Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **519,** 5271–5287 (2023)
Advance Access publication 2022 December 27

https://doi.org/10.1093/mnras/stac3801

The ASAS-SN catalogue of variable stars X: discovery of 116 000 new variable stars using *G*-band photometry

C. T. Christy [1,2★] T. Jayasinghe [2,3,4†] K. Z. Stanek,[2,3] C. S. Kochanek,[2,3] T. A. Thompson [2,3] B. J. Shappee,[5] T. W.-S. Holoien [6‡] J. L. Prieto,[7] Subo Dong[8] and W. Giles[9]

[1]*Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721, USA*
[2]*Department of Astronomy, The Ohio State University, 140 West 18th Avenue, Columbus, OH 43210, USA*
[3]*Centre for Cosmology and Astroparticle Physics, The Ohio State University, 191 W. Woodruff Avenue, Columbus, OH 43210, USA*
[4]*Department of Astronomy, University of California Berkeley, Berkeley, CA 94720, USA*
[5]*Institute for Astronomy, University of Hawai'i, 2680 Woodlawn Drive, Honolulu, HI 96822, USA*
[6]*The Observatories of the Carnegie Institution for Science, 813 Santa Barbara St., Pasadena, CA 91101, USA*
[7]*Núcleo de Astronomía de la Facultad de Ingeniería y Ciencias, Universidad Diego Portales, Av. Ejército 441, Santiago, Chile*
[8]*Kavli Institute for Astronomy and Astrophysics, Peking University, Yi He Yuan Road 5, Hai Dian District, China*
[9]*ASC Technology Services, Mendenhall Laboratory, 433, 125 Oval Dr S, Columbus, OH 43210, USA*

Accepted 2022 December 19. Received 2022 November 18; in original form 2022 May 9

ABSTRACT

The All-Sky Automated Survey for Supernovae (ASAS-SN) is the first optical survey to monitor the entire sky, currently with a cadence of $\lesssim 24$ h down to $g \lesssim 18.5$ mag. ASAS-SN has routinely operated since 2013, collecting $\sim 2\,000$ to over 7 500 epochs of *V*- and *g*-band observations per field to date. This work illustrates the first analysis of ASAS-SN's newer, deeper, and higher cadence *g*-band data. From an input source list of ∼55 million isolated sources with $g < 18$ mag, we identified 1.5×10^6 variable star candidates using a random forest (RF) classifier trained on features derived from *Gaia*, 2MASS, and AllWISE. Using ASAS-SN *g*-band light curves, and an updated RF classifier augmented with data from Citizen ASAS-SN, we classified the candidate variables into eight broad variability types. We present a catalogue of ∼116 000 new variable stars with high-classification probabilities, including ∼111 000 periodic variables and ∼5 000 irregular variables. We also recovered ∼263 000 known variable stars.

Key words: catalogues – surveys – stars: variables: general – binaries: eclipsing – stars: rotation.

1 INTRODUCTION

Variable stars are an important and dynamic area of modern astronomical research. Variability provides extra observational information (periods, amplitudes, etc.), which can be used to determine physical parameters such as mass, radius, luminosity, and rotation rates (e.g. Percy 2007). For example, δ Scuti variables allow us to study the scaling relations between stellar parameters (effective temperature, surface gravity, density, etc.) and astroseismology (e.g. Hasanzadeh, Safari & Ghasemi 2021). Eclipsing binaries allow us to accurately measure stellar parameters, such as mass and radius (e.g. Torres, Andersen & Giménez 2009). RR Lyrae variables have been used to derive the structural parameters of the inner halo and thick disc of the Milky Way (e.g. Mateu & Vivas 2018). The period-luminosity relationship of Cepheid variables allow distance estimates on cosmic scales (e.g. Leavitt 1908). In short, variable stars are used to solve astrophysical problems, especially those requiring knowledge of distances, stellar structure, and populations (e.g. Feast & Whitelock 2014).

Continuity of coverage and continuing surveys are particularly important for finding rare systems that can become 'Rosetta stones' for stellar processes. In the modern era, large surveys such as the All-Sky Automated Survey (ASAS; Pojmanski 2002), the All-Sky Automated Survey for SuperNovae (ASAS-SN; Shappee et al. 2014; Kochanek et al. 2017; Jayasinghe et al. 2018, 2021), the Asteroid Terrestrial-impact Last Alert System (ATLAS; Heinze et al. 2018; Tonry et al. 2018a), the Catalina Real-Time Transient Survey (CRTS; Drake et al. 2009), EROS (Derue et al. 2002), *Gaia* (Prusti et al. 2016; Brown et al. 2018), MACHO Alcock et al. 2000, the Northern Sky Variability Survey (NSVS; Wozniak et al. 2004), the Optical Gravitational Lensing Experiment (OGLE; Udalski 2004), and the Zwicky Transient Facility (ZTF; Bellm 2014) have rapidly advanced the collection of known variables to over $\sim 2.1 \times 10^6$ examples based on the American Association of Variable Star Observers (AAVSO) catalogue (Watson, Henden & Price 2006).

ASAS-SN was originally designed to study bright supernovae and other transients such as tidal disruption events, cataclysmic variables, AGN, and stellar flares (Holoien et al. 2016). ASAS-SN data are also well-suited for the cataloguing, classification, and study of variable stars. For the initial *V*-band catalogue of variables, the

★ E-mail: collinchristy@arizona.edu
† NHFP Hubble Fellow
‡ NHFP Einstein Fellow

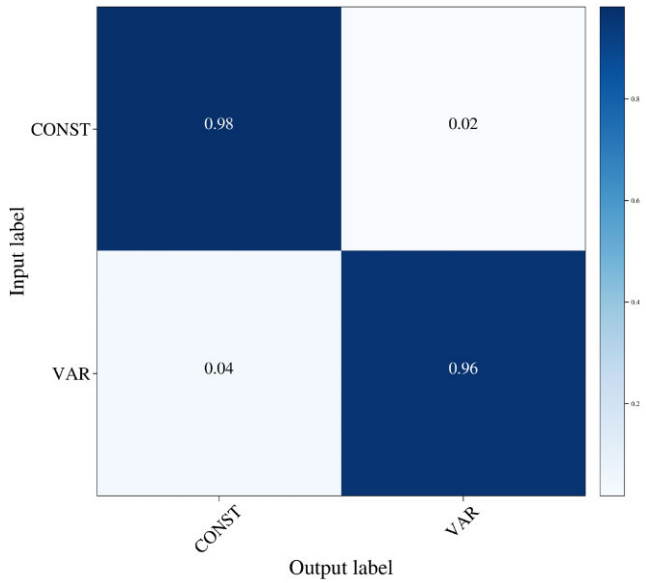

Figure 1. Confusion matrix for the VAR/CONST classifier trained on features from *Gaia* EDR3, 2MASS, and AllWISE.

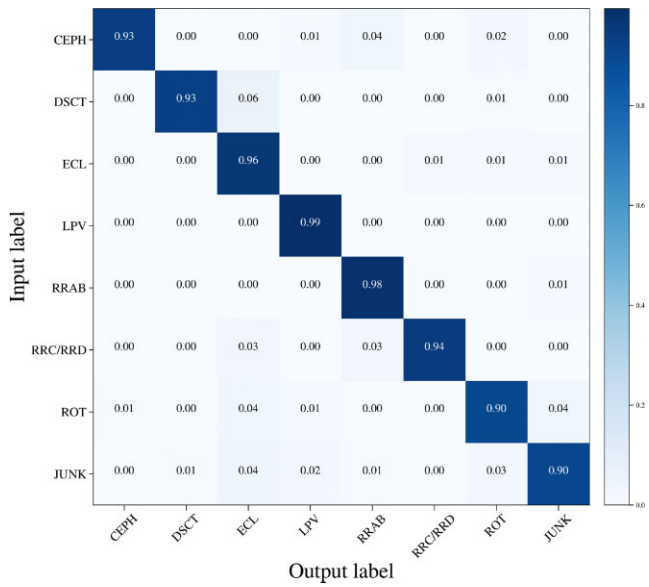

Figure 2. Confusion matrix for the updated classifier from Christy et al. (2022). In the updated version used in this work, we included the 'JUNK' class to identify light curves with spurious variability.

light curves for ∼60 million stars were classified through machine learning techniques, resulting in a catalogue of ∼426 000 variables, of which ∼220 000 were new discoveries (Jayasinghe et al. 2021). We are now using citizen science through the Citizen ASAS-SN project hosted on the Zooniverse[1] (Christy et al. 2021) to identify and classify variables in the *g*-band data. The ASAS-SN citizen science campaign has also already begun to identify a host of new variable stars in our data (Christy et al. 2022).

[1]Zooniverse: https://www.zooniverse.org/

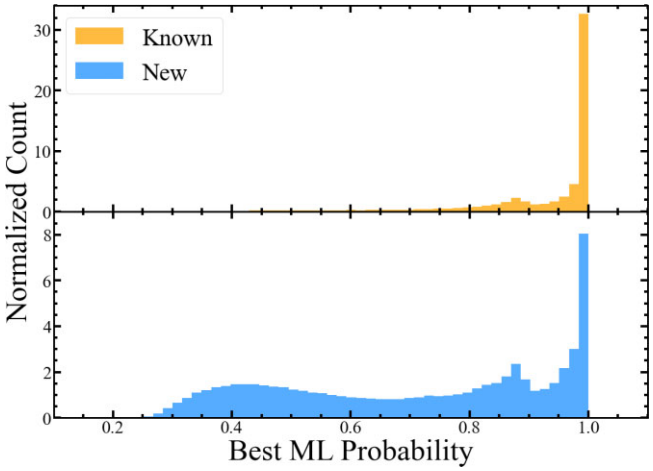

Figure 3. Machine learning probability distribution for the full input list of ∼755 000 variables separated into known and new sources.

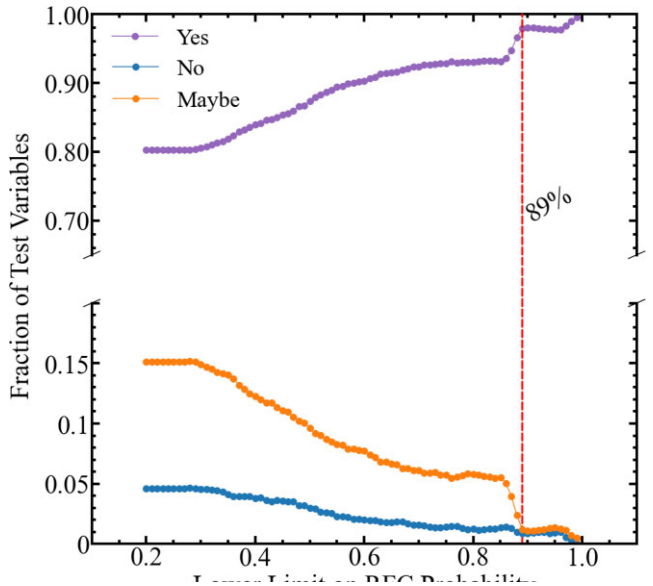

Figure 4. Integral Yes/Maybe/No fraction of test variables in our quality control sample as a function of the RFC classification probability.

In Paper I (Jayasinghe et al. 2018), we discovered ∼66 000 new variables that were flagged during the search for supernovae and homogeneously analysed ∼412 000 known variables from the VSX catalogue in Paper II (Jayasinghe et al. 2019b). In Paper III (Jayasinghe et al. 2019a), we characterized the variability of ∼1.3 million sources in the southern Transiting Exoplanet Survey Satellite (TESS; Ricker et al. 2015) continuous viewing zone and identified ∼11 700 variables, including ∼7 000 new discoveries. In Paper IV (Pawlak et al. 2019), we have also explored the synergy between ASAS-SN and large-scale spectroscopic surveys using data from APOGEE (Holtzman et al. 2015) with the discovery of the first likely non-interacting binary composed of a black hole with a field red giant (Thompson et al. 2019). In Paper V, we identified ∼220 000 variable sources with $V < 17$ mag in the southern hemisphere, of which ∼88 300 were new discoveries (Jayasinghe et al. 2019c). In Paper VI, we derived period–luminosity relationships

Table 1. ML Classification breakdown of the variables from this search.

