

Persistent and occasional: Searching for the variable population of the ZTF/4MOST sky using ZTF Data Release 11^{★,★★}

P. Sánchez-Sáez^{1,2}, J. Arredondo³, A. Bayo^{1,4}, P. Arévalo^{4,5}, F. E. Bauer^{6,7,2,8},
G. Cabrera-Vives^{2,9,10}, M. Catelan^{6,2,7}, P. Coppi¹¹, P. A. Estévez^{2,12}, F. Förster^{13,2,14,15},
L. Hernández-García^{2,4}, P. Huijse^{15,2}, R. Kurtev^{4,2}, P. Lira^{16,5},
A. M. Muñoz Arancibia^{2,14}, and G. Pignata^{17,2}

¹ European Southern Observatory, Karl-Schwarzschild-Strasse 2, 85748 Garching bei München, Germany
e-mail: pasanchezsaez@gmail.com; paula.sanchezsaez@eso.org

² Millennium Institute of Astrophysics (MAS), Nuncio Monseñor Sotero Sanz 100, Of. 104, Providencia, Santiago, Chile

³ Departamento de Ingeniería Informática, Universidad de Santiago de Chile, Av. Ecuador 3659, Santiago, Chile

⁴ Instituto de Física y Astronomía, Facultad de Ciencias, Universidad de Valparaíso, Gran Bretaña No. 1111, Playa Ancha, Valparaíso, Chile

⁵ Millennium Nucleus on Transversal Research and Technology to Explore Supermassive Black Holes (TITANS), 4030000 Concepción, Chile

⁶ Instituto de Astrofísica, Facultad de Física, Pontificia Universidad Católica de Chile, Casilla 306, Santiago 22, Chile

⁷ Centro de Astroingeniería, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, 7820436 Macul, Santiago, Chile

⁸ Space Science Institute, 4750 Walnut Street, Suite 205, Boulder, Colorado 80301, USA

⁹ Department of Computer Science, Universidad de Concepción, Concepción, Chile

¹⁰ Data Science Unit, Universidad de Concepción, Edmundo Larenas 310, Concepción, Chile

¹¹ Department of Astronomy, Yale University, PO Box 208101, New Haven, CT 06520-8101, USA

¹² Department of Electrical Engineering, Universidad de Chile, Av. Tupper 2007, Santiago 8320000, Chile

¹³ Data and Artificial Intelligence Initiative (ID&IA), University of Chile, Santiago, Chile

¹⁴ Center for Mathematical Modeling (CMM), Universidad de Chile, Beauchef 851, Santiago 8320000, Chile

¹⁵ Instituto de Informática, Facultad de Ciencias de la Ingeniería, Universidad Austral de Chile, General Lagos 2086, Valdivia, Chile

¹⁶ Departamento de Astronomía, Universidad de Chile, Casilla 36D, Santiago, Chile

¹⁷ Departamento de Ciencias Físicas, Universidad Andres Bello, Avda. Republica 252, Santiago, Chile

Received 3 February 2023 / Accepted 9 April 2023

ABSTRACT

Aims. We present a variability-, color-, and morphology-based classifier designed to identify multiple classes of transients and persistently variable and non-variable sources from the Zwicky Transient Facility (ZTF) Data Release 11 (DR11) light curves of extended and point sources. The main motivation to develop this model was to identify active galactic nuclei (AGN) at different redshift ranges to be observed by the 4MOST Chilean AGN/Galaxy Evolution Survey (ChANGES). That being said, it also serves as a more general time-domain astronomy study.

Methods. The model uses nine colors computed from CatWISE and Pan-STARRS1 (PS1), a morphology score from PS1, and 61 single-band variability features computed from the ZTF DR11 g and r light curves. We trained two versions of the model, one for each ZTF band, since ZTF DR11 treats the light curves observed in a particular combination of field, filter, and charge-coupled device (CCD) quadrant independently. We used a hierarchical local classifier per parent node approach, where each node is composed of a balanced random forest model. We adopted a taxonomy with 17 classes: non-variable stars, non-variable galaxies, three transients (SNIa, SN-other, and CV/Nova), five classes of stochastic variables (lowz-AGN, midz-AGN, highz-AGN, Blazar, and YSO), and seven classes of periodic variables (LPV, EA, EB/EW, DSCT, RRL, CEP, and Periodic-other).

Results. The macro-averaged precision, recall, and F1-score are 0.61, 0.75, and 0.62 for the g -band model, and 0.60, 0.74, and 0.61, for the r -band model. When grouping the four AGN classes (lowz-AGN, midz-AGN, highz-AGN, and Blazar) into one single class, its precision, recall, and F1-score are 1.00, 0.95, and 0.97, respectively, for both the g and r bands. This demonstrates the good performance of the model in classifying AGN candidates. We applied the model to all the sources in the ZTF/4MOST overlapping sky ($-28 \leq \text{Dec} \leq 8.5$), avoiding ZTF fields that cover the Galactic bulge ($|gal_b| \leq 9$ and $gal_l \leq 50$). This area includes 86 576 577 light curves in the g band and 140 409 824 in the r band with 20 or more observations and with an average magnitude in the corresponding band lower than 20.5. Only 0.73% of the g -band light curves and 2.62% of the r -band light curves were classified as stochastic, periodic, or transient with high probability ($P_{\text{init}} \geq 0.9$). Even though the metrics obtained for the two models are similar, we find that, in general, more reliable results are obtained when using the g -band model. With it, we identified 384 242 AGN candidates (including low-, mid-, and high-redshift AGN and Blazars), 287 156 of which have $P_{\text{init}} \geq 0.9$.

Key words. galaxies: active – stars: variables: general – supernovae: general – surveys – methods: statistical – methods: data analysis

* Tables containing the classifications and features for the ZTF g and r bands, the labeled set, and the master catalog used to create the labeled set are available at the CDS via anonymous ftp to cdsarc.cds.unistra.fr (130.79.128.5) or via <https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/675/A195>.

** Individual catalogs per class and band, as well as the labeled set catalogs, can be downloaded from Zenodo via [10.5281/zenodo.7826045](https://doi.org/10.5281/zenodo.7826045)

1. Introduction

We are approaching an era in astronomy in which data sets are too large to be processed efficiently with traditional tools and methods. In the last few decades, a majority of astronomers have used tools that require only the computing power of a simple laptop to obtain insights from a given data set. However, with the advent of surveys such as the Zwicky Transient Facility (ZTF; Bellm et al. 2019) and the upcoming *Vera Rubin* Observatory Legacy Survey of Space and Time (LSST; Ivezić et al. 2019), we need to develop new tools to extract the most out of the data and manipulate the huge samples observed by these surveys. Recently, several community brokers have begun to process the ZTF public alert stream, including the Automatic Learning for the Rapid Classification of Events (ALeRCE; Förster et al. 2021), Alert Management, Photometry and Evaluation of Lightcurves (AMPEL; Nordin et al. 2019), Arizona-NOAO Temporal Analysis and Response to Events System (ANTARES; Narayan et al. 2018), Fink (Möller et al. 2021), and Lasair (Smith et al. 2019). These brokers make use of different statistical and machine learning techniques to process the large amounts of data generated by the ZTF survey, which amounts to over 300 000 alerts per night¹.

