

## A ZTF Look at Optical Variability of Young Stellar Objects in the North America and Pelican Nebulae Complex

LYNNE A. HILLENBRAND,<sup>1</sup> THADDAEUS J. KIKER,<sup>2,3</sup> MILES GEE,<sup>2</sup> OWEN LESTER,<sup>2</sup> NOAH L. BRAUNFELD,<sup>2</sup>  
LUISA M. REBULL,<sup>4</sup> AND MICHAEL A. KUHN<sup>1</sup>

<sup>1</sup>*Department of Astronomy, California Institute of Technology, Pasadena, CA 91125, USA*

<sup>2</sup>*Harriton High School, Bryn Mawr, PA 19010, USA*

<sup>3</sup>*Current address: Sunny Hills High School, Fullerton, CA 92833, USA*

<sup>4</sup>*Infrared Science Archive (IRSA), IPAC, 1200 E. California Blvd., California Institute of Technology, Pasadena, CA 91125, USA*

(Received 27 October 2021, UT)

Submitted to The Astronomical Journal

### ABSTRACT

We present a study of 323 photometrically variable young stellar objects that are confirmed members of the North America and Pelican (NAP) nebulae star forming region. To do so, we utilize over two years of data in the  $g$  and  $r$  photometric bands from the Zwicky Transient Facility (ZTF). We first investigate periodic variability, finding 46 objects ( $\sim 15\%$  of the sample) with significant periods that phase well, and can be attributed to stellar rotation. We then use the quasi-periodicity ( $Q$ ) and flux asymmetry ( $M$ ) variability metrics to assign morphological classifications to the remaining aperiodic light curves. Another  $\sim 38\%$  of the variable star sample beyond the periodic sources are also flux-symmetric, but with a quasi-periodic or stochastic nature. Concerning flux-asymmetric sources, our  $Q - M$  analysis reveals  $\sim 14\%$  bursters and  $\sim 29\%$  dippers. We also investigate the relationship between variability slopes in the  $g$  vs  $g - r$  color-magnitude diagram, and the light curve morphological classes. Burster-type objects have shallow slopes, while dipper-type variables tend to have higher slopes that are consistent with extinction driven variability. Our work is one of the earliest applications of the  $Q$  and  $M$  metrics to ground-based data. We therefore contrast the  $Q$  values of high-cadence and high-precision space-based data, for which these metrics were designed, with  $Q$  determinations resulting from degraded space-based lightcurves that have the cadence and photometric precision characteristic of ground-based data.

**Keywords:** Young star clusters(1833); Pre-main-sequence stars (1290); T Tauri stars (1681); Irregular variable stars(865); Stellar rotation(1629); Time domain astronomy(2109)

### 1. INTRODUCTION

The Zwicky Transient Facility (ZTF; [Kulkarni 2018](#)) has been used prolifically in contemporary transient detection and characterization efforts. Astrophysical transients observed and/or discovered by ZTF include: tidal disruption events (e.g. [van Velzen et al. 2021](#)), type Ia supernovae (e.g. [Yao et al. 2019](#); [Miller et al. 2020](#); [Bulla et al. 2020](#)), dwarf novae (e.g. [Soraisam et al. 2021](#)), candidate electromagnetic counterparts to gravitational

wave events (e.g. [Graham et al. 2020](#)), active galactic nuclei (e.g. [Frederick et al. 2019](#); [Ward et al. 2021](#)), AM CVN systems ([van Roestel et al. 2021a](#)), general stellar variables (e.g. [van Roestel et al. 2021b](#); [Roulston et al. 2021](#)), and solar system bodies such as comets and asteroids (e.g. [Ye et al. 2020a,b](#)).

While photometric variability has long been known as a ubiquitous characteristic of young stellar objects ([Joy 1945](#)), no study to date has utilized ZTF for investigating the general variability characteristics of young stars using a significantly large sample. ZTF data has, however, contributed to investigations of young stellar objects (YSOs) in a few studies focused on individual sources (e.g. [Dahm & Hillenbrand 2020](#); [Hodapp et al.](#)

Corresponding author: Lynne A. Hillenbrand  
[lah@astro.caltech.edu](mailto:lah@astro.caltech.edu)

56 2020; Hillenbrand et al. 2021). The plethora of multi-  
 57 year, nearly nightly cadence, and multi-color ZTF ob-  
 58 servations is ideally suited for studying the variability of  
 59 optically visible YSOs.

60 Our study targets a sample of YSOs in the vicinity of  
 61 the North America and Pelican (hereafter NAP) nebu-  
 62 lae (Reipurth & Schneider 2008). Initial interest in this  
 63 region as a laboratory for studying the evolution and  
 64 variability of YSOs was piqued after Herbig (1958) iden-  
 65 tified and characterized numerous YSOs based on H $\alpha$   
 66 emission lines. More recently, Guieu et al. (2009) and  
 67 Rebull et al. (2011) identified more than two thousand  
 68 candidate YSOs in the region based on infrared excess  
 69 emission using data from Spitzer and 2MASS (Skrutskie  
 70 et al. 2006). Focussing on the L935 dark cloud region,  
 71 Armond et al. (2011) identified new H $\alpha$  emitters and  
 72 Herbig-Haro flows, and Damiani et al. (2017) studied  
 73 x-ray emitters. These catalogs have been of interest in  
 74 recent years for other studies investigating various as-  
 75 pects of the young stars in the region, including kine-  
 76 matics (Kuhn et al. 2020) and HR diagrams (Fang et al.  
 77 2020).

78 Considering previous variability studies, Findeisen  
 79 et al. (2013) identified 43 large-amplitude aperiodic vari-  
 80 ables with distinct bursting or fading behavior, which  
 81 is believed to be driven by variable accretion and extinc-  
 82 tion processes, respectively (e.g. Herbst et al. 2002;  
 83 Cody & Hillenbrand 2018). The results of long-term  
 84 photometric monitoring observations of the region have  
 85 been reported by Poljančić et al. (2014), Ibryamov et al.  
 86 (2015), and Ibryamov et al. (2018), providing detailed  
 87 analyses of small samples of stars based on decades of  
 88 multi-color data from numerous observatories. Bhard-  
 89 waj et al. (2019) found 56 periodic variables in the Pel-  
 90 ican Nebula region, visually classifying an additional 11  
 91 variables as non-periodic while Froebrich et al. (2021)  
 92 identified 59 periodic variables, many of which overlap  
 93 the Bhardwaj et al. (2019) sample.

94 In this paper, we utilize the flux asymmetry and quasi-  
 95 periodicity metrics of Cody et al. (2014) to quantita-  
 96 tively describe variability classes of YSOs in the region,  
 97 greatly expanding upon the results of previous variabil-  
 98 ity studies. In doing so, we present the largest single  
 99 sample of YSOs classified by the flux asymmetry and  
 100 quasi-periodicity metrics to date. We connect the light  
 101 curve classifications with analysis of color-magnitude  
 102 variability trends, which helps to provide a more com-  
 103 prehensive understanding of the nature of the variability.

104 The remainder of this paper is structured as follows.  
 105 We describe our stellar sample in Section 2 and the ob-  
 106 servations used as well as data selection and handling  
 107 in Section 3. The identification of variability and the

108 search for periodicities are discussed in Section 4. In  
 109 Section 5 we present our light curve classification and  
 110 color-magnitude results. In Section 6 we consider rela-  
 111 tionships of the variability patterns to the presence of  
 112 circumstellar disks. We discuss our results in the context  
 113 of other recent studies in Section 7, and summarize our  
 114 findings in Section 8. Several appendices present: the  
 115 full light curve set, the results of testing performance  
 116 of one of the variability metrics (quasi-periodicity) with  
 117 data cadence and precision, and our findings on the pe-  
 118 riodograms for two particularly interesting sources.

## 119 2. SAMPLE SELECTION

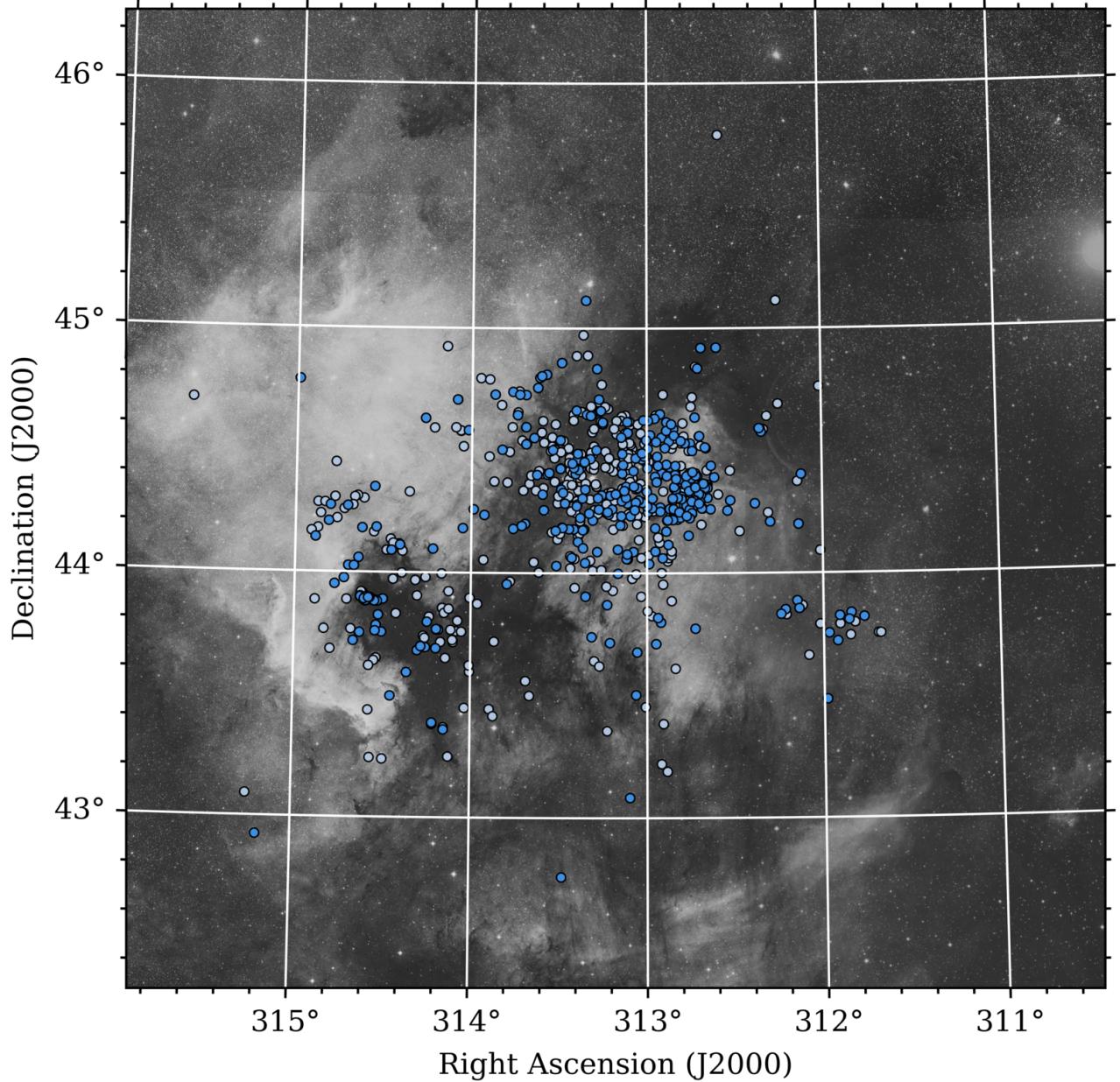
120 Our starting sample consists of YSOs in the North  
 121 America and Pelican Nebulae that have been promoted  
 122 as members based on kinematic, spectroscopic, and H-  
 123 R diagram selection criteria. The starting list for this  
 124 investigation was created by cross-matching all objects  
 125 from Table 3 in Fang et al. (2020) with all objects from  
 126 Table 2 in Kuhn et al. (2020) using a match radius of  
 127 0.5'', including every source in either of those tables.  
 128 Figure 1 shows the initial sample, as well as the reduced,  
 129 final sample discussed below.

## 130 3. ZTF PHOTOMETRY

131 ZTF (Bellm et al. 2019) employs the 48-inch Schmidt  
 132 telescope at Palomar Observatory and possesses a cam-  
 133 era with a 47 deg<sup>2</sup> field of view. It efficiently surveys  
 134 the northern sky, cataloging more than a billion objects  
 135 since first light in 2018 (Masci et al. 2019). ZTF serves  
 136 as the current standard for rapid cadence, ground based  
 137 photometric surveys as well as a test bed for techniques  
 138 to be used in upcoming next generation surveys such as  
 139 LSST (Graham et al. 2019).

140 In this paper, we analyze data from the fourth pub-  
 141 lic ZTF data release (ZTF DR4), which corresponds to  
 142 more than two years of data taken in the NAP field be-  
 143 tween 28 March 2018 and 6 June 2020 ( $58205 \leq \text{MJD} \leq 59028$ ). We focus our analysis on the ZTF *g*-band  
 144 ( $\lambda_{\text{eff}} = 4722 \text{ \AA}$ ) and *r*-band ( $\lambda_{\text{eff}} = 6340 \text{ \AA}$ ) data, as the  
 145 long-term *i*-band survey mainly covers high galactic lat-  
 146 itude fields. We refer the reader to Bellm et al. (2019)  
 147 and Masci et al. (2019) for in depth descriptions of the  
 148 ZTF observing and image processing systems. The pho-  
 149 tometry is calibrated to the PanSTARRS photometric  
 150 system (with *g* having  $\lambda_{\text{eff}} = 4811 \text{ \AA}$  and *r* having  
 151  $\lambda_{\text{eff}} = 6156 \text{ \AA}$ ) and reported in AB magnitudes.