RF Classification	Description	N_{Known}	N_{New}	N_{New}/N_{Known}
CWA	W Virginis type variables with P > 8 d	153	–	–
CWB	W Virginis type variables with P < 8 d	73	1	0.01
DCEP	δ Cephei-type classical Cepheid variables	432	2	<0.01
DCEPS	First overtone Cepheid variables	109	3	0.03
DSCT	δ Scuti type variables	848	1 547	1.82
EA	Detached Algol-type binaries	17 447	4 480	0.26
EB	β Lyrae-type binaries	11 820	1 551	0.13
EW	W Ursae Majoris type binaries	42 737	4 833	0.11
GCAS	γ Cassiopeiae variables	–	–	–
HADS	High-amplitude δ Scuti type variables	1 725	506	0.29
L	Irregular variables	9 152	4 786	0.52
M	Mira variables	6 287	363	0.06
ROT	Spotted Variables with rotational modulation	14 755	38 414	2.60
RRAB	Fundamental mode RR Lyrae variables	18 455	580	0.03
RRC	First overtone RR Lyrae variables	6 518	450	0.07
RRD	Double mode RR Lyrae variables	482	30	0.06
RVA	RV Tauri variables (Subtype A)	3	–	–
SR	Semiregular variables	131 479	57 925	0.44
YSO	Young stellar objects	209	147	0.70
VAR	Variable star of unspecified type	150	409	2.73
Total		262 834	116 027	0.44

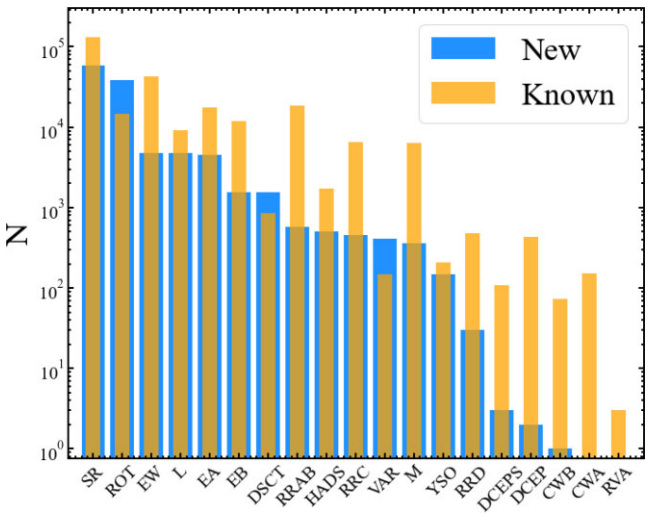

Figure 5. Number distribution of machine learning classifications after probability cuts.

for δ Scuti stars (Jayasinghe et al. 2020c). We studied contact binaries in Paper VII (Jayasinghe et al. 2020b). In Paper VIII, we identified 11 new 'dipper' stars in the Lupus star-forming region (Bredall et al. 2020). In Paper IX, we used spectroscopic information from *LAMOST*, GALAH, RAVE, and APOGEE to study the physical and chemical properties of these variables (Jayasinghe et al. 2021).

In this paper, we present the first all-sky catalogue of variables detected in the newer, deeper, higher cadence *g*-band ASAS-SN data. The complete list of the crossmatched variables and the ASAS-SN discoveries along with their *g*-band light curves are provided online at the ASAS-SN Variable Stars Database (https://asas-sn.osu.edu/variables) and have been reported to the AAVSO. Section 2 details the data and methods used to identify and classify the variable star candidates. In Section 3 we discuss the results, and in Section 4 we present our conclusions.

2 OBSERVATIONS AND METHODS

2.1 Data

In 2014, ASAS-SN began surveying the entire sky in the *V* band with a limiting magnitude of $V \lesssim 17$ mag and a $\sim$2 – 3 d cadence using 8 telescopes on two mounts in Chile and Hawaii (Shappee et al. 2014; Kochanek et al. 2017). Since 2018, ASAS-SN has shifted to using a *g*-band filter and expanded to 20 cameras on 5 mounts, adding new units in South Africa, Texas, and Chile (Jayasinghe et al. 2018). All of the ASAS-SN telescopes are hosted by the Las Cumbres Observatory (LCO; Brown et al. 2013). When compared to the *V*-band data, the *g*-band data have an improved depth ($g \lesssim 18.5$ mag), cadence ($\lesssim$ 24 h in the *g* band versus $\sim$2 – 3 d in the *V* band), and reduced diurnal aliasing due to the longitudinal spread of the ASAS-SN units. The ASAS-SN *V*-band observations were made by the 'Brutus' (Haleakala, Hawaii) and 'Cassius' (CTIO, Chile) quadruple telescopes between 2013 and 2018. Our *g*-band observations added data from the 'Payne' (Sutherland, South Africa), 'Bohdan' (CTIO, Chile), and 'Leavitt' (McDonald, Texas) quadruple telescopes starting in 2018. Each ASAS-SN camera takes 3 images with 90 s exposures for each epoch. The field of view of an ASAS-SN camera is 4.5 deg^2, the pixel scale is 8.″0, and the FWHM is typically $\sim$2 pixels. ASAS-SN saturates at $g \sim 11 - 12$ mag, but we attempt to correct the light curves of saturated sources for the bleed trails (see Kochanek et al. 2017). The *g*-band light curves were extracted using image subtraction (Alard & Lupton 1998; Alard 2000) and aperture photometry on the subtracted images with a 2-pixel radius aperture. We rejected images taken during poor weather conditions, images that were out of focus with FWHM >2 pixels, images that had poor astrometric solutions, and images where the variable source is within 0.2° of a detector edge (Jayasinghe et al. 2018). We corrected

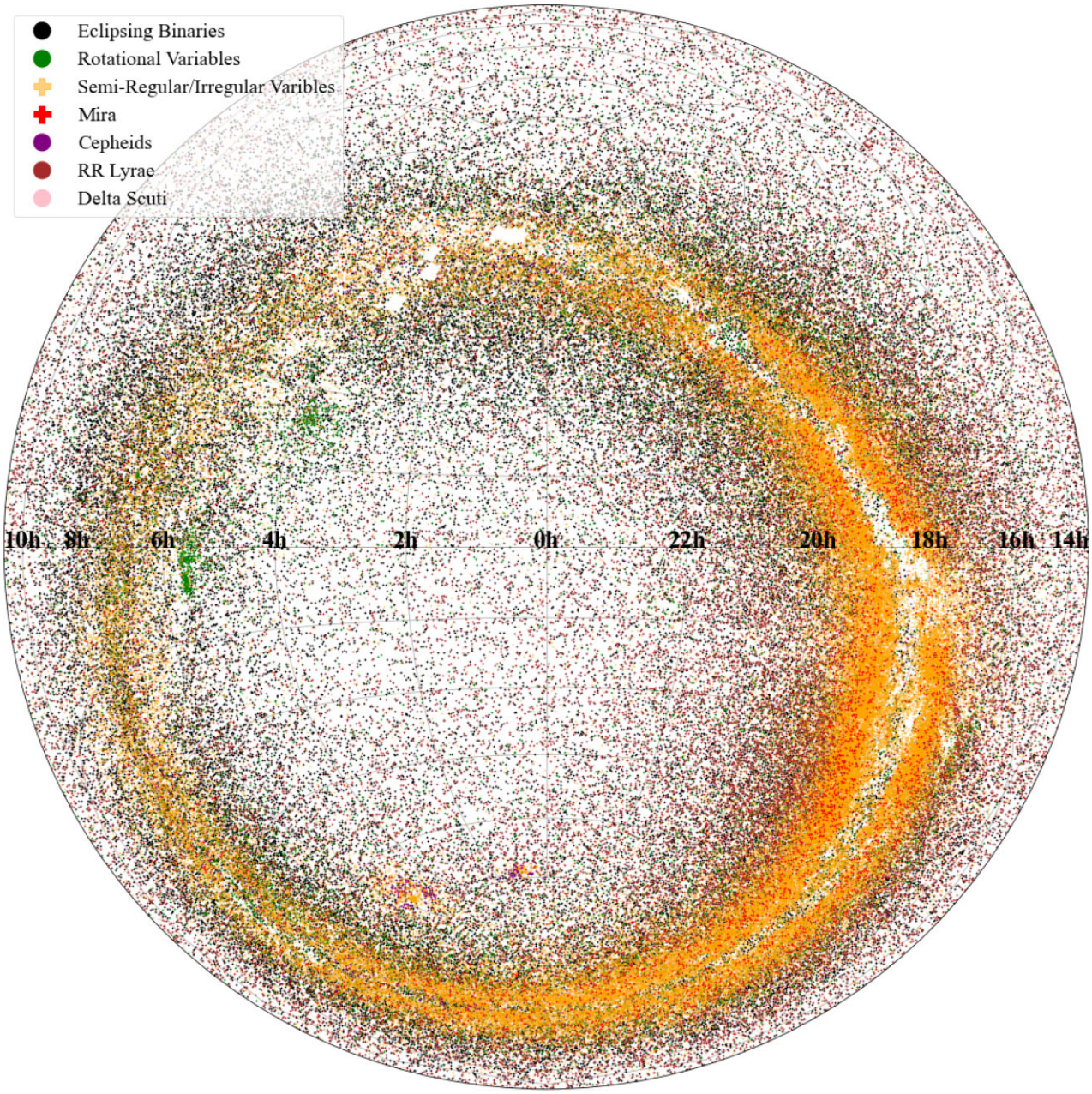

Figure 6. Equatorial distribution of the ∼263 000 known variables we recovered in equatorial coordinates.

the zero point offsets between the different cameras as described in Jayasinghe et al. (2018). The photometric errors were recalculated as described in Jayasinghe et al. (2019a).

2.2 Identifying variable star candidates

We started with the ATLAS All-Sky Stellar Reference Catalogue `refcat2` (Tonry et al. 2018b) as our input source catalogue. The `refcat2` catalogue includes *g*, *r*, *i*, and *z* photometry for sources from PanSTARRS DR1, the ATLAS Pathfinder photometry project, the ATLAS reflattened APASS data, SkyMapper DR1, APASS DR9, Tycho-2, and the Yale Bright Star Catalogue (Tonry et al. 2018b). We selected ∼54.8 million `refcat2` sources with $g < 18$ mag and `r1`$>30''$, where `r1` is the radius at which the cumulative *G* flux in the aperture exceeds the flux of the source being considered and is a measure of the crowding around a star. We use the limit on `r1` to reduce the number of heavily blended sources.

The production of 55 million of light curves is computationally expensive, so we used external photometry from *Gaia* EDR3, 2MASS, and AllWISE to identify likely variable sources rather than simply generating light curves for every source. In particular, the photometric uncertainties in *Gaia* EDR3 encodes information about the photometric variability of sources (see for e.g. Andrew, Swihart & Strader 2021). At a fixed *G* magnitude, variable stars in *Gaia* have larger photometric uncertainties than constant stars and therefore, we can use the photometric uncertainties available in *Gaia* EDR3 as a feature to identify stellar variability.