The use of alert streams is optimal for studies of transient objects, such as supernovae (SNe), tidal disruption events, or kilonovae, as demonstrated by several works (e.g., van Velzen et al. 2021; Sánchez-Sáez et al. 2021b; Carrasco-Davis et al. 2021; Leoni et al. 2022; Miranda et al. 2022). However, no information is provided by this stream when sources do not vary with amplitudes large enough to produce an alert, or it can suffer from strong miscalibrations due to changes in the reference images used to generate the difference images. On the other hand, the study of persistent and/or low-amplitude variable objects can benefit from a denser sampling of their light curves, where all epochs are considered without an absolute threshold in their relative variations. This is the case of the ZTF data release (ZTF DR) light curves, which are constructed from the point spread function (PSF) photometry over all the ZTF science images, for objects detected in the ZTF reference images (for details, see Masci et al. 2019). For extended sources, ZTF DR light curves are sensitive to seeing variations and can show spurious variations; thus, they are not optimal for variability studies. Active galactic nuclei (AGN), young stellar objects (YSOs), and binary stars are among the persistent variable objects whose analysis benefits the most from the use of DR light curves, as their alert light curves can be very incomplete and biased. An example of this is the work presented by Sánchez-Sáez et al. (2021a), where ZTF DR5 light curves were used to identify changing-state active galactic nucleus (CSAGN; sources that change their classification as type 1 or type 2 AGN) candidates. Sánchez-Sáez et al. (2021a) find that several of their promising CSAGN candidates present only a few or even no alerts.

Chen et al. (2020) present one of the first attempts to classify ZTF DR light curves. They used ZTF DR2 light curves to classify 781 602 periodic variable stars into 11 classes. They selected variable stars from the DR2 light curves using the period (computed using the Lomb-Scargle periodogram; Lomb 1976; Scargle 1982) and the variability amplitude, as well as other statistics, and applied feature cuts to classify their candidates, with a misclassification rate of 2%. van Roestel et al. (2021) present the

ZTF project light curve classifier, which was designed to classify all persistent point sources in the ZTF DR time series. They present two models, one that uses variability features and a gradient-boosted decision tree classifier and another that uses deep neural networks applied to magnitude-time histograms (*dmdt*; Mahabal et al. 2017), as well as variability features. To test their selection technique, they applied their models to ten pairs of ZTF fields, obtaining precision and recall scores higher than 0.8 in most of the classes included in their taxonomy. More recently, Aleo et al. (2022) used ZTF DR4 light curves to search for missed transients in the ZTF alert stream, using variability features computed with the *light-curve* package (Malanchev et al. 2021) and applying a *k*-dimensional tree algorithm (Bentley 1975). From this model, they were able to identify 11 missing transients.

In this work we present a ZTF DR light curve classifier that has been designed to work with the PSF light curves of both extended and point sources, constructed from their ZTF science images. As previously mentioned, these light curves are not optimal for variability studies of extended sources, but in this work we demonstrate that they can still be used to separate non-variable and variable extended sources, as well as to identify transients whose host is detected directly in the science images. The model separates sources into 17 different classes of persistently variable objects, transient events, and non-variable galaxies and stars. The main purpose of this classifier is to identify AGN candidates at different redshifts to be observed by the 4-m Multi-Object Spectroscopic Telescope (4MOST; de Jong et al. 2019) as part of the 4MOST Chilean AGN/Galaxy Evolution Survey (ChANGES; F. Bauer et al., in prep.). To this end, we focus on the classification of objects located in the region of overlap between ZTF and 4MOST skies (ZTF/4MOST sky), excluding the Milky Way bulge. The total size of the ZTF DR11 light curves in this area amounts to 139.5 GB in the *g* band and 484.5 GB in the *r* band, including 86 576 577 and 140 409 824 light curves with 20 or more observations and an average magnitude lower than 20.5, respectively.

The classifier presented here uses 61 variability features computed using the ZTF DR11 (Masci et al. 2019; Bellm et al. 2019) light curves, nine colors (when available) obtained from Pan-STARRS1 (PS1; Chambers et al. 2016) and CatWISE (Eisenhardt et al. 2020; Marocco et al. 2021), and one morphology score from PS1 (Tachibana & Miller 2018). The classifier is inspired by the ALeRCE broker alert light curve classifier from Sánchez-Sáez et al. (2021b, hereafter *SS21b*). However, several modifications have been made to this original model, as we are dealing now with ZTF DR light curves: due to the large number of sources considered in this work, and the presence of non-variable classes, we changed the taxonomy tree (see Fig. 1) and used a fully hierarchical approach for the classifier; we excluded the non-detection features presented in *SS21b*; each ZTF light curve with a unique DR ID is treated independently (implicitly meaning there are no multiband features); and, in order to have an estimate of the redshift of our AGN candidates, we separated the AGN classes into three broad redshift bins.

The paper is organized as follows. In Sect. 2 we describe the data used for this work. In Sect. 3 we describe the taxonomy considered by our model and the construction of the labeled set (LS). In Sect. 4 we define the set of features used by the model. In Sect. 5 we explain the machine learning classifier used in this work and evaluate its performance. In Sect. 6 we show the results obtained when the model is applied to the ZTF light curves from the ZTF/4MOST sky. In Sect. 7 we discuss our findings and

¹ ZTF produces an alert when there is a 5σ detection in the difference image of a particular object. This detection can be produced by the flux variations of persistently variable and transient objects, as well as by moving targets, or bogus detections.

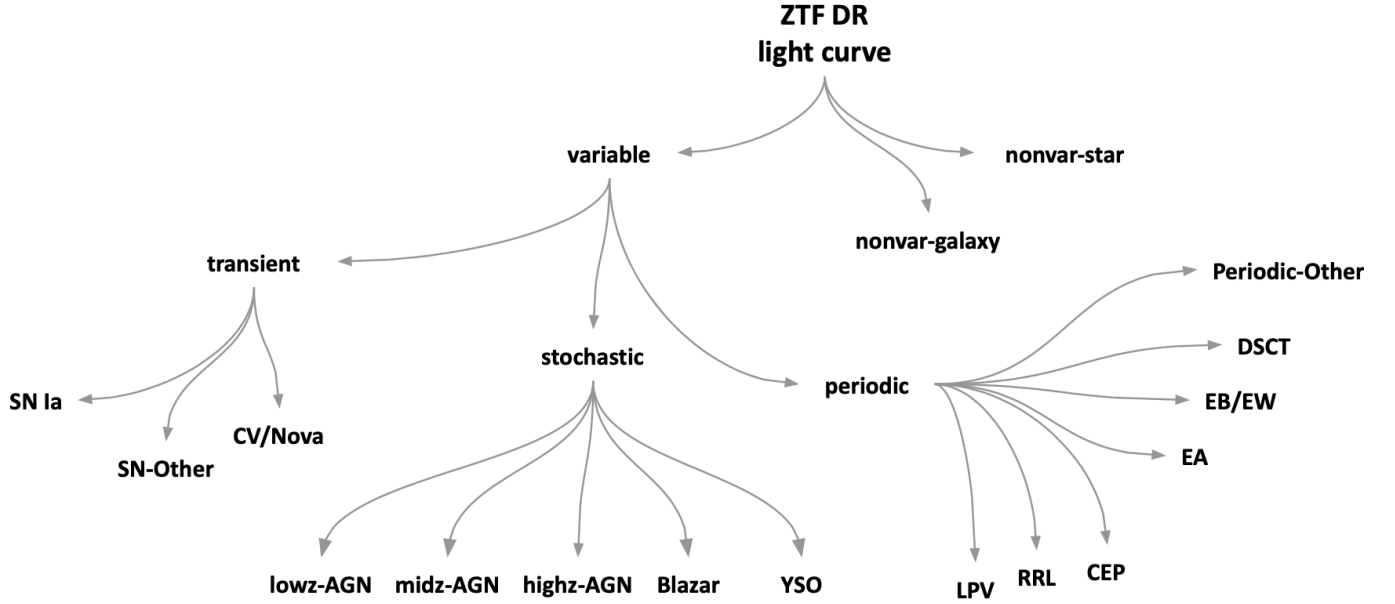


Fig. 1. Hierarchical taxonomy used in this work. See Sect. 3.1 for details

compare our results with previous works. Finally, in Sect. 8 we conclude and summarize the paper.