152 The median cadence of the ZTF light curves in the  
 153 NAP region is  $\sim 1$  day. The observations were gen-  
 154 erally taken nightly, apart from seasonal gaps of  $\sim 30$   
 155 days plus additional short gaps due to adverse weather.  
 156 We ignore observations affected by clouds and/or the



**Figure 1.** Full sample of 694 selected NAP nebulae complex members plotted over a DSS2-R red photographic image of the region. Our final sample of 323 variable sources is in darker blue, with additional members sources not meeting all of the applied ZTF photometric cuts in lighter blue.

moon (these observations have catflag = 32768). We also ignore the observations between 58448 and 58456 MJD as these epochs encompass the ZTF high-cadence experiments (Kupfer et al. 2021), which have some noticeably poor photometry in the NAP region due to bad weather, which adversely impacts period searches and color-magnitude analysis.

For our analysis, we restrict the sample to objects with mean  $g < 20.8$  mag or mean  $r < 20.6$  mag over the entire time series. We further require at least 30 observations

**Table 1.** Summary of remaining sample size with the application of various selection criteria.

Criterion	# in Sample
Initial sample of NAP members	694
Brightness $g < 20.8$ or $r < 20.6$ mag	410
Epochs $> 30$ for both $g$ and $r$	392
Variability level $\nu > 15$ th percentile	323

in both the  $g$ -band and  $r$ -band (after the application of a

169  $5\sigma$  clip of potential outliers from the median magnitude  
 170 of the light curve). A summary of the cuts applied to  
 171 our initial sample and its reduction to our final sample  
 172 for light curve analysis is provided in Table 1. Figure 2  
 173 shows the photometric precision of the  $r$ -band observa-  
 174 tions as a function of median  $r$  magnitude for objects in  
 175 our final sample.

#### 176 4. PHOTOMETRY ANALYSIS

177 The  $r$ -band light curves,  $g$  vs  $g - r$  color-magnitude  
 178 trends, and periodogram results are illustrated for a set  
 179 of representative sources in Appendix A; the full sample  
 180 is available in an online figure set.

181 The results of our ZTF light curve analysis as de-  
 182 scribed below appear in Table 2. Photometric variability  
 183 amplitude, and timescale or period (as appropriate) are  
 184 discussed in this section. Light curve asymmetry and  
 185 waveform repeatability are presented in the next sec-  
 186 tion.

##### 187 4.1. Variability Search

188 We utilize the normalized peak-to-peak variability  
 189 metric

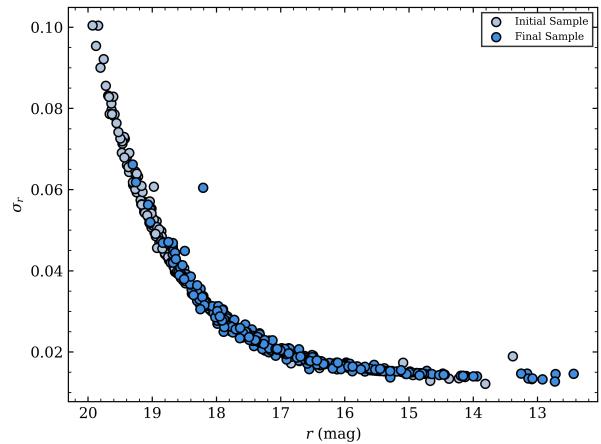
$$\nu = \frac{(m_i - \sigma_i)_{\max} - (m_i + \sigma_i)_{\min}}{(m_i - \sigma_i)_{\max} + (m_i + \sigma_i)_{\min}}, \quad (1)$$

190 featured in Sokolovsky et al. (2017) to quantify vari-  
 191 ability amplitude for objects in our YSO sample. In  
 192 this formula,  $m_i$  represents a magnitude measurement  
 193 and  $\sigma_i$  represents the corresponding measurement error,  
 194 with the minimum and maximum determined from the  
 195 full light curve. We eliminated possible outliers with  
 196 the application of a  $5\sigma$  clip to all light curves, before  
 197 calculating  $\nu$ , which is necessary for this metric to be  
 198 considered a sensitive variability indicator. For each  $r$ -  
 199 band light curve in our sample that meets the candidacy  
 200 measures described in Section 3, we calculate  $\nu$ .

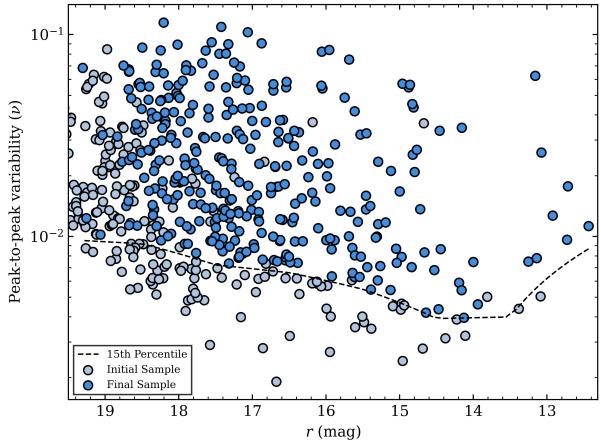
201 Our final sample for variability analysis includes ob-  
 202 jects with  $\nu$  greater than the 15th percentile of  $\nu$  as a  
 203 function of magnitude. Although not a rigorously justi-  
 204 fied cutoff, this criterion ensures that we are study-  
 205 ing the fractionally larger amplitude variables at each  
 206 brightness level. Figure 3 depicts the selection of sources  
 207 based on  $\nu$ , and Table 1 shows the effect of the  $\nu$  cut on  
 208 our final sample size.

##### 209 4.2. Period Search

210 We use the Astropy implementation of the Lomb-  
 211 Scargle periodogram (VanderPlas et al. 2012; Vander-  
 212 Plas & Ivezić 2015) to search for periodic signals in the  
 213  $\sim 825$  day ZTF data stream. Possible periods between  
 214 0.5 and 250 days are considered.



**Figure 2.** Mean  $1\sigma$  photometric uncertainty as a function of mean  $r$  magnitude, calculated from the ZTF DR4 data products. Objects in our final variable star sample are shown in darker blue, and those not meeting all photometric cuts in lighter blue. The outlier at  $r \approx 18$  and  $\sigma_r \approx 0.06$  is GDR1 2162947420052773760, a source that undergoes an abrupt dimming event of  $> 4.5$  mag approximately halfway through the observing window.



**Figure 3.** Normalized peak-to-peak variability metric  $\nu$  as a function of mean  $r$  magnitude for our sample of 392 NAP YSOs that meet the cuts on magnitude and minimum number of observations. Objects in darker blue above the dashed 15th percentile line are the 323 sources included in our variable star sample, while those not included are in lighter blue.

215 We compute the periodogram for every object in  
 216 the variable sample, flagging periods between 0.5–0.51,  
 217 0.98–1.02, 1.96–2.04, and 26–30 days for further analysis  
 218 to avoid the most common aliases associated with the  
 219 solar and sidereal days, as well as the lunar cycle (fol-  
 220 lowing Rodriguez et al. 2017; Ansdel et al. 2018). In

**Table 2.** Properties of variable stars in our NAP YSO sample; see text for column definitions. The primary variability class corresponds to the dominant behavior of an object and any secondary class describes the subordinate behavior. Objects with a reported timescale but a  $Q - M$  class other than “P” correspond to objects with a significant Lomb-Scargle period in either  $r$  or  $g$ , but a high phase dispersion and thus high  $Q$ . CMD slope angles are provided if their derived error is  $< 10^\circ$ . A full, machine-readable, version of this table is available electronically.

Name	RA (deg)	Dec (deg)	Timescale or Period (d)	Q	M	$\nu$	$\langle r \rangle$ (mag)	Primary Var. Class	Secondary Var. Class	CMD Angle (deg)
2MASS J20470481+4349114	311.77007	43.81984	-	0.89	0.07	0.0318	19.03	S	-	63.9
2MASS J20475256+4345137	311.96890	43.75372	5.936	0.3	0.25	0.0096	18.24	P	-	44.8
2MASS J20475581+4329019	311.98253	43.48387	1.014	0.93	0.02	0.0723	17.68	S	-	76.6
2MASS J20483292+4351190	312.13717	43.85530	-	0.64	-0.75	0.0685	19.31	B	-	-
2MASS J20505039+4450115	312.70997	44.83654	222.6	0.48	0.35	0.0905	16.88	QPD	-	86.2
2MASS J20505363+4424305	312.72346	44.40845	5.340	0.86	0.78	0.0286	17.0	APD	-	67.9
2MASS J20505543+4417461	312.73100	44.29609	0.902	0.81	0.18	0.0138	17.8	QPS	-	52.6
2MASS J20510029+4424364	312.75124	44.41009	7.009	0.24	0.12	0.0207	18.33	P	-	61.1
2MASS J20510663+4421407	312.77764	44.36131	1.397	0.42	-0.16	0.0088	15.65	P	L	68.6
2MASS J20510895+4416420	312.7873	44.27831	1.190	0.78	0.58	0.0171	18.23	QPD	-	48.0

particular, the window function shows that we have the potential for extremely strong signals at frequencies corresponding to 1.000 and 0.997 day periods. To account for additional potentially aliased periods, we flag periods with half or double multiples that fail at least one of the aforementioned alias checks. We then compute the 90%, 95%, and 99% confidence false alarm probability levels, and consider the period corresponding to the highest power peak greater than the 99% confidence false alarm probability ( $FAP < 0.01$ ) to be significant (following Messina et al. 2010). We repeat the above procedure with  $g$  band data for objects lacking a significant period in  $r$ , and if a significant period is found in  $g$ , we return this period for the source.

For each object, the full light curve was phased at the four most significant peaks, and visually examined to assess the dispersion across phase. We also marked periods corresponding to beats of the  $\sim 1$  day sampling interval to assess the expected Fourier patterns. In the vast majority of cases, the highest periodogram peak was retained as correct. In some cases, however, a longer period with a slightly less prominent peak was retained as the most likely true period.

Objects with significant Lomb-Scargle periodogram peaks that passed our vetting procedures are reported in Table 2. We have labelled the column as containing both timescales and true periods. In cases where the value of  $Q$  (quasi-periodicity metric; see next section) is high, the periodogram peaks are more appropriately thought of as timescales, whereas when  $Q$  is low, the

light curve is truly periodic in the sense of strictly repeating, with low residual dispersion in the phased light curve. In this latter case, there is also a “P” in the column labelled as the primary variability classification. A few sources have multiple significant peaks in the periodogram, and are given “MP” classifications; see Table 3 for additional information.

The distribution of periods for the sources identified as strictly periodic is shown in Figure 4. There are 46 such stars, constituting  $\sim 15\%$  of the variable sample with reported timescales/periods in Table 2. The period distribution has a bimodal peak, as well as a skew towards longer periods. Both features are consistent with the populations of fast and slow rotators discussed in previous young star literature, e.g. Gallet & Bouvier (2013).

Figure 4 also shows the full range of variability timescales, which exhibits a strong peak on the same few week timescales covered by the purely rotationally variable stars. The peak is significantly higher though, indicating a large quasi-periodic and aperiodic population with additional physical processes that are occurring on similar timescales. The timescale distribution is relatively flat out to longer times. The longer variability patterns, from weeks to months, can be seen in the light curves themselves. We emphasize that, even though these timescales were detected and measured using Lomb-Scargle periodogram techniques, they really are timescales and not true periods. Other methods for quantifying timescales in aperiodic light curves are discussed in Findeisen et al. (2015).

**Table 3.** Stars with complex periodogram peaks. FHK 176 and FHK 286 are two cases with unusually long periods, and are discussed in Appendix C. FHK 127 and FHK 576 have myriad significant periodogram peaks, with several that are real and others that are harmonically related. FHK 408 and FHK 511 each seem to have several genuine independent peaks.

Name	Primary Period (d)	Other Periods (d)	Interpretation
FHK 176	0.972	35.314	beat
	38.990	1.026	alias of 0.972 plus beat
FHK 286	0.975	43.01	beat
	38.656	1.024	alias of 0.975 plus beat
FHK 127	1.357	0.575, 1.714, 0.63, 2.400	beat, real, beat, beat
FHK 408	1.41	0.77, 3.44	all real
FHK 511	4.961	11.863, 11.498, 11.181	all real?
FHK 576	1.65	0.82, 0.62, 4.74, 9.621	harmonic, beat, real, real

281 One particular source with a long period or timescale  
 282 is FHK 36, which is classified as a quasi-periodic dipper  
 283 (QPD) with a timescale of 31 days. In this case, the  
 284 phased light curve shows substantially variable depth  
 285 dips that are more consistent with material orbiting in  
 286 a disk, than with e.g. a stellar rotation period.

287 We also performed a Lomb-Scargle period search on  
 288 the  $g - r$  color curves, with colors created from the  
 289 individual  $g$  and  $r$  time series as described below in §5.2.  
 290 Few objects were found with significant periodicity in  
 291 their colors. Among them, FHK 216 and FHK 442 are  
 292 examples with periodic color variations. While their  
 293 color curves do not phase particularly well, the color pe-  
 294 riods are the same as the single-band periods, reported  
 295 in Table 2. FHK 216 has a color-magnitude slope sim-  
 296 ilar to the reddening vector while FHK 442 is slightly  
 297 shallower.