We built a variability classifier based on a random forest (RF) model with `scikit-learn` (Pedregosa et al. 2012). The goal was to first divide the stars into two groups: CONST (constant stars) and

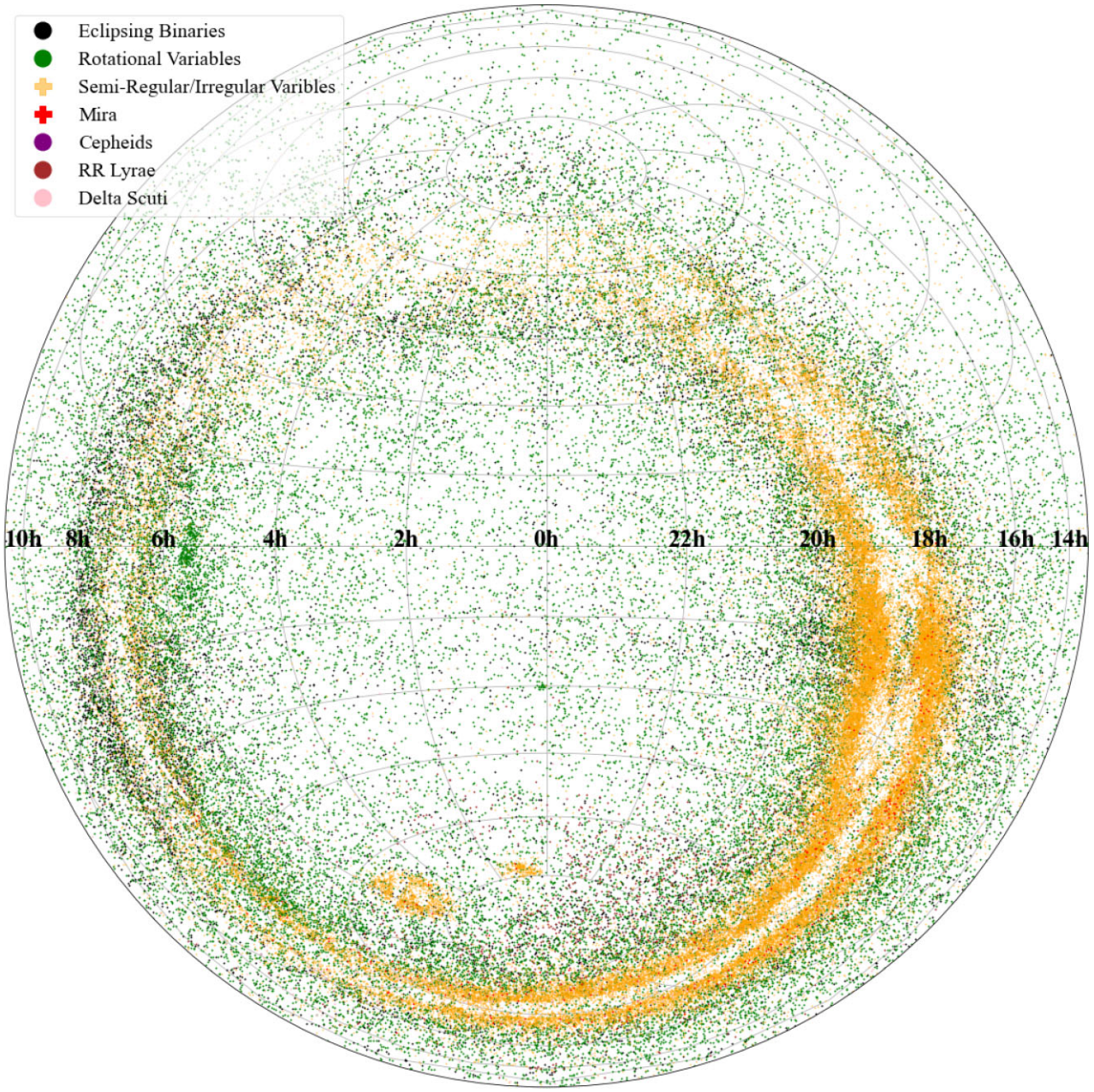

Figure 7. Equatorial distribution of the ∼116 000 new variables we recovered in equatorial coordinates.

VAR (potential variables). The variable star candidates will then be analysed in detail, so it is more important not to lose real variables than to accidentally include non-variables. For the training set, we used the ∼204 000 known variables used to train the RFC variability classifier in Christy et al. (2022) and the variables identified in Citizen ASAS-SN DR1. For the constant sources, we used a set of ∼250 000 non-variable sources identified in Citizen ASAS-SN DR1. We used 16 features from *Gaia* EDR3, 2MASS, and AllWISE. These include the EDR3 G, BP, RP magnitudes and the associated uncertainties, the $BP - RP$ colour, the $BP - RP$ excess factor, the signal-to-noise ratios in G and BP, the renormalized unit weight error (RUWE), the $J - K_s$ colour, the absolute W_{RP} magnitude and the absolute W_{JK} magnitude. The EDR3 signal-to-noise ratios are essentially the ratio of the observed flux divided by the error in the flux. As noted earlier, the EDR3 photometric uncertainties and flux errors encode information about the photometric variability of stars. We also used the absolute, 'reddening-free' Wesenheit magnitudes (Madore 1982; Lebzelter et al. 2018)

$$W_{RP} = M_{\rm RP} - 1.3(BP - RP)\,, \tag{1}$$

and

$$W_{JK} = M_{\rm K_s} - 0.686(J - K_s) \tag{2}$$

and the probabilistic EDR3 distances from Bailer-Jones et al. (2021).

The parameters of the RF model were optimized using cross-validation to maximize the overall F_1 score of the classifier. The number of decision trees in the forest was initialized to `n_estimators=1200`. We also limited the maximum depth of the decision trees to `max_depth=16` in order to miti-

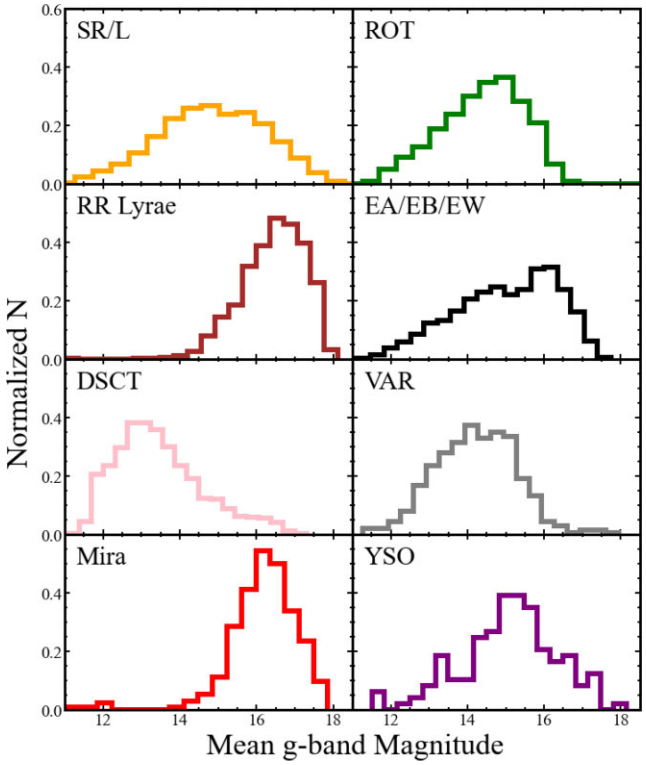

Figure 8. Distribution of the mean g-band magnitudes by variable class.

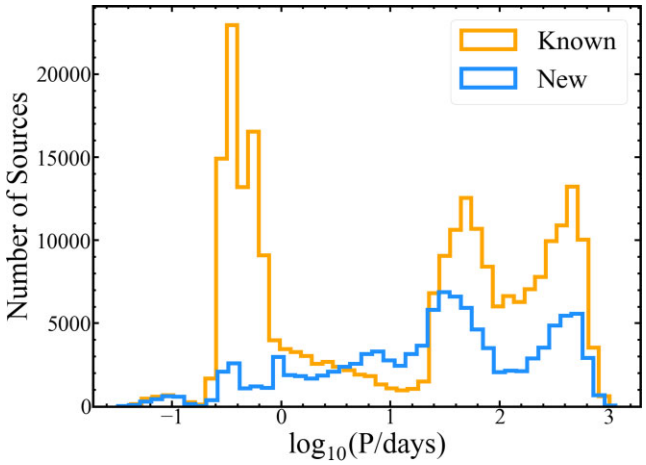

Figure 9. Number distribution of the variability periods for the known and new sources.

gate over-fitting, set the number of samples needed to split a node as `min_samples_split=10` and set the number of samples at a leaf node as `min_samples_leaf=5`. To further minimize over-fitting, we also assigned weights to each class with `class_weight='balanced_subsample'`. For any given source, the RF classifier assigns classification probabilities Prob(Const) and Prob(Var) = 1 − Prob(Const). The output classification of the RF classifier is the class with the highest probability. We split the training sample, using 90 per cent for training and 10 per cent for testing, in order to evaluate the performance of the RF classifier. The confusion matrix for the trained RF model is shown in Fig. 1. The greatest confusion (4 per cent) arises from input variable sources that are subsequently classified

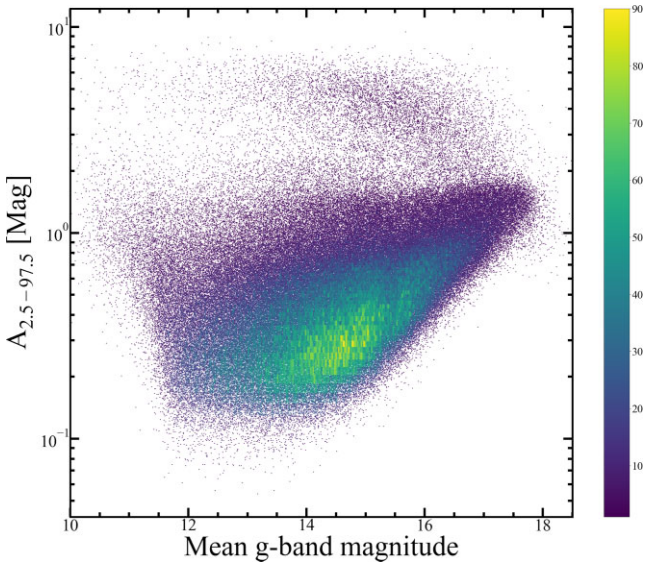

Figure 10. Distribution of variability amplitude with mean g-band magnitude. This includes both the new and known variables.

as CONST. The overall precision, recall and F_1 parameters for the classifier are 97.3 per cent, 97.1 per cent, and 97.2 per cent, respectively.

We then classified the ∼54.8 million `refcat2` sources using the trained RF classifier, and we identified ∼1.48 million variable star candidates (∼2.7 per cent). We extracted ASAS-SN g-band light curves of these candidates and determined periods using the `ASTROPY` implementation of the generalized Lomb–Scargle (GLS, Scargle 1982; Zechmeister & Kürster 2009) periodogram over the range $0.025 \leq P \leq 1\,000$ d. To minimize computational costs, we chose to use only the GLS periodogram. The GLS periodogram can sometimes fail for more complex light curves, including detached eclipsing binaries, and therefore, such sources may not be selected as periodic variables. For the eclipsing binaries, GLS tends to measure a period that is half of the true period (see Jayasinghe et al. 2018, 2019b). To correct for this, we have adapted the automated period doubling routine from Jayasinghe et al. (2019b) for the eclipsing binaries, which significantly reduces the number of binaries with periods corresponding to half the true period. Future work will also use the Box Least Squares (BLS; Kovács, Zucker & Mazeh 2002) periodogram to improve the detection of detached eclipsing binaries.