The code used to process ZTF DR11 light curves, compute features, and cross-match ZTF DR11 objects with CatWISE and PS1 catalogs is publicly available², as is the code used to train and evaluate this classifier and to match external catalogs with our ZTF/4MOST classification catalog³. Catalogs that contain the LS for each band, the original master catalog used to cross-match with the ZTF objects, and the features and classifications for the ZTF/4MOST sky sources are available in electronic form at the CDS and Zenodo.

2. Data

2.1. ZTF DR11

In this work we used data from ZTF DR11, which includes observations taken from March 2018 to March 2022. We used the bulk download option described in Sect. 12.c of the ZTF DR11 documentation⁴ to recover the light curves of the full ZTF sky. Each ZTF field spans $\approx 7^\circ \times 7^\circ$, and each file corresponds to a field/chip/quadrant/filter combination. The files include the PSF fit-based optical light curves in the g , r , and i bands. The light curves contain the heliocentric-based modified Julian date (MJD), which corresponds to the middle observing date of each exposure (with integration time ≥ 30 s), the PSF magnitude in a given band (calibrated for a source with color $g-r=0$ in the AB photometric system), the magnitude error, a linear color coefficient (not used in this work as it requires previous knowledge of the simultaneous $g-r$ color of each source) and a quality score (catflags).

ZTF DR11 provides unique object IDs for unique combinations of RA, Dec, filter, field, charge-coupled device (CCD), and CCD quadrant. Therefore, a single astronomical source may

be associated with more than one light curve per band. This makes the position-matching of light curves from different bands or observed with different CCDs computationally expensive, especially when dealing with several million real astronomical sources. Hence, we decided to treat each light curve independently. This implies that some astronomical objects will have several object IDs associated with them.

Since the aim of this work is to implement a variability-based classifier, we only considered light curves with 20 or more observations ($N_{\text{obs}} \geq 20$). This minimum number of observations is more conservative than the one used by SS21b and is selected to ensure the proper identification of intrinsic variations and proper computation of variability features from the DR light curves, which are noisier and prone to show spurious variations compared to the ZTF alert light curves. We also decided to exclude the i band from the analysis, as it only covers two years of data, with fewer epochs, in the public ZTF DR11. There are 704 388 801 unique object IDs in the g band and 1 251 889 824 in the r band with $N_{\text{obs}} \geq 20$. We call this set of sources with $N_{\text{obs}} \geq 20$ in at least one band the “long-lc-ZTF-DR11” sample. We note that we avoided computing internal matches per band for this sample, as it is computationally expensive (although this was done for small samples, as described in Sects. 3.2.4 and 7.2). Therefore, from this work, a real astronomical object can have more than one set of features and classifications in a given band.

2.2. Additional photometric data

As mentioned in Sect. 2.1, sources with unique IDs are treated independently, and thus we do not compute colors from the ZTF DR11 light curves. However, SS21b demonstrated the importance of including average optical colors, as well as near-infrared colors, as features for the classification of variable and transient objects. Therefore, in this work, we also included (when available) optical photometry in the g , r , and i bands from PS1, and W1 and W2 photometry from CatWISE. To associate a ZTF DR11 object ID with a PS1 or CatWISE source, we used a radius of $1.5''$ (which is the default matching radius used by ZTF) and $2''$ (which is the standard matching radius adopted by

² https://github.com/alercebroker/ztf_dr

³ <https://github.com/PaulaSanchezSaez/ZTF-DR11-classifier>

⁴ https://irsa.ipac.caltech.edu/data/ZTF/docs/releases/dr11/ztf_release_notes_dr11.pdf

ALeRCE for CatWISE data, considering the larger pixel scale of the CatWISE images compared to ZTF), respectively.

3. Classification taxonomy and labeled set

3.1. Classification taxonomy

In this work, we use a classification taxonomy inspired by the one presented in [SS21b](#); although we decided to change some of the classes presented there and add new ones. In particular, we are working with DR light curves, which include every observation performed on a source, therefore significantly differing from the alert light curves. Thus, it becomes necessary to modify the taxonomy presented in [SS21b](#) and for instance to include non-variable sources. In addition, the light curves of transient objects in the DRs look very different from the alert ones; thus, we decided to modify the taxonomy tree presented in [SS21b](#), and use one that is more suitable for the DR light curves. For instance, the class CV/Nova is included in the transient subclass (in [SS21b](#) it was in the stochastic subclass), since in DR light curves they share more similarities with SNe light curves. In total, we consider 17 classes, including persistent variable objects, transients, and non-variable classes. As in [SS21b](#), we divide the taxonomy in a hierarchical fashion; however, we include an additional hierarchy level in the taxonomy tree, as shown in Fig. 1. The division is done as follows:

1. Non-variable star: sources classified as stars that do not show variable behavior according to external catalogs (nonvar-star).

2. Non-variable galaxy: sources classified as passive or star-forming galaxies, or obscured (type 2) AGN (nonvar-galaxy).

3. Variable object: sources classified as persistent variable objects or transients (variable). This class is subdivided into the following:

- (a) Transient: sources that are not persistently variable. We include here the classes Type Ia supernova (SNIa); all the other SN types (SN-other), including Ibc SN, type II SN, or super luminous SN; and cataclysmic variable or nova (CV/Nova).

- (b) Stochastic: sources that show persistent stochastic variability. We include here AGN with $z \leq 0.5$ (lowz-AGN); AGN with $0.5 < z \leq 3$ (midz-AGN); AGN with $z > 3$ (highz-AGN); beamed jet-dominated AGN (Blazar); and YSO (class dominated by sources at evolutionary stages where the stochastic variations should dominate over the periodic ones).

- (c) Periodic: stars with periodic variable signal. We consider here long-period variable (LPV; including regular, semi-regular, and irregular variable stars); RR Lyrae (RRL); Cepheid (CEP); detached eclipsing binary (EA); semidetached and contact variable (EB/EW); δ Scuti (DSCT); and other periodic variable stars (Periodic-other; this includes variable stars classified as miscellaneous, rotational, or RS Canum Venaticorum-type systems).

3.2. Labeled set construction

We created a set of sources for training and testing (LS) by cross-matching the sources of the long-lc-ZTF-DR11 sample with different catalogs of objects with known labels, obtained from multiple photometric and spectroscopic surveys. For this, we followed a priority strategy similar to that presented in [Förster et al. \(2021\)](#). However, since in this case we are dealing with several million objects, we decided to create first a master catalog of known objects, which was then cross-matched

with the long-lc-ZTF-DR11 sample. In the following sections we describe how this master catalog was created, and how we dealt with the cross-match of catalogs with a large number of sources.

3.2.1. Catalogs of non-variable stars

Catalogs of non-variable stars can be tricky to create, as a star can be constant in a given survey and variable in another one with better photometric precision. Thus, in order to generate a catalog of constant stars in ZTF, we decided to use surveys with photometric precision equal to or better than the ZTF precision.

To create a master catalog of non-variable stars, we used two different surveys, namely the Sloan Digital Sky Survey (SDSS; [York et al. 2000](#)) and the *Gaia* early Data Release 3 (*Gaia* eDR3; [Gaia Collaboration 2016, 2021](#)). We first used the latest release of the SDSS Stripe 82 Standard Star Catalog⁵ (v4.2; [Ivezić et al. 2007](#); [Thanjavur et al. 2021](#)), which contains 999 472 stars with at least 4 observations in each *gri*, and that are considered non-variable from their measured χ^2 per filter. In order to reduce the contamination by variable stars in the sample, we only kept stars with 10 or more total observations (N_{tot} in the original catalog) and with a mean magnitude per band brighter than 21 magnitudes, with a root-mean-square scatter (rms) lower than 0.16 (to reduce any possible contamination from variable stars), without classification of variable star in the Set of Identifications, Measurements and Bibliography for Astronomical Data (SIMBAD) database ([Wenger et al. 2000](#)), and that are not classified as variable star in the LS used by [SS21b](#). After all these filters, we ended up with a sample of 407,926 stars (SDSS_calibstar catalog).