## 298 5. LIGHT CURVE CLASSIFICATION

### 299 5.1. $Q$ and $M$ Variability Metrics

300 Our light curve classification scheme follows the quasi-  
 301 periodicity and flux asymmetry statistics first developed  
 302 by Cody et al. (2014), and further refined by Cody &  
 303 Hillenbrand (2018). Quasi-periodicity is defined as

$$304 Q = \frac{\sigma_{\text{resid}}^2 - \sigma_{\text{phot}}^2}{\sigma_m^2 - \sigma_{\text{phot}}^2}, \quad (2)$$

305 where  $\sigma_{\text{phot}}$  is the measurement uncertainty,  $\sigma_m^2$  is the  
 306 variance of the original light curve, and  $\sigma_{\text{resid}}^2$  is the variance  
 307 of the residual light curve after the smoothed dominant  
 308 periodic signal has been subtracted. We take  $\sigma_{\text{phot}}$   
 309 as the mean photometric error for all observations in an  
 310 object’s light curve, multiplied by a factor of 1.25 to ac-  
 311 count for an initial compression of  $Q$  values between 0  
 312 and  $\sim 0.6$  that we noticed early in our investigation. We  
 313 attribute this compression to two effects: 1) the likely  
 314 underestimate of reported photometric errors, and 2)  
 315 the cadence dependence of the  $Q$  metric, which we dis-  
 316 cuss in Appendix B. For  $\sigma_{\text{resid}}^2$ , we calculate the residual

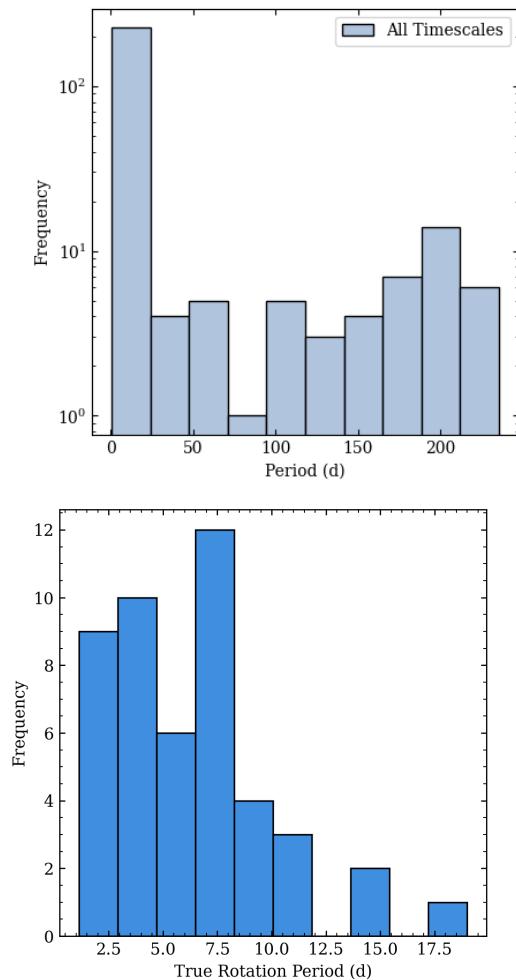
316 light curve by adopting a modified version of the peri-  
 317 odicity routine described in Section 4.2. The light curve  
 318 is phased to the period with the highest power, after po-  
 319 tentially aliased periods have been rejected. Then, fol-  
 320 lowing Bredall et al. (2020), we mitigate for edge effects  
 321 by concatenating three cycles of the period and convolv-  
 322 ing with a boxcar window of width 25% of the period,  
 323 effectively smoothing the light curve by a factor of four.  
 324 Finally, we subtract the middle cycle of this model light  
 325 curve from the folded light curve and compute  $\sigma_{\text{resid}}$  as  
 326 the standard deviation of this residual curve. The re-  
 327 sulting  $Q$  values are expected to fall between 0 (strong  
 328 periodicity with low residual noise) and 1 (completely  
 329 stochastic behavior).

330 Flux asymmetry is defined as

$$M = \frac{\langle m_{10\%} \rangle - m_{\text{med}}}{\sigma_m} \quad (3)$$

331 where  $\langle m_{10\%} \rangle$  is the mean of all magnitude mea-  
 332 surements in the top and bottom deciles of the light curve,  
 333  $m_{\text{med}}$  is the median magnitude measurement, and  $\sigma_m$   
 334 is the standard deviation of the light curve (Cody &  
 335 Hillenbrand 2018). Objects that have  $M$  values  $< 0$   
 336 are predominately brightening, objects with  $M$  values  
 337  $> 0$  are predominantly dimming, and objects with  $M$   
 338 values near 0 have symmetric light curves. In practice,  
 339 we follow previous literature and assign  $M < -0.25$  as  
 340 bursters,  $M > 0.25$  as dippers, and the intermediate  $M$   
 341 values as relatively symmetric lightcurves.

342 We calculate  $Q$  and  $M$  based on the  $r$ -band light  
 343 curve data. Figure 5 provides illustrative light curves  
 344 along the  $Q$  and  $M$  sequences. We then distinguish  
 345 YSO light curves as populating nine categories, follow-  
 346 ing Cody et al. (2014), and present these classification  
 347 results in Table 2. In general, the objects are classified  
 348 based on their  $Q$  and  $M$  values and then validated by  
 349 visual examination, but there are cases where we adjust  
 350 the classification to defer to the human eye for definitive



**Figure 4.** Distribution of significant Lomb-Scargle periodogram peaks. Top panel shows shows all timescales from Table 2, regardless of light curve classification. Bottom panel shows only stars labelled as “P” in Table 2, which are those that phase well to their periodogram peaks and likely have their origin in spot pattern modulation due to stellar rotation. As expected, the true stellar rotation periods are concentrated towards the shorter time scales, with values of a few days to a few weeks.

categorization. This is especially prevalent in cases near the morphological boundaries in the  $Q - M$  plane.

Periodic variables (P) are defined as objects with significant periods from Section 4.2 that have  $-0.25 < M < 0.25$  and  $Q < 0.45$ . We discuss our rationale for using a higher  $Q$  boundary for periodic objects than adopted in previous publications (e.g. Cody & Hillenbrand 2018; Bredall et al. 2020), in Section 7.2. The variability of the strictly periodic objects is most commonly interpreted as driven by rotational modulation due to the presence of star spots, and the periods de-

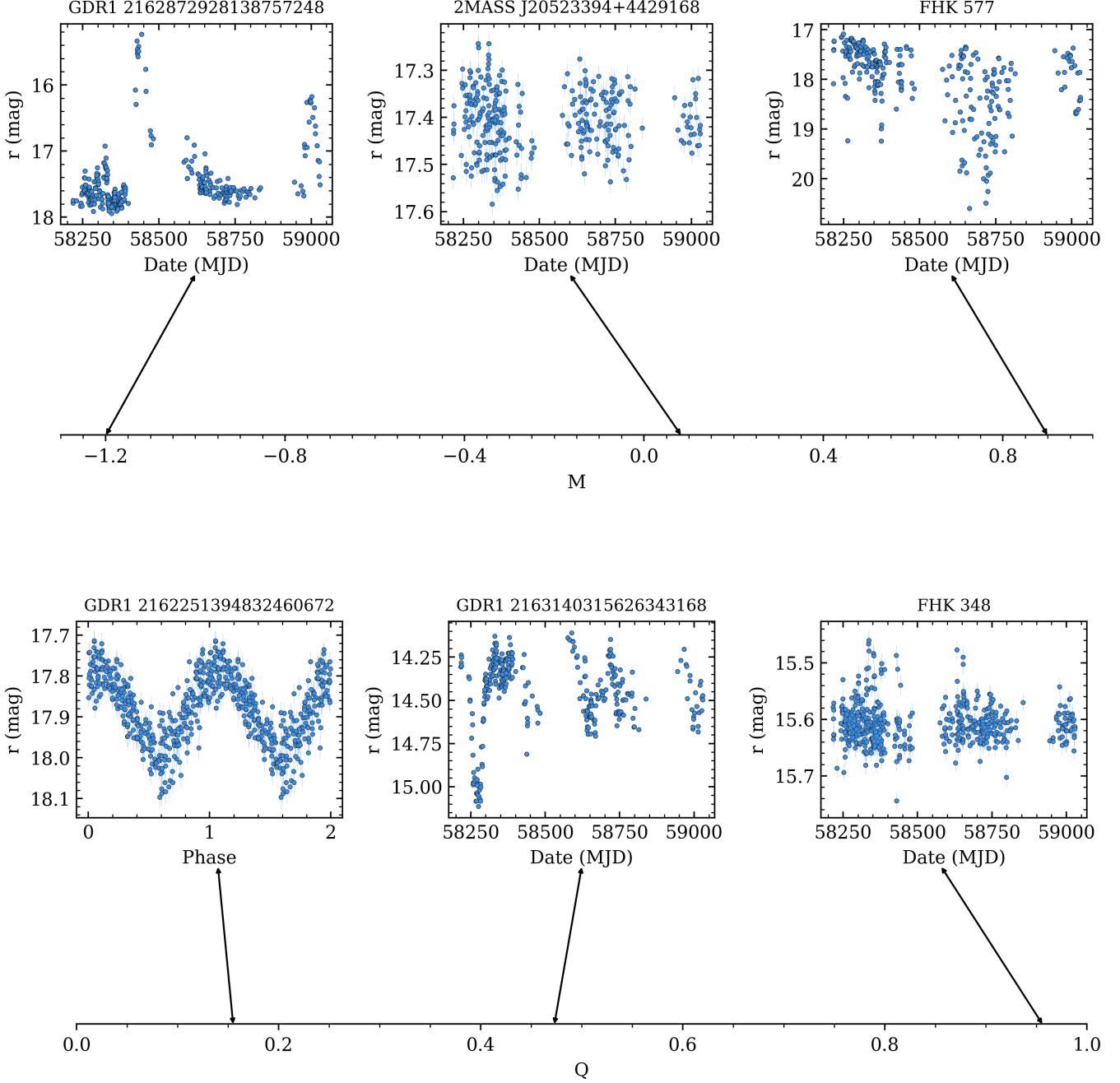
rived are thus considered measures of stellar rotation (e.g. Rebull et al. 2008). Objects with higher  $Q$  but that phase well with low dispersion can still be labelled as periodic, while others are designated as quasi-periodic with an origin that is not dominated by starspot modulation, or is otherwise obfuscated. Multi-periodic (MP) objects exhibit more than one distinct period, that is not a beat or alias. One common explanation for these is star spot modulation in binary pairs (e.g. Stauffer et al. 2018; Maxted & Hutcheon 2018).

Quasi-periodic symmetric (QPS) variables are defined as having  $-0.25 < M < 0.25$  and  $0.45 < Q < 0.87$ . The variability of these objects is believed to have two possible sources. The first is a combination of purely periodic spot behavior with longer timescale aperiodic changes (e.g. variable accretion), and the second is a single variability process that is unstable from cycle to cycle (Cody et al. 2014). Quasi-periodic dipper (QPD) variables are categorized based on  $M > 0.25$  and  $0.45 < Q < 0.87$ . The variability of these objects is believed to stem from variable extinction caused by time-variable occultation by circumstellar material (e.g. Alencar et al. 2010; Morales-Calderón et al. 2011; Ans-dell et al. 2016). Following previous literature, we do not define a quasi-periodic burster category as distinct from the highly stochastic bursters.

Burster (B) variables are defined as having  $M < -0.25$ , and correspond to objects with erratic but distinct accretion bursts that are generally short-lived (Cody et al. 2017). These bursts are distinct from larger amplitude and longer timescale outbursts from EX Lup and FU Ori type sources, which are more eruptive (Hartmann & Kenyon 1996; Herbig 2008). Aperiodic dipper variables (APD) are defined as having  $M > 0.25$  and  $Q > 0.87$ . These objects experience variable extinction, and it has been proposed by Turner et al. (2014) that their variability stems from inner disk scale-height changes induced by magnetic turbulence. Stochastic (S) variables are defined as having  $-0.25 < M < 0.25$  and  $Q > 0.87$ .

Long timescale (L) objects exhibit variability on timescales  $\gtrsim 100$  day timescales and have  $Q > 0.87$ . Finally, we assign 2 objects to the unclassifiable (U) category because we were unable to calculate their  $Q$  due to  $\sigma_{\text{phot}}^2$  being larger than  $\sigma_{\text{resid}}^2$  and/or  $\sigma_m^2$ .

The results of this classification scheme are reported in Table 2 for individual sources. Additionally, all stars in the variable sample are plotted in the  $Q - M$  plane in Figure 6, where they are differentiated in color by the assigned morphological class. We note that some objects have secondary classifications in Table 2 when more than one light curve type is evident in the data, but they



**Figure 5.** Example ZTF data at low, intermediate, and high values of the asymmetry parameter  $M$  (upper panels) and the quasi-periodicity parameter  $Q$  (lower panels). Negative values of  $M$  indicate flaring or bursting type behavior, while positive values of  $M$  signify dipping or fading behavior. Low values of  $Q$  indicate more strictly periodic light curves, while high values of  $Q$  are characteristic of stochastic light curves.

**Table 4.** Distribution of primary and secondary light curve morphology classifications for objects in our variable sample. Most sources have a single dominant variability type, but  $\sim 14\%$  of the variable sample merited secondary classifications.

Variability Class	Primary %	Secondary %
Periodic	13.6	0.9
Multi-periodic	1.5	-
Quasi-periodic symmetric	27.2	0.3
Quasi-periodic dipper	19.2	-
Aperiodic dipper	9.6	-
Burster	14.2	-
Long timescale	2.8	12.7
Stochastic	11.1	-
Unclassifiable	0.6	-
Total	99.8	13.9

are plotted according to the primary classification. The distribution of both primary and secondary variability classes is summarized in Table 4.

Secondary classifications are typically long-timescale (L), representing objects that exhibit a strong primary behavior such as short timescale periodicity, quasi-periodic behavior, or bursting, along with a pronounced long-term trend e.g. brightening or dimming, or lasting changes in variability amplitude, etc. Two objects are designated as periodic in their secondary classification; both have robust periods, but the  $M$  metric reveals that the full amplitude of the light curve is caused by bursty behavior. For GDR1 2162127974652045056, also known as V1716 Cyg, the behavior over most of the light curve is plainly sinusoidal, but for a little over one year of our time series, there is continuous bursting at the 1 mag level. The source further distinguishes itself by having a W UMa-like phased lightcurve, even though it is a fairly secure kinematic and spectroscopic member of the NAP region. For the other object, FHK 496, bursting behavior is ubiquitous throughout the light curve, and causes scatter at the 0.4 mag level above the clear periodic signal.