2.3 Classifying variable stars

We retrained the g-band RF classifier described in Christy et al. (2022) to include a new category for spurious 'JUNK' variables. This classifier uses features derived from both the light curves and *Gaia* EDR3, 2MASS, and AllWISE photometry. The features derived from the light curves include log(P), R_{41}, R_{31}, R_{21}, A, IQR, $A_{\rm HL}$, Skew, rms, σ, and MAD from Table 2 in (Jayasinghe et al. 2019b), and the updated features $\delta(t, P)$, $\delta(P, 2P)$ from Christy et al. (2022). We also use the $BP - RP$, $J - K_s$, $J - H$ colours, and the M_{Ks} absolute magnitude derived from photometry. The training set for the JUNK class was the ∼12 000 'JUNK' variables identified by citizen scientists in the first data release from Citizen ASAS-SN (Christy et al. 2022). These 'JUNK' variables include light curves dominated by systematics (e.g. saturation artifacts from nearby bright stars), and those that have poor signal-to-noise (see Christy

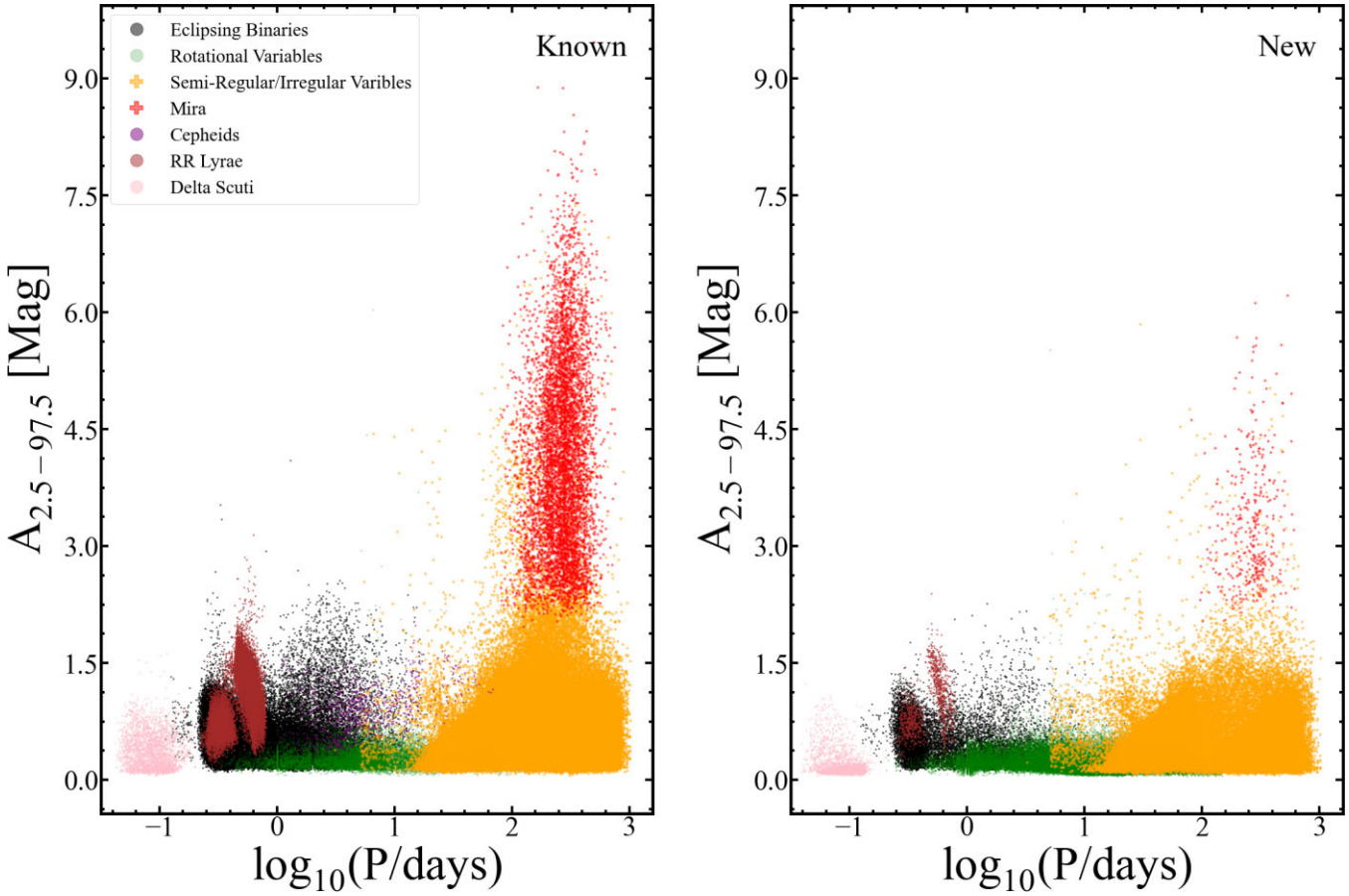

Figure 11. The *g*-band variability amplitude versus $\log_{10}$(P/days) relationship for the known variables (Left) and new variables (Right) in our catalogue using labels given by the *g*-band machine learning classifier.

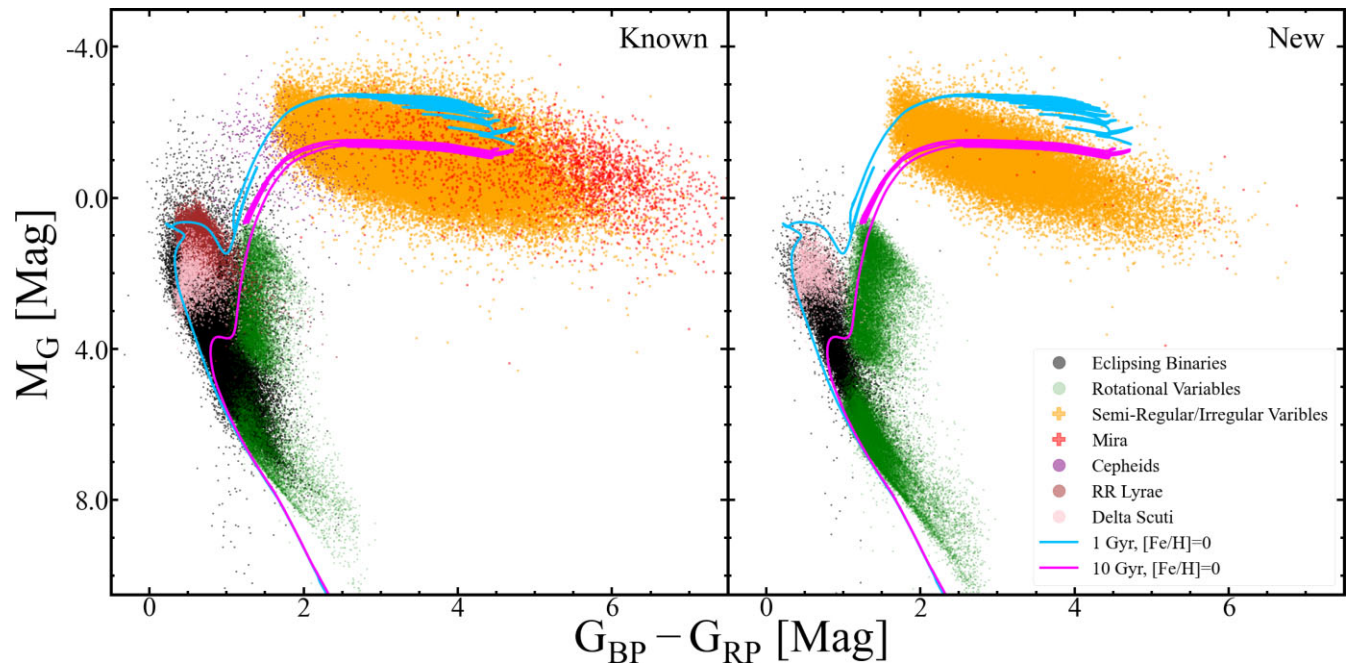

Figure 12. The *Gaia* EDR3 M_G versus $G_{BP} - G_{RP}$ colour-magnitude diagram for the known variables (Left) and new variables (Right) in our catalogue using labels given by the *g*-band machine learning classifier, including 1 Gyr and 10 Gyr [Fe/H]=0 MIST isochrones.

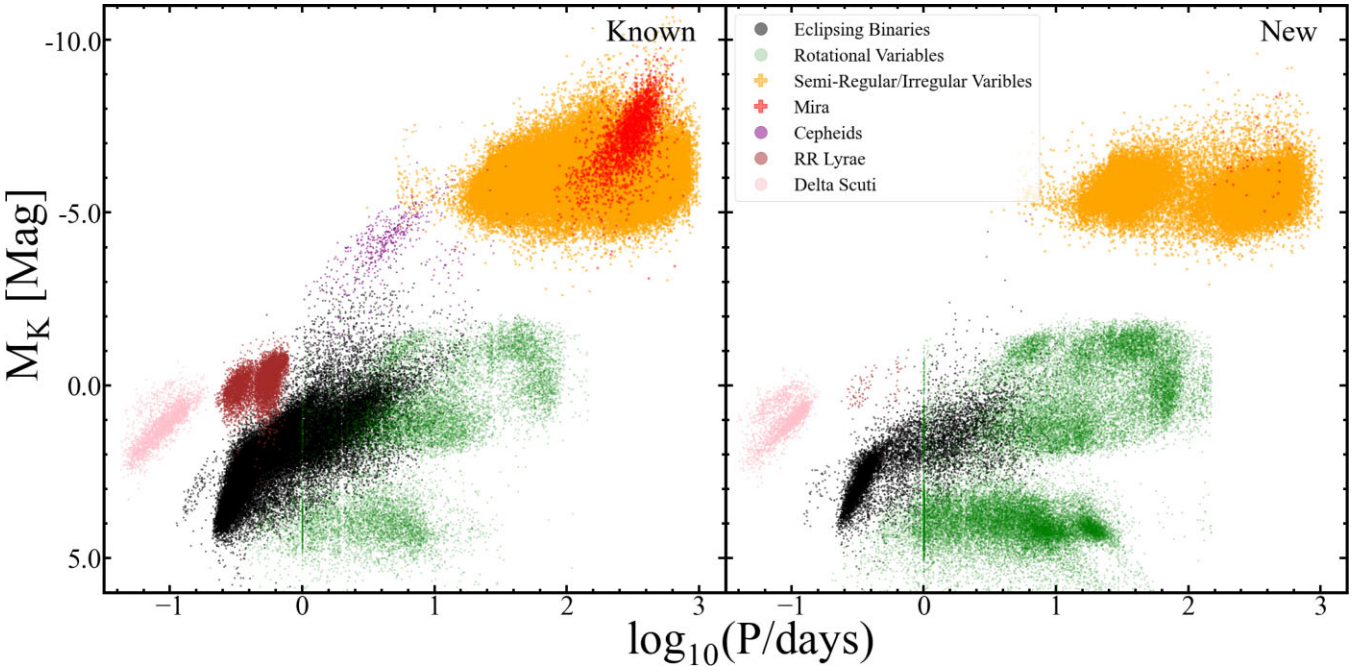

Figure 13. The M_K versus $\log_{10}$(P/days) period-luminosity diagram for the known variables (Left) and new variables (Right) in our catalogue using labels given by the *g*-band machine learning classifier.

et al. 2022). Training sets for variable star classification tend to be imbalanced due to the different occurrence rates of various variable types. To reduce over-fitting, weights were assigned to each class by using `class_weight='balanced_subsample'`. We set the number of decision trees in the forest as `n_estimators=1000`, pruned the trees at a maximum depth of `max_depth=16`, set the number of samples needed to split a node as `min_samples_split=10` and set the number of samples at a leaf node as `min_samples_leaf=5`. The updated RF classifier classifies sources into eight broad classes (CEPH, DSCT, ECL, LPV, RRAB, RRc/RRd, ROT, and JUNK) which are subsequently refined into sub-classes (see Jayasinghe et al. 2019b). The overall precision, recall and F_1 parameters for the updated RF classifier are 91.6 per cent, 94.2 per cent, and 92.8 per cent, respectively. The confusion matrix for the trained RF model is shown in Fig. 2.

We applied the updated variability classifier to the ∼1.48 million variable star candidates. To reduce the number of false positives, we imposed a probability cut of Prob > 0.95 for variables with periods close to diurnal and lunar aliases (e.g. 1, 2, 29, and 60 d). Additionally, to eliminate false positives caused by spurious variability, we used a probability cut of Prob > 0.8 for short period DSCT variables. Following this step, we are left with ∼755 000 variables.

We cross-matched the list of ∼755 000 variables with the AAVSO VSX (Watson et al. 2006), OGLE III (Poleski et al. 2012), and OGLE IV (Kozlowski et al. 2013; Soszyński et al. 2014; Udalski, Szymański & Szymański 2015; Soszyński et al. 2015, 2016; Udalski et al. 2018; Pietrukowicz et al. 2020; Soszyński et al. 2021) catalogues using a matching radius of 16 arcsec and found 359 265 previously known variables. The VSX catalogue contains all the variables previously identified by many wide field surveys, including ASAS-SN (Jayasinghe et al. 2021), ATLAS (Heinze et al. 2018), WISE (Chen et al. 2018), and ZTF (Chen et al. 2020). After excluding the known variables, there are 395 494 new variables in our list.

2.4 Quality control

In Fig. 3, we show the distribution of the highest probability assigned to each variable candidate's classification for the known variables and the new candidates. The probability distribution for the known variables is more concentrated towards unity than the new candidates. The differences indicated a need for additional sample restrictions.