In addition, to include a sample of non-variable stars located at different positions in the sky, we used *Gaia* eDR3, following the strategy presented in [Mowlavi et al. \(2021\)](#). We used the following advanced query (ADQL) in the *Gaia* eDR3 database:

```
SELECT *
FROM gaiaedr3.gaia_source
WHERE dec>Y AND dec<X
AND abs(pmra)<1 AND abs(pmdec)<1
AND phot_g_mean_mag>14 AND phot_g_mean_mag<20
AND phot_g_n_obs>20
```

Here, X and Y correspond to the minimum and maximum DEC for the query. We searched for sources in the ZTF sky using five different ranges of declination, querying 3 million objects per range, ending up with 15 million objects. To clean the sample and ensure that extended sources such as galaxies, unresolved AGNs, or sources with poor photometry are not included in the final set, we used the corrected *Gaia* blue (BP) and red (RP) flux excess factor C^* (see Eq. (6) and Table 2 in [Riello et al. 2021](#)). [Riello et al. \(2021\)](#) show in Fig. 21 that galaxies and unresolved AGNs tend to present values of C^* larger than 2; thus, we decided to only include sources with $-0.15 < C^* \leq 1$, a range that is expected to be dominated by stars. We also decided to remove from the sample all the sources with fewer than 60 observations in both BP and RP, to ensure that the final sample is as pure as possible. This filtering left us with 718,983 sources.

Then, following [Mowlavi et al. \(2021\)](#), we measured a proxy of the scatter in each *Gaia* eDR3 band by using their Eq. (2): $A_{\text{proxy}} = \sqrt{N} \frac{\epsilon(I)}{I}$, where N corresponds to the number of observations in the band, I to the weighted mean flux in the band, and $\epsilon(I)$ to the error in the mean flux. We used A_{proxy} to remove

⁵ <https://faculty.washington.edu/ivezic/sdss/catalogs/stripe82.html>

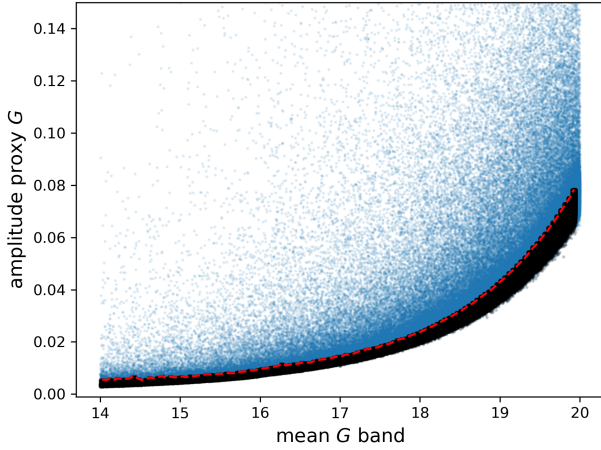


Fig. 2. Amplitude proxy in the *Gaia* *g*-band magnitude versus the mean *g*-band magnitude for a sample of potential non-variable stars. In blue we show the sources that are removed from the sample, and in black the sources included in the final sample of non-variable stars. The dashed red line shows the mean A_{proxy} per bin of magnitude in the *g* band (A_{mean}).

potential variable sources by measuring the mean A_{proxy} per band and per magnitude bin of size 0.05 mag (A_{mean}), and then removing the sources from the sample with $A_{\text{proxy}} < A_{\text{mean}}$. Figure 2 shows an example of this procedure for the *g* band. The final sample of *Gaia* eDR3 non-variable stars corresponds to the sources that satisfy this requirement in the three bands. After this, we ended up with a sample of 43 584 non-variable stars (*Gaia*_nonvarstar catalog). None of these sources is located in the Stripe 82 area.

3.2.2. Catalogs of non-variable galaxies

Galaxies without AGN activity, as well as obscured or type 2 AGN, are not expected to show persistent optical variability. Thus, to create a sample of non-variable galaxies, we used catalogs of known galaxies and known type 2 AGN.

Optical emission lines have been used for years to separate star-forming galaxies from objects with AGN activity. In particular, the Baldwin, Phillips, and Terlevich (BPT) diagram, which consists of indexes that relate the strengths of several emission lines ($H\beta$, $[\text{O III}]\lambda 5007$, $H\alpha$, and $[\text{N II}]\lambda 6583$) as proxies for the state of ionization of a region (Baldwin et al. 1981), has been widely used for this purpose. We used the Portsmouth emission line flux measurements (Thomas et al. 2013) to select a sample of 238 773 sources classified as star-forming galaxies from their location on the BPT diagram. Then, we removed from the sample all the sources with public alerts reported by ZTF (including data obtained until November 2021), as some of these galaxies could be hosts of SNe events, or have a weak AGN in their nuclei. After this, we ended up with 221 805 star-forming galaxies (SDSS_galaxy catalog).

Furthermore, we use the Million Quasars Catalog (MILLIQUAS Catalog v7.4c; Flesch 2019) to select a sample of type 2 AGN. We selected 42 639 sources classified as type 2 narrow-line core-dominated (K) or as type 2 Seyferts/host-dominated (N) in MILLIQUAS v7.4c.

3.2.3. Catalogs of variable and transient sources

To create the sample of variable and transient sources, we used the same catalogs described in Förster et al. (2021) and SS21b:

the ASAS-SN catalog of variable stars (ASASSN; Jayasinghe et al. 2018, 2019b,a, 2020), the Catalina Surveys Variable Star Catalogs (CRTS; Drake et al. 2014, 2017), the LINEAR catalog of periodic light curves (LINEAR; Palaversa et al. 2013), the *Gaia* Data Release 2 sample of LPVs and other variable stars (LPV_GAIA and GAIADR2VS, respectively; Mowlavi et al. 2018; Rimoldini et al. 2019), the Transient Name Server (TNS) database (only SN classes were included)⁶, the Roma-BZCAT Multi-Frequency Catalog of Blazars (ROMABZCAT; Massaro et al. 2015), the MILLIQUAS catalog (version 7.4c), the New Catalog of Type 1 AGN (Oh2015; Oh et al. 2015), and the SIMBAD database (SIMBAD_variables; Wenger et al. 2000). Some additional CV labels were obtained from different catalogs (including Ritter & Kolb 2003), compiled by Abril et al. (2020, called JAbri1).

In addition, we included two new catalogs: the new catalog of 78 hydrogen-poor super-luminous SNe observed by ZTF (SLSN_ZTF; Chen et al. 2023), and a set of YSO catalogs curated by the ALerCE collaboration (YSO_ALerCE). The YSO compilation set consisted of several primary sources of targets that were then matched against the ZTF DR. These primary sources include catalogs that are not necessarily based on variability features and cover large portions of the sky, such as Torres et al. (2006), Elliott et al. (2016), Vioque et al. (2020), Kuhn et al. (2021), or Zúñiga-Fernández et al. (2021). But also others fully motivated by variability, such as the American Association of Variable Star Observers (AAVSO) database, or more concrete studies focused on specific star forming regions. Among the later we covered a large range of ages, environments and wavelengths at which variability was confirmed, namely: NGC 1333 (Rebull et al. 2015), GGD12-15 (Wolk et al. 2015), the cluster IRAS 20050+2720 (Poppenhaeger et al. 2015), Lynds 1688 (Günther et al. 2014), the Cep OB3b OB association (Sergison et al. 2020), the Pelican Nebula (Bhardwaj et al. 2019), the Cygnus OB2 association (Roquette et al. 2020), Praesepe (Rebull et al. 2017), Upper Scorpius and ρ Ophiuchus (Ansdell et al. 2018; Rebull et al. 2018), and Taurus (Rodríguez et al. 2017; Rebull et al. 2020).