## 5.2. Color-Magnitude Analysis

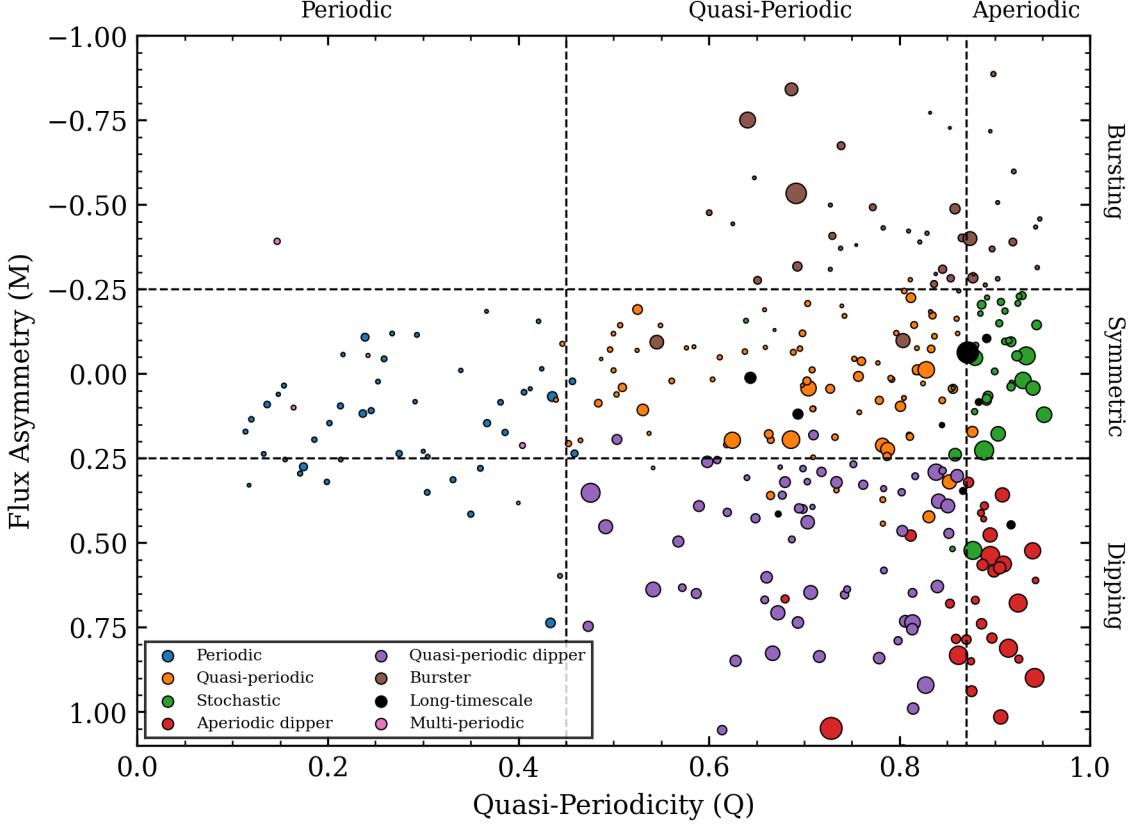
As mentioned above in Section 5.1, there are many different physical mechanisms that can cause photometric variability in young stars. Color time series data can be a helpful tool in distinguishing them. Among others, Carpenter et al. (2001); Günther et al. (2014); Wolk et al. (2015) showed how near-infrared and mid-infrared multi-color photometry can distinguish extinction-related and accretion-related mechanisms. Optical photometry is even more sensitive to

these behaviors, due to the steep rise in extinction from the infrared to the optical, and because of the hot nature of the accretion process.

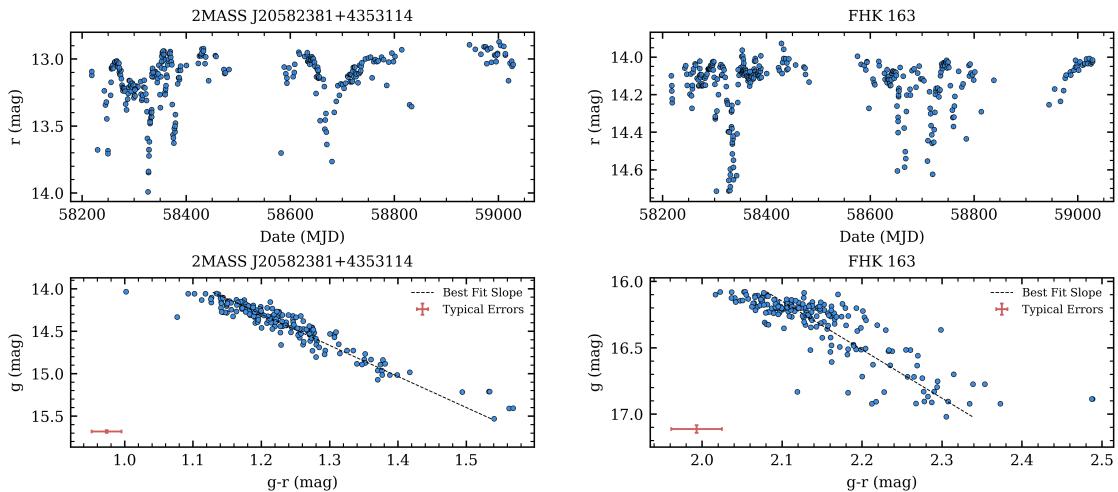
The ZTF  $g$ -band and  $r$ -band observations are often taken close to one other, within a few hours up to a few days, but they are not simultaneous. Color-magnitude diagrams were thus constructed as follows. First, we partitioned the  $g$ -band light curves into segments in which the maximum separation between any two subsequent  $g$  observations is less than three days. We adopted this maximum observation separation constraint because linearly interpolating between  $g$  observations taken further apart created spurious values in the color-magnitude plane. In practice, within the defined  $<3.0$  day intervals, approximately 20.5% of the  $g$ -band observations occur within 0.25 days of the prior  $g$ -band observation, 87.0% within 1.25 days, 93.4% within 2.0 days, and 97.5% occur within 2.25 days. For each  $g$ -band value that was paired with an  $r$  band observation, we then linearly interpolated between the date bounds of the defined  $g$ -band intervals, in order to estimate  $g-r$  colors. Errors on the color values were calculated as the Pythagorean sum of the photometric uncertainties in the  $r$ -band observation, and the  $g$ -band observation that is closest in brightness to the interpolated  $g$ -band value.

The time series color-magnitude diagrams exhibit a variety of morphologies, with some sources showing large color and magnitude excursions while others have little color variation, and are essentially constant within the errors. Figures 7 and 8 show examples from among the dipper ( $M > 0.25$ ) and burster ( $M < -0.25$ ) categories, which have distinct slope angles. A selection of sources representing different  $Q$  values is presented in Appendix A, where a range of CMD slope angles can also be seen. Lightcurves and CMDs for the full sample are available in an associated online figure set.

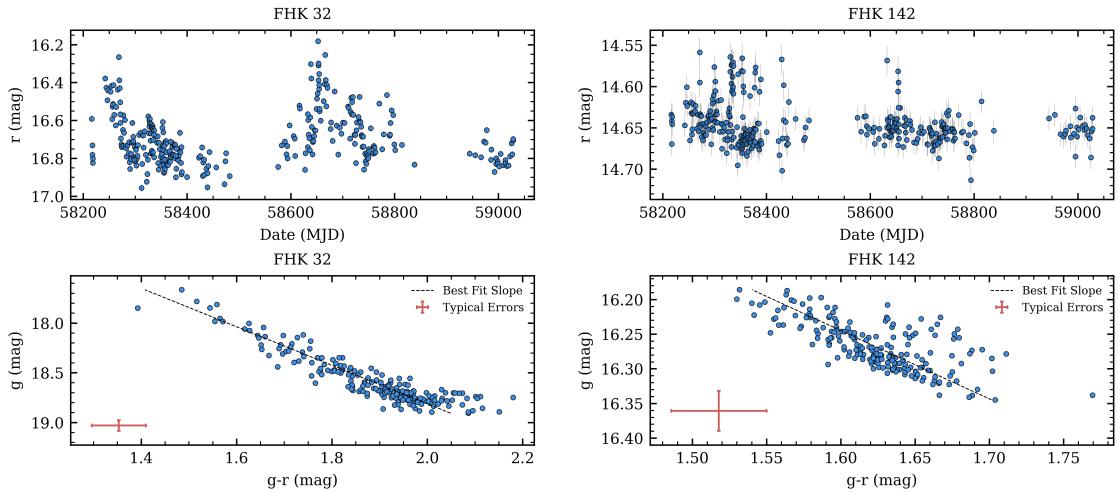
For each member of the variable star set, we performed a linear fit to its color-magnitude diagram (CMD). To do so, we applied least-squares linear orthogonal fits using the python package `scipy.odr` (Boggs & Rogers 1990; Virtanen et al. 2020). This method was chosen following Poppenhaeger et al. (2015), because it can account for the significant and partially correlated errors in both axes. The regression assumes that points lie along a line in the  $g$  vs.  $g-r$  plane, with Gaussian errors perpendicular to the line. While the condition is not strictly true, the assumption is reasonable given that there is no clear independent and dependent variable in this situation. and this model describes the expected behavior of extinction. To alleviate the effect of outliers, we performed the fits on the middle 95% of the CMD spans in  $g$  and  $g-r$ .



**Figure 6.** Quasi-periodicity versus flux asymmetry for our sample of variables, color coded by light curve morphological type as reported in Table 2. The size of each point is proportional to the normalized peak-to-peak variability metric  $\nu$  raised to the 1.25th power. Note that some objects, especially periodic sources with high values of  $Q$ , can fall outside of the defined boundaries between the morphological classes in the  $Q - M$  plane. However, the  $Q - M$  classification is fairly robust with only a minor number of sources deserving adjustment to their numerically assigned classes.

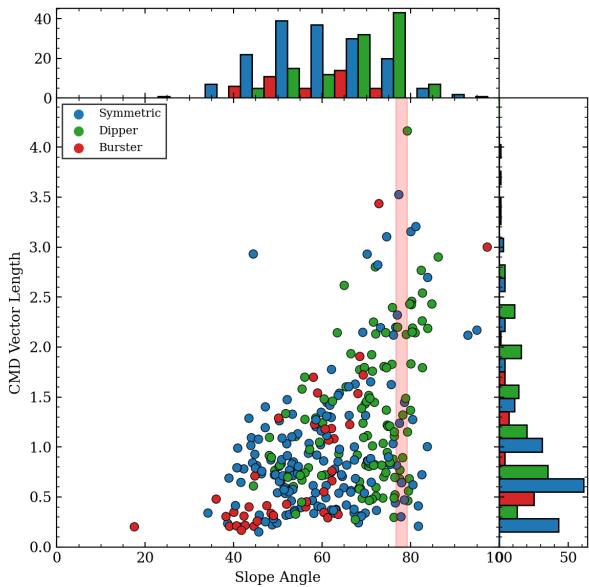


**Figure 7.** Two exemplary dipper objects with  $r$ -band light curves shown in the top row, and the corresponding  $g$  vs  $g - r$  CMDs presented in the bottom row. Both sources have CMD slope angles consistent with their photometric variability being caused by extinction changes, likely along a line of sight that passes through circumstellar material.



**Figure 8.** Same as Figure 7, but for two burster objects. Both sources have CMD slope angles much shallower than the expectations from extinction changes.

After computing the best fit slopes in  $g$  vs  $g - r$ , we define slope angles as the inverse tangent, with the angles in degrees increasing clockwise from  $0^\circ$  (corresponding to color variability with no associated  $g$ -band variability) to  $90^\circ$  (colorless  $g$ -band variability). To compute errors on the best fit angles, we adopt the pYSOVAR (Günther et al. 2014; Poppenhaeger et al. 2015) function `fit_twocolor_odr` and consider slope angles with errors less than  $10^\circ$  to be significant. These values are reported in Table 2. We note that a few sources even exceed  $90^\circ$  slopes: FHK 101 at  $93^\circ$ , GDR1 2161843098061857280 at  $95^\circ$ , and FHK 473 at  $97.3^\circ$ . Each of these sources is a large-amplitude variable, with a Pythagorean vector length in the range of 2-3, and high  $Q$  as well as  $M < 0$ .



**Figure 9.** Fitted CMD slope angle versus Pythagorean vector length, differentiated by flux asymmetry ( $M$ ) class. The vast majority of sources have slope angles much shallower than the expectations from reddening (shaded band), indicating that the observed variability is dominated by accretion effects rather than by extinction effects.

In Figure 9, all significant CMD slope angles are plotted against their CMD vector lengths, following Poppenhaeger et al. (2015), and differentiated by their flux asymmetry ( $M$ ) classes. There is a sizable span in the CMD vector lengths, with the longest spans limited to the largest slope angles (indicating the largest amplitude and color changes). A wide range of CMD slope angles is also represented, spanning over  $45^\circ$ . The outlier at small slope angle is GDR1 2162932993257907584, which varies over  $\sim 0.2$  mag in  $g - r$  color with very little variation in  $g$  brightness; in  $r$ -band it is a low-level burster. The outlier with vector length  $\sim 3$  is GDR1

2162966146109606528, which varies over 2 mag in  $g - r$  color and 3 mag in  $g$  brightness; it is classified as quasi-periodic symmetric. Finally, the extreme vector length  $\sim 4$  source is GDR1 2162939208074527616, an aperiodic dipper. We also note GDR1 2162872928138757248, the out-of-family burster with vector length  $\sim 3.5$ ; this source was also featured in Figure 5.

Small slopes,  $< 35^\circ$ , are essentially unpopulated, implying a lack of colorless variability. While likely an astrophysical reality, this could also indicate a bias in our slope determination methods. For example, stars with little to no statistical correlation between variability in magnitude, and variability in color, generally have uncertain slope angles. Such sources are not included in the plot. However, the full set of slope and vector length determinations includes only 4 (1.2%) more sources than the set with  $< 10^\circ$  error that we are plotting.