We addressed this by building a private test workflow on the Zooniverse platform. We randomly selected 2 000 light curves from the new candidates. The light curves were presented in the same fashion as in ASAS-SN's citizen science project Citizen ASAS-SN (see Christy et al. 2022) which include the light curves phased using the best period, phased using twice the best period, and the observed light curve. The user was asked 'Is this a variable?' with a multiple choice response of Yes, No, and Maybe. After the test set was fully classified we found that ∼80 per cent received a 'Yes' indicating that variability was clear while ∼5 per cent received a 'No' and ∼15 per cent received a 'Maybe'. Fig. 4 shows how their Yes/Maybe/No classification scaled with the RFC classification probability. We decided to set the lower limit cutoff at $P \geq 0.89$, and for this limit the relative numbers are 98 per cent Yes, 1 per cent No, and 1 per cent Maybe. Spot checks of some of the high probability light curves voted as No and Maybe showed that these were likely user classification mistakes. Applying this cut to our full variable sample reduced the number of known and new candidates by 27 per cent and 71 per cent, respectively to leave in ∼263 000 known variables and ∼116 000 new variables.

3 DISCUSSION

The complete catalogue of 378 861 variables has been added to the ASAS-SN Variable Stars Database (https://asas-sn.osu.edu/variables) along with their *g*-band light-curve data and will be publicly catalogued in the AAVSO VSX catalogue. The ASAS-SN catalogue

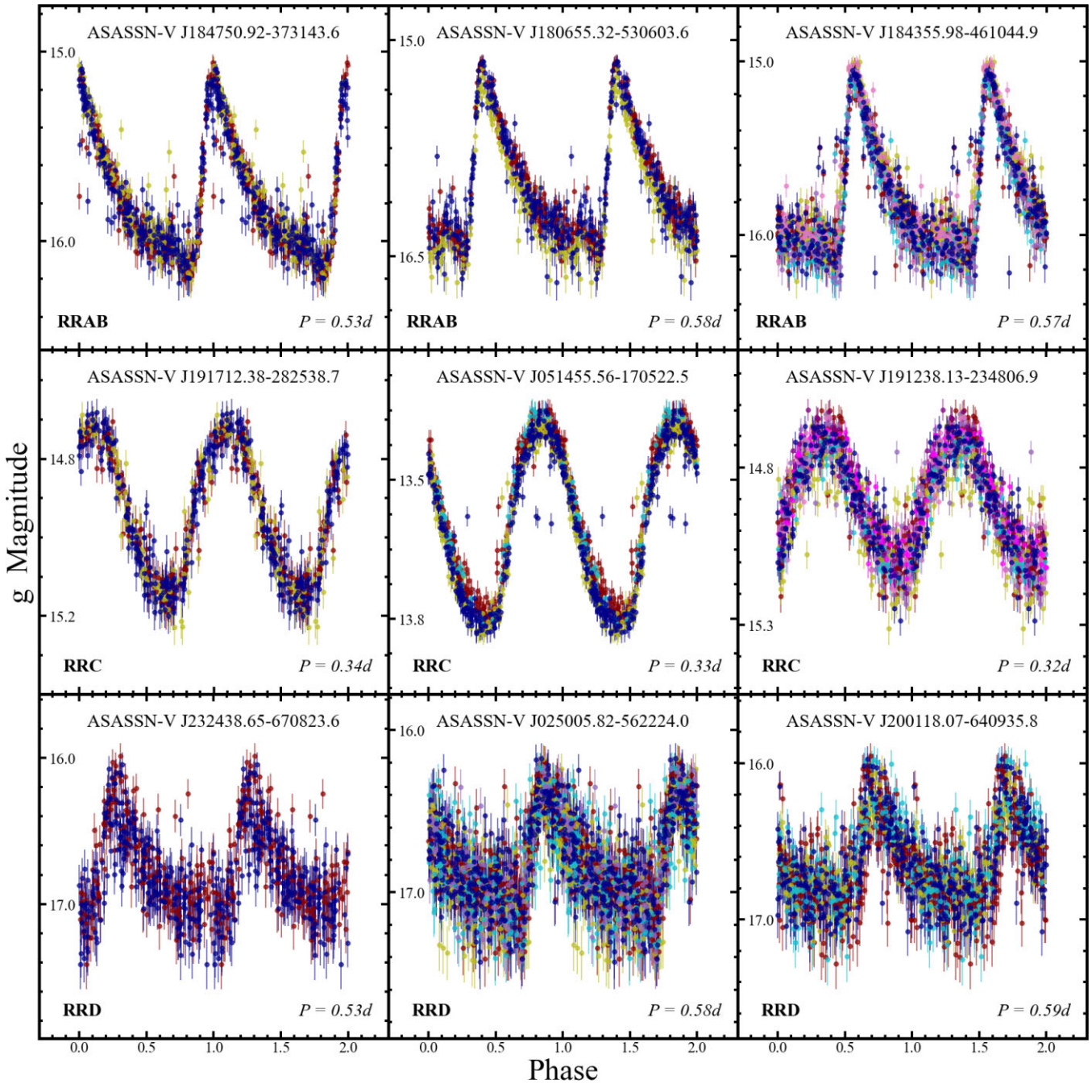

Figure 14. Phased light curves for examples of the newly discovered RR Lyrae variables. The light curves are scaled by their minimum and maximum *g*-band magnitudes. Different coloured points correspond to data from the different ASAS-SN cameras. The different variability types are defined in Table 1.

can be downloaded in its entirety at the ASAS-SN Variable Stars Database.

In Table 1 and Fig. 5 we show the number of sources for each variability type from this search. The two most common known variables found are Semiregular (SR) and W Ursae Majoris type binaries (EW). Of the new discoveries, the most common are Semiregular (SR) and spotted rotational variables (ROT). The number of known variable candidates was greater than the number of new candidates for all except for the ROT, DSCT, and generic VAR classes. The VAR class describes systems that display clear variability, but could not be categorized by the machine learning classifier. The large number of rotating spotted stars (ROT) among our new discoveries seems to be a property of using the *g* band instead of the *V* band. The larger rotational variability signal in the *g* band is likely a combination of it being a bluer band and that it contains the strong Ca, H, and K absorption feature, whose strength varies significantly with activity. We found no new W Virginis type variables (CWA), RV Tauri variables (RVA), or γ Cassiopeiae (GCAS) type variables. There were candidates classified as GCAS variables, but they all systematically had low classification probabilities and were dropped after the cuts. The full sky distribution for several common classes of the known and new variables are shown in Figs 6 and 7. The

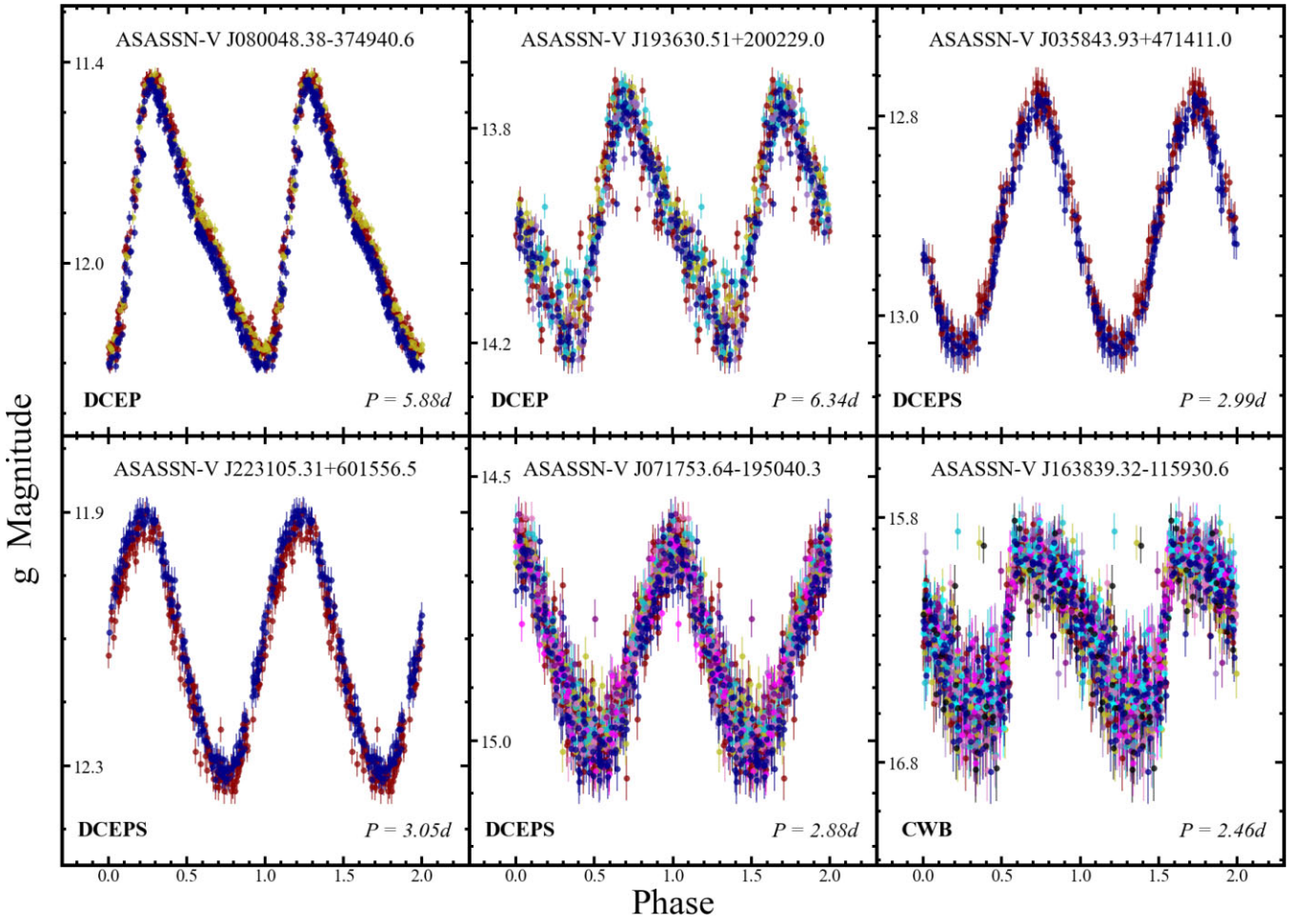

Figure 15. Light curves for the newly discovered Cepheid variables. The format is the same as for Fig. 14.

differences illustrate the large number of new ROT compared to the known ones. The different sky distributions are simply the differences between dwarfs and giants. The ROT variables are predominantly dwarfs and so have a relatively isotropic distribution. The SR/L variables are giants, so they trace the large-scale structure of the Galaxy.

Fig. 8 shows the distribution of the most common new variables in magnitude. Not surprisingly, most are concentrated towards fainter magnitudes since the *g*-band data are deeper than the *V* band, with higher amplitude variables (e.g. RR Lyrae and Mira variables) being found at fainter magnitudes than lower amplitude variables (e.g. spotted stars and the generic VAR class). The two classes which do not conform to this pattern are the DSCT and SR/L variables. The DSCT variables have very short periods (hours) and their identification benefits greatly from the significantly cleaner window function for identifying short period variables in the *g*-band data. Since the ASAS-SN *g*-band light curves are the first ever to provide this level of time sampling, it is not surprising that we would identify large number of DSCT variables even at bright magnitudes. The identification of larger numbers of intermediate magnitude SR/L variables is likely a consequence of the steadily increasing total time span of the ASAS-SN light curves. We probably do not see the same thing for the Miras because their variability amplitudes are so large that they are very difficult to miss.

In Fig. 9, we show the period distributions for the variables. The structures at long periods are fairly similar, and are simply due to finding different relative numbers of Mira, semiregular, and irregular variables. The lack of the large peak for periods shorter than a day is due to the low yield of RR Lyrae and EW type binaries among the new variables. These variables are comparatively easy to detect with relatively large amplitudes and easily sampled periods, so it is not surprising that the completeness of existing samples is high. The comparable numbers at very short periods comes from the many new DSCT variables. The larger numbers near 10 d come from the large number of ROT.