3.2.4. Master catalog generation and cross-match with ZTF DR11 sources

As mentioned in Sect. 2, ZTF DR11 contains hundreds of millions of light curves per band, making the generation of a label set a computationally expensive task. Thus, to generate an LS, instead of cross-matching each catalog with the ZTF sample and then combining the results, as done in Förster et al. (2021), we first created a master catalog that was then cross-matched against the long-lc-ZTF-DR11 sample. To avoid having repeated objects with different labels in the master catalog, we used a priority order, similar to the one presented in Förster et al. (2021), which includes the new catalogs considered in this work. These priorities were defined considering the methods used to generate each catalog and the level of disagreement in the classifications provided by them.

The master catalog was generated by concatenating the different catalogs presented in Sects. 3.2.1, 3.2.2, and 3.2.3, in the following priority order (from highest to lowest): JAbri1, SLSN_ZTF, RomaBZCAT, Oh2015, MILLIQUAS, TNS, YSO_ALerCE, SIMBAD_variables, CRTSnorth, CRTSsouth, LINEAR, LPV_GAIA, GAIADR2VS, ASASSN, SDSS_calibstar, GAIA_nonvarstar, SDSS_galaxy. There are sources that can

⁶ <https://www.wis-tns.org/>

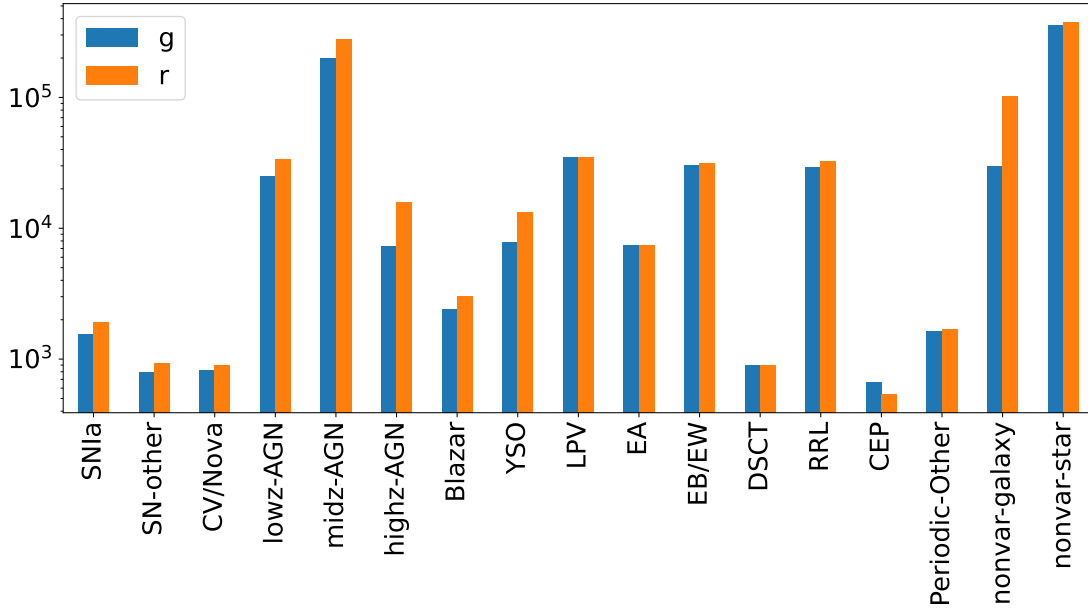


Fig. 3. Number of sources per class in the LS of the g and r bands (note the log scaling on the y -axis).

appear in more than one catalog; therefore, we used the internal match available in TOPCAT (Taylor 2005) to remove all but the first element of a group of sources within a $3''$ radius (and keeping the order of priorities). This master catalog contains 1 907 096 sources in total (1 903 799 when only including the classes considered in this work), and it can be used as an LS in future works. The catalog can be downloaded from CDS or Zenodo.

To generate the LS for ZTF DR11, we cross-matched the master catalog with the long-lc-ZTF-DR11 sample using a radius of $1.5''$. There are more than 700 million sources and more than 1.2 billion sources in the long-lc-ZTF-DR11 sample with g -band light curves and r -band light curves, respectively; therefore, to cross-match the catalogs, we used the ASTROIDE extension (Brahem et al. 2018) for Apache Spark (Zaharia et al. 2016), which is designed to deal with large-volume astronomical data sets and includes a cross-match tool. However, instead of using ASTROIDE with Spark, we used a Python wrapper of ASTROIDE developed by the ALerCE broker, which is publicly available⁷. Only sources with more than 20 observations, after removing epochs with $\text{catflags} > 0$, are considered for the cross-match. Once the variability features were computed (see Sect. 4.1, we also removed from the long-lc-ZTF-DR11 sample all light curves with an average magnitude greater than 20.5; $\bar{m} > 20.5$). After cross-matching, in order to avoid having repeated light curves in a given band for a given target, we kept the longest light curve associated with each source in each bandpass. This internal match was done using TOPCAT, and was possible due to the small size of the LS. From this, we ended up with 741 263 and 936 145 sources in the LS of the g band and the r band, respectively. Figure 3 shows the number of unique astronomical targets per class in each band. Table 1 presents the number of sources from each of the catalogs presented in Sects. 3.2.1–3.2.3, included in the master catalog, and in the LS of the g band and the r band. Catalogs containing the LS for each band can be downloaded from CDS or Zenodo.

Table 1. Number of objects (#) per catalog.

Catalog	# master catalog	# LS g -band	# LS r -band
CVsJavierAbril	1140	494	548
SLSN_ZTF	78	14	22
RomaBZCAT	3443	1803	2173
Oh2015	5533	4046	4610
MILLIQUAS	868 952	247 561	349 640
TNS	13 215	2386	2890
YSO_ALerCE	14 800	6037	9989
Simbad_variables	12 887	1769	3234
CRTSnorth	42 947	32 451	34 801
CRTSsouth	34 066	3316	3420
LINEAR	3044	2116	2438
LPV_GAIA	75 560	20448	20 659
GAIDR2VS	105 294	31 718	33 535
ASASSN	56 006	16 002	14 945
SDSS_calibstar	407 874	32 3488	33 8572
GAIA_nonvarstar	43 583	34 777	36 997
SDSS_galaxy	215 377	12 837	77 672

4. Features

The main purpose of this work is to classify objects in the ZTF/4MOST sky for ChANGES; therefore, we restricted the computation of features to the ZTF fields located in this region. ZTF divides the sky into two grids. In this work, we focus on the primary grid fields (637 in total), selecting the fields centered in the sky region with $-28^\circ \leq \text{Dec.} \leq 8.5^\circ$ (from field 245 to field 498). We also excluded fields in the ZTF/4MOST area with Galactic longitude $0^\circ \leq \text{gal}_l \leq 50^\circ$, and Galactic latitude in the range $-9^\circ \leq \text{gal}_b \leq 9^\circ$, in order to avoid the extremely dense Milky Way bulge, where there is a high probability of having blended sources, and the high-extinction regions of the Milky Way disk. The features were computed for all the light curves from the long-lc-ZTF-DR11 sample located in this sky region.

⁷ https://github.com/alercebroker/minimal_astroide

In contrast, we computed the features of the sources in the LS, regardless of their location in the ZTF sky. For each individual light curve, a total of 61 variability features, nine colors, and one morphology feature are included per source, giving a total of 71 features per target.