The entirety of the populated parameter space in Figure 9 is represented by variables that are relatively symmetric in flux (blue). The dipper-type variables (green) tend to dominate the population at the longest vector lengths and highest slope angles; their median slope angle is  $73.8^\circ$ . The burster-type variables (red), conversely, are concentrated towards the shorter vector lengths and populate mainly the lower slope angles, almost exclusively  $< 70^\circ$ , with median slope angle  $47.7^\circ$ . Beyond just median values, the differences between the morphological dipper and burster classes, and dipper and symmetric classes are very statistically significant based on pairwise K-S tests, with  $p << 10^{-11}$ . The burster and symmetric difference is also statistically significant, though only with  $p = 0.013$ .

We can assess the CMD slope angles in the context of standard reddening due to dust. We adopt the extinction curve of Fitzpatrick (1999), as translated by Schlafly & Finkbeiner (2011) into the PanSTARRS photometric system (to which ZTF data are calibrated). The expected slopes in  $g$  vs.  $g - r$  CMDs for different reddening laws are: 3.52 for  $R_V = 3.1$ , 4.39 for  $R_V = 4.1$ , and 5.24 for  $R_V = 5.1$ , with steeper slopes corresponding to larger dust grain sizes, as might be expected in molecular clouds and circumstellar disks. Considering a range of possibilities for the extinction law, we expect objects with time-variable CMD behavior dominated by extinction effects to have slope angles between  $74.1^\circ$  and  $79.2^\circ$  in the  $g$  vs  $g - r$  diagram.

Examination of Figure 9 shows, however, that the majority of slopes are much shallower than the expectations from purely reddening variations. We interpret the distribution of slope angles as indicating that much of the observed variability is at least partially related to accretion effects, with extinction effects (likely occurring out-

side of the accretion zone) perhaps playing some role. We acknowledge that other morphological light curve types can have slope angles in the CMD consistent with reddening, that is, not all sources that look like they are experiencing variable extinction also exhibit identifiable dipping behavior.

## 6. CORRELATION OF VARIABILITY PROPERTIES WITH INFRARED EXCESS

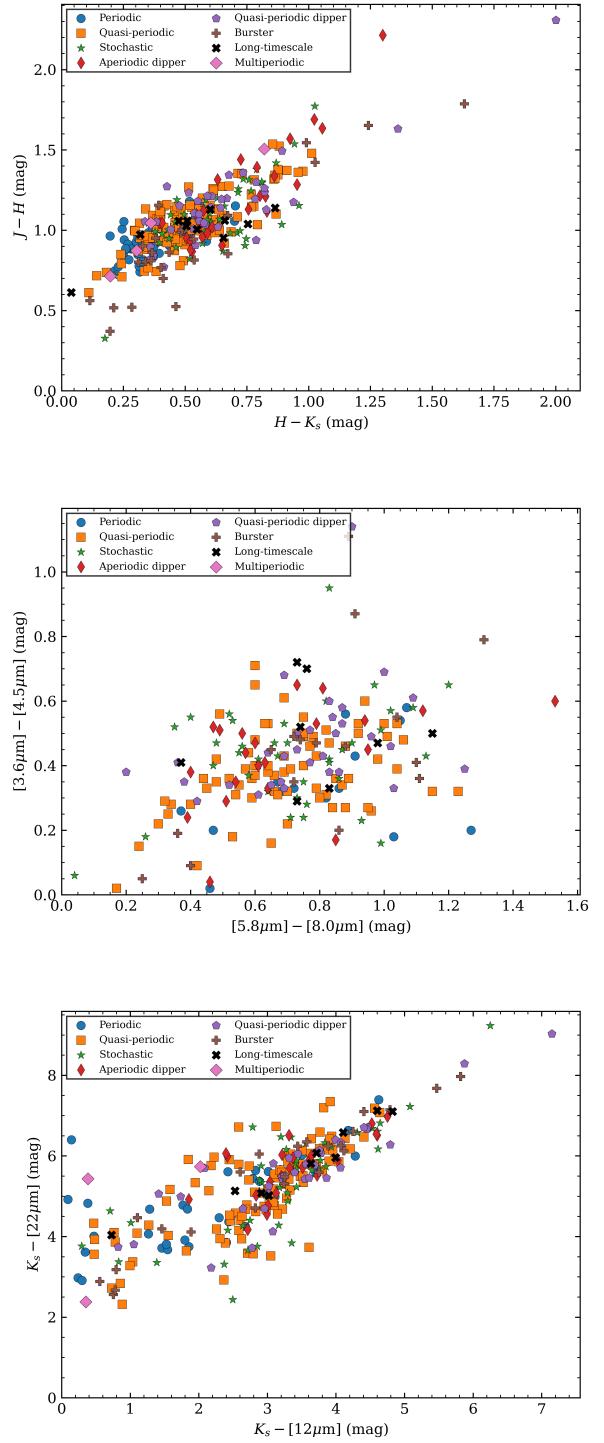
There is some expectation that the light curve morphological classes could correlate with infrared excess. Using  $K2$  data sets, several previous studies, namely Cody & Hillenbrand (2018) in Oph/Sco, Cody et al. (2022) in Taurus, and Venuti et al. (2021) in the Lagoon Nebula all found that both the periodic (P) and the quasi-periodic symmetric (QPS) sources tend to have lower values of infrared excess on average, and in fact dominate the source population at small infrared color values. This is consistent with a relatively clean line of sight to the spotted stellar photosphere, and a lack of accretion effects in the light curve. Sources with higher  $Q$ , indicating quasi-periodic and stochastic light curves, as well as the dippers and bursters, on the other hand, had larger values of infrared excess in these previous studies.

To explore whether these correlations can be seen using light curve morphologies assigned based on ZTF data, we cross-matched our variable star sample with photometry from 2MASS (Cutri et al. 2003), WISE (Wright et al. 2010; Cutri & et al. 2012), and Spitzer (Rebull et al. 2011). Figure 10 shows several color-color diagrams with the light curve morphologies from Table 2 distinguished. Figure 11 shows one infrared color as a function of our normalized variability amplitude metric.

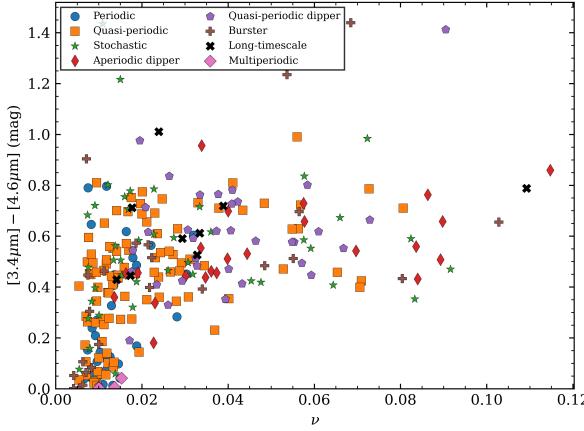
There is a familiar trend for the periodic sources to populate the lower infrared excess values. As in Venuti et al. (2021) for example, the sources with essentially photospheric colors at  $H - K < 0.3$  are dominated by the P and QPS types. However, some of the other previously identified trends between light curve morphology and infrared excess are not clearly present in our analysis. Of note is that our ground-based data has lower cadence and precision than  $K2$  data, and we are thus less sensitive to the smaller-scale variability patterns that can be discerned in the exquisite  $K2$  light curves. Another factor may be that, relative to the variability morphology vs infrared color analysis in Cody & Hillenbrand (2018) studying Upper Sco and Cody et al. (2022) studying Taurus, we may be more subject here in the NAP region to the effects of reddening.

## 7. DISCUSSION

Our discussion focuses on three aspects of our analysis. In §7.1 we compare our periods to those found in



**Figure 10.** Infrared color-color diagrams with objects having different  $Q - M$  classifications distinguished. Except for the bluest 5-10% of the sample, the sources have colors consistent with those of Class II YSOs. The objects classified with periodic light curves tend to have small infrared colors, while those with other behavior tend to have larger infrared colors, indicative of excesses consistent with the presence of circumstellar dust.



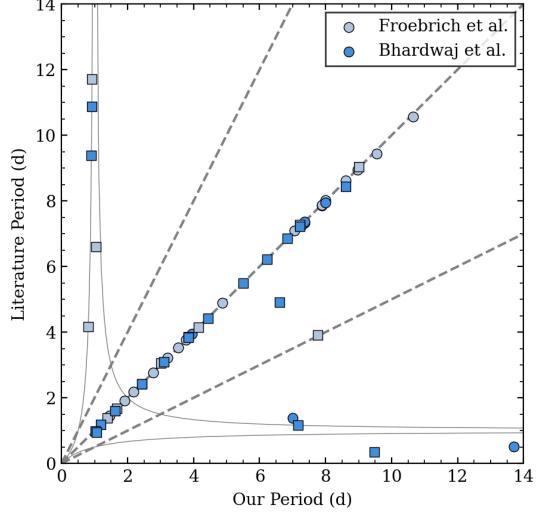
**Figure 11.** Infrared color vs  $\nu$  metric, measuring normalized amplitude of the optical variability. Colors redder than 0.2 mag in these bandpasses indicate young stars with infrared excess. Essentially all sources with  $\nu > 0.02$  appear to have excess infrared color, indicating young stars surrounded by circumstellar dust.

previous literature. In §7.2 we philosophize about the broad applicability of the quasi-periodicity  $Q$  metric, in particular regarding the need to tailor its boundaries for use in ground-based data sets. Finally in §7.3, we assess how the distribution of sources in the  $Q - M$  plane for the NAP region compares to what has been found in other star forming regions.

### 7.1. Period Recovery for Strictly Periodic Variables

In order to test the quality of our derived periods, we compared them to the periods published in Bhardwaj et al. (2019) and Froebrich et al. (2021), both considering objects in the Pelican Nebula.

Of the 40 periodic variables identified in Froebrich et al. (2021), 32 are within 2'' of a source in our sample, of which 20 we classify as periodic in  $Q - M$  space. All 20 have periods within 5% of our derived periods. Additionally, among the 56 objects found to have strong periodic variations in Bhardwaj et al. (2019), 32 variables match within 2'' of a source in our sample and we classify 5 as periodic in  $Q - M$  space. In this sub-sample, we recover 3 periods within 5% of their Bhardwaj et al. (2019) values. We compared the phase-folded light curves of the two objects with discrepant periods and found that forcing them to the period in Bhardwaj et al. (2019) either introduces pronounced additional scatter or provides no improvement in phase dispersion. Among the additional 17 stars in common, 13 have periods in agreement within 10%, though we have labelled these periodogram peaks as timescales in Table 2 rather than periods, due to poor phasing and/or high  $Q$  values.



**Figure 12.** Comparison of periods for objects that overlap between our work and previous literature. Lines indicate the 1:1  $P$  (period) relationship, the  $P/2$  and  $2P$  harmonics, and the  $(|1/P \pm 1|)^{-1}$  beats with the  $\sim 1$  day sampling. Circles are sources we find to exhibit periodic behavior, whereas squares are sources that we designate as having timescales rather than strictly periodic behavior, based on their high  $Q$  values.

Overall, between these two samples our period recovery rate is 92%. These period comparisons are presented in Figure 12.

### 7.2. Comments on the Quasi-Periodicity Metric $Q$

The quasi-periodicity metric  $Q$  has proven to be an extremely powerful tool for classifying the variable behaviour of YSOs (Cody et al. 2014; Cody & Hillenbrand 2018). In this work, we have shown that the metric is sufficiently applicable to ZTF data, affirming its utility for ground-based as well as space-based photometric data sets. However, recent studies have revealed issues in translating  $Q$  results between various sources of data (Cody & Hillenbrand 2018; Bredall et al. 2020).

When comparing objects studied with different space-based telescopes, systematic effects present in the precision photometry from different observatories, or those caused by slight wavelength dependencies of the astrophysical phenomena, could produce systematically different  $Q$  values for sources with similar intrinsic behavior (Cody et al. 2014; Cody & Hillenbrand 2018). Notably, the morphological boundaries in  $Q$  were adjusted going from analysis of CoRoT data in Cody et al. (2014) to K2 data in Cody & Hillenbrand (2018). For the CoRoT data, Cody et al. (2014) set  $Q$  boundaries at 0.11 and 0.61, with  $M$  boundaries at  $\pm 0.25$ . For K2 data, in Cody & Hillenbrand (2018) the adopted  $Q$  boundaries were 0.15 and 0.85, with  $M$  remaining at  $\pm 0.25$ .

In contrast to this adjusting of the  $Q$  metric, Bredall et al. (2020) did not adjust the morphological boundaries of  $Q$  in their ground based study based on ASAS-SN data, instead opting to simply apply the morphological boundaries defined for K2 data in Cody & Hillenbrand (2018). These authors did note the discrepancies that appeared in their classification scheme, such as objects having been classified in previous variability studies as QPDs, assessed in their numerical analysis as strictly periodic objects.

Indeed, prior to tailoring our  $Q$  routine to the ZTF data set, we noticed a similar compression and shift of our objects towards the periodic end (low  $Q$ ) of the quasi-periodicity scale, with an absence of objects having  $Q > 0.65$ . This would be a very different distribution from the results of Cody et al. (2014) and Cody & Hillenbrand (2018), as in both of these studies, the range from  $Q=0$  to  $Q=1$  is densely populated (in fact, exceeding these bounds in Cody et al. 2014).

Given the substantially larger errors and the lower observational cadence associated with ground-based data, significant discrepancies in  $Q$  values might be expected when comparing objects studied with space-based versus ground-based telescopes. In Appendix B we investigate in detail the effects that cadence and photometric uncertainty have on  $Q$ .

For our analysis, we opted to follow the methodology of Cody & Hillenbrand (2018), tailoring both our morphological boundaries as well as our procedure for calculating  $Q$  to the nuances of our particular ZTF data set. As stated in §5.1, our adopted  $Q$  boundaries are 0.45 and 0.87. We selected these particular values after visually inspecting all variable object light curves, both unfolded and phase-folded, and comparing them to light curves from both Cody et al. (2014) and Cody & Hillenbrand (2018). We assigned the  $Q$  boundaries to occur at the most obvious behavioral transitions we noticed in the light curves when ordered by  $Q$ , and to minimize the number of edge-case objects that we end up re-classifying by eye. After this visual inspection, approximately 11% of objects were re-classified.

We believe that such a flexible treatment of the morphological boundaries indicated by  $Q$  is the best way to ensure that variables classified in the  $Q - M$  parameter space are behaviorally representative of the original classes as defined by Cody et al. (2014). In other words, there is an unquantified error inherent in the calculation of the  $Q$  and  $M$  metrics, and it is more useful for furthering our understanding of the origins of young star variability, and for comparing sources observed with different instruments, if the quantitative procedures and morphological boundaries are tailored to the instruments

used in each analysis. While objects of the same intrinsic morphological class could appear in different  $Q$  ranges in different studies, they should be properly grouped with their light curve peers.

Regarding the full range spanned by our derived  $Q$  values (about 0.11-0.96), we believe there are reasonable explanations why the full zero-to-one parameter space is not fully covered. On the low side, none of our objects have  $Q$  values between 0.0 and 0.1 partially due to our choice to represent the estimated photometric uncertainty with the mean photometric error for each object, and our further inflation of the photometric errors during the  $Q$  calculation process. This imposes an artificial minimum to  $Q$ . There is also the likely reality that no objects in our sample have negligible phase dispersion (including photometric uncertainty) relative to the amplitude of a hypothetical periodic oscillation. On the high side of  $Q$ , the maximum at 0.96 instead of 1.00 may indicate simply that we have not inflated our errors enough to produce a truly stochastic  $Q$  metric. On the other hand, there is no rationale for fine tuning so as to have any particular number of objects with the theoretical maximum value of  $Q = 1.00$ .

Regarding the range spanned by the  $M$  values (about -1.2-1.1) for our aperiodic objects, there is again no expectation that any given variable star sample should span, or exist constrained within, the theoretical range from -1 to +1; although we observe our results seem to appropriately populate the parameter space in almost all cases. We note that, as the procedure for calculating the flux asymmetry  $M$  metric is more agnostic to the instrumental effects than the procedure for calculating  $Q$ , we were comfortable adopting the same flux asymmetry boundaries used in previous studies to designate dippers, symmetries, and bursters. We note that the periodic objects at  $Q < 0.45$  occupy a more restricted range in  $M$ , closer to the symmetry line of  $M = 0$ .

### 7.3. Comparison to $Q - M$ Distributions in Other Star Forming Regions

In Table 4, we summarized the fraction of objects in the member sample we have assembled for the North America and Pelican Nebula region, in the different variability classes. We can compare our relative distribution to those resulting from previous uses of the  $Q - M$  parameter space to study young stars.

To do so, we refer to Table 3 in Cody et al. (2022) which includes the Taurus, Oph, and NGC 2264 regions studied with *K2* and *CoRoT* data taken from space-based platforms. We focus our comparisons on the young regions with comparable ages to the NAP. The only other region with such information, Upper Sco, has

variability properties that are distributed somewhat differently, perhaps because Upper Sco is somewhat older. We do not consider the results of [Bredall et al. \(2020\)](#) from ground-based data in Lupus, due to the apparent distortion in their  $Q$  distribution towards low values; this may be a sign of the issue we faced in this study (§5.1) where inflating our photometric errors by 25% recovered more of the expected range in  $Q$  for a young star variable sample. There is also the issue of whether ground-based and space-based data sets are comparable. As mentioned earlier, we assess the effects of data precision and data cadence on  $Q$  in Appendix B.

Relative to the previously studied star forming regions, we find a similar fraction of periodic sources, including the periodic (P) and multi-periodic (MP) classes, at around 15% of the sample. This is interesting given that our sample is designed to include all members and not just disk-bearing sources that comprised the samples studied previously. The result can be understood if we accept that our sample is more biased towards periodic sources, but at the same time we are less capable of detecting them as such, given the lower quality of our data set. Another commonality is that the quasi-periodic symmetric (QPS) category is our largest, and similar to the other star-forming regions at about 27%. Our stochastic fraction at 11% also matches well to the previous results. The quasi-periodic and aperiodic dipper (QPD and APD) groups are also similar, cumulatively around 29% of our sample. There is a range in these fractions among the other star-forming regions, of a factor of two, with our value in the middle. Our burster fraction is 14%, on the low side of the other regions (which range from about 13-21%). An explanation here is that much of the burster activity identified from *K2* is short-lived (e.g. [Cody et al. 2017](#)) and stands out relative to what could be detected from the ground.

Summarizing, our ground-based data set has matched the variability properties of space-based studies of comparably-aged stars rather well, with any differences plausibly attributed to our lower sensitivity from the ground to narrow bursts and accretion flaring activity.

## 8. SUMMARY

In this work we have studied the variability of young stellar objects in the North America and Pelican Nebulae region using multi-color observations from  $> 800$  days of Zwicky Transient Facility measurements. Our primary results are as follows:

- We have searched for periodicity in the light curves of all variable objects in our sample and compared our derived periods with those previously published in the literature.

839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851

- We have classified 323 stars using the  $Q - M$  variability plane, to quantify flux asymmetry and waveform repeatability in the light curves. Of these, 44 objects are periodic variables (with an additional 2 variables showing bursting behavior with very strong periodic behavior), several objects are multi-periodic variables, 88 quasi-periodic symmetric variables, 62 objects quasi-periodic dippers, 31 aperiodic dippers, 46 bursters, 36 stochastic variables, and 9 long-timescale variables. We additionally designated 45 objects with secondary classifications where more than one behavior is exhibited.

852  
853  
854  
855  
856  
857  
858  
859  
860

- We have compared light curve morphology classes with the variability exhibited in CMDs to investigate the physical drivers of variability. There is a clear distinction in the distribution of variability slope angles in the CMD for dippers and bursters, with dippers closer to the expectations from extinction-driven variability phenomena, and bursters exhibiting much flatter slopes suggestive of accretion-related variability drivers.

861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875

- We have investigated and discussed various subtleties of the quasi-periodicity metric  $Q$ , and recommend methods for its use in future studies. Specifically, as acknowledged in previous literature, the metric is sensitive to the accuracy of the photometric errors and further, different data sets necessitate flexible treatment of the light curve morphology boundaries when applying the  $Q$  metric. We have also shown that, while photometric uncertainty is important, and the lower precision of ground-based data causes us to miss some low-level variability that is detectable in data from K2, the lower cadence of ZTF has a more pronounced effect on  $Q$ , substantially reducing it for certain categories of variability.

876  
877  
878

- The python code used to calculate the  $Q$  and  $M$  variability metrics is available at <https://github.com/HarringtonResearchLab/NAPYSOs>.

## ACKNOWLEDGMENTS

We thank Leah Seignourel, Pietro Romussi, and Korgan Atillasoy for their participation in early discussions about this work. We also thank John Bredall, Katja Poppenhaeger, and Hans Moritz Günther for assistance with various Python questions. Ann Marie Cody provided the K2 light curves that form the basis of Appendix B, as well as valuable advice concerning  $Q$ . This work is based on data from the Zwicky Transient Facility, which is supported by the National Science Foundation under Grant No. AST-1440341 and a collaboration including Caltech, IPAC, the Weizmann Institute for Science, the Oskar Klein Center at Stockholm University, the University of Maryland, the University of Washington, Deutsches ElektronenSynchrotron and Humboldt University, Los Alamos National Laboratories, the TANGO Consortium of Taiwan, the University of Wisconsin at Milwaukee, and Lawrence Berkeley National Laboratories.

879     *Facility:* P48:ZTF, IRSA

880     *Software:* NumPy (Harris et al. 2020), Matplotlib  
 881     (Hunter 2007), Pandas (McKinney 2009), Astropy (Astropy Collaboration et al. 2013), and SciPy (Virtanen  
 882     et al. 2020).

884

## 9. REFERENCES

### REFERENCES

- 885     Alencar, S. H. P., Teixeira, P. S., Guimarães, M. M., et al.  
 886     2010, A&A, 519, A88, doi: [10.1051/0004-6361/201014184](https://doi.org/10.1051/0004-6361/201014184)
- 887     Ans dell, M., Gaidos, E., Rappaport, S. A., et al. 2016, ApJ,  
 888     816, 69, doi: [10.3847/0004-637X/816/2/69](https://doi.org/10.3847/0004-637X/816/2/69)
- 889     Ans dell, M., Oelkers, R. J., Rodriguez, J. E., et al. 2018,  
 890     MNRAS, 473, 1231, doi: [10.1093/mnras/stx2293](https://doi.org/10.1093/mnras/stx2293)
- 891     Armond, T., Reipurth, B., Bally, J., & Aspin, C. 2011,  
 892     A&A, 528, A125, doi: [10.1051/0004-6361/200912671](https://doi.org/10.1051/0004-6361/200912671)
- 893     Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2019,  
 894     PASP, 131, 018002, doi: [10.1088/1538-3873/aaecbe](https://doi.org/10.1088/1538-3873/aaecbe)
- 895     Bhardwaj, A., Panwar, N., Herczeg, G. J., Chen, W. P., &  
 896     Singh, H. P. 2019, A&A, 627, A135,  
 897     doi: [10.1051/0004-6361/201935418](https://doi.org/10.1051/0004-6361/201935418)
- 898     Boggs, P. T., & Rogers, J. E. 1990, Contemporary  
 899     Mathematics, 112, 183
- 900     Bredall, J. W., Shappee, B. J., Gaidos, E., et al. 2020,  
 901     MNRAS, 496, 3257, doi: [10.1093/mnras/staa1588](https://doi.org/10.1093/mnras/staa1588)
- 902     Bulla, M., Miller, A. A., Yao, Y., et al. 2020, ApJ, 902, 48,  
 903     doi: [10.3847/1538-4357/abb13c](https://doi.org/10.3847/1538-4357/abb13c)
- 904     Carpenter, J. M., Hillenbrand, L. A., & Skrutskie, M. F.  
 905     2001, AJ, 121, 3160, doi: [10.1086/321086](https://doi.org/10.1086/321086)
- 906     Cody, A. M., & Hillenbrand, L. A. 2018, AJ, 156, 71,  
 907     doi: [10.3847/1538-3881/aacead](https://doi.org/10.3847/1538-3881/aacead)
- 908     Cody, A. M., Hillenbrand, L. A., David, T. J., et al. 2017,  
 909     ApJ, 836, 41, doi: [10.3847/1538-4357/836/1/41](https://doi.org/10.3847/1538-4357/836/1/41)
- 910     Cody, A. M., Hillenbrand, L. A., & Rebull, L. R. 2022,  
 911     submitted
- 912     Cody, A. M., Stauffer, J., Baglin, A., et al. 2014, AJ, 147,  
 913     82, doi: [10.1088/0004-6256/147/4/82](https://doi.org/10.1088/0004-6256/147/4/82)
- 914     Curtis, J. L., Agüeros, M. A., Matt, S. P., et al. 2020, ApJ,  
 915     904, 140, doi: [10.3847/1538-4357/abbf58](https://doi.org/10.3847/1538-4357/abbf58)
- 916     Cutri, R. M., & et al. 2012, VizieR Online Data Catalog,  
 917     II/311
- 918     Cutri, R. M., Skrutskie, M. F., van Dyk, S., et al. 2003,  
 919     2MASS All Sky Catalog of point sources.
- 920     Dahm, S. E., & Hillenbrand, L. A. 2020, AJ, 160, 278,  
 921     doi: [10.3847/1538-3881/abbf2](https://doi.org/10.3847/1538-3881/abbf2)
- 922     Damiani, F., Pillitteri, I., & Prisinzano, L. 2017, A&A, 602,  
 923     A115, doi: [10.1051/0004-6361/201629402](https://doi.org/10.1051/0004-6361/201629402)
- 924     Fang, M., Hillenbrand, L. A., Kim, J. S., et al. 2020, ApJ,  
 925     904, 146, doi: [10.3847/1538-4357/abba84](https://doi.org/10.3847/1538-4357/abba84)
- 926     Findeisen, K., Cody, A. M., & Hillenbrand, L. 2015, ApJ,  
 927     798, 89, doi: [10.1088/0004-637X/798/2/89](https://doi.org/10.1088/0004-637X/798/2/89)
- 928     Findeisen, K., Hillenbrand, L., Ofek, E., et al. 2013, ApJ,  
 929     768, 93, doi: [10.1088/0004-637X/768/1/93](https://doi.org/10.1088/0004-637X/768/1/93)
- 930     Fitzpatrick, E. L. 1999, PASP, 111, 63, doi: [10.1086/316293](https://doi.org/10.1086/316293)
- 931     Frederick, S., Gezari, S., Graham, M. J., et al. 2019, ApJ,  
 932     883, 31, doi: [10.3847/1538-4357/ab3a38](https://doi.org/10.3847/1538-4357/ab3a38)
- 933     Froebrich, D., Derezea, E., Scholz, A., et al. 2021, arXiv  
 934     e-prints, arXiv:2107.08524.  
 935     <https://arxiv.org/abs/2107.08524>
- 936     Gallet, F., & Bouvier, J. 2013, A&A, 556, A36,  
 937     doi: [10.1051/0004-6361/201321302](https://doi.org/10.1051/0004-6361/201321302)

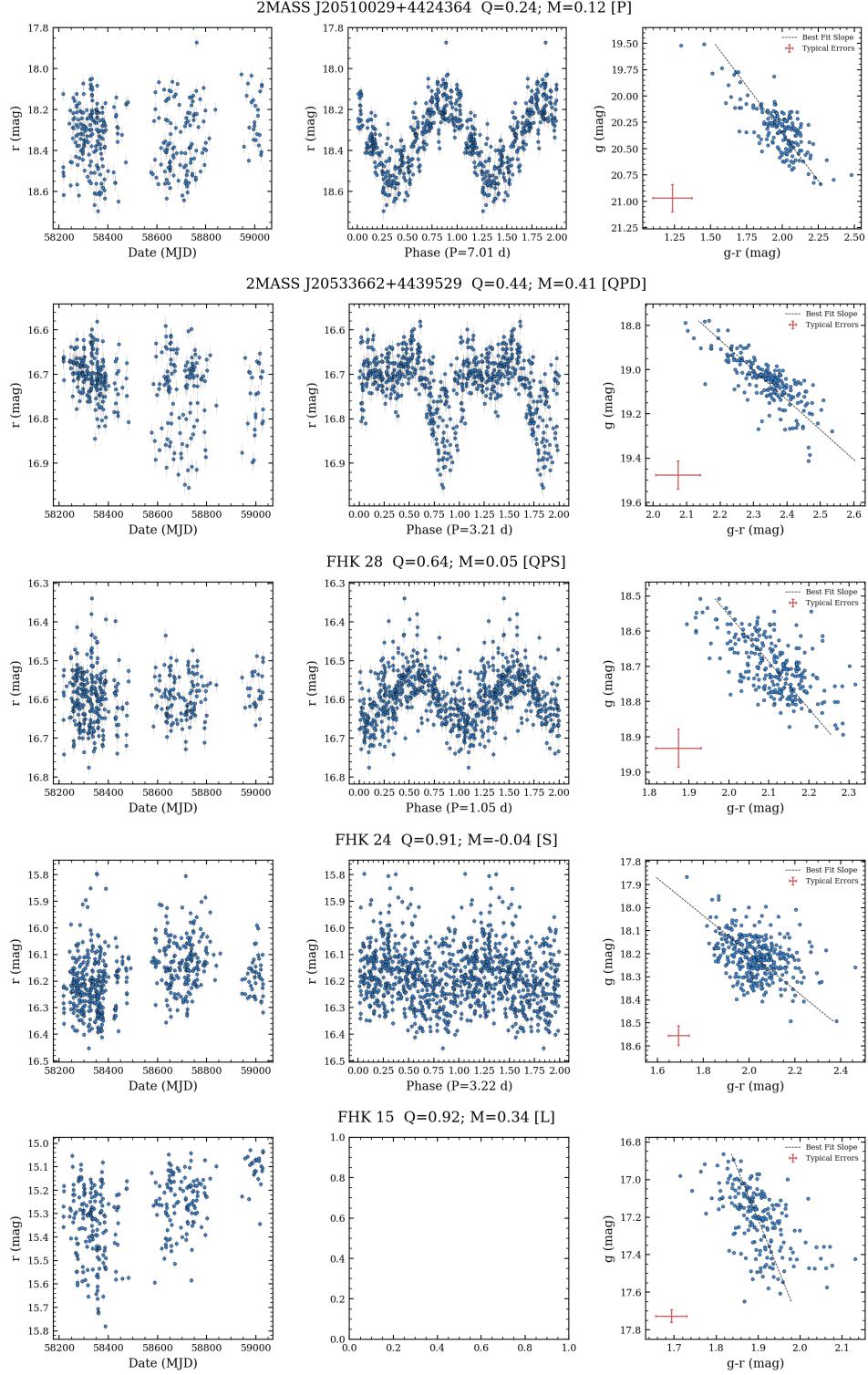
- 938 Graham, M. J., Kulkarni, S. R., Bellm, E. C., et al. 2019,  
 939 PASP, 131, 078001, doi: [10.1088/1538-3873/ab006c](https://doi.org/10.1088/1538-3873/ab006c)
- 940 Graham, M. J., Ford, K. E. S., McKernan, B., et al. 2020,  
 941 PhRvL, 124, 251102,  
 942 doi: [10.1103/PhysRevLett.124.251102](https://doi.org/10.1103/PhysRevLett.124.251102)
- 943 Guieu, S., Rebull, L. M., Stauffer, J. R., et al. 2009, ApJ,  
 944 697, 787, doi: [10.1088/0004-637X/697/1/787](https://doi.org/10.1088/0004-637X/697/1/787)
- 945 Günther, H. M., Cody, A. M., Covey, K. R., et al. 2014,  
 946 AJ, 148, 122, doi: [10.1088/0004-6256/148/6/122](https://doi.org/10.1088/0004-6256/148/6/122)
- 947 Hartmann, L., & Kenyon, S. J. 1996, ARA&A, 34, 207,  
 948 doi: [10.1146/annurev.astro.34.1.207](https://doi.org/10.1146/annurev.astro.34.1.207)
- 949 Herbig, G. H. 1958, ApJ, 128, 259, doi: [10.1086/146540](https://doi.org/10.1086/146540)  
 950 —. 2008, AJ, 135, 637, doi: [10.1088/0004-6256/135/2/637](https://doi.org/10.1088/0004-6256/135/2/637)
- 951 Herbst, W., Bailer-Jones, C. A. L., Mundt, R.,  
 952 Meisenheimer, K., & Wackermann, R. 2002, A&A, 396,  
 953 513, doi: [10.1051/0004-6361:20021362](https://doi.org/10.1051/0004-6361:20021362)
- 954 Hillenbrand, L. A., Isaacson, H., Rodriguez, A. C., et al.  
 955 2021, arXiv e-prints, arXiv:2108.05406.  
<https://arxiv.org/abs/2108.05406>
- 956 Hodapp, K. W., Denneau, L., Tucker, M., et al. 2020, AJ,  
 957 160, 164, doi: [10.3847/1538-3881/abad96](https://doi.org/10.3847/1538-3881/abad96)
- 958 Ibryamov, S., Semkov, E., Milanov, T., & Peneva, S. 2018,  
 959 Research in Astronomy and Astrophysics, 18, 137,  
 960 doi: [10.1088/1674-4527/18/11/137](https://doi.org/10.1088/1674-4527/18/11/137)
- 961 Ibryamov, S. I., Semkov, E. H., & Peneva, S. P. 2015,  
 962 PASA, 32, e021, doi: [10.1017/pasa.2015.21](https://doi.org/10.1017/pasa.2015.21)
- 963 Joy, A. H. 1945, ApJ, 102, 168, doi: [10.1086/144749](https://doi.org/10.1086/144749)
- 964 Kuhn, M. A., Hillenbrand, L. A., Carpenter, J. M., &  
 965 Avelar Menendez, A. R. 2020, ApJ, 899, 128,  
 966 doi: [10.3847/1538-4357/aba19a](https://doi.org/10.3847/1538-4357/aba19a)
- 967 Kulkarni, S. R. 2018, The Astronomer's Telegram, 11266, 1
- 968 Kupfer, T., Prince, T. A., van Roestel, J., et al. 2021,  
 969 MNRAS, 505, 1254, doi: [10.1093/mnras/stab1344](https://doi.org/10.1093/mnras/stab1344)
- 970 Masci, F. J., Laher, R. R., Rusholme, B., et al. 2019,  
 971 PASP, 131, 018003, doi: [10.1088/1538-3873/aae8ac](https://doi.org/10.1088/1538-3873/aae8ac)
- 972 Maxted, P. F. L., & Hutcheon, R. J. 2018, A&A, 616, A38,  
 973 doi: [10.1051/0004-6361/201732463](https://doi.org/10.1051/0004-6361/201732463)
- 974 Messina, S., Desidera, S., Turatto, M., Lanzafame, A. C., &  
 975 Guinan, E. F. 2010, A&A, 520, A15,  
 976 doi: [10.1051/0004-6361/200913644](https://doi.org/10.1051/0004-6361/200913644)
- 977 Miller, A. A., Yao, Y., Bulla, M., et al. 2020, ApJ, 902, 47,  
 978 doi: [10.3847/1538-4357/abb13b](https://doi.org/10.3847/1538-4357/abb13b)
- 979 Morales-Calderón, M., Stauffer, J. R., Hillenbrand, L. A.,  
 980 et al. 2011, ApJ, 733, 50,  
 981 doi: [10.1088/0004-637X/733/1/50](https://doi.org/10.1088/0004-637X/733/1/50)
- 982 Poljančić, B. I., Jurdana-Šepić, R., Semkov, E. H., et al.  
 983 2014, A&A, 568, A49, doi: [10.1051/0004-6361/201322965](https://doi.org/10.1051/0004-6361/201322965)
- 984 Poppenhaeger, K., Cody, A. M., Covey, K. R., et al. 2015,  
 985 AJ, 150, 118, doi: [10.1088/0004-6256/150/4/118](https://doi.org/10.1088/0004-6256/150/4/118)
- 987 Rebull, L. M., Stauffer, J. R., Wolff, S. C., & Strom, S. E.  
 988 2008, in Astronomical Society of the Pacific Conference  
 989 Series, Vol. 384, 14th Cambridge Workshop on Cool  
 990 Stars, Stellar Systems, and the Sun, ed. G. van Belle, 327
- 991 Rebull, L. M., Guieu, S., Stauffer, J. R., et al. 2011, ApJS,  
 992 193, 25, doi: [10.1088/0067-0049/193/2/25](https://doi.org/10.1088/0067-0049/193/2/25)
- 993 Reipurth, B., & Schneider, N. 2008, Star Formation and  
 994 Young Clusters in Cygnus, ed. B. Reipurth, Vol. 4, 36
- 995 Rodriguez, J. E., Ansdell, M., Oelkers, R. J., et al. 2017,  
 996 ApJ, 848, 97, doi: [10.3847/1538-4357/aa8c78](https://doi.org/10.3847/1538-4357/aa8c78)
- 997 Roulston, B. R., Green, P. J., Toonen, S., & Hermes, J. J.  
 998 2021, arXiv e-prints, arXiv:2105.00037.  
<https://arxiv.org/abs/2105.00037>
- 999 Schlaufy, E. F., & Finkbeiner, D. P. 2011, ApJ, 737, 103,  
 1000 doi: [10.1088/0004-637X/737/2/103](https://doi.org/10.1088/0004-637X/737/2/103)
- 1001 Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ,  
 1002 131, 1163, doi: [10.1086/498708](https://doi.org/10.1086/498708)
- 1003 Sokolovsky, K. V., Gavras, P., Karampelas, A., et al. 2017,  
 1004 MNRAS, 464, 274, doi: [10.1093/mnras/stw2262](https://doi.org/10.1093/mnras/stw2262)
- 1005 Soraismam, M. D., DeSantis, S. R., Lee, C.-H., et al. 2021,  
 1006 AJ, 161, 15, doi: [10.3847/1538-3881/abc5af](https://doi.org/10.3847/1538-3881/abc5af)
- 1007 Stauffer, J., Rebull, L. M., Cody, A. M., et al. 2018, AJ,  
 1008 156, 275, doi: [10.3847/1538-3881/aae9ec](https://doi.org/10.3847/1538-3881/aae9ec)
- 1009 Turner, N. J., Fromang, S., Gammie, C., et al. 2014, in  
 1010 Protostars and Planets VI, ed. H. Beuther, R. S. Klessen,  
 1011 C. P. Dullemond, & T. Henning, 411,  
 1012 doi: [10.2458/azu\\_uapress\\_9780816531240-ch018](https://doi.org/10.2458/azu_uapress_9780816531240-ch018)
- 1013 van Roestel, J., Creter, L., Kupfer, T., et al. 2021a, AJ,  
 1014 162, 113, doi: [10.3847/1538-3881/ac0622](https://doi.org/10.3847/1538-3881/ac0622)
- 1015 van Roestel, J., Duev, D. A., Mahabal, A. A., et al. 2021b,  
 1016 AJ, 161, 267, doi: [10.3847/1538-3881/abe853](https://doi.org/10.3847/1538-3881/abe853)
- 1017 van Velzen, S., Gezari, S., Hammerstein, E., et al. 2021,  
 1018 ApJ, 908, 4, doi: [10.3847/1538-4357/abc258](https://doi.org/10.3847/1538-4357/abc258)
- 1019 VanderPlas, J., Connolly, A. J., Ivezić, Z., & Gray, A.  
 1020 2012, in Proceedings of Conference on Intelligent Data  
 1021 Understanding (CIDU, 47–54,  
 1022 doi: [10.1109/CIDU.2012.6382200](https://doi.org/10.1109/CIDU.2012.6382200)
- 1023 VanderPlas, J. T., & Ivezić, Ž. 2015, ApJ, 812, 18,  
 1024 doi: [10.1088/0004-637X/812/1/18](https://doi.org/10.1088/0004-637X/812/1/18)
- 1025 Venuti, L., Cody, A. M., Rebull, L. M., et al. 2021, arXiv  
 1026 e-prints, arXiv:2105.11588.  
<https://arxiv.org/abs/2105.11588>
- 1027 Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020,  
 1028 Nature Methods, 17, 261, doi: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2)
- 1029 Ward, C., Gezari, S., Frederick, S., et al. 2021, ApJ, 913,  
 1030 102, doi: [10.3847/1538-4357/abf246](https://doi.org/10.3847/1538-4357/abf246)
- 1031 Wolk, S. J., Günther, H. M., Poppenhaeger, K., et al. 2015,  
 1032 AJ, 150, 145, doi: [10.1088/0004-6256/150/5/145](https://doi.org/10.1088/0004-6256/150/5/145)
- 1033 Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al.  
 1034 2010, AJ, 140, 1868, doi: [10.1088/0004-6256/140/6/1868](https://doi.org/10.1088/0004-6256/140/6/1868)
- 1035

- <sup>1037</sup> Yao, Y., Miller, A. A., Kulkarni, S. R., et al. 2019, ApJ,  
<sup>1038</sup> 886, 152, doi: [10.3847/1538-4357/ab4cf5](https://doi.org/10.3847/1538-4357/ab4cf5)
- <sup>1039</sup> Ye, Q., Masci, F. J., Ip, W.-H., et al. 2020a, AJ, 159, 70,  
<sup>1040</sup> doi: [10.3847/1538-3881/ab629c](https://doi.org/10.3847/1538-3881/ab629c)
- <sup>1041</sup> Ye, Q., Kelley, M. S. P., Bolin, B. T., et al. 2020b, AJ, 159,  
<sup>1042</sup> 77, doi: [10.3847/1538-3881/ab659b](https://doi.org/10.3847/1538-3881/ab659b)

1043

## APPENDIX

1044 A. INDIVIDUAL OBJECT LIGHT CURVES, PHASED LIGHT CURVES, AND COLOR-MAGNITUDE  
1045 DIAGRAMS1046 In the main text, Figures 7 and 8 presented exemplar objects from the “dipper” and “burster” categories, re-  
1047 spectively. Figure 13 now shows examples from the other morphological categories, including sources classified as  
1048 “periodic”, “quasi-periodic dipper”, “quasi-periodic symmetric”, “stochastic”, and “long timescale”. An online sup-  
1049 plemental figure set provides similar figures for every object in the variable star sample.1050 For each source, we plot the light curve as well as the light curve folded to the time scale value reported in Table 2,  
1051 if one exists, and the  $g$  vs  $g - r$  color-magnitude diagram. If no period was detected by the Lomb-Scargle algorithm  
1052 with power greater than the 99% False Alarm Probability, then the folded light curve subplot is left blank. If the slope  
1053 angle error is less than ten degrees, then the best fit CMD slope appears in the CMD subplots.

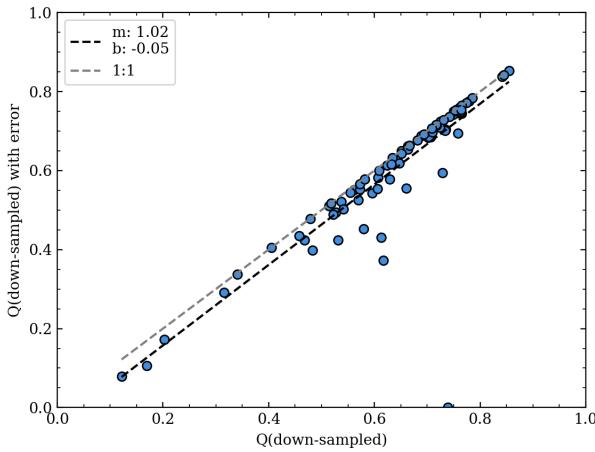


**Figure 13.** Examples from the online figure set, illustrating objects from different parts of the  $Q - M$  plane. The object name along with the  $Q$  and  $M$  values, the assigned light curve morphological class, and the reported timescale, are provided in the plot labelling. One object from each of the periodic (P), quasi-periodic dipper (QPD), quasi-periodic symmetric (QPS), stochastic (S), and long-timescaled (L) groups is shown.

1054      B. THE INFLUENCE OF CADENCE AND PHOTOMETRIC PRECISION ON Q

1055      The Q and M metrics were developed to quantify stellar variability in high-quality, regularly spaced, photometric  
 1056      time series data taken from space-based platforms. In this work, we have applied them to lower-precision light curves  
 1057      sampled at irregular intervals, taken from the ground. We assess in this appendix the transferability of these metrics  
 1058      from space-based to ground-based data. To do so, we explore how  $Q$  behaves on a  $K2$  data set – specifically the  $\sim 100$   
 1059      young stars in the Taurus region with light curves reported in [Cody et al. \(2022\)](#), and versions of these data that have  
 1060      been degraded to the cadence and precision typical of our ground-based ZTF data set.

1061      For each light curve in the Taurus sample observed by  $K2$ , we began with a baseline  $Q$  value that was calculated for  
 1062      the Kepler/K2 data set as in §5 for the ZTF data. Then, we down-sampled the observational cadence to match the  
 1063      actual time series of the ZTF data set, and recalculated the  $Q$  metric. We then also included error bars appropriate  
 1064      to the ground-based data, sampling from the empirical error distribution in Figure 2. We assumed errors of 0.015 and  
 1065      0.036 mag, which represent the 10th and 90th percentile errors for the ZTF time series data analyzed in this paper.  
 1066      We then recalculated the  $Q$  metric based on the down-sampled and error-adjusted Kepler light curves.

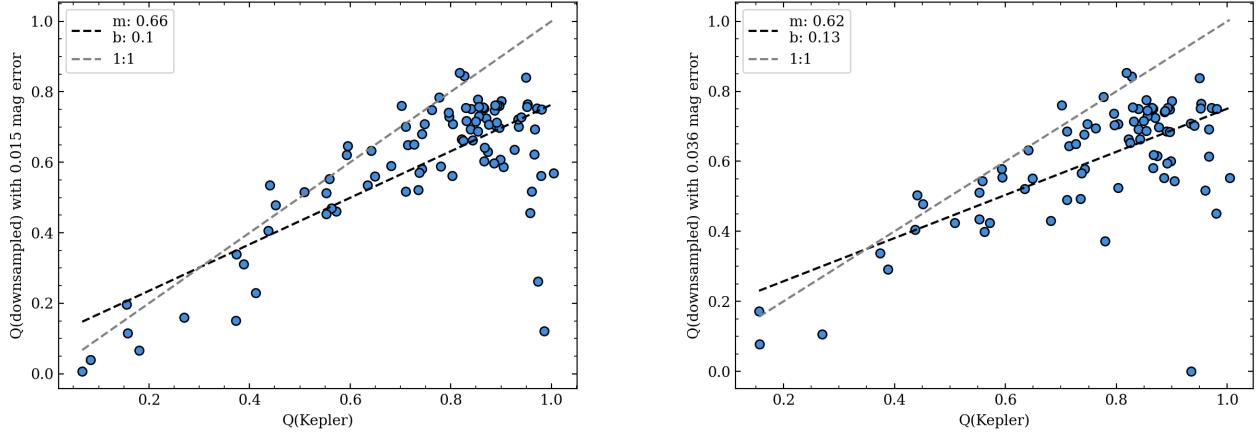


**Figure 14.** Effect of light curve precision on the  $Q$  metric. Abscissa shows  $Q$  values calculated in the absence of error, while ordinate shows  $Q$  values for the same data but including realistic ground-based errors appropriate to our ZTF observations (0.036 mag). Lines indicate the relationships shown in the legend. Photometric error reduces  $Q$  by an amount that likely depends on light curve amplitude relative to the photometric error, and light curve morphology. While the shift is systematic, many stars are unaffected and the slope remains close to unity.

1067      Considering only the influence of photometric uncertainty on the derived  $Q$ , Figure 15 illustrates for the down-  
 1068      sampled data, that the introduction of error lowers the  $Q$  values. This is the expected behavior since the photometric  
 1069      error is subtracted from both the numerator and the denominator in Eq. 2. However, there are a few cases in which the  
 1070      variance of the residuals is so small, that the numerator goes negative. Similarly, the actual variance of the light curve  
 1071      can be smaller than the mean photometric error, causing  $Q$  to be greater than 1. We have excluded these unphysical  
 1072      values in our linear fit.

1073      In the simulated ground-based data, including both the precision and the cadence adjustments from space to ground,  
 1074      the range of  $Q$  values extends from -3.18 to 2.20, compared to 0.068 - 1.01 for the original high-cadence light curves.  
 1075      For the analysis below, we consider only  $Q$  between the nominal 0-1 range. In Figure 14, we compare  $Q$  from the  
 1076      mock ground-based data set (down-sampled, with added errors) to  $Q$  from the original, fully sampled and essentially  
 1077      error-free Kepler data set. There is a clear positive association remaining. However, the relationship is not one-to-one  
 1078      as above when only error was considered. Instead, the relationship flattens out as  $Q$  became higher. This can be  
 1079      attributed mostly to the Lomb Scargle algorithm detecting periods that are not actually present, which then affects  
 1080      the calculation of the light curve residual in the  $Q$  formalism (Eq. 2).

1081      We note that the downsampling of the Kepler data stream takes into account only a fraction of the number of data  
 1082      points that were available for calculating  $Q$  in the full ZTF time series. A major consideration in this experiment  
 1083      is that the Kepler objects were measured over only about 80 days, while the ZTF objects were observed over 300



**Figure 15.** Effect of light curve cadence on the  $Q$  metric, illustrated for two different values of the light curve precision. Abscissa shows  $Q$  values calculated at the native cadence of  $K2$ , while ordinate shows  $Q$  values for the same  $K2$  data but down-sampled to the true cadence of our ZTF observations over the same limited time interval. Lines indicate the relationships shown in the legend. Cadence has a large effect on  $Q$ , with more scatter towards higher  $Q$ . The effects of error are apparent, with overestimated errors leading to a drop in  $Q$ . The values go negative for some stars, particularly in the higher error (right) panel, causing them to fall outside the bounds of the plot.

days (including data through ZTF DR4). Thus, the actual ZTF periodograms are based on more data than the down-sampled Kepler periodogram analysis. Therefore, the ZTF periodograms and hence the residual light curves are probably a more accurate representation of the period-subtracted residual plot for objects with long timescales. We found that the  $Q$  values are fairly accurate when they are low, but became increasingly more inaccurate as the actual Kepler  $Q$  increased. This finding should translate over to the actual ZTF objects, i.e. lower  $Q$  values are more accurate than higher  $Q$  values. Additionally, all of the down-sampled  $Q$  values were slightly underestimated, which again is probably true of the ZTF objects as well.

Overall, this experiment reveals that there are constraints on data quality for assessing  $Q$  in various photometric data sets. Specifically, while photometric uncertainty is important, lower cadence has a more pronounced affect on  $Q$ , substantially reducing it for certain categories of variability.

### C. A LESSON REGARDING TRUE PERIODS NEAR THE OBSERVING CADENCE

Two high-confidence NAP members are classified as strictly periodic (P) sources, but have periodograms with strong peaks in an unusual pattern. In each case, of the four peaks, several are related as beat aliases, but it is not clear which is the real period, which two are the beats, and which one is the odd extra period. Two of the peaks are just above and just below 1 day. In addition, there are two more peaks around 40 days, corresponding to unexpectedly slow stellar rotation rates; these long sinusoids are visible – by eye – in the light curves.

The two sources are FHK 176 and FHK 286, with period amplitudes of  $\sim 0.1$  and  $\sim 0.2$  mag respectively, suggestive of significantly spotted photospheres. Their time series and phased light curves are illustrated in Figure 16. Based on the information in Fang et al. (2020), one star has spectral type K8.5 and the other M0; both have Li I absorption and H $\alpha$  emission, plus securely pre-main sequence locations in the H-R diagram. Neither star has identifiable infrared excess.

If real, the slow rotation rates of these two stars are far out on the tail of the period distribution (Figure 4) for this cluster – a full factor of 10 above the median rotation rate – and anomalously slow for members of a star-forming region. Indeed, it is unusual for late K-type stars of any age to rotate this slowly, although some such slow rotators are known among the field star population and among  $> 5$  Gyr old cluster members (Curtis et al. 2020).

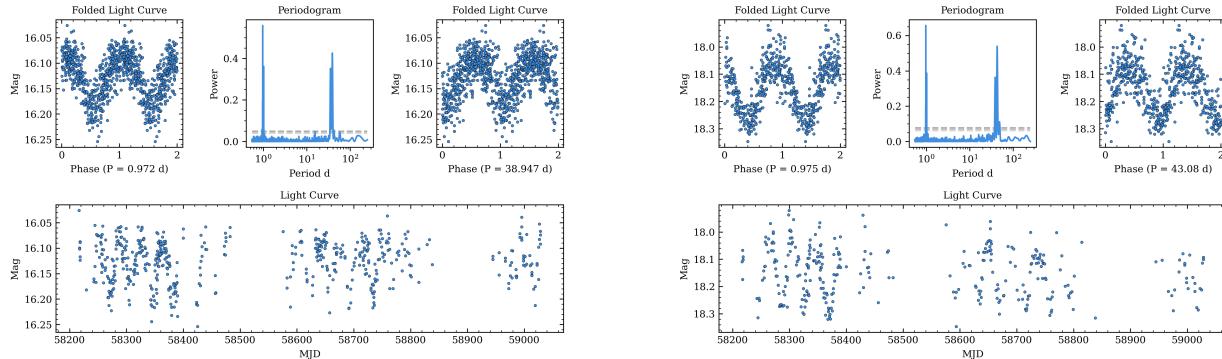
On interpretation is that the multi-periodicity could indicate that these two sources are each binary systems, in which both components of the binary are individually detected as spotted rotators. In this scenario, the two short periods might be supposed as real, with the two long periods beat aliases. However, the relative periodogram peak heights do not seem consistent with this. Furthermore, a recent Keck/HIRES spectrum shows no sign of a spectroscopic binary

in cross correlation analysis. If, instead, one of the short and one of the longer periods is real, then we would need to explain how angular momentum was removed so quickly from just one component of a binary. We ultimately reject the short+long period binary scenario on the grounds that subtracting just one of these signals from the light curve effectively removes both peaks in the periodogram.

In terms of beat periods where  $1/P_{beat} = 1/P_{true} \pm 1$ , the taller peak around 1 day has a beat at the shorter peak around 40 days. Similarly, the taller peak around 40 days has a beat at the shorter peak around 1 day. Although we can explain one of the longer periods as a beat, we can not explain both long periods in that manner through simple approximations.

The situation resolves itself when we consider that the strongest peak for each of the two stars is just under one day, and quite close to the data sampling interval. Simulating the short period at the actual ZTF cadence does produce both short-period and both long-period peaks in the observed ratio of their powers. Specifically, injecting a pure sinusoid with period 0.972 days (FK 176) and 0.975 days (FK 286) using the actual MJD values from the ZTF data stream, results in the long period peaks that can be seen by eye in Figure 16. The long period aliases of the true short periods weaken substantially only when the sampling times are randomized beyond the current staggered ZTF sampling in this particular field, with a standard deviation exceeding 4 hours.

As a check on this conclusion, we consulted the ZTF high cadence data (Kupfer et al. 2021) available for this field on MJD = 58448, 58451, and 58455 days. Although stated earlier as having generally poor photometry, one of these nights is better than the other two, and for FHK 176 the data support the tallest periodogram peak as the true period. Systematic changes in the star's brightness can be seen during the single night, suggesting that the short  $\sim 1$  day period is the correct one. For FHK 286, the situation is more ambiguous but we believe the same conclusion applies.



**Figure 16.** Light curves, periodograms, and phased light curves at the two strongest peaks for FHK 176 and FHK 286. Each of these sources has a short period just below one day that appears to be the real period. However, there is complex periodogram structure that includes a double-peak around  $\sim 1$  day and another double-peak around  $\sim 40$  day. The latter periodic signal can be seen directly in the unphased light curves. See text for how the signal can be deciphered.