We examined the discrete period spikes close to the expected aliased periods of ∼1 d (diurnal alias), ∼14.8 d (half lunar alias), and ∼29.5 d (lunar alias). We spot checked a random sample of variables located near each spike and found that those with the lunar and half lunar aliased periods showed poorer signs of variability and were systematically dimmer than the sources with periods at ∼1 d. We suspect the dimmer sources are more prone to suffer from a low signal to noise causing the GLS periodogram to find the wrong period. To combat this, we imposed a magnitude cutoff of $g < 16$ for variables with potentially aliased periods which removed the spikes at ∼14.8 and ∼29.5 d. This cut did not remove the discrete spike near $P \sim 1$ d as most of these variables were brighter and showed clear signs of variability. We suspect that the excess of $P \sim 1$ d period variables comes from relaxing the discrimination against

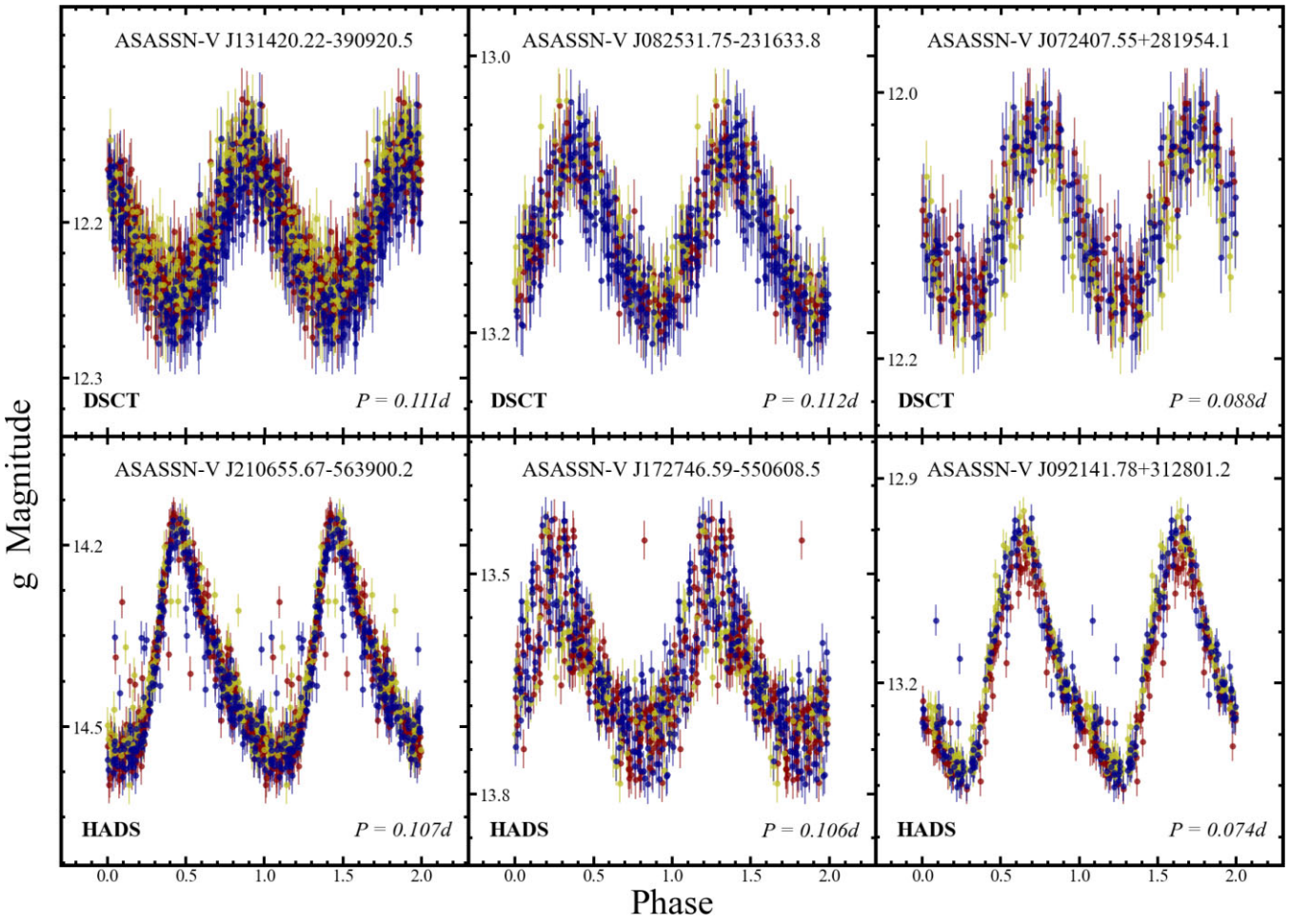

Figure 16. Light curves for examples of the newly discovered δ Scuti variables. The format is the same as for Fig. 14.

diurnal aliases because of the better sampling of the *g*-band data. Citizen ASAS-SN also provided us with many examples of false positives at the diurnal aliases which went into training for the JUNK classification.

Fig. 10 shows the distribution of variability amplitude as a function of the mean *g*-band magnitude. The amplitude $A_{2.5-97.5}$ is the 2.5 per cent to the 97.5 per cent percentile range of the data. For bright stars, we identify variables down to amplitudes of order $\sim$0.05 mag, and then require progressively higher amplitudes beyond $g \sim 16$ mag. We show the period-amplitude distribution for the known and new sources in Fig. 11. We find relatively few bright, high-amplitude pulsators, which is another indicator that searches for such variables are relatively complete.

Fig. 12 shows the *Gaia* EDR3 M_G and $G_{BP} - G_{RP}$ colour-magnitude diagram for the known and new sources. We also overlay 1 Gyr and 10 Gyr MIST isochrones for [Fe/H]=0 to point out various evolutionary stages (Paxton et al. 2018). These tracks illustrate a lack of instability strip variables among the new variables and the abundance of spotted giants, spotted main-sequence stars, and semiregular AGB stars. We found that a higher percentage of the new ROT were located on the main sequence compared to the known rotators. Fig. 13 shows their distribution in M_K and $\log_{10}$(P/days). The two populations of semiregular/irregular variables are clearly seen for the new variables and have been extensively observed by the OGLE survey (see Soszyński et al. 2007). As outlined in Paper VI, the period-luminosity diagram displays two distinct populations of DSCT stars: the fundamental mode pulsators and the more luminous overtone pulsators. We found that our new *g*-band observations were able to more efficiently identify overtone DSCTs. We also found that with our larger sample of ROT, certain regions of the period-luminosity diagram were preferentially occupied. A more thorough analysis of the rotators will be left for a future work. The behaviour of the different broad variable types and where they occupy agrees with the distribution found in the ASAS-SN *V*-band catalogue (Jayasinghe et al. 2019b).

In Fig. 14, we show a selection of light curves for new RR Lyrae variables. In Fig. 15, we show light curves for the new Cepheid variables. This variable class made up the smallest fraction of the new discoveries, with only six new sources. Our observations did not yield any new RV Tauri variables (subtype A) or W Virginis variables with periods greater than 8 d. In Fig. 16, we show a selection of the new δ Scuti type variables. We show light curves for the new eclipsing binaries in Fig. 17, which were collectively the third most common new variables. In Fig. 18, we show the phased and observed light curves for the ROT class. These stars were the second most common type we found and many of their observed light-curves display amplitude modulation due to changes of the stars' spots with time. In Figs 19 and 20, we show the light curves for the new long period and irregular variables. Lastly, we show light curves for new variables that were given the generic VAR class label in Fig. 21.

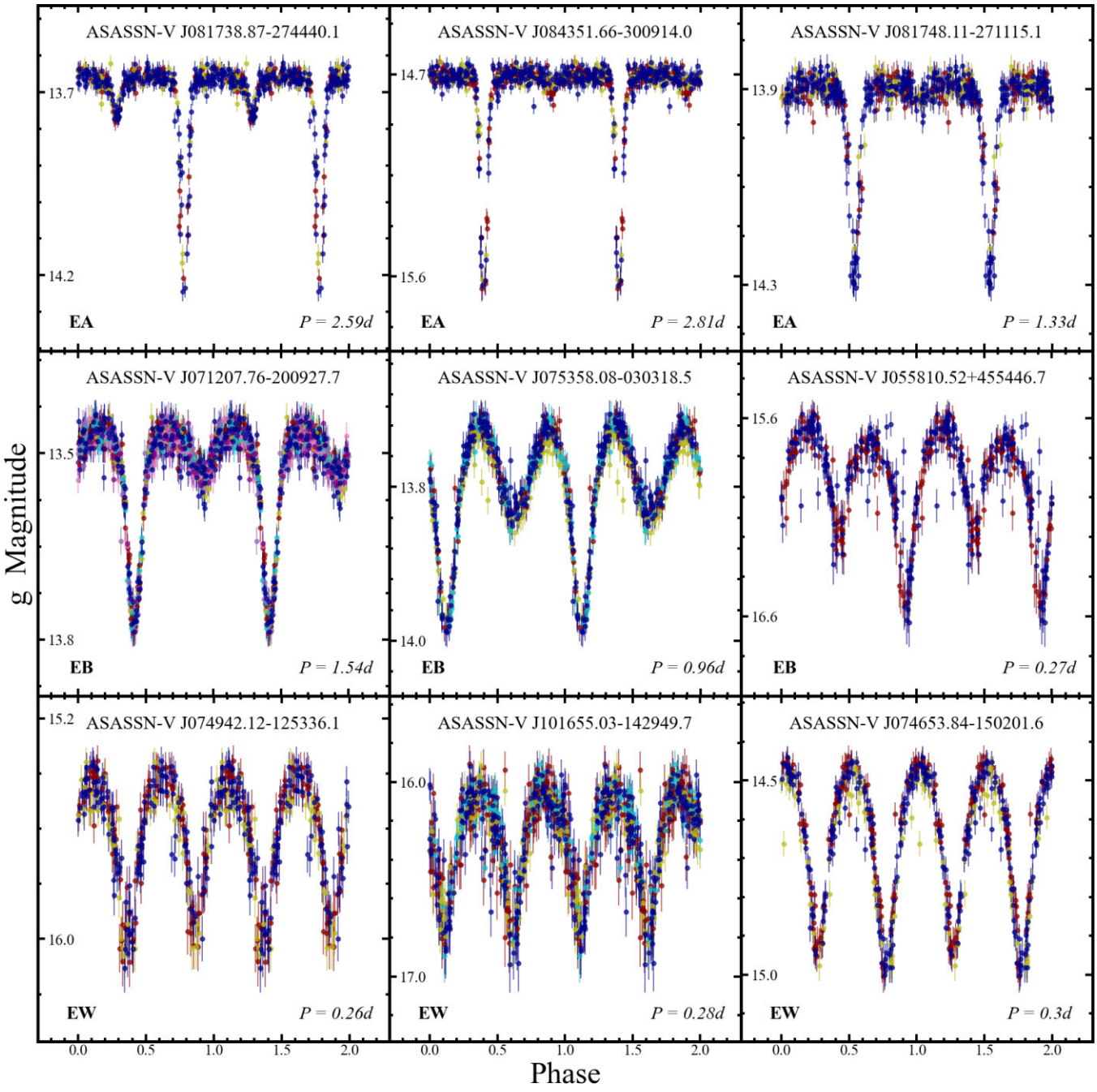

Figure 17. Light curves for examples of the newly discovered eclipsing binary variables. The format is the same as for Fig. 14.

These variables often display a clear signature of variability but the classifier was not able to assign a more specific classification.

4 CONCLUSION

In this paper, we provide an initial catalogue of variable stars using the ASAS-SN *g*-band light curves. The complete list of the crossmatched variables and the ASAS-SN discoveries along with their *g*-band light curves are provided online at the ASAS-SN Variable Stars Database (https://asas-sn.osu.edu/variables). From an input catalogue of ∼54.8 million stars, we identified ∼1.48 million variable candidates based on information from *Gaia* EDR3, 2MASS, and AllWISE. We then analysed the light curves of these sources and found 378 861 variables, of which 262 834 are known variables and 116 027 are new discoveries. We generally recovered more known variables of each type, with the exception of the ROT, DSCT, and generic VAR classes. The most common new variables were the semiregular and ROT. We found an excess of new ∼1 d period variables because the higher cadence and longitude spread of the *g*-band ASAS-SN configuration gives better control of the diurnal aliasing and so allows searches at these periods with fewer false positives. We also find that rotational modulations are stronger in the *g* band, leading to many new ROT.