We computed features for 86 576 577 and 140 409 824 individual light curves in the g and r bands, respectively, in the ZTF/4MOST area. Due to the large size of these samples, we did not perform internal cross-matches to remove repeated objects; thus, a given astronomical object can have more than one set of features associated per band.

4.1. Variability features

In this work we used several variability features taken from SS21b. In particular, we used the “Detection Features” presented in Table 2 of SS21b, excluding those that are computed using both the g and r bands at the same time since we treat these two bands independently. The features are computed for all the g and r -band light curves, as described in Sect. 4. The following 61 features are computed for each light curve: MHPS_ratio, MHPS_low, MHPS_high, SPM_A, SPM_t0, SPM_gamma, SPM_beta, SPM_tau_rise, SPM_tau_fall, SPM_chi, Amplitude, AndersonDarling, Autocor_length, Beyond1Std, Con, Eta_e, Gskew, MaxSlope, Meanvariance, MedianAbsDev, MedianBRP, PairSlopeTrend, PercentAmplitude, Q31, Rcs, Skew, SmallKurtosis, Std, StetsonK, Pvar, ExcessVar, SF_ML_amplitude, SF_ML_gamma, IAR_phi, LinearTrend, GP_DRW_sigma, GP_DRW_tau, Period, PPE, Power_rate_1/4, Power_rate_1/3, Power_rate_1/2, Power_rate_2, Power_rate_3, Power_rate_4, Psi_CS, Psi_eta, Harmonics_mag_n (with $n = 1..7$), Harmonics_phase_n (with $n = 2..7$), and Harmonics_mse. A detailed description of these features is available in SS21b and the references therein.

All the features were computed as described in SS21b, except for the period – which is computed here using the single-band multi-harmonic analysis of variance (MHAOV) periodogram available in the P4J Python package (Huijse et al. 2018)⁸ – and the features obtained from the supernova parametric model (SPM) – which were originally based on the model presented in Villar et al. (2019). We used a version of the SPM model that can properly deal with DR light curves, which can include a non-variable component associated with the host galaxy. This SPM model adds a constant value (C_{flux}) to the modeled flux:

$$F = C_{\text{flux}} + \frac{A(1 - \beta' \frac{t-t_0}{t_1-t_0})}{1 + \exp(-\frac{t-t_0}{\tau_{\text{rise}}})} \left[1 - \sigma\left(\frac{t-t_1}{3}\right) \right] + \frac{A(1 - \beta') \exp(-\frac{t-t_1}{\tau_{\text{fall}}})}{1 + \exp(-\frac{t-t_0}{\tau_{\text{rise}}})} \left[\sigma\left(\frac{t-t_1}{3}\right) \right]. \quad (1)$$

This C_{flux} parameter, however, is not included in the classifier as a feature, to avoid any bias in the LS magnitude distribution.

The large number of light curves for which features have to be calculated (several tens of millions), and the diversity of target densities in different regions of the ZTF/4MOST sky, make the computation of features a challenging task. For this aim, we used the super-computing infrastructure of the National Laboratory for High Performance Computing (NLHPC) in Chile, which allowed us to compute features using 300 independent cores,

with the “general” partitions⁹, each with an allocated memory of 4.3 Gigabytes.

4.2. Color and morphology features

As mentioned in Sect. 2.2, colors can improve the classification of variable and transient objects. Therefore, we used PS1 and CatWISE photometry to compute the following colors: $g - r$, $r - i$, $g - W1$, $g - W2$, $r - W1$, $r - W2$, $i - W1$, $i - W2$, and $W1 - W2$. We note that these colors are not corrected for Galactic extinction. We also used a morphological star/galaxy score (ps_score) obtained from Tachibana & Miller (2018), which used PS1 data to separate stars from galaxies.

5. Classification algorithm

5.1. The hierarchical balanced random forest

Taking into account the hierarchical structure of the taxonomy presented in Sect. 3.1, we decided to construct a hierarchical classifier, where each ZTF band is treated independently. There are various approaches to performing a hierarchical classification (e.g., Silla & Freitas 2011; Naik & Rangwala 2018); in this work, we used a local classifier per parent node approach, where each node of the hierarchical classifier corresponds to a multi-class classifier. The model follows the hierarchy presented in Fig. 1. In the first level of the model, we classify each source as nonvar-star, nonvar-galaxy, or variable (node_init). All the sources classified as variable go to the second level, where we classify the sources as transient, stochastic, or periodic (node_variable). Finally, in the third level, we further classify the transient sources into SN Ia, SN-other, or CV/Nova (node_transient); the stochastic sources into lowz-AGN, midz-AGN, highz-AGN, Blazar, or YSO (node_stochastic); and the periodic sources into LPV, RRL, CEP, EA, EB/EW, DSCT, or Periodic-other (node_periodic).

Considering the high class imbalance of our LS, as shown in Fig. 3, we decided to use a balanced random forest (BRF¹⁰; Chen et al. 2004) in each node of the hierarchical classifier. Therefore, we call this model the hierarchical balanced random forest (HBRF). In particular, we used the BRF implementation available in the Imbalanced-learn Python package (Lemaître et al. 2017). We trained independent BRF classifiers for each node; thus, the HBRF of each band is composed of five BRF classifiers.

We used 80% of the LS as a training set and the other 20% as a testing set. In order to maintain the original percentage of samples per class, we used the train_test_split method available in scikit-learn (Pedregosa et al. 2011), in a stratified fashion. For instance, for the r band, in the training set there are 1545 SNIa, 223 680 midz-AGN, and 27 772 LPV, and in the testing set there are 386 SNIa, 55 920 midz-AGN, and 6943 LPV.

We optimized the hyperparameters of each BRF classifier independently. We used the K-fold randomized cross-validation procedure available in scikit-learn, with $k = 5$ folds (from the training set) and using the “F1-macro” as target score. We searched for the optimal number of trees in the forest (number of estimators: [100, 200, 300, 400, 500]), the fraction of features to consider in each split (max features: [0.2, 0.4, auto, sqrt]), and the maximum depth of each tree (max depth: [10, 30, 50, 70, 90, None]). Table 2 shows the optimized hyper-parameters

⁹ <https://www.nlhpc.cl/infraestructura/>

¹⁰ For further details about the BRF classifier, see SS21b.

⁸ <https://github.com/phuijse/P4J>

Table 2. Hyper-parameters of the HBRF.

Band	Node	Number of estimators	Max features	Max depth
<i>g</i>	node_init	500	0.4	None
	node_variable	400	0.4	30
	node_transient	300	0.4	10
	node_stochastic	300	0.4	30
	node_periodic	300	0.4	None
<i>r</i>	node_init	300	0.4	70
	node_variable	500	0.4	70
	node_transient	200	auto	10
	node_stochastic	200	0.4	70
	node_periodic	400	0.4	50

used by each node of the HBRF classifier. The rest of the parameters were not optimized; we used *entropy* as the split criterion and *balanced_subsample* for the class_weight parameter; for the rest, we used the default values.

From this model, we classify objects among 17 classes, which correspond to the final leaves of each node, and we provide the probabilities obtained by the corresponding node (P). For instance, for SN Ia, we report the probability according to the HBRF of the node_transient; for YSOs we report the probability of the node_stochastic; for LPVs, we report the probability of the node_periodic; and for nonvar-stars, we report the probability of the node_init. For the sources classified as variable in the node_init, we also report the probabilities and classifications obtained from this node (P_{init}), and the probabilities and classifications of the node_variable.

5.2. Classification performance

To evaluate the performance of the HBRF, we applied the model to the testing set, which corresponds to the 20% of the LS not used during training. In particular, we used the `classification_report` method available in `scikit-learn`, to recover the precision, recall, and F1-score for each of the 17 classes independently, and for the full testing set (macro-averaged scores). These scores are defined as

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i}, \quad (2)$$

$$\text{Recall}_i = \frac{TP_i}{TP_i + FN_i}, \quad (3)$$

$$\text{F1-score}_i = 2 \times \frac{\text{Precision}_i \times \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}, \quad (4)$$

where i corresponds to a particular class, and TP_i , FP_i , FN_i are the corresponding numbers of true positives, false positives, and false negatives, respectively.

From these scores per class, we compute the macro-averaged scores as

$$\text{Precision}_{\text{macro}} = \frac{1}{n_{\text{cl}}} \sum_{i=1}^{n_{\text{cl}}} \text{Precision}_i, \quad (5)$$

$$\text{Recall}_{\text{macro}} = \frac{1}{n_{\text{cl}}} \sum_{i=1}^{n_{\text{cl}}} \text{Recall}_i, \quad (6)$$

Table 3. HBRF classifier metrics per band and class, and macro-averaged scores.

Band	Class	Precision	Recall	F1-score
<i>g</i>	SN Ia	0.72	0.38	0.50
	SN-other	0.34	0.35	0.35
	CV/Nova	0.24	0.86	0.37
	lowz-AGN	0.47	0.53	0.50
	midz-AGN	0.99	0.79	0.88
	highz-AGN	0.22	0.89	0.36
	Blazar	0.12	0.56	0.20
	YSO	0.88	0.43	0.58
	LPV	0.98	0.98	0.98
	EA	0.76	0.90	0.82
	EB/EW	0.95	0.82	0.88
	DSCT	0.39	0.83	0.53
	RRL	0.98	0.90	0.94
	CEP	0.30	0.86	0.45
	Periodic-other	0.25	0.65	0.36
	nonvar-galaxy	0.71	0.99	0.83
	nonvar-star	0.99	1.00	0.99
	macro-av	0.61	0.75	0.62
<i>r</i>	SN Ia	0.62	0.35	0.44
	SN-other	0.29	0.32	0.30
	CV/Nova	0.19	0.80	0.31
	lowz-AGN	0.48	0.51	0.49
	midz-AGN	0.99	0.80	0.88
	highz-AGN	0.32	0.89	0.47
	Blazar	0.10	0.50	0.16
	YSO	0.91	0.58	0.71
	LPV	0.98	0.99	0.98
	EA	0.73	0.89	0.80
	EB/EW	0.95	0.81	0.88
	DSCT	0.30	0.84	0.44
	RRL	0.97	0.88	0.92
	CEP	0.18	0.83	0.30
	Periodic-other	0.27	0.68	0.39
	nonvar-galaxy	0.86	0.99	0.92
	nonvar-star	0.99	1.00	0.99
	macro-av	0.60	0.74	0.61

$$\text{F1-score}_{\text{macro}} = \frac{1}{n_{\text{cl}}} \sum_{i=1}^{n_{\text{cl}}} \text{F1-score}_i, \quad (7)$$

where n_{cl} is the total number of classes (17 in this work). Table 3 shows the scores obtained per class and per band for the final classification (final leaves of each node). From the table, we can see that the classification of nonvar-star has high precision, recall, and F1-score, in both ZTF bands. The same is observed for nonvar-galaxies. The lowest scores are for SN-other, Blazar, CEP, and Periodic-other.

The low scores observed in some classes can be explained by the confusion with classes that show similar variability behaviors. To better illustrate this, we show in Fig. 4 the confusion matrices for each band. We can see in both bands that SNe classes tend to be miss-classified as nonvar-galaxy; the same is observed for lowz-AGN and Blazar. On the other hand, YSO are confused with nonvar-stars and Periodic-other. Periodic-other are confused with binary classes. These confusions between persistent variable and transient classes are expected and have been

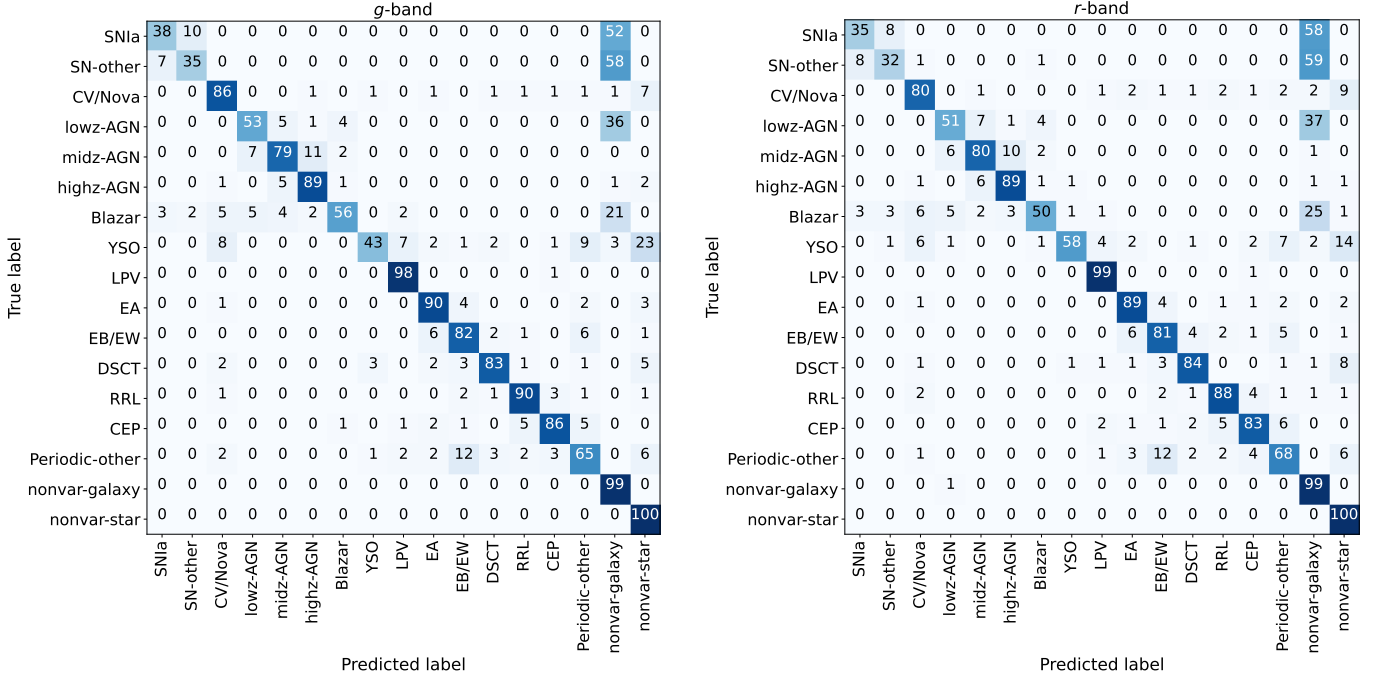


Fig. 4. Confusion matrices of the final 17 classes obtained by using the HBRF in the testing set. The matrix for the g band is shown in the left panel, and for the r band in the right panel. To normalize the confusion matrix results as percentages, we divided each row by the total number of objects per class with known labels, and we rounded these percentages to integer values.

observed before in other works that use ZTF light curves (e.g., Sánchez-Sáez et al. 2021b; van Roestel et al. 2021). They can be explained by the similarities between classes and the lack of precise variability features when measured from the sparse and irregularly sampled DR light curves. For the case of YSOs the confusion with periodic classes can be explained by the diversity of variability behaviors of YSOs, which can present stochastic and periodic signals (Lakeland & Naylor 2022). The confusion with non-variable classes can be explained by the fact that we are using DR light curves, which can hinder variations typically seen in the light curves of extended sources (like lowz-AGN), or in low-amplitude point sources located in crowded fields (like YSOs). Moreover, the confusion between SNe classes and the nonvar-galaxy class is expected, since a transient event does not necessarily occur in the nucleus of the host, and thus, the DR light curves do not necessarily include the transient signal.

Regardless of these classification confusions, the fraction of non-variables classified as variables is close to zero. Among nonvar-galaxies, 1% of the testing set are classified as AGN, but this confusion is expected since the class nonvar-galaxy includes type 2 AGN, which can, in fact, correspond to misidentified type 1 AGN (López-Navas et al. 2023), or that can correspond to former type 2 AGN that have transitioned to type 1 (CSAGN; e.g., López-Navas et al. 2022). We also expect a small fraction of nonvar-galaxies to be classified as SNe when they are the latter's hosts.

Moreover, when grouping the four AGN classes into one single class (lowz-AGN, midz-AGN, highz-AGN, and Blazar), its precision, recall, and F1-score are 1.00, 0.95, and 0.97, respectively, in both g and r bands. Therefore, the low scores measured for some of the AGN classes, can be explained by confusion within these classes. This confusion is expected since these four classes intrinsically represent the same astrophysical class

(AGN) but are observed under different conditions and with different properties (such as the redshift or the presence or not of a relativistic jet).

These results demonstrate that we can use ZTF DR light curves to recover clean samples of variables and transients objects, and particularly of AGN, which is encouraging, considering that we are using PSF fit-based light curves, which are very sensitive to seeing and sky condition variations when dealing with extended sources.

5.3. Feature ranking

The HBRF is composed of five different BRF classifiers. Each of these classifiers will have different feature rankings, since the classes they consider are different. In Table 4 we show the top 15 features per node and per band, ordered by relevance. From the table, we can see that the PS1 and CatWISE colors are relevant in all the classifiers. The PS1 morphology score (ps_score) is relevant for all classifiers except the node_periodic; this is expected, as this is the only classifier that considers purely point sources. The most relevant variability features remain distinct for each classifier; among the most important ones, we highlight the following: features related to the amplitude of the variability (ExcessVar, GP_DRW_sigma, Meanvariance, Amplitude, among others); features related to the timescale of the variations or the level of autocorrelation (GP_DRW_tau, Autocor_length, and IAR_phi); the Mexican hat power spectrum (MHPS) features; and the SPM features. For the case of the node_periodic, the Period is the most relevant feature, which is expected from previous works (e.g., SS21b).

The differences observed in the ranking of features are produced by the complexity of the model and the diversity of classes considered. Therefore, we decided to keep all the features described in Sect. 4 for the final model used in this work.

Table 4. Feature ranking (top 15) for each node of the HBRF and for each band, ordered by relevance.

Band	node_init	node_variable	node_transient	node_stochastic	node_periodic
<i>g</i>	ps_score	W1 – W2	ps_score	W1 – W2	Period
	<i>i</i> – W2	SPM_chi	SPM_chi	<i>g</i> – <i>r</i>	Skew
	W1 – W2	Psi_eta	MHPS_high	<i>r</i> – W1	IAR_phi
	PPE	PPE	SPM_A	<i>g</i> – W1	Gskew
	<i>i</i> – W1	ps_score	<i>g</i> – W1	ps_score	MedianAbsDev
	Psi_eta	<i>i</i> – W2	<i>g</i> – <i>r</i>	<i>i</i> – W2	Q31
	<i>r</i> – W2	SPM_A	SPM_tau_fall	<i>r</i> – <i>i</i>	GP_DRW_tau
	IAR_phi	<i>g</i> – W1	<i>r</i> – <i>i</i>	SPM_chi	<i>g</i> – W1
	GP_DRW_tau	<i>g</i> – <i>r</i>	Skew	<i>r</i> – W2	MedianBRP
	ExcessVar	<i>r</i> – <i>i</i>	Gskew	<i>i</i> – W1	Beyond1Std
	SPM_chi	Skew	PercentAmplitude	ExcessVar	MHPS_ratio
	Autocor_length	<i>g</i> – W2	PPE	<i>g</i> – W2	<i>r</i> – <i>i</i>
	GP_DRW_sigma	<i>r</i> – W2	<i>g</i> – W2	MHPS_low	<i>g</i> – W2
	<i>g</i> – <i>r</i>	MHPS_low	<i>i</i> – W1	IAR_phi	<i>i</i> – W2
	<i>r</i> – W1	Harmonics_mse	GP_DRW_tau	Amplitude	SPM_chi
<i>r</i>	ps_score	W1 – W2	ps_score	<i>r</i> – <i>i</i>	Period
	<i>i</i> – W2	SPM_chi	<i>i</i> – W1	W1 – W2	Skew
	W1 – W2	Psi_eta	MHPS_high	<i>g</i> – <i>r</i>	Gskew
	<i>i</i> – W1	ps_score	SPM_A	<i>r</i> – W1	<i>g</i> – W1
	GP_DRW_sigma	SPM_A	SPM_chi	ps_score	GP_DRW_tau
	Psi_eta	Eta_e	<i>g</i> – W1	SPM_chi	MHPS_ratio
	PPE	<i>i</i> – W2	<i>r</i> – <i>i</i>	<i>r</i> – W2	<i>g</i> – W2
	<i>r</i> – W2	Harmonics_mse	Harmonics_mse	<i>i</i> – W2	<i>i</i> – W1
	Meanvariance	<i>g</i> – W2	<i>g</i> – <i>r</i>	<i>i</i> – W1	IAR_phi
	Q31	<i>g</i> – W1	<i>g</i> – W2	<i>g</i> – W1	<i>i</i> – W2
	<i>g</i> – <i>r</i>	<i>r</i> – <i>i</i>	SPM_tau_fall	IAR_phi	Beyond1Std
	GP_DRW_tau	<i>r</i> – W2	PercentAmplitude	ExcessVar	<i>r</i> – W1
	Harmonics_mag_1	<i>g</i> – <i>r</i>	<i>r</i> – W2	Eta_e	MedianBRP
	Eta_e	Amplitude	<i>r</i> – W1	Amplitude	SPM_chi
	SPM_chi	PPE	<i>i</i> – W2	<i>g</i> – W2	MedianAbsDev

5.4. Calibration of the BRF probabilities

Random forest models tend to provide uncalibrated probabilities (Niculescu-Mizil & Caruana 2005). For binary classification, techniques such as Platt scaling (Platt 2000) or the isotonic method (Zadrozny & Elkan 2001, 2002) can be used to calibrate the output probabilities. However, for multi-class, and/or hierarchical, and highly imbalanced classification, the use of these techniques is not straightforward.

To understand how well calibrated the output probabilities of each node of the HBRF classifier are, we computed the reliability diagrams (Guo et al. 2017) and the expected calibration error (ECE; Naeini et al. 2015) for each node of the classifier independently, using the testing set for each band. To construct the reliability diagram, we used ten bins of probabilities (each one with a width of 0.1) and measured the positive fraction (PF; the number of correctly classified samples versus the total of samples; equivalent to the micro-averaged precision) per bin of probability and the average predicted probability of the bin (confidence). The reliability diagram corresponds to the comparison of the PF versus the confidence of each bin. When this comparison is close to the identity function, we can say that the probabilities are well calibrated. Then we measured the gap, which corresponds to the absolute value of the difference between the PF and the confidence in each bin ($\text{gap} = |\text{PF} - \text{confidence}|$), and

calculated the ECE as

$$\text{ECE} = \sum_{m=1}^M \frac{N_{\text{bin}}}{N} \text{gap},$$

where N corresponds to the total of samples, N_{bin} to the total of samples in the probability bin, and M to the total number of probability bins. When the ECE score is close to zero, we can say that the model is well calibrated.

Figures 5 and 6¹¹ present the reliability diagrams and the ECE scores for each node of the HBRF, for the *g*-band and the *r*-band testing sets, respectively. For each bin, we plot the PF (in black) and the gap (in red) versus the confidence. From these figures, we can see that the node_init, node_stