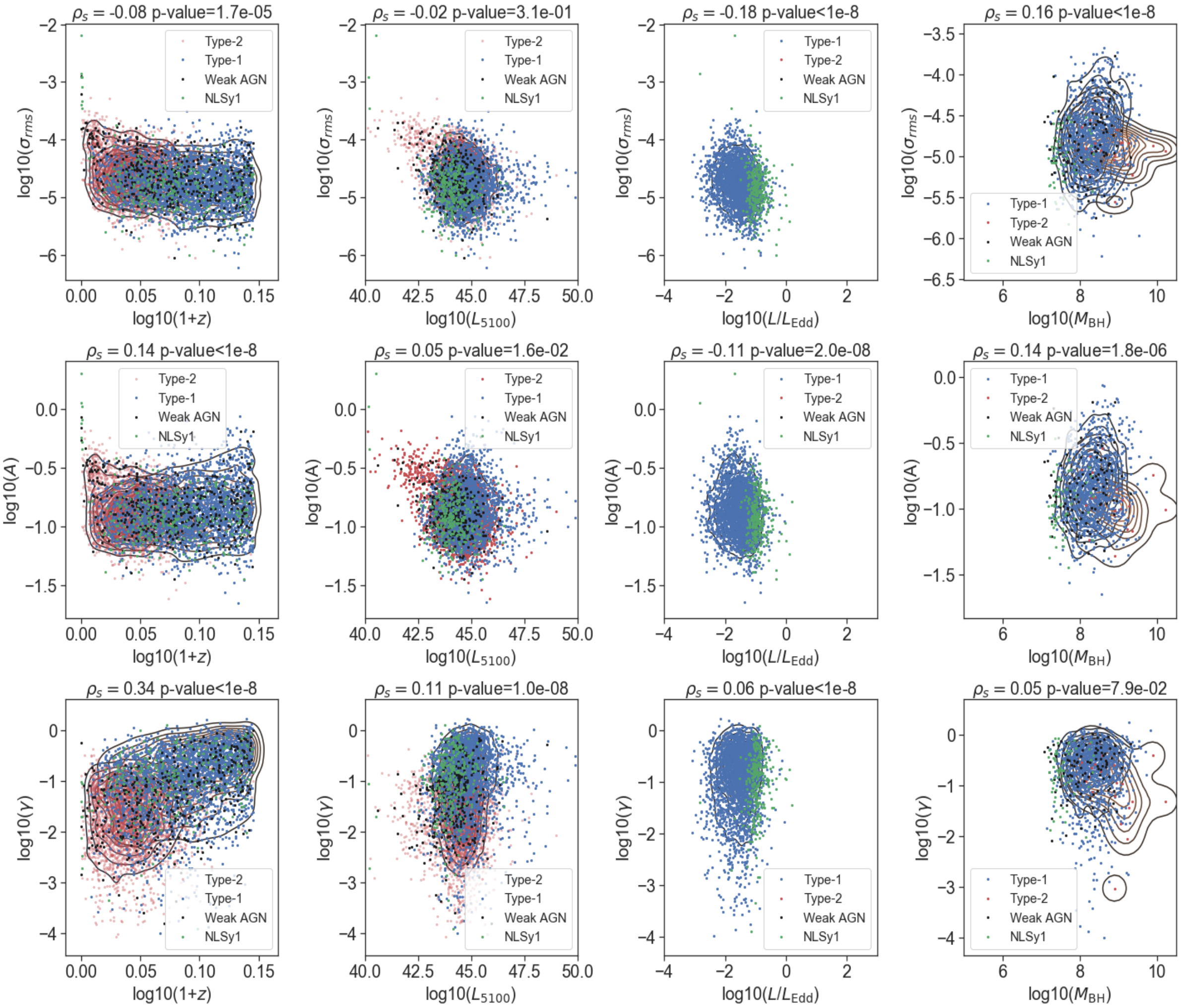


Figure 5.9: Correlations of the variability parameters from the r-band lightcurves with the spectral properties of AGN. The Spearman correlation coefficients and their respective p-values are given as reference at the top of each sub-plot.

5.6 Chapter Discussion and Conclusions

In this chapter, we have looked more closely at the variability of moderate luminosity Type 1 AGN ($L_{5100} \sim 10^{42.5-45.5}\mathrm{erg\,s^{-1}}$ with most of them having $L_{5100} < 10^{44.5}\mathrm{erg\,s^{-1}}$). These AGN have redshifts < 0.4 which introduces the uncertainty of host contamination especially in the case of SF analysis (e.g., Kimura et al. 2020). We find in section 5.5 that the estimated A values from the r-band lightcurves have lower mean and median values compared to those from g-band lightcurves, which confirms that the host contamination is higher at longer wavelengths. The mean and median values of the amplitude (A) for ZTF g-band lightcurves, estimated in section 4.6, are similar to the value found in previous studies (i.e., $\sim$ 0.2 mag, e.g., MacLeod et al. 2012, Sánchez-Sáez et al. 2018). We, however, argue that the values in section 4.6, but also the respective values for the r-band in section 5.5, show no significant difference from the A distribution of the Type 2 objects. As many of the lightcurves have timespans close to or less than around one year in the rest-frame, we will need longer lightcurves to investigate further the non-detectable difference in the A distribution between the Type 1 and Type 2 objects. We will also need a large sample of AGN with well-sampled lightcurves in both ZTF bands (g,r) and data from other surveys for a complete characterisation of this dependency.

We find, however, in sections 5.4 and 5.5 that our results are in agreement with key results from recent studies (e.g., Sánchez-Sáez et al. 2018) with regards to the anti-correlation between the amplitude measurements (A, σ_{rms}) and the Eddington ratio. Our values are slightly lower than the aforementioned survey (i.e., -0.23 and -0.22 for A-L/L_{Edd} and σ_{rms}-L/L_{Edd} respectively) which can result from the fact that our sample covers a narrower range of redhshift (i.e., ~0.01-0.4) than Sánchez-Sáez et al. (2018) (i.e., ~0.7-2.1). The host contamination is more likely to affect the A parameter, which is extracted from the SF, rather than the σ_{rms} parameter, given our results (fig. 5.5 and fig. 5.9). The slightly lower correlation, however, can be also interpreted as a result of the different ways the Eddington ratios are calculated in this and the Sánchez-Sáez et al. (2018) study. In our study, we have used the Hα line to estimate the Eddington ratio (see section 5.3), whereas Sánchez-Sáez et al. (2018) uses the average of the black hole masses calculated with all the available useful lines (e.g., Hβ , Mg II) and from there they calculate the Eddington ratio. The anti-correlation between the amplitude of variability and the Eddington ratio has also been reported in previous studies of AGN (e.g., Wilhite et al. 2008, MacLeod et al. 2010, Simm et al. 2016).

There are in the literature many different interpretations of the anti-correlation between the amplitude of variability and the Eddington ratio. This observational results challenge the accretion disk models, which have also tried to explain the AGN optical variability. Many of the previous studies have discussed the possible interpretations but to date no explanation is self consistent with all the results (e.g., Wilhite et al. (2008), MacLeod et al. 2010, Simm et al. 2016, Sánchez-Sáez et al. 2018). The standard accretion model proposed by Shakura & Sunyaev (1973), predicts that sources with higher Eddington ratio have hotter disks. Global changes in the accretion rate (assuming a standard α-disk as proposed by Shakura & Sunyaev 1973) can predict an anti-correlation between the variability amplitude and the Eddington ratio. However, as Simm et al. (2016) point out, this predicted relationship is flatter than the observed one. Additionally, the timescales of optical variability are much shorter than the time needed for global changes of the accretion rate in the whole disk, which is associated with the sound-crossing and viscous timescales (e.g., Ruan et al. 2014).

The anti-correlation of the amplitude of variability and Eddington ration has also been interpreted as a reflection of the dependence on the wavelength, which depends on the radius of the disk. As we said, a standard accretion disk predicts that a higher Eddington ratio means a hotter disk, which would mean that the optical flux originates at a larger radius. Further, the anti-correlation of the amplitude of variability and the wavelength would lead to an anti-correlation with the Eddington ratio (MacLeod et al. 2010). In MacLeod et al. (2010), however, they find that neither the predicted thermal timescale or the viscous timescale (using thin disk assumptions) can match the observations in this study. Local fluctuations have also been proposed as a way to explain the optical variability using inhomogeneous models such as the one proposed by Dexter & Agol (2011), although, this model fails to explain the flux-flux correlation between optical bands as explained in Simm et al. (2016) and Kokubo (2015).

In earlier studies (e.g., Hopkins et al. 2005, Wilhite et al. 2008), it was argued that the Eddington ratio could be used as a proxy for the age of the AGN. This means that for sources of similar black hole mass, optical luminosity can indicate AGN age, where younger objects are more luminous due to a greater available gas/fuel supply. On the other hand, for similar luminosity larger black hole masses would mean an older AGN, where the lower Eddington ratio would indicate that the source has exhausted its fuel supply. Variability due to variations in the fuel supply, however, would need much longer timescales (i.e. viscous timescale $\sim 10^5 - 10^7$days which control the large scale changes in the accretion rate; see also Sánchez-Sáez et al. 2018) for the amplitude found in that study compared to the $\sim$ 1year found in this study. We note again, however, that the amplitude numbers in this study are subject to the time length of lightcurves and the host contamination.

Additionally, we find no correlation between the amplitude of variability (neither A or σ_{rms}) with the luminosity, which is in agreement with Sánchez-Sáez et al. (2018), but disagrees with previous studies that do report an anti-correlation between these two AGN properties (e.g., MacLeod et al. 2010, Simm et al. 2016, Caplar et al. 2017). Kimura et al. (2020) found for a sample of 491 variability-selected AGN (up to a redshift of 4.26) that the amplitude of variability anti-correlates with AGN luminosity and that the host contamination is also inversely correlated with the luminosity which shows that the amplitude of variability could be an indicator of the host contamination. This remains to be investigated further with longer lightcurves to evaluate the current caveats for the amplitude (i.e., similar numbers for Type 1 and Type 2 AGN).

In our low-redshift, moderate luminosity sample, we find that the logarithmic gradient (γ) does not correlate significantly with any of the AGN physical parameters, which agrees with the results of Sánchez-Sáez et al. (2018) results. That the logarithmic gradient does not correlate with the black hole mass and/or Eddington ratios for the sources in our sample, together with the fact that we find much lower γ values than the ones expected from the DRW model, disagrees with the findings of Kelly et al. (2009), who interpret their findings in terms of optical variability arising as a result of thermal fluctuations related to underlying stochastic processes (e.g., turbulent magnetic fields). The aforementioned study focuses on higher luminosity AGN. Other studies have described the limitations and caveats related to using the DRW to describe AGN lightcurves and measure their variability properties (e.g., Kasliwal et al. 2015, Kozłowski 2017) for both ground-based and space surveys. The current ground based surveys (e.g., ZTF, GOTO) are used to test the DRW model via well-sampled lightcurves and to unveil the connections between AGN variability and other AGN parameters prior to the LSST, which will provide time series data for millions of AGN. Finally, an even bigger challenge for astronomers will be combining the data from the LSST with other ground-based surveys which would allow well-sampled lightcurves that would span for over 20 years for hundreds of thousands or millions of AGN to study both short term (days to months) and long term (years to decades) AGN variability.

Chapter 6

Tidal Disruption Event discovery with GOTO

6.1 Chapter Introduction

The goal of this chapter is to investigate the current and future prospects of using GOTO as a discovery machine for Tidal Disruption Events (TDEs). To investigate these capabilities, we use difference imaging from GOTOphoto and forced photometry from the LSST stack run on GOTO images, since both methods can reveal changes in the flux of the host due to any transient located in it. The chapter will focus on the lessons learnt from the observations of known TDEs during the 4-UT phase of GOTO survey to assess whether it is possible to develop a pipeline that would recover these known TDEs. I also attempt to estimate the number of candidates such a pipeline would deliver, as well as the level of contamination from other types of transients (e.g. supernovae). In the second part (section 6.3), I consider the future setup of GOTO and use simulated TDE lightcurves to determine how the planned future developments will improve the rapid identification of TDE candidates.

6.2 Detection of known TDEs

GOTO started to collect data in July 2017 although, as in previous chapters, we only use the data taken after the major hardware upgrades in February 2019. Since then, the GOTO collaboration has put a lot of effort into software development, which, due to the primary science goal of GOTO (i.e., detection of optical counterparts of GW events) is mostly focused on difference imaging. This involves aligned and PSF-matched template images being subtracted from nightly science images. In theory, difference imaging reveals those sources that have changed in brightness since the template image was obtained. The task of finding transients via difference imaging comes with a number of challenges. First, it is important that the template and nightly science images are of sufficiently high quality to avoid introducing any spurious detections. Even in cases, where the templates are of sufficiently high quality, however, we still need to distinguish between real astrophysical changes and those due to, for example, misalignment, hot pixels, cosmic rays, or other artefacts.[1]

A standard way of separating real from spurious detections is through the development of a classifier, which ranks a detection based on how "real" it thinks the source is. Such a ranking allows us to make a first pass at reducing the, potentially, hundreds of thousands of detections per night to just a few, highly ranked, detections. Such a classifier has been developed for GOTO which attributes each sources with a `realbogus` label, which I use as a potential means to identify true TDEs. Another piece of information from GOTO's difference imaging analysis is the distance of the transient from the nucleus of the host which may help to distinguish nuclear transients (which include TDEs) from non-nuclear transients. There are two main ways of measuring the distance of the transient from the host: one can measure the distance of the transient from the center of the host on the template image after the astrometric alignment or one can measure the

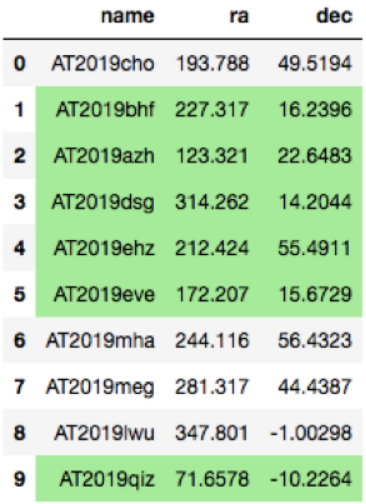
---

[1] Real changes can also come from variable stars.

distance of the transient from the center of the host as reported in an external catalogue such as SDSS or PanSTARRS.

Aside from the `realbogus` value and the separation from the host nucleus, there are other candidate selection criteria based on the photometric properties of the transient and the host galaxy. The nature of the host galaxy can be indicative of the transient event since the environment helps to inform on the nature of the transient event(e.g., core-collapse supernovae prefer star forming galaxies; Kelly & Kirshner 2012, TDEs prefer early type galaxies; Arcavi et al. 2014). The type of host (i.e., if it is starforming or early type) can be identified using the colour of the host galaxy. The selection of the host galaxy based on its type introduces a bias against TDEs hosted by AGN or "bluer" galaxies. Aside from the colour of the host galaxy, it is also important to know the color of the transient as, for example, TDEs are found to have bluer colours (e.g., Arcavi et al. 2014, Hung et al. 2018). Since GOTO performs its survey in the L-band during its standard mode, it is important to have an early identification of candidates so we can trigger multi-band observations with GOTO (i.e., with Baader R,G and B filters), or with other facilities so we can obtain colour information. Finally, information obtained from the lightcurve of the transient can also be used to identify TDEs. Transients can be characterised, in part, by their rising and fading times, as well as their colour evolution, all of which differ between different types of transient.

As mentioned previously, using known TDEs discovered during the studied period of the GOTO survey allows us to examine whether GOTO was able to detect them and whether the values of various measured parameters (e.g., `realbogus`, distance from host centre) would have meant we would have identified them as TDEs. I also investigate the quality of their GOTO lightcurves and examine fitting them with TDE models. From ZTF survey data, van Velzen et al. (2020) reported 17 new spectroscopically confirmed TDEs discovered between March 2018 and September 2019. Since I only consider GOTO data from February 2019 onwards, I focus on the 10 TDEs, which were discovered in



	name	ra	dec
0	AT2019cho	193.788	49.5194
1	AT2019bhf	227.317	16.2396
2	AT2019azh	123.321	22.6483
3	AT2019dsg	314.262	14.2044
4	AT2019ehz	212.424	55.4911
5	AT2019eve	172.207	15.6729
6	AT2019mha	244.116	56.4323
7	AT2019meg	281.317	44.4387
8	AT2019lwu	347.801	-1.00298
9	AT2019qiz	71.6578	-10.2264

Figure 6.1: The 10 TDE discovered in 2019 by the ZTF survey and reported in van Velzen et al. (2020). The objects highlighted with green are objects are TDEs with peak r-band magnitude brighter than 19, which suggests that they should have been detected by GOTO.

2019. The table in Figure 6.1 shows the ten TDEs including their Transient Name Server (TNS) names and their position in RA and Dec in degrees. Six of these TDEs have peak magnitude brighter than 19 in ZTF r-band, which indicates that they should have been detected as candidates by the GOTO survey. Of these ten TDEs, nine (i.e., except AT2019iwu) have host galaxies that are contained within the GOTOphoto photometry table (i.e., they are detected by GOTOphoto). Four of these nine are in the table of candidate transients identified via difference imaging (AT2019azh, AT2019dsg, AT2019ehz, AT2019qiz)[2]. Five (i.e., these of AT2019azh, AT2019bhf, AT2019cho, AT2019ehz and AT2019qiz) of the original ten TDEs have hosts that are contained within the forced photometry database produced by the LSST stack. This reduced number is due to the limited coverage of the reference catalogue.

In Figure 6.2, I present the lightcurves of four TDEs (i.e., AT2019azh, AT2019bhf, AT2019ehz and AT2019qiz). These lightcurves have been generated by GOTOphoto's blind photometry (left side plots) and the LSST stack's forced photometry measure-

---

[2] There are already cuts applied for the sources in the candidate table.

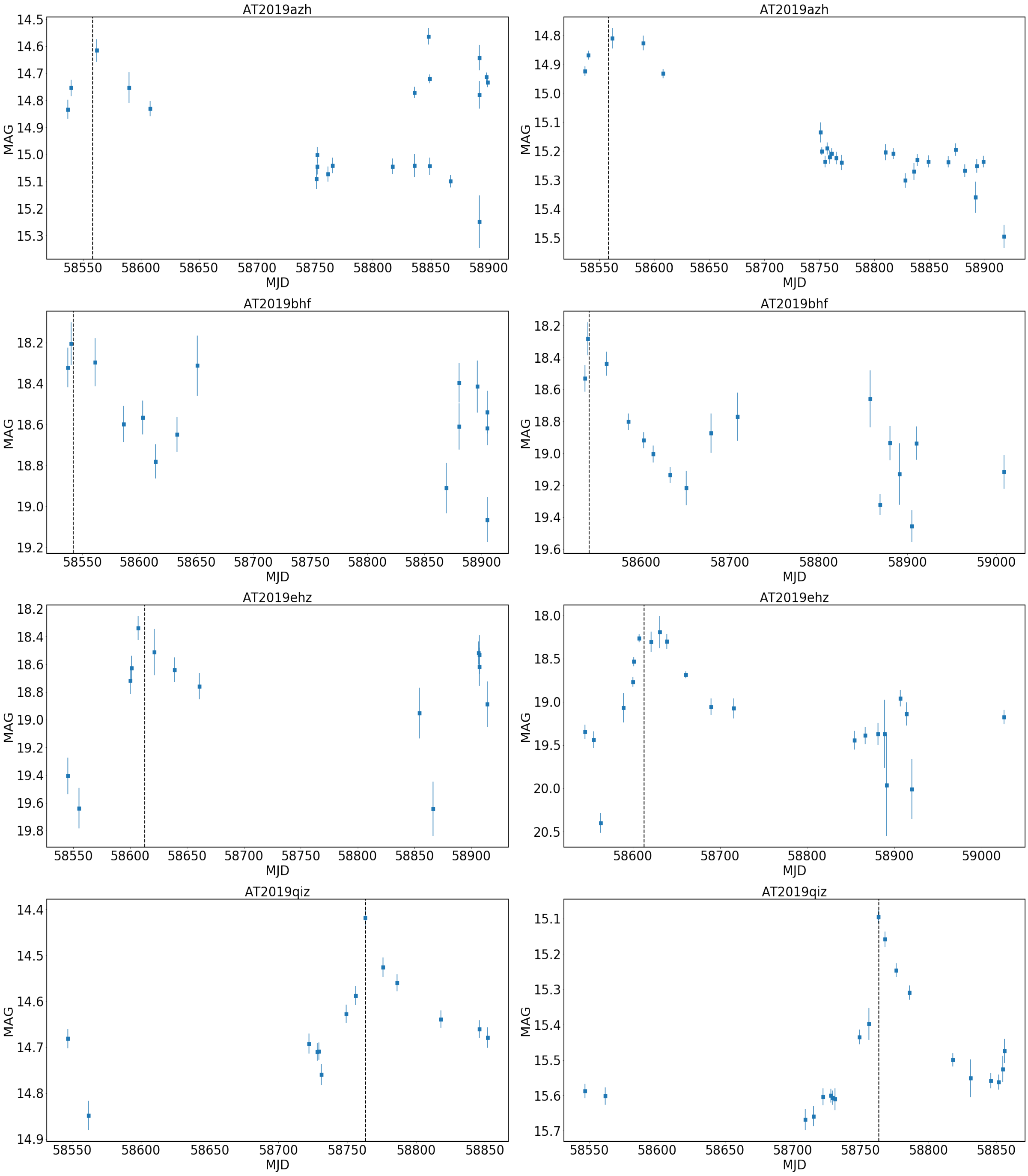


Figure 6.2: The GOTOphoto (left) and GOTO/LSST stack (right) lightcurves of 4 of the TDEs discovered by van Velzen et al. (2020). The dotted line represents the time of peak reported in the van Velzen et al. (2020) paper.

ments (right side plots). For the LSST stack-generated lightcurves, I have used aperture photometry for those sources whose mean L-band magnitudes are brighter than 17.5 (AT2019azh, AT2019qiz), and PSF photometry for the remainder (AT2019bhf, AT2019cho, AT2019ehz). In all of the lightcurves presented in Figure 6.2, there is an indication of some flaring. As expected, this flaring is clearer for brighter objects which are not as close to the flux limit of GOTO. As the brightness of the host and transient approaches the flux limit, the measurement noise introduces larger uncertainties for the photometry measurements. The position of the peak in GOTO mostly coincides with the peak reported by van Velzen et al. (2020), although there is one (possibly two) interesting cases where GOTO appeared to catch the TDE closer to its peak brightness (i.e., AT2019qiz, and possiibly AT2019bhf).

Figure 6.2 clearly shows that GOTO measured flaring events that coincide with the reported TDEs, and thus it is reasonable to say that GOTO "detected" these TDEs. However, the goal of this study is to test whether it would have been possible to identify these events as TDEs prior to gaining this information from other sources (in this case, van Velzen et al. 2020). It is only if that is the case that GOTO can be used as a means of notifying us of TDEs. In order to investigate this, I consider whether we are able to uniquely distinguish these sources from all the other candidate transient sources (e.g., supernovae) in GOTO's database by using only data available at the time of the TDE (i.e., GOTO data plus freely-available archival data). Since we expect that all TDEs that are detectable by GOTO will be hosted in resolved (i.e., extended) galaxies, I first identify all candidates that are positionally associated with extended sources.

To identify candidates that are located in extended sources, I first cross-match a catalogue of 1,892,979 (i.e., the number of detections in the candidate table[3] from 2019-02-22 to 2020-04-20) with the PanSTARRS catalogue using a matching radius of 2". I chose the PanSTARRS catalogue since any TDE bright enough to be detected by

---

[3] The table of the GOTOphoto database populated with the detections from difference imaging

GOTO will likely be close enough for its host galaxy to be detected and resolved in PanSTARRS. After cross-matching, we find 1,512,939 GOTO candidates associated with PanSTARRS sources. Then, we define as extended those sources which have `rMeanPSFMag-rKronMag`≥ 0, which separates stars from galaxies with accuracy 99.9% (Tachibana & Miller 2018). After applying this cut, we find that there are 546,934 detections in the candidates table that are associated with extended sources. We note that these are detections and not unique sources, which means that a few of these detections may be associated with the same transient source (e.g., repeated detections of supernova in a galaxy). In order to compare the information between known extragalactic transients, I have also cross-matched this list of candidates with Transient Name Server (TNS) with matching radius 2.5" and found 1,182 detections associated with known transients which corresponds to 628 unique sources.

After performing the above cross-matches, I first consider the values of the `realbogus` measurement and the distance from the centre of the host galaxy. In this work, I analyse the data from the early GOTO `realbogus` classifier. This version employed a Random Forest (RF) model which matched closely the one presented in Bloom et al. (2012)[4]. As Figure 6.3 shows, GOTO's `realbogus` classifier assigned low values (< 0.5; meaning it considered them likely bogus) to all of the four TDEs within the GOTOphoto candidate list. We find, however, that this is a general problem for TNS-reported transients, not just for TDEs - the GOTO `realbogus` classifier currently assigns low values (< 0.5) to 91% of the 628 TNS-reported sources in its candidate list. The GOTO collaboration is currently developing a new classifier that addresses, among others, this issue. Given the inability of the realbogus classifier to identify our TDEs, I turn to the measurement of the distance of the transient from the centre of the host galaxy as a possible means

---

[4] Amuch improved model is under development (Killestein et al. submitted), using instead a convolutional neural network (CNN) to analyse the pixel-level data (in contrastto extracting human-selected "features" of the detections, asis required for the RF approach). This model, however, was not fully implemented while writing this thesis. For the data used in this thesis only the early classifier was available.

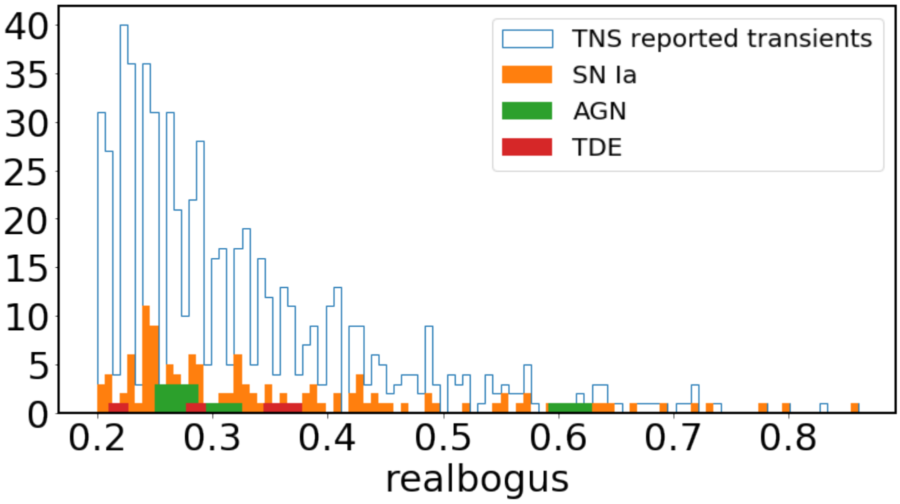


Figure 6.3: Realbogus values for TNS detected transients. The plot also shows the respective histograms for SN Ia, AGN and TDE.

to identify TDEs. Such a spatial cut is important to separate nuclear transients from other extragalactic transients like supernovae. I have investigated two different methods of measuring the distance from the host. The first is to measure the distance from the centre of the host on the template image and the second one is by measuring the distance from the centre of the host as found in the PanSTARRS catalogue.

In the top panel of Figure 6.4, we plot the distribution of offsets from the host nucleus of all the GOTO candidates with TNS counterparts, highlighting those that are TDEs and Type Ia supernovae. This plot shows that the TDEs are not measured to be systematically closer to their hosts' nuclei than other types of transients. In the bottom panel of Figure 6.4, we show the same, but this time the distance is measured from the nuclear position reported in the PanSTARRS catalogue. Again, this also shows that the TDEs are measured to not be systematically closer to the hosts' nuclei. For both methods, a separation of 1.75" from the nucleus would be necessary to currently retrieve all the known TDE. Such a cut would, however, result in a large number of possible TDE candidates. For example, of the 546,934 candidates associated with galaxies, 509,667 (i.e., 93%) are within 1.75 " of the centre of their host galaxy. We also note that the

1.75” is based on small number statistics (i.e, 4).

With regards to the number of detections of a transient, to consider something as a real transient it would, ideally, be required to have more than one detection (i.e., from different UTs or more than one epochs) in the candidate table (i.e. the result from difference imaging) associated with the same source in the previous few months. With the present data, however, this would imply that we would not identify two of the four genuine TDEs within the GOTOphoto candidate list (AT2019ahz, AT2019dsg) as TDEs. The other two have, however, five detections (three epochs; AT2019ehz) and 5 detections (5 epochs; AT2019qiz). Of the 546,934 detections in the GOTOphoto candidate table (considering data up to April, 2020) that are associated with extended sources, 393,045 are distinct sources, of which 70,303 are detected more than once. In the case of the 1,182 detections reported by the TNS and considered in this study, 628 are unique sources, and 231 have been detected in more than one epoch. In the next section, I investigate how an improved cadence will affect the identification of TDEs from lightcurves. Since we aim to use the GOTO lightcurve data, multiple epochs are also important in this respect.

At present, it is not possible to use the data stored within the GOTO database to identify TDEs. The GOTO analysis pipeline, however, continues to be developed and improved, and so it is expected that the `realbogus` classifier will become more attuned to identifying nuclear transients. Despite that, it is likely that we will still need other ways reduce the number of possible TDE candidates. As some recent studies have shown, one way of doing this may be to consider the properties of the host galaxy. It has been known for some time that TDEs preferentially occur in early type (i.e., red) galaxies (e.g., Arcavi et al. 2014, French et al. 2016). With this in mind Hung et al. (2018) included a selection step when searching for TDE candidates in iPTF data, selected nuclear events hosted in red galaxies. They defined red galaxies (using SDSS colours) as galaxies that have $u - g > 1$ mag and $g - r > 0.5$ mag. For transients outside of the SDSS footprint,

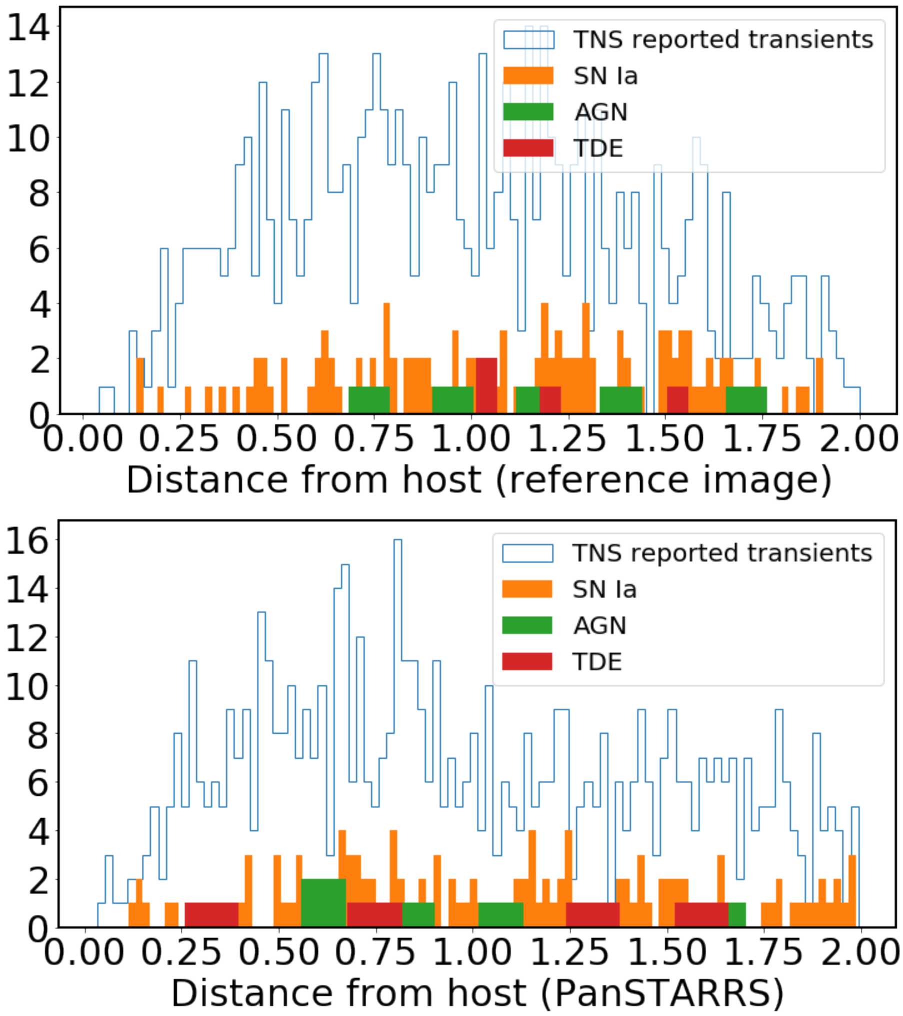


Figure 6.4: Distance from the centre of the host galaxy of TNS reported transients measured as the distance of the transient from the centre of the host on the template image (top panel) and as the distance of the transient from the centre of the host reported in the PanSTARRS catalogue.

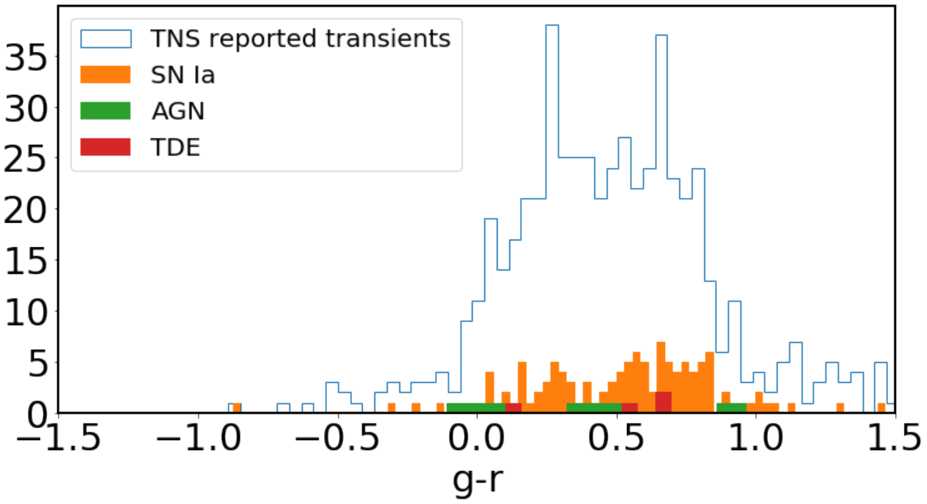


Figure 6.5: The g-r of host galaxies for TNS detected transients. The plot also shows the respective histograms for SN Ia, AGN and TDE.

they used $g - r > 0.5$ mag based PanSTARRS photometry, since the u filter is absent in PanSTARRS. To study our sample, I have used the PanSTARRS catalogue which is more complete in terms of GOTO sky coverage. We find, however, that only three of the four TDEs in the GOTO candidate list satisfy this host galaxy colour cut, which is consistent with the findings of van Velzen et al. (2020), who showed that such colour cuts tend to miss TDEs hosted in “greener” galaxies.

In the case of the TNS reported transients in this study, 289 out of the 628 (∼45%) sources reside in galaxies with $g - r > 0.5$. Using this cut therefore results in a 55% decrease in the number of candidates, yet still delivers 75% of TDEs, implying that this would, indeed, be beneficial. We note, however, that this is based on small number statistics. Similarly, of the 393,045 candidates in the GOTO catalogue, 183,602 (∼47%) have $g - r > 0.5$. It is usually common in the studies which look for TDEs in archival data to exclude galaxies that host AGN. This is because AGN are more common than TDEs and for the most of them it is very hard to distinguish a flare due to the AGN variability from a flare due to a TDE. There are a few TDE candidates that have been found in galaxies that include some kind of AGN activity (e.g., Blanchard et al. 2017, Tadhunter et al. 2017), which indicates that selections can indeed be biased when AGN hosts are excluded from TDE studies. French et al. (2020) in their review, examine TDE hosts and identify some in hosts with on-going gas accretion by a SMBH. They urge caution when interpreting their results due to the selection biases and uncertainties associated with identifying low-luminosity AGN. In the case of the 546,934 detections in the candidate list, there are 1,233 that are associated with sources in the Million Quasars Catalogue (Flesch 2015, v6.3) and an additional 46 after cross-matching them with the WISE AGN R90 catalogue (Assef et al. 2018). The low number is not sufficient to allow us to extract safe conclusions on the number of sources we could eliminate using this selection. Avoiding known AGN, however, would be useful to select nuclear transients that are most likely due to a TDE.

An additional important parameter is the colour of the transient whereby TDEs are found to be bluer than e.g. supernovae and thus has been used in previous studies to select TDE candidates (e.g., Arcavi et al. 2014, Hung et al. 2018). Although GOTO does not run its standard survey in another band, it is equipped with Baader RGB filters which are available in case an interesting transient is identified. The previously mentioned steps would, however, need to be made more efficient (in terms of reducing the number of potential candidates) to be able to trigger follow-up observations with other filters.

Within this section, I have identified major problems associated with identifying TDEs in current GOTO data. The first is that the realbogus classifier fails to give a high "real" score to known TDEs (or even most known transients). The second is that the measured distance from the nucleus is, in most cases, comparable to other types of extragalactic transient (e.g., AGN, SN Ia). The GOTO collaboration is currently in the process of developing an improved `realbogus` classifier, and early results show that this will be significantly better at identifying nuclear transients. The problems associated with the nuclear separation represent a more significant challenge, not least because they may stem from one or more of a number of causes. For example, imperfect astrometry solutions, GOTO's large pixel scale, a varying PSF etc. could all contribute to this problem. It is hoped, however, that improvements to the overall processing pipeline will help.

To end this section, I summarise the steps which we suggest could lead to the selection of TDE candidates within GOTO data:

- First, it is important to flag if the detection is found in an extended source which will remove detections from variable stars. In the case of the studied sample, there are 1,892,979 detections, there are 546,934 detections in which host is an extended source according to the `rMeanPSFMag-rKronMag`$\geq$0 criterion.
- If we also select detections in "redder" galaxies according to the criterion $g-r$ >0.5, it would return 384,341 detections.
- AGN hosts exclusion (using the Million Quasar Catalogue and the WISE R90 catalogue) eliminates only 1,279 sources.

Although, the selection of extended and red hosts work well to exclude sources, the realbogus value and the distance from the center of the host currently fail to separate real from fake sources and nuclear from non-nuclear detections in an efficient way. This leaves us with large numbers of possible TDE candidates. In the next section, I investigate whether the GOTO lightcurve data can help in the identification of TDE candidates.

6.3 TDE candidates identification through GOTO lightcurves

In this section, I investigate whether GOTO lightcurve data can be used to help identify TDE candidates using photometric data. I also explore whether it would be possible to identify TDEs, pre-peak, using GOTO lightcurves. Throughout this thesis, we have introduced and studied various lightcurves obtained using different techniques i.e. photometry from "blind" detection on GOTO images, difference image photometry and forced photometry using the LSST stack. In this section, I focus on the difference imaging lightcurves because in this case the host contribution is already subtracted. In the future, however, forced photometry will also be explored as additional tool when using lightcurve data to select TDE candidates.

In order to study TDEs, I fit the lightcurves using the Modular Open Source Fitter for Transients (MOSFiT; Guillochon et al. 2018), a `Python` based package which uses Monte Carlo methods for lightcurve fitting and produces the associated Bayesian parameter posteriors. While the main purpose of the package is to fit lightcurve data, it can be also used to generate mock lightcurves of transients (e.g., TDE, SN, kilonovae). For TDEs, the MOSFiT package uses the model of Mockler et al. (2019), which uses hydrodynamical simulations of disruptions of $1\,\mathrm{M}_{\odot}$stars by a $10^6\,\mathrm{M}_{\odot}$SMBH to model the mass fallback rate. Then, it uses scaling relations and interpolations for a range of black hole masses, star masses and nine other parameters i.e. the scaled impact parameter, b, which determines whether the star was partially or fully disrupted; the efficiency, ϵ, of converting accreted mass to energy; the normalisation and power-law index, $R_{ph,0}$ and l_{ph}, connecting the radius of the photosphere to the instantaneous luminosity; the viscous delay time, T_{ν}, (the time taken for matter to circularise and/or move through the accretion disk), the time of first fallback, t_0, the extinction, proportional to the hydrogen column density, n_H, in the host galaxy; and a white noise parameter, σ (Mockler et al. 2019, Nicholl et al. 2020).

First, I investigate whether it is possible to use MOSFiT to fit GOTO lightcurves in the archival data in order to identify clear TDE candidates from photometric data. In order to investigate this, I choose to fit the difference imaging lightcurve of the best case (i.e., most epochs) in our sample of known TDEs, i.e. AT2019qiz. This event has been studied by Nicholl et al. (2020) and I compare our results from fitting the current GOTO data against the results reported in that study. As mentioned in previous chapters, we use the L-band[5] photometry of GOTO which was used to conduct the standard survey mode. The GOTO difference imaging photometry of AT2019qiz contains 5 detected epochs and since these are measured from difference imaging the host contribution is already subtracted. The respective ZTF r and g band lightcurves presented in Nicholl et al. (2020) have ∼30 detected epochs[6]. We can expect that since there are 10 fitted parameters for our model and only 5 data points in our lightcurve, there will be large uncertainties in the posterior distributions.

In Figure 6.6, I present the 18 "highest" score (>15; which is still very low) realizations which corresponds to the models with the best fitting. The low scores (i.e., highest score is ∼23) are expected from the small number of data points, which also cause high uncertainties in the fitted parameters. In the Table 6.1, I present the posterior parameter values from the fitting, together with their errors which represent the 16-84 percentiles and of the posterior probability distributions. From the table, we can see that even though there are few epochs in the GOTO lightcurve, the hydrogen column density of the galaxy is predicted quite well from fitting the current GOTO data against the results reported in that study Nicholl et al. (2020) ($\log(N_H)= 20.03^{+0.26}_{-0.48}\mathrm{cm}^{-2}$; therefore within uncertainties). The black hole mass reported by Nicholl et al. (2020) is $\log(\mathrm{M}_{BH}\ /\mathrm{M}_{\odot})= 5.89^{+0.05}_{-0.06}$, which is outside the range of our uncertainties. We note, however, that we gave a slightly more relaxed range of prior black hole masses ([5,8.7] instead of [5,8]). The star mass and the other parameters have larger uncertainties and deviate from the values predicted from the aforementioned work. For example, from the GOTO lightcurve fitting, the star mass is $0.57^{+0.33}_{-0.47}$ while Nicholl et al. (2020) predicts a solar mass star with low uncertainties. Also, we infer a date of first fallback of MJD

---

[5] The wavelength information is also used by MOSFiT when fitting the lightcurves.

[6] The transient was early identified as interesting by ZTF (i.e., pre-peak) for photometric and spectroscopic follow up which partly explains the large number of epochs.

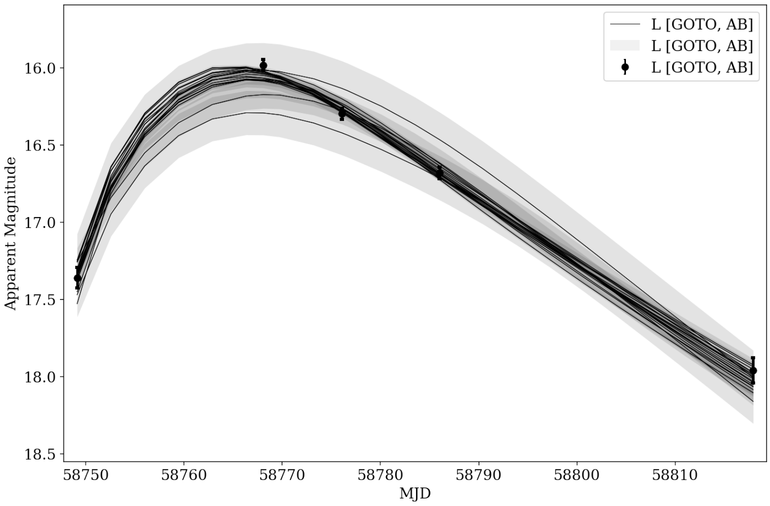


Figure 6.6: Fitting to the L-band lightcurve (from difference imaging) using the TDE model in MOSFiT.

58740, while the same value in Nicholl et al. (2020) is 58729. We conclude that even though the host properties could be estimated to some degree, the low sampling affects the results even if the full (i.e., rising and fading), lightcurve exists, as is the case for AT2019qiz. Another important drawback is the low scores of the fitting for which a high number of iterations (i.e., 55,000 and also 200 walkers) was used. We do note, however, that larger samples of TDEs will need to be tested in order to extract further conclusions on whether GOTO lightcurves from archival data can be used in order to identify strong TDE candidates photometrically. As we shall see in the final chapter of this thesis, this is part of our planned future work.

6.3.1 Mock lightcurves

In this section, I try to evaluate whether lightcurve fitting can be used in order to identify TDE candidates. In this sub-section, I focus on the output of the first tests to examine whether it will be possible to pick TDE candidates for spectroscopic follow-up when they are close to their peak. This is investigated in terms of cadence but also I investigate whether they could be separated from other transients, specifically SN, using the rising

Parameter	Prior	Posterior	Units
log(M_{BH})	[5,8.7]	$6.25^{+0.06}_{-0.11}$	$\mathrm{M}_{\odot}$
M_*	[0.01,100]	$0.57^{+0.33}_{-0.47}$	$\mathrm{M}_{\odot}$
b	[0,2]	$0.85^{+0.21}_{-0.12}$	
$\log(\epsilon)$	[-4,-0.4]	$-2.52^{+0.62}_{-0.78}$	
$\log(R_{ph,0})$	[-4,4]	$1.37^{+0.49}_{-0.41}$	
l_{ph}	[0,4]	$1.97^{+0.18}_{-0.41}$	
$\log(T_\nu)$	[-3,5]	$-1.43^{+1.32}_{-0.71}$	days
t_0	[-500,0]	$-10.11^{+1.52}_{-2.03}$	days
$\log(n_{H,host})$	[16,23]	$18.45^{+1.44}_{-1.78}$	cm^{-2}
$\log(\sigma)$	[-3,2]	$-2.11^{+0.90}_{-0.67}$	

Table 6.1: Priors and marginalised posteriors for the MOSFiT TDE model. Priors are flat within the stated ranges, except for M_*, which uses a Kroupa initial mass function. The quoted results are the median of each distribution, and error bars are the 16th and 84th percentiles.

part of the lightcurve and MOSFiT to fit it.

While MOSFiT's main purpose is to provide the models to fit different transients, it can, however, be used to generate mock lightcurves of different transient events (e.g., supernovae, TDE, kilonovae, etc.). To generate simulated lightcurves with MOSFiT the user can simply pass the model of the transient to the code.[7] The user can also adapt the parameters of the model to generate samples of transient events. Our goal is to generate mock lightcurves of TDEs with a different sampling each time and investigate whether we are able to recover the type of the transients (i.e., TDE), while comparing it with the fitting of supernovae models. SN are the main source of contamination of the TDE candidate samples and so we focus on these.

In order to simulate L-band lightcurves, I provide the MOSFiT code with the filter throughput information and generate mock GOTO lightcurves. To simulate the GOTO lightcurves, I also use the limiting magnitude option from MOSFiT where I use 20.0 magnitude as the limiting magnitude with 0.5 as its variance. In this case MOSFiT sets

---

[7] In the case of SN there are various models.

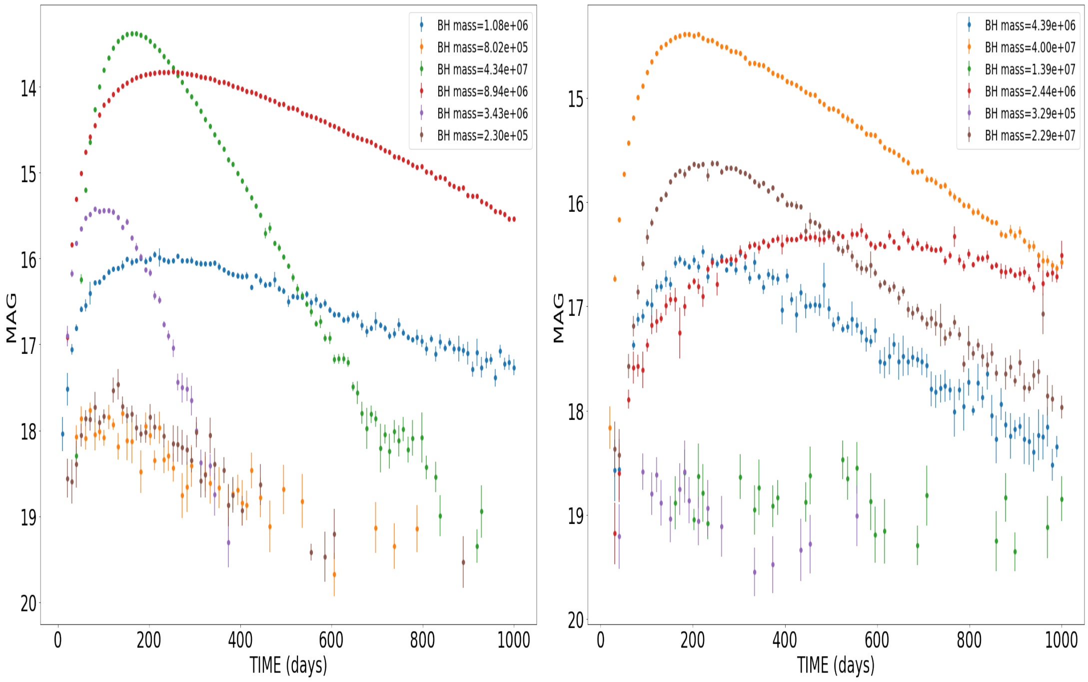


Figure 6.7: Examples of mock L-band TDE lightcurves with MOSFiT at redshift 0.05 (left) and 0.1 (right) for different black hole masses disrupting an 1 solar mass star. The shape of the lightcurve depends on the black hole mass, if the star is fully or partially disrupted, the mass of the disrupted star and the structure of the star, the accretion process and the time of the disruption. In our case, although I have used a complete disruption of 1 solar mass stars in all cases, I allowed the black hole mass and the parameters associated with the structure of the star to vary which results different shaped TDE lightcurves.

the limiting magnitude for the instrument and then randomly draws observations based upon the flux error implied by that limiting magnitude. In Figure 6.7 I present different realizations of model TDEs in which I have varied the redshift and mass of the black hole. I also explored different values for the time of the first fallback and the scaled impact parameter in order to test how this affects the simulations. In the future, we will also investigate different values for the mass of the disrupted star. In the plots presented in Figure 6.7 the $t_{\mathrm{first-fallback}}$ spans the range [0,50], b=1, and the black hole mass spans the range [10^5, 10^9] $\mathrm{M}_\odot$.

MOSFiT provides the user with many tools to study and predict transient observations. Here, however, I focus on investigating how different cadences affect how confidently we can predict a TDE (i.e., whether we will be able to use the pre-peak lightcurve to select TDE candidates). Figures 6.8 and 6.9 show two different cases of TDE mock lightcurves. The first shows examples where I have only modelled three data points pre-peak (10 days cadence), while the second shows more well-sampled lightcurves (1 day cadence). I run a code for 200 walkers and 3000 iterations for each model. In addition to fitting the TDE model to these mock TDE lightcurves, I also attempt to fit three supernovae models in order to ascertain whether we are able to distinguish between TDEs and supernovae.

Figures 6.8 and 6.9 show the cases of the low-sampled lightcurve and the well sampled lightcurve with the aforementioned fitted models. It is obvious that cadence affects the score (i.e., how well the model describes the lightcurve) of the models as better statistics provide tighter constrains on the model parameters. Different lightcurves also show that in some cases the differences of the scores between the models can be large and in other cases there is slight difference in the median score of the models (e.g., Figure 6.9) where we find that the `tde` model has insignificantly larger median score than the other models. MOSFiT, however, can provide an estimate of the fitting even when the lightcurve is still on the rise as the examples presented here. Other methods (e.g., fitting of power law; van Velzen et al. 2019) can be used to further select TDE candidates from large sample of transients. With current tests we find that using the rising part of the lightcurve and MOSFiT fitting, it is not possible to distinguish between TDEs and SN even for the well-sampled cases. By using a larger sample of ZTF and Gaussian and exponential fits for the rising and the fading part of their lightcurves, van Velzen et al. (2019) and van Velzen et al. (2020), show that they are able to distinguish between TDEs and SN using the rising time, fading time and colour evolution. We find, however, than even from their study, it is easier to separate the two from their fading time and colour evolution rather than the difference in their rising times. In the future, we aim to generate more mock lightcurves with different prior parameters and also use larger samples of real data to evaluate our conclusions using larger statistics.

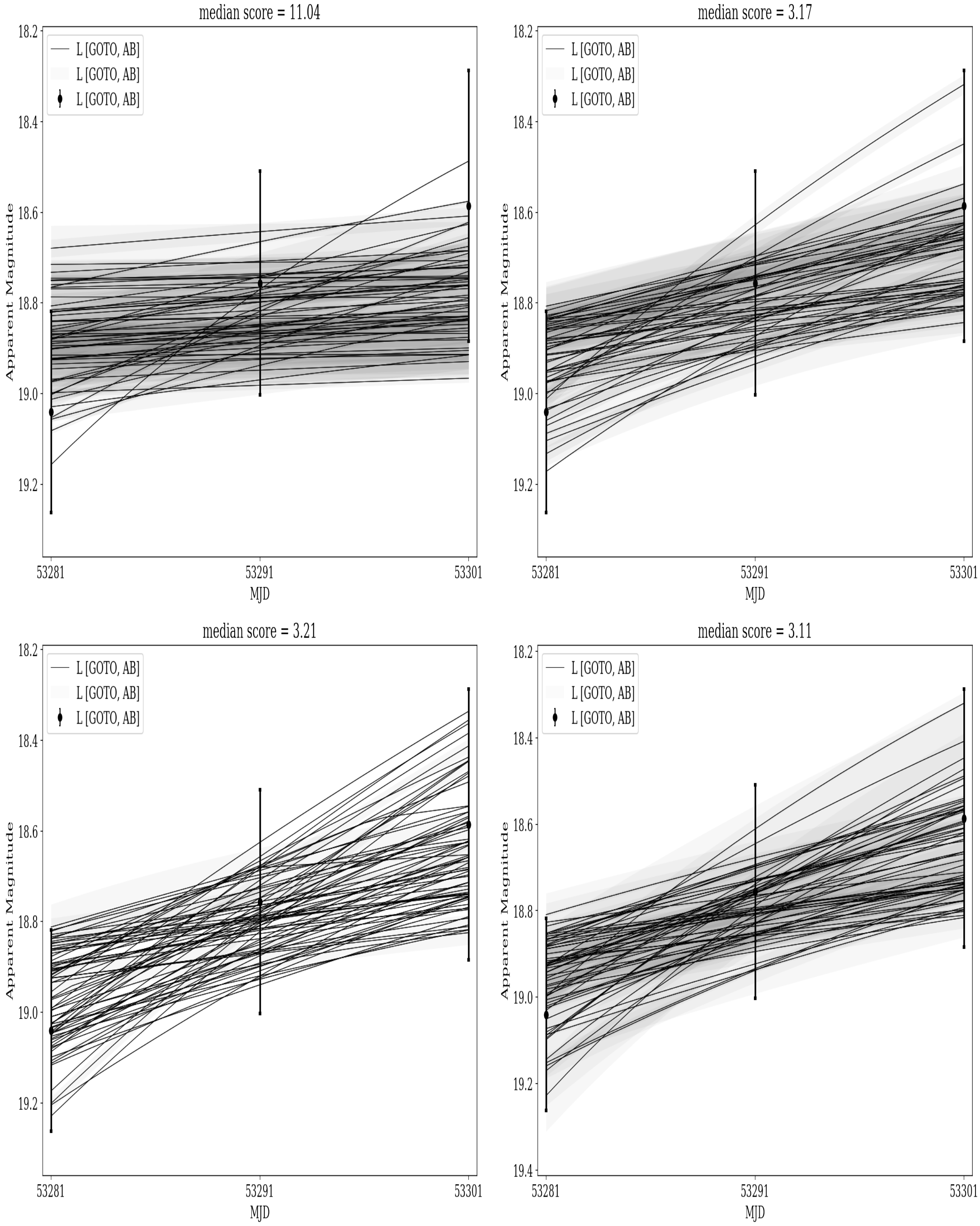


Figure 6.8: Low-sampling (10 days cadence) lightcurve and its fitting using the `tde` model (top left), the `default` model (top right), the `csm` model (bottom left) and the `slsn` model (bottom right). The median scores of the model are found at the top of each plot.

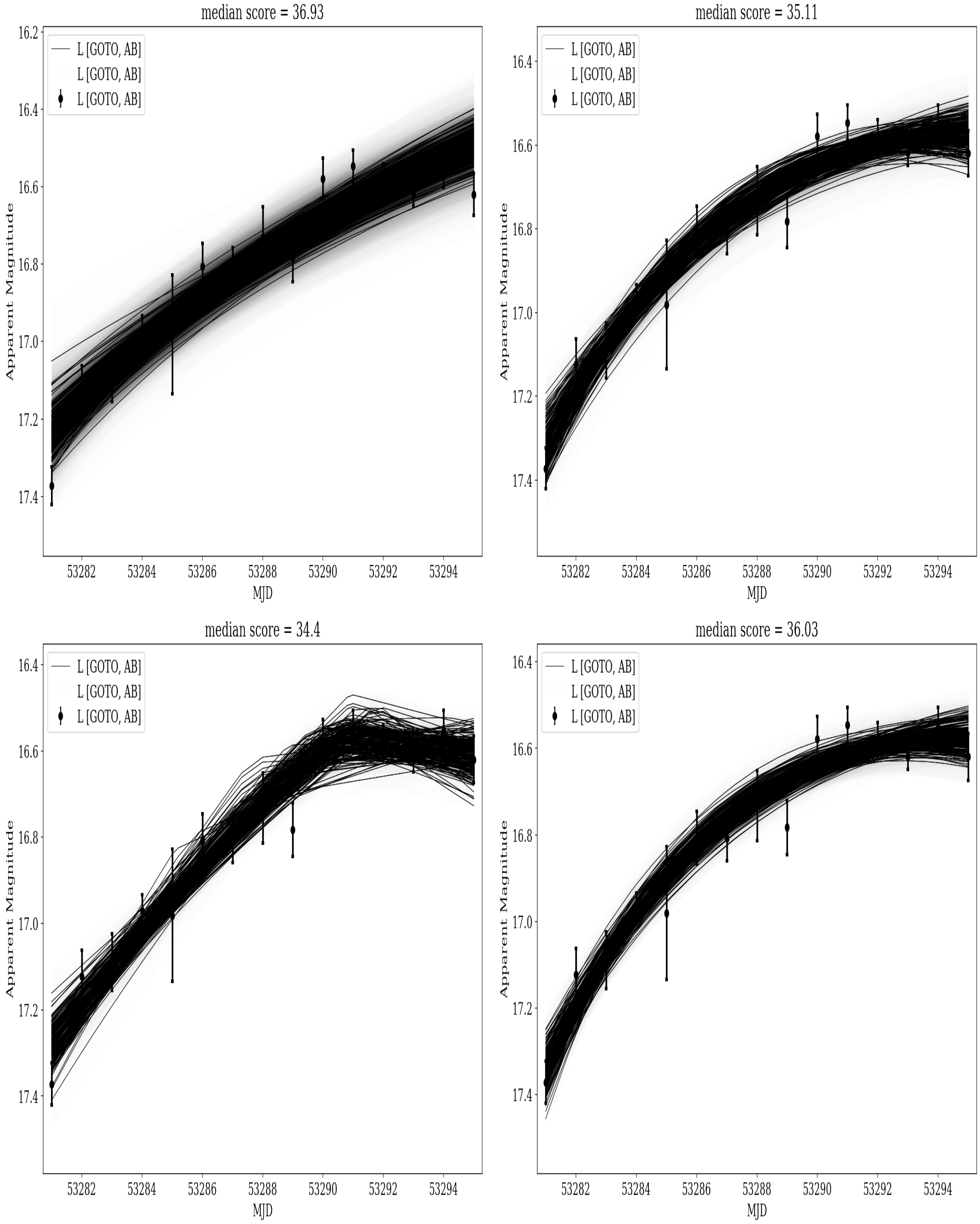


Figure 6.9: Well-sampled (1 day cadence) lightcurve and its fitting using the `tde` model (top left), the `default` model (top right), the `csm` model (bottom left) and the `slsn` model (bottom right). The median scores of the model are found at the top of each plot.

6.4 Chapter conclusions

From Equation 7 of van Velzen et al. (2011), we estimate that GOTO will potentially be able to detect ∼10-20 TDEs per year, with the given the rate of TDEs presented in that paper and GOTO's L-band magnitude limit of ∼20 for the full GOTO-N and GOTO-S set up. To be able to recover all these TDEs pre-peak, GOTO will need to use both spatial and temporal information efficiently to sift them from other non-TDE transients. In this chapter I presented the various tools that can be used in the selection process i.e., `realbogus` value, distance from host, host properties (section 6.2). We find that, at present, we obtained very high numbers of candidates when we tried to recover a small number (i.e., 4) known TDEs using the selection criteria outlined in sections 6.2. Also, we find that pre-peak TDE lightcurves are currently difficult to be distinguished from SNe lightcurves with high confidence even with higher cadence observations, which however, needs to be tested for larger samples. Higher cadences, however, provide tighter constrains on the model parameters and will also be useful for full (i.e., rising-fading) lightcurves. As we shall see in the next section future work will include testing the selection method for data processed with GOTO's new classifier, which is currently under development. Larger samples of mock and real lightcurves will also be tested to further demonstrate if pre-peak lightcurves can be used, complementary to the other selection criteria, to pick TDE candidates for spectroscopic and photometric follow-up near their peak flux.

Chapter 7

Conclusions

The study of SMBHs at the centres of the galaxies has proven a fascinating and challenging area of research with observations and theory trying to find a common ground. The discovery of the most luminous AGN (i.e quasars; Schmidt 1963) and the further realisation that these objects must be powered by an extremely compact engine provided the first indirect evidence of SMBHs. Soon after the discovery of quasars, it was found that these objects vary with time (Matthews & Sandage 1963) over various timescales i.e. days to decades (e.g., Fahlman & Ulrych 1975, Edelson et al. 1996). This property became one of the most powerful means of developing our understanding of the central engines of AGN (i.e., the SMBH surrounded by an accretion disk).

Just a few years ago, the study of AGN variability suffered severely from low number statistics, which many studies focusing on just a handful of AGNs. Recently, however, wide field surveys (e.g., SDSS, PTF) have dramatically changed this. Wide-field, high-cadence surveys are now leading the way in our understanding of AGN variability (e.g., Vanden Berk et al. 2004, MacLeod et al. 2010, Caplar et al. 2017). These surveys have also enabled, through their delivery of large amounts of data, the discovery of phenomena that had previously only been predicted by theory. One such example is TDEs which, despite having been predicted since the 1970s (Hills 1975), only received optical confirmation in the last 10 years (e.g., van Velzen et al. 2011, Arcavi et al. 2014).

Current (e.g., GOTO, ZTF) and future (e.g., LSST) surveys will provide the astronomical community with well-sampled lightcurves spanning multiple years (and eventually decades). Such data resources will allow us to study AGN variability and TDEs in greater detail to provide greater insights into SMBHs and their immediate environments. In this work, I have investigated the prospects of using GOTO to study these two phenomena. In the near future, the GOTO survey will be conducted by observatories in both the northern and southern hemispheres, each equipped with 16 UTs to achieve a cadence of around 2 days. These characteristics make GOTO, at least in theory, ideal for studies that require well-sampled lightcurves of large numbers of sources. It was our goal at the outset of this project to assess whether GOTO, in its current guise, can be used to study AGN variability and TDEs and, if not, what must be improved to reach this goal.

With GOTO's primary objective being the identification of optical counterparts to gravitational-wave events, its primary processing pipeline is focused around difference imaging, since this is recognised as the best method so far to identify so-called transient sources. By contrast, variability studies can instead benefit from forced photometry. As such, it was deemed necessary for us to develop the means to perform forced photometry on nightly GOTO observations. In order to do this, we decided to use the LSST stack, which already contains forced photometry modules. To use the LSST stack, however, I had to adapt it for the GOTO telescope by developing the necessary package (i.e., `obs_goto`) that works as an interface between the software and the data (see chapter 2). After successfully adapting the LSST stack for GOTO we were able to generate the reference catalogue for approximately 50% of the GOTO observable sky by coadding images obtained in February-March 2019. Then, I performed forced photometry on nightly GOTO data from February 2019 to June 2020. In Chapter 3, we characterise the results for the reference catalogue and the forced photometry data (that cover the period February 2019 - October 2019) by comparison with external catalogue (i.e., PanSTARRS and GOTOphoto) and also calculating the RMS of non-variable sources. I find that photometry measurements obtained by the LSST-stack on GOTO data agrees with PanSTARRS photometry to within 10 mmag (1σ) in the case of bright sources (m_L ∼14), and to within 200 mmag in the case of faint sources (m_L ∼18). I obtain similar results when comparing to blind source photometry measurements delivered by GOTO's in-house pipeline, GOTOphot. The RMS for the period we used for the characterisation was found to be typically better than ∼20mmag for sources brighter than 16 mag.

In Chapter 4, I investigated AGN variability using GOTO data from its prototype phase (4 UTs). I used the fact that our sample contains both Type 1 and Type 2 AGN to assess whether GOTO is able to measure AGN variability. This is based on the prior knowledge that Type 1 AGN show much higher levels of variability than Type 2 AGN (e.g., Cartier et al. 2015). We found, however, that we were unable to distinguish between Type 1 and Type 2 AGN via their GOTO lightcurves (whether from the LSST stack or GOTOphoto processed data), which I interpreted as meaning that we are currently unable to study AGN variability using GOTO data. This was confirmed when I performed the same analysis on ZTF lightcurves of the same parent sample of AGN, thus indicating the problem arises within the GOTO data (or processing thereof) and not the methods used to analyse the lightcurves or the nature of the parent sample. We noted, however, that this may also be affected by the lower statistics for observations separated by shorter (1-5 days) time lags in GOTO data compared to ZTF data. Whilst from simulations we predict that this situation may improve GOTO's ability to study AGN variability, at present we suspect that its main flaw resides in poorly-understood systematic uncertainties.

Being unable to use the GOTO data to study AGN variability and realising that with ZTF lightcurves we were able to distinguish the Type 1 and Type 2 AGN in the A-γ parameter space, I decided to investigate the variability of our AGN sample using ZTF data. I analysed the g-band and the r-band lightcurves using SF fitting and also the excess variance as an additional way to quantify the amplitude of variability (Chapter 5). I then investigated how A, γ and σ_{rms} correlate with luminosity, Eddington ratio, redshift and black hole mass. We recover the weak anti-correlation between the amplitude of variability and the Eddington ratio previously found for samples of higher luminosity and/or higher redshift studies (e.g., MacLeod et al. 2010, Sánchez-Sáez et al. 2018). I do not, however, find any dependence on luminosity, which is in agreement with Sánchez-Sáez et al. (2018) but disagrees with previous studies based on higher luminosity AGN (e.g., Caplar et al. 2017). Our results for A seem to be further affected by the host contribution as our sample contains more moderate luminosity AGN at lower redshifts ($z < 0.4$). Evidence for host contamination also arises in the form of lower A values from the r-band lightcurves compared to those from the g-band, since the host is expected to be more prominent at longer wavelengths (e.g., Kimura et al. 2020). The γ values do not correlate with any of the physical properties in this study which further challenges previous models based on higher luminosity AGN samples (e.g., Kelly et al. 2009). Another finding of this study is the low γ values, which are unaffected of the wavelength, compared to those expected from the DRW model. Longer lightcurves, however, will be needed to confirm this result.

Finally in Chapter 6 I investigated the prospects of using GOTO to detect TDEs and study their lightcurves. For this study I have used known TDEs found in the ZTF survey. I identified the problems arising from both the current `realbogus` flags and the current measured distance from the host nucleus, as measured by the GOTOphoto pipeline. The first assigns very low values (< 0.5) to real transient events while the second one does not allow us, at present, to distinguish between nuclear and non-nuclear transients. A new classifier is under development and aims to address these issues. I also investigated if pre-peak lightcurves can help with the identification of TDE candidates for fast spectroscopic and photometric follow-up. We find, however, that using only the rising part of the lightcurves (and in the absence of colour information) provides too little information to confidently distinguish between TDEs and SNe. We do note, however, that the TDE study is in its early stages and further studies with larger samples will need to be processed in order to extract more statistically significant conclusions.

As with many wide-field optical surveys, GOTO provides a means of preparing for the LSST. The thesis allowed us to gain valuable lessons and experience in terms of software, data and methodology from GOTO as one of the precursors of LSST. In terms of software, the use of the LSST stack gave us the chance to communicate obstacles and results to the LSST community, giving feedback on the various tools used within this thesis (e.g. PSF modelling, coaddition, forced photometry). The science drivers of this thesis are also two of the science drivers of the LSST (i.e. AGN variability and TDEs). The method used to confirm the identification of AGN variability (i.e. discrepancy between Type 1 and Type 2 AGN) can be used to evaluate the data from the first year of LSST (we note that we obviously expect far less systematics). Forced photometry has also proved to be a powerful tool for variability studies and in this thesis I evaluated the LSST stack task used to conduct this. Finally, the TDE study is part of the framework that many groups contribute to in order to prepare LSST for the best way of "hunting" for these events. In this work, I identified the issues associated with this task and the importance of developing pipelines that will eliminate most of the contamination, which prevents us from selecting a viable sample of TDE candidates for follow-up. I have also highlighted the significance of the performance of the software (e.g. image difference, realbogus classifiers) in order to make the selection more efficient.

7.1 Future Work

Future work is focused on the problems and challenges identified throughout this thesis. First, in the case of the LSST stack work, we aim to coadd data for the whole GOTO observable sky choosing the best quality (i.e., based on PSF FWHM) data to coadd. This will provide us with a more complete reference catalogue for forced photometry. For the forced photometry, we we intend to focus on improving the photometric calibration, especially in terms of deriving more robust calibration errors via an improved selection of calibration sources. We also intend to focus more on improving the PSF modelling and the subsequent improvement in PSF photometry, especially since the latter is better for fainter sources (which make up the vast majority of sources in any survey).

As the age of the GOTO survey increases, we hope to investigate whether the increasing timespan of our lightcurves improves its ability to study AGN variability. After developments in both pipelines the goal is to re-process the whole GOTO data since February 2019 and test the uncertainties of the GOTO lightcurves. As the studied sample contains low-redshift AGN we also wish to investigate the best way to remove the host contamination to improve the results for the amplitude of variability which is affected by this. Ultimately, we aim to combine different optical wavelength data (SDSS, PTF, ZTF, GOTO) in the best way to generate decades long lightcurves for the sample used in this thesis to study short and long term variability.

With regards to the TDE study, as mentioned before, we aim to study larger samples of TDEs and also use the data from the new pipeline and classifier development to ascertain if it does indeed address the problems identified in this work. We also aim to investigate if forced photometry can be used as an additional means to select TDE candidates from photometric data along with difference imaging. Identifying strong TDE candidates from photometric data will enable the study of TDE samples in terms of host properties and lightcurves. Ultimately, we aim to find the best possible means to identify events near their peak luminosity to enable follow-up spectral observations leading to far greater insights into the nature of TDEs.

Bibliography

Aihara H., AlSayyad Y., Ando M., Armstrong R., Bosch J., Egami E., et al., 2019, Publications of the Astronomical Society of Japan, 71, 114

Aihara H., Armstrong R., Bickerton S., et al., 2018, Publications of the Astronomical Society of Japan, 70, S8

Aird J., Nandra K., Laird E. S., Georgakakis A., et al., 2010, MNRAS, 401, 2531

Antonucci R. R. J., 1983, Nature, 303, 158

Arcavi I., Gal-Yam A., Sullivan M., Pan Y.-C., Cenko S. B., Horesh A., et al., 2014, ApJ, 793, 38

Assef R. J., Stern D., Noirot G., Jun H. D., et al., 2018, ApJS, 234, 23

Bade N., Komossa S., Dahlem M., 1996, A&A, 309, L35

Baldassare V. F., Geha M., Greene J., 2018, ApJ, 868, 152

Baldassare V. F., Geha M., Greene J., 2019, arXiv e-prints, p. arXiv:1910.06342

Baldwin J. A., Phillips M. M., Terlevich R., 1981, PASP, 93, 5

Bauer A., Baltay C., Coppi P., Ellman N., Jerke J., Rabinowitz D., Scalzo R., 2009, ApJ, 696, 1241

Becker A., , 2015, HOTPANTS: High Order Transform of PSF ANd Template Subtraction

Bellm E. C., Kulkarni S. R., Graham M. J., et al., 2019, PASP, 131, 018002

Bennert N., Jungwiert B., Komossa S., Haas M., Chini R., 2006, A&A, 456, 953

Bernstein G. M., Jarvis M., 2002, AJ, 123, 583

Bertin E., 2011, in Evans I. N., Accomazzi A., Mink D. J., Rots A. H., eds, Astronomical Data Analysis Software and Systems XX Vol. 442 of Astronomical Society of the Pacific Conference Series, Automated Morphometry with SExtractor and PSFEx. p. 435

Bertin E., Arnouts S., 1996, A&A Suppl., 117, 393

Blagorodnova N., Cenko S. B., Kulkarni S. R., Arcavi I., Bloom J. S., Duggan G., et al., 2019, ApJ, 873, 92

Blanchard P. K., Nicholl M., Berger E., Guillochon J., et al., 2017, ApJ, 843, 106

Blandford R. D., Königl A., 1979, ApJ, 232, 34

Bloom J. S., Richards J. W., Nugent P. E., Quimby R. M., Kasliwal M. M., et al., 2012, PASP, 124, 1175

Bosch J., Armstrong R., Bickerton S., Furusawa H., Ikeda H., Koike M., Lupton R., Mineo S., Price P., et al., 2018, Publications of the Astronomical Society of Japan, 70, S5

Bricman K., Gomboc A., 2020, ApJ, 890, 73

Burtscher L., Meisenheimer K., Tristram K. R. W., Jaffe W., Hönig S. F., Davies R. I., Kishimoto M., Pott J. U., Röttgering H., Schartmann M., Weigelt G., Wolf S., 2013, A&A, 558, A149

Caplar N., Lilly S. J., Trakhtenbrot B., 2017, ApJ, 834, 111

Carter B., Luminet J. P., 1983, A&A, 121, 97

Cartier R., Lira P., Coppi P., Sánchez P., Arévalo P., Bauer F. E., Rabinowitz D., Zinn R., Muñoz R. R., Meza N., 2015, ApJ, 810, 164

Chambers K. C., Magnier E. A., Metcalfe N., Flewelling H. A., Huber M. E., Waters C. Z., et al., 2016, arXiv e-prints, p. arXiv:1612.05560

Clavel J., Reichert G. A., Alloin D., Crenshaw D. M., et al., 1991, ApJ, 366, 64

Collier S., Peterson B. M., 2001, ApJ, 555, 775

Czerny B., 2004, arXiv e-prints, pp astro–ph/0409254

de Vries W. H., Becker R. H., White R. L., Loomis C., 2005, AJ, 129, 615

Dexter J., Agol E., 2011, ApJL, 727, L24

Dyer M. J., 2020, PhD thesis, University of Sheffield

Dyer M. J., Dhillon V. S., Littlefair S., Steeghs D., et al., 2018 Vol. 10704 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, A telescope control and scheduling system for the Gravitational-wave Optical Transient Observer (GOTO). p. 107040C

Edelson R. A., Alexander T., Crenshaw D. M., Kaspi S., et al., 1996, ApJ, 470, 364

Emmanoulopoulos D., McHardy I. M., Papadakis I. E., 2013, MNRAS, 433, 907

Esquej P., Saxton R. D., Komossa S., Read A. M., Freyberg M. J., Hasinger G., García-Hernández D. A., Lu H., Rodriguez Zaurín J., Sánchez-Portal M., Zhou H., 2008, A&A, 489, 543

Fahlman G. G., Ulrych T. J., 1975, ApJ, 201, 277

Fath E. A., 1909, Lick Observatory Bulletin, 149, 71

Flesch E. W., 2015, PASA, 32, e010

French K. D., Arcavi I., Zabludoff A., 2016, ApJL, 818, L21

French K. D., Wevers T., Law-Smith J., Graur O., Zabludoff A. I., 2020, Space Science Reviews, 216, 32

Gezari S., Chornock R., Rest A., Huber M. E., et al., 2012, Nature, 485, 217

Gezari S., Heckman T., Cenko S. B., Eracleous M., Forster K., Gonçalves T. S., Martin D. C., Morrissey P., Neff S. G., Seibert M., Schiminovich D., Wyder T. K., 2009, ApJ, 698, 1367

Gezari S., Hung T., Cenko S. B., Blagorodnova N., et al., 2017, ApJ, 835, 144

Gibson R., Anderson S. F., Ballantyne D. R., et al., 2011, in American Astronomical Society Meeting Abstracts #217 Vol. 217 of American Astronomical Society Meeting Abstracts, AGN Science with LSST. p. 252.17

Giveon U., Maoz D., Kaspi S., Netzer H., Smith P. S., 1999, MNRAS, 306, 637

Górski K. M., Hivon E., Banday A. J., Wand elt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759

Graham M. J., Djorgovski S. G., Drake A. J., Mahabal A. A., Chang M., Stern D., Donalek C., Glikman E., 2014, MNRAS, 439, 703

Graham M. J., Kulkarni S. R., Bellm E. C., et al., 2019, PASP, 131, 078001

Greene J. E., Ho L. C., 2005, ApJ, 630, 122

Guillochon J., Nicholl M., Villar V. A., Mockler B., Narayan G., Mandel K. S., Berger E., Williams P. K. G., 2018, ApJS, 236, 6

Gunn J. E., Stryker L. L., 1983, ApJS, 52, 121

Hawkins M. R. S., 1993, Nature, 366, 242

Hawkins M. R. S., 2002, MNRAS, 329, 76

Hills J. G., 1975, Nature, 254, 295

Hodapp K. W., Siegmund W. A., Kaiser N., Chambers K. C., Laux U., Morgan J., Mannery E., 2004, in Oschmann Jacobus M. J., ed., Ground-based Telescopes Vol. 5489 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Optical design of the Pan-STARRS telescopes. pp 667–678

Holoien T. W. S., Kochanek C. S., Prieto J. L., Stanek K. Z., Dong S., Shappee B. J., Grupe D., et al., 2016, MNRAS, 455, 2918

Hook I. M., McMahon R. G., Boyle B. J., Irwin M. J., 1994, MNRAS, 268, 305

Hopkins P. F., Hernquist L., Cox T. J., Di Matteo T., Martini P., Robertson B., Springel V., 2005, ApJ, 630, 705

Hughes P. A., Aller H. D., Aller M. F., 1992, ApJ, 396, 469

Hung T., Gezari S., Blagorodnova N., Roth N., Cenko S. B., Kulkarni S. R., Horesh A., Arcavi I., McCully C., Yan L., Lunnan R., Fremling C., Cao Y., Nugent P. E., Wozniak P., 2017, ApJ, 842, 29

Hung T., Gezari S., Cenko S. B., van Velzen S., Blagorodnova N., Yan L., Kulkarni S. R., Lunnan R., Kupfer T., Leloudas G., Kong A. K. H., Nugent P. E., Fremling C., Laher R. R., Masci F. J., Cao Y., Roy R., Petrushevska T., 2018, ApJS, 238, 15

Ivezić Ž., Kahn S. M., Tyson J. A., Abel B., Acosta E., Allsman R., Alonso D., AlSayyad Y., et al., 2019, ApJ, 873, 111

Ivezić Ž., Smith J. A., Miknaitis G., et al., 2007, AJ, 134, 973

Jurić M., Kantor J., Lim K. T., Lupton R. H., et al., 2017, The LSST Data Management System. p. 279

Kasliwal V. P., Vogeley M. S., Richards G. T., 2015, MNRAS, 451, 4328

Kaspi S., Maoz D., Netzer H., Peterson B. M., Vestergaard M., Jannuzi B. T., 2005, ApJ, 629, 61

Kawaguchi T., Mineshige S., 1999, in Terzian Y., Khachikian E., Weedman D., eds, Activity in Galaxies and Related Phenomena Vol. 194 of IAU Symposium, AGN Variability. p. 356

Kawaguchi T., Mineshige S., Umemura M., Turner E. L., 1998, ApJ, 504, 671

Kelly B. C., Bechtold J., Siemiginowska A., 2009, ApJ, 698, 895

Kelly P. L., Kirshner R. P., 2012, ApJ, 759, 107

Kimura Y., Yamada T., Kokubo M., Yasuda N., Morokuma T., Nagao T., Matsuoka Y., 2020, ApJ, 894, 24

Kokubo M., 2015, MNRAS, 449, 94

Komossa S., 2015, Journal of High Energy Astrophysics, 7, 148

Kozłowski S., 2016, ApJ, 826, 118

Kozłowski S., 2017, A&A, 597, A128

Krolik J. H., Horne K., Kallman T. R., Malkan M. A., Edelson R. A., Kriss G. A., 1991, ApJ, 371, 541

Kron R. G., 1980, ApJS, 43, 305

LaMassa S. M., Cales S., Moran E. C., Myers A. D., Richards G. T., Eracleous M., Heckman T. M., Gallo L., Urry C. M., 2015, ApJ, 800, 144

Lang D., Hogg D. W., Mierle K., Blanton M., Roweis S., 2010, AJ, 139, 1782

Lanzuisi G., Ponti G., Salvato M., Hasinger G., et al., 2014, ApJ, 781, 105

Law N. M., Kulkarni S. R., Dekany R. G., Ofek E. O., Quimby R. M., et al., 2009, PASP, 121, 1395

Law-Smith J., Ramirez-Ruiz E., Ellison S. L., Foley R. J., 2017, ApJ, 850, 22

Lupton R., Gunn J. E., Ivezić Z., Knapp G. R., Kent S., 2001, in Harnden F. R. J., Primini F. A., Payne H. E., eds, Astronomical Data Analysis Software and Systems X Vol. 238 of Astronomical Society of the Pacific Conference Series, The SDSS Imaging Pipelines. p. 269

MacLeod C. L., Ivezić Ž., Kochanek C. S., Kozłowski S., Kelly B., Bullock E., Kimball A., Sesar B., Westman D., Brooks K., Gibson R., Becker A. C., de Vries W. H., 2010, ApJ, 721, 1014

MacLeod C. L., Ivezić Ž., Sesar B., de Vries W., Kochanek C. S., Kelly B. C., Becker A. C., Lupton R. H., Hall P. B., Richards G. T., Anderson S. F., Schneider D. P., 2012, ApJ, 753, 106

MacLeod C. L., Ross N. P., Lawrence A., Goad M., Horne K., Burgett W., Chambers K. C., Flewelling H., Hodapp K., Kaiser N., Magnier E., Wainscoat R., Waters C., 2016, MNRAS, 457, 389

Magnier E. A., Chambers K. C., Flewelling H. A., Hoblitt J. C., Huber M. E., Price P. A., Sweeney W. E., Waters C. Z., Denneau L., Draper P., Hodapp K. W., Jedicke

R., Kaiser N., Kudritzki R. P., Metcalfe N., Stubbs C. W., Wainscoast R. J., 2016, arXiv e-prints, p. arXiv:1612.05240

Maksym W. P., Ulmer M. P., Eracleous M., 2010, ApJ, 722, 1035

Matthews T. A., Sandage A. R., 1963, ApJ, 138, 30

Mejía-Restrepo J. E., Trakhtenbrot B., Lira P., Netzer H., Capellupo D. M., 2016, MNRAS, 460, 187

Miyazaki S., Komiyama Y., Kawanomoto S., et al., 2018, Publications of the Astronomical Society of Japan, 70, S1

Mockler B., Guillochon J., Ramirez-Ruiz E., 2019, ApJ, 872, 151

Mullaney J. R., Alexander D. M., Fine S., Goulding A. D., Harrison C. M., Hickox R. C., 2013, MNRAS, 433, 622

Mullaney J. R., Makrygianni L., Dhillon V., Littlefair S., Ackley K., Dyer M., et al., 2021, PASA, 38, e004

Netzer H., Trakhtenbrot B., 2007, ApJ, 654, 754

Nicholl M., Wevers T., Oates S. R., Alexand er K. D., Leloudas G., Onori F., Jerkstrand A., Gomez S., Campana S., Arcavi I., et al., 2020, arXiv e-prints, p. arXiv:2006.02454

Pereyra N. A., Vanden Berk D. E., Turnshek D. A., Hillier D. J., Wilhite B. C., Kron R. G., Schneider D. P., Brinkmann J., 2006, ApJ, 642, 87

Peterson B. M., Berlind P., Bertram R., Bochkarev N. G., Bond D., Brotherton M. S., Busler J. R., et al., 1994, ApJ, 425, 622

Pickles A. J., 1998, PASP, 110, 863

Rau A., Kulkarni S. R., Law N. M., et al., 2009, PASP, 121, 1334

Rees M. J., 1988, Nature, 333, 523

Ruan J. J., Anderson S. F., Dexter J., Agol E., 2014, ApJ, 783, 105

Salpeter E. E., 1964, ApJ, 140, 796

Samus' N. N., Kazarovets E. V., Durlevich O. V., Kireeva N. N., Pastukhova E. N., 2017, Astronomy Reports, 61, 80

Sánchez P., Lira P., Cartier R., Pérez V., et al., 2017, ApJ, 849, 110

Sánchez-Sáez P., Lira P., Cartier R., Mirand a N., Ho L. C., Arévalo P., Bauer F. E., Coppi P., Yovaniniz C., 2019, ApJS, 242, 10

Sánchez-Sáez P., Lira P., Mejía-Restrepo J., Ho L. C., Arévalo P., Kim M., Cartier R., Coppi P., 2018, ApJ, 864, 87

Sartori L. F., Trakhtenbrot B., Schawinski K., Caplar N., Treister E., Zhang C., 2019, ApJ, 883, 139

Schmidt K. B., Marshall P. J., Rix H.-W., Jester S., Hennawi J. F., Dobler G., 2010, ApJ, 714, 1194

Schmidt M., 1963, Nature, 197, 1040

Sesar B., Ivezić Ž., Lupton R. H., Jurić M., Gunn J. E., et al., 2007, AJ, 134, 2236

Shakura N. I., Sunyaev R. A., 1973, A&A, 500, 33

Shen Y., Richards G. T., Strauss M. A., et al., 2011, ApJS, 194, 45

Simm T., Salvato M., Saglia R., Ponti G., Lanzuisi G., Trakhtenbrot B., Nandra K., Bender R., 2016, A&A, 585, A129

Stone N. C., Metzger B. D., 2016, MNRAS, 455, 859

Suberlak K., Ivezić Ž., MacLeod C. L., Graham M., Sesar B., 2017, MNRAS, 472, 4870

Suganuma M., Yoshii Y., Kobayashi Y., Minezaki T., Enya K., Tomita H., Aoki T., Koshida S., Peterson B. A., 2006, ApJ, 639, 46

Tachibana Y., Miller A. A., 2018, PASP, 130, 128001

Tadhunter C., Spence R., Rose M., Mullaney J., Crowther P., 2017, Nature Astronomy, 1, 0061

Tonry J. L., Denneau L., Flewelling H., Heinze A. N., Onken C. A., Smartt S. J., Stalder B., Weiland H. J., Wolf C., 2018, ApJ, 867, 105

van Velzen S., Farrar G. R., Gezari S., Morrell N., Zaritsky D., Östman L., Smith M., Gelfand J., Drake A. J., 2011, ApJ, 741, 73

van Velzen S., Gezari S., Cenko S. B., Kara E., Miller-Jones J. C. A., Hung T., Bright J., Roth N., Blagorodnova N., et al., 2019, ApJ, 872, 198

van Velzen S., Gezari S., Hammerstein E., et al., 2020, arXiv e-prints, p. arXiv:2001.01409

Vanden Berk D. E., Wilhite B. C., Kron R. G., Anderson S. F., Brunner R. J., Hall P. B., Ivezić Ž., Richards G. T., Schneider D. P., York D. G., Brinkmann J. V., Lamb D. Q., Nichol R. C., Schlegel D. J., 2004, ApJ, 601, 692

Vaughan S., Edelson R., Warwick R. S., Uttley P., 2003, MNRAS, 345, 1271

Wilhite B. C., Brunner R. J., Grier C. J., Schneider D. P., vanden Berk D. E., 2008, MNRAS, 383, 1232

York D. G., Adelman J., Anderson John E. J., et al., 2000, AJ, 120, 1579

Zacharias N., Finch C. T., Girard T. M., Henden A., Bartlett J. L., Monet D. G., Zacharias M. I., 2013, AJ, 145, 44

Zel'dovich Y. B., 1964, Soviet Physics Doklady, 9, 195