We plan to incorporate these variables, including the lower probability candidates, into our Citizen Science initiative to help refine our classifications and improve our machine learning techniques.

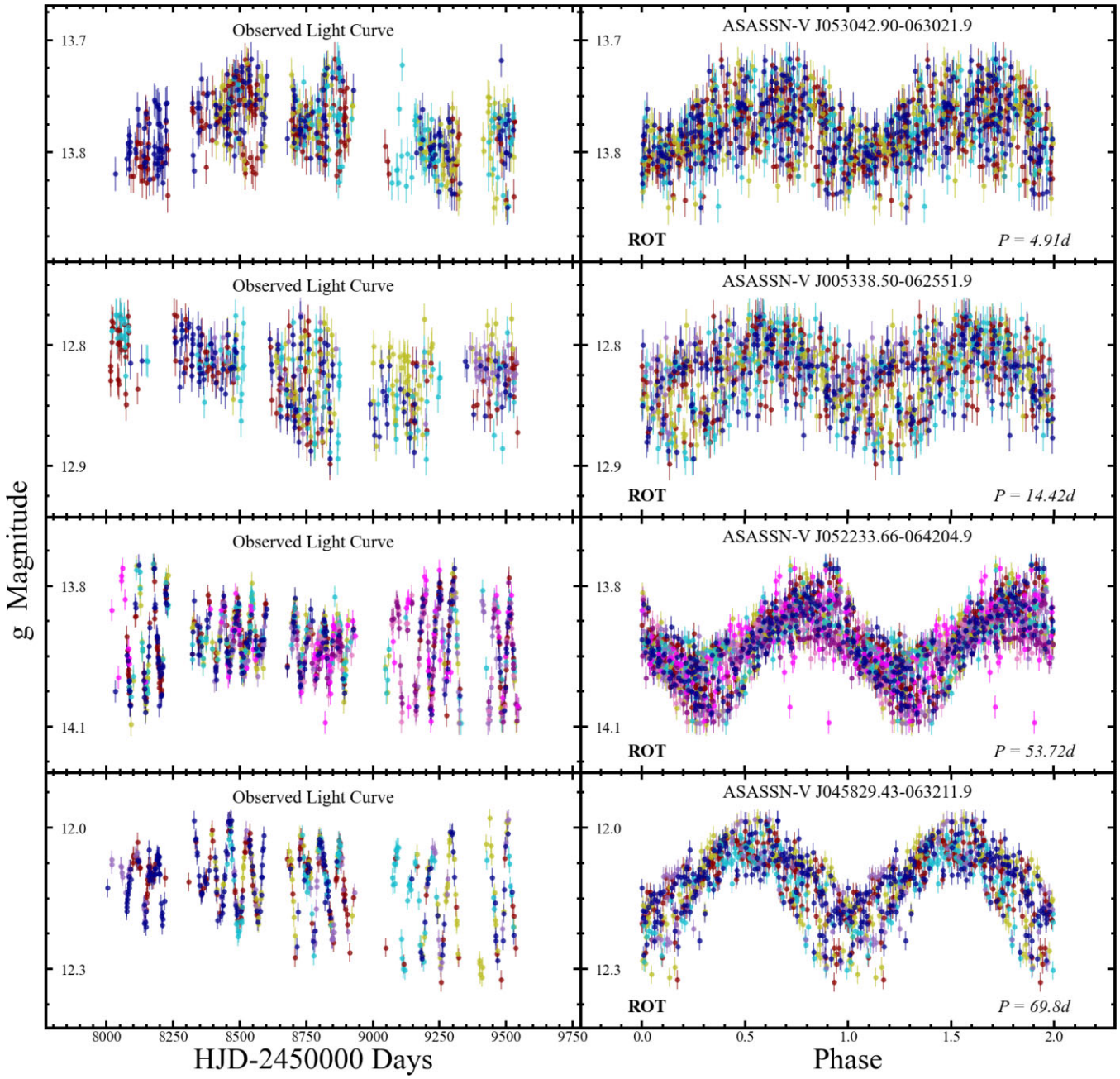

Figure 18. Light curves for examples of the newly discovered highly spotted ROT. The format is the same as for Fig. 14.

The citizen scientists outperformed our present machine learning classifier in identifying spurious variables (Christy et al. 2022). Further building the JUNK training set should lead to steady improvements in the machine learning classifier. The citizen scientists also excelled at identifying unusual or extreme variable candidates. Looking forward to the *Gaia* DR3 catalogue, many of the variables in this catalogue will be bright enough to have radial velocity measurements. *Gaia*'s on-board radial velocity spectrometer (RVS) can collect radial velocities for stars brighter than $G_{\rm RVS} = 14$ mag (Seabroke et al. 2021). With this limiting magnitude, we expect many of the new discoveries in our catalogue to eventually have RV measurements from *Gaia*.

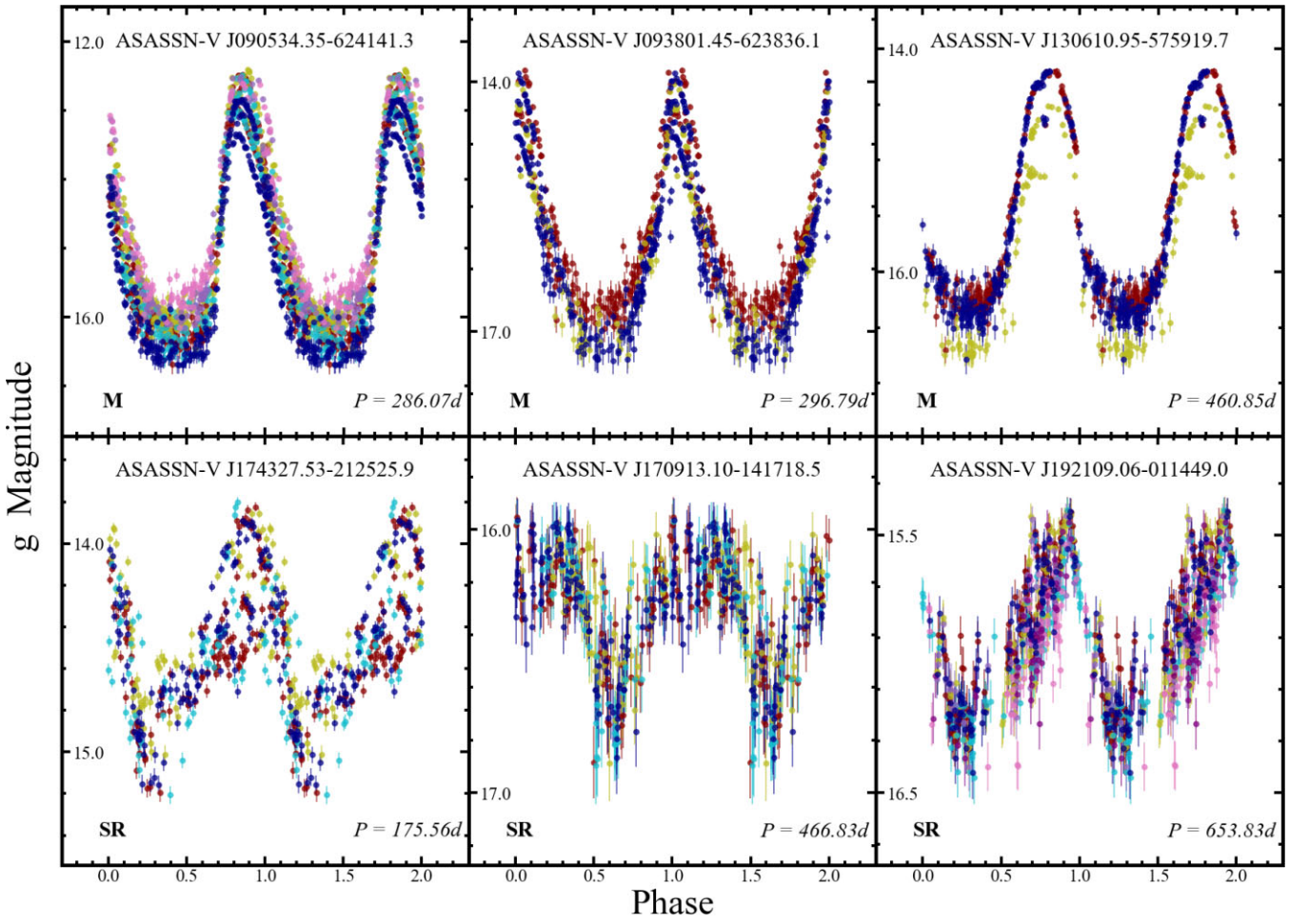

Figure 19. Light curves for examples of the newly discovered long-period variables. The format is the same as for Fig. 14.

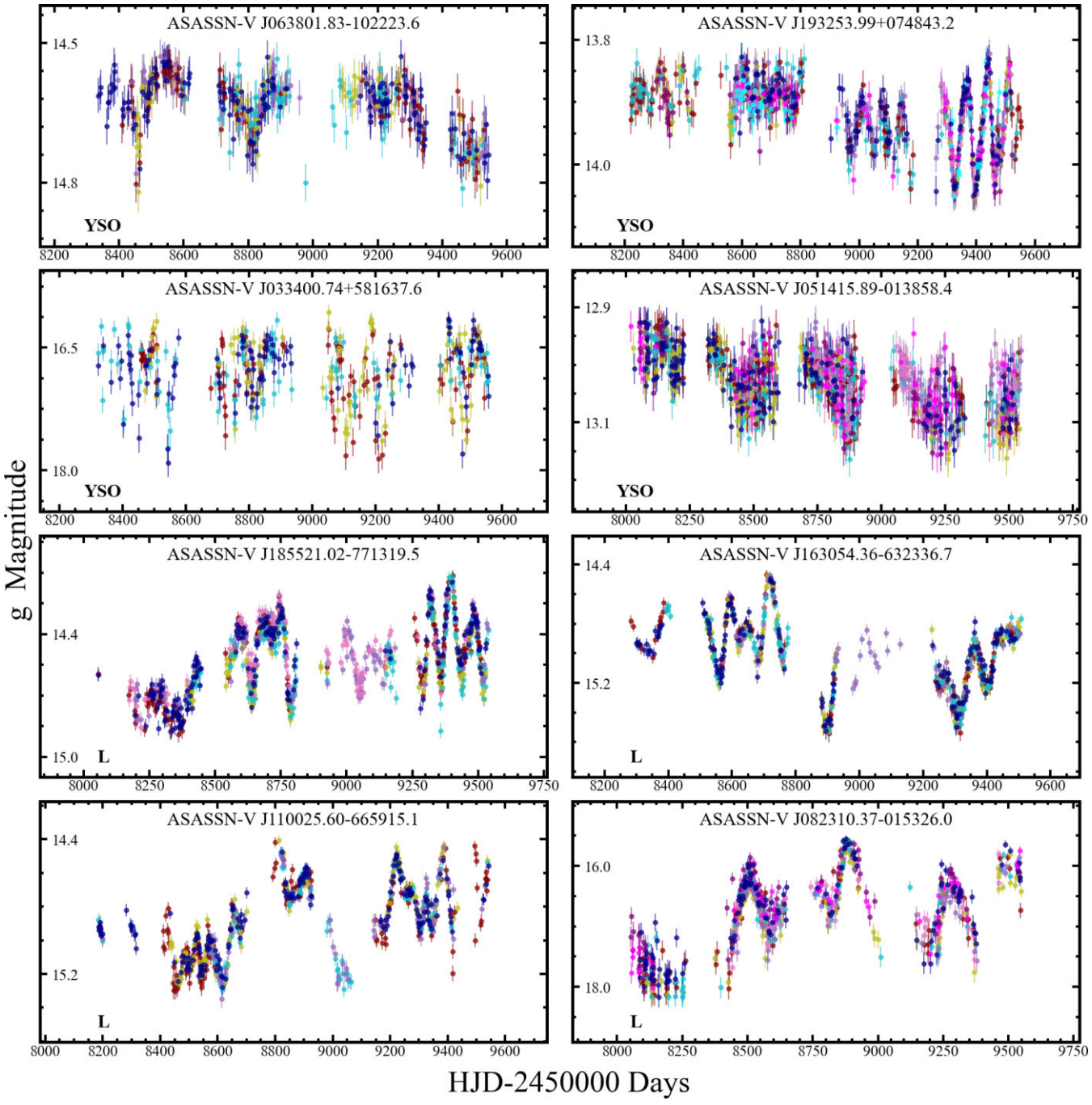

Figure 20. Light curves for examples of the newly discovered irregular variables. The format is the same as for Fig. 14.

ACKNOWLEDGEMENTS

We thank the Las Cumbres Observatory and their staff for its continuing support of the ASAS-SN project. We also thank the Ohio State University College of Arts and Sciences Technology Services for helping us set up and maintain the ASAS-SN variable stars database.

ASAS-SN is funded in part by the Gordon and Betty Moore Foundation through grants number GBMF5490 and GBMF10501 to the Ohio State University, and also funded in part by the Alfred P. Sloan Foundation grant number G-2021-14192. KZS and CSK are supported by NSF grants number AST-1814440 and AST-1908570. Support for TJ was provided by NASA through the NASA Hubble Fellowship grant number HF2-51509 awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract number NAS5-26555. BJS is supported by NASA grant number 80NSSC19K1717 and NSF grants number AST-1920392 and AST-1911074. Support for TWSH were provided by NASA through the NASA Hubble Fellowship grant number HST-HF2-51458.001-A awarded by the Space Telescope Science Institute (STScI), which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract number NAS5-26555. SD acknowledges Project number 12133005 sup-

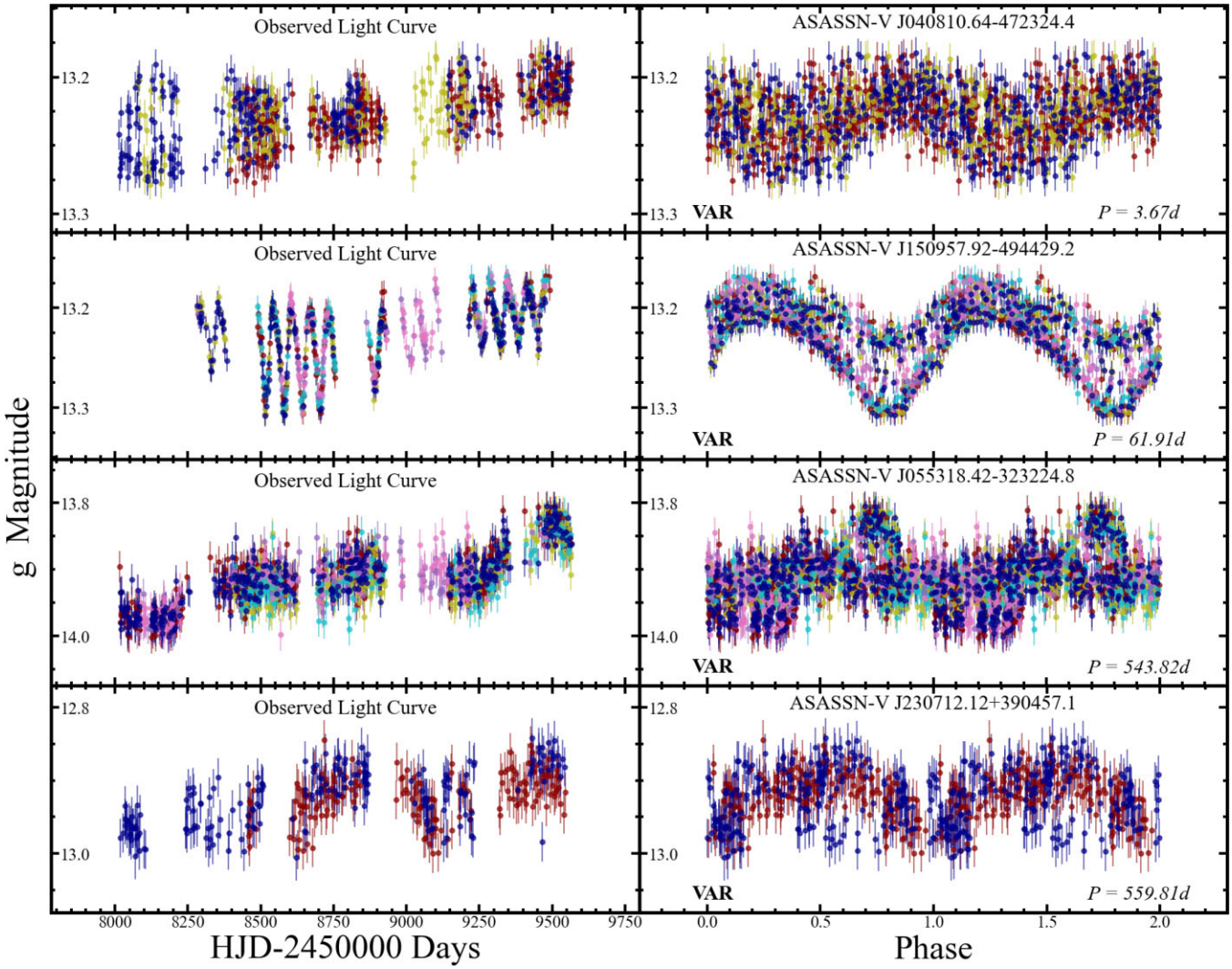

Figure 21. Light curves for examples of the newly discovered generic VAR class variables. The format is the same as for Fig. 14.

ported by National Natural Science Foundation of China (NSFC).

Development of ASAS-SN has been supported by NSF grant number AST-0908816, the Mt. Cuba Astronomical Foundation, the Centre for Cosmology and AstroParticle Physics at the Ohio State University, the Chinese Academy of Sciences South America Centre for Astronomy (CAS- SACA), the Villum Foundation, and George Skestos. TAT is supported in part by Scialog Scholar grant number 24216 from the Research Corporation. Support for JLP is provided in part by FONDECYT through the grant number 1151445 and by the Ministry of Economy, Development, and Tourism's Millennium Science Initiative through grant number IC120009, awarded to The Millennium Institute of Astrophysics, MAS.

This work has made use of data from the European Space Agency (ESA) mission *Gaia* (https://www.cosmos.esa.int/gaia), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement.

This research has made use of the VizieR catalogue access tool, CDS, Strasbourg, France. The original description of the VizieR service was published in A&AS 143, 23.

This research made use of Astropy, a community-developed core Python package for Astronomy (Astropy Collaboration 2013).

DATA AVAILABILITY

The variables are publicly catalogued with the AAVSO and the ASAS-SN light curves can be obtained using the ASAS-SN Sky Patrol (https://asas-sn.osu.edu). The catalogue of variables and the associated light curves are available on the ASAS-SN variable stars database (https://asas-sn.osu.edu/variables). The external photometric data underlying this article were accessed from sources in the public domain: *Gaia* (https://www.cosmos.esa.int/gaia), 2MASS (https://old.ipac.caltech.edu/2mass/overview/access.html), AllWISE (http://wise2.ipac.caltech.edu/docs/release/allwise/), and GALEX (https://archive.stsci.edu/missions-and-data/galex-1/).

REFERENCES

Alard C., 2000, A&AS, 144, 363
Alard C., Lupton R. H., 1998, ApJ, 503, L325
Alcock C. et al., 2000, ApJ, 542, L281
Andrew S., Swihart S. J., Strader J., 2021, ApJ, 908, L180

Astropy Collaboration, 2013, A&A, 558, A33
Bailer-Jones C. A. L., Rybizki J., Fouesneau M., Demleitner M., Andrae R., 2021, AJ, 161, 147
Bellm E. C., 2014, preprint (arXiv:1410.8185)
Bredall J. W. et al., 2020, MNRAS, 496, 3257
Brown A. G. A. et al., 2018, A&A, 616, 1
Brown T. M. et al., 2013, PASP, 125, 1031
Chen X., Wang S., Deng L., de Grijs R., Yang M., 2018, ApJS, 237, 28
Chen X., Wang S., Deng L., de Grijs R., Yang M., Tian H., 2020, ApJS, 249, 18
Christy C. T. et al., 2021, Research Notes of the American Astronomical Society, 5, 38
Christy C. T. et al., 2022, PASP, 134, 024201
Derue F. et al., 2002, A&A, 389, 149
Drake A. J. et al., 2009, ApJ, 696, L870
Feast M., Whitelock P. A., 2014, Proc. IAU Symp. 9, Variable Stars and Galactic Structure. Cambridge Univ. Press, Cambridge, p. 40
Hasanzadeh A., Safari H., Ghasemi H., 2021, MNRAS, 505, 1476
Heinze A. N. et al., 2018, AJ, 156, 241
Holoien T.-S. et al., 2016, MNRAS, 464, 2672
Holtzman J. A. et al., 2015, AJ, 150, 148
Jayasinghe T. et al., 2018, MNRAS, 477, 3145
Jayasinghe T. et al., 2019a, MNRAS, 485, 961
Jayasinghe T. et al., 2019b, MNRAS, 486, 1907
Jayasinghe T. et al., 2019c, MNRAS, 491, 13
Jayasinghe T. et al., 2020b, MNRAS, 493, 4045
Jayasinghe T. et al., 2020c, MNRAS, 493, 4186
Jayasinghe T. et al., 2021, MNRAS, 503, 200
Kochanek C. S. et al., 2017, PASP, 129, 104502
Kovács G., Zucker S., Mazeh T., 2002, A&A, 391, 369
Kozlowski S. et al., 2013, AcA, 63, 1
Leavitt H. S., 1908, Annals of Harvard College Observatory, 60, 87
Lebzelter T., Mowlavi N., Marigo P., Pastorelli G., Trabucchi M., Wood P. R., Lecoeur-Taïbi I., 2018, A&A, 616, 13
Madore B. F., 1982, ApJ, 253, L575
Mateu C., Vivas A. K., 2018, MNRAS, 479, 211
Pawlak M. et al., 2019, MNRAS, 487, 5932
Paxton B. et al., 2018, ApJS, 234, 34
Pedregosa F. et al., 2012, preprint (arXiv:1201.0490)
Percy J. R., 2007, Understanding Variable Stars. Cambridge Univ. Press, Cambridge
Pietrukowicz P. et al., 2020, AcA, 70, 241
Pojmanski G., 2002, AcA, 52, 397
Poleski R., Soszynski I., Udalski A., Szymanski M. K., Kubiak M., Pietrzynski G., Wyrzykowski L., Ulaczyk K., 2012, AcA, 62, 1
Prusti T. et al., 2016, A&A, 595, 1
Ricker G. R. et al., 2015, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 014003
Scargle J. D., 1982, ApJ, 263, L835
Seabroke G. M. et al., 2021, A&A, 653, 160
Shappee B. J. et al., 2014, ApJ, 788, L48
Soszyński I. et al., 2007, AcA, 57, 201
Soszyński I. et al., 2014, AcA, 64, 177
Soszyński I. et al., 2015, AcA, 65, 297
Soszyński I. et al., 2016, AcA, 66, 405
Soszyński I. et al., 2021, AcA, 71, 189
Thompson T. A. et al., 2019, Science, 366, 637
Tonry J. L. et al., 2018a, PASP, 130, 064505
Tonry J. L. et al., 2018b, ApJ, 867, L105
Torres G., Andersen J., Giménez A., 2009, Astron. Astrophys. Rev., 18, 67
Udalski A., 2004, AcA, 53, 291
Udalski A., Szymański M. K., Szymański G., 2015, AcA, 65, 1
Udalski A. et al., 2018, AcA, 68, 315
Watson C. L., Henden A. A., Price A., 2006, Society for Astronomical Sciences Annual Symposium, 25, 47
Wozniak P. R. et al., 2004, AJ, 127, 2436
Zechmeister M., Kürster M., 2009, A&A, 496, 577

This paper has been typeset from a TEX/LATEX file prepared by the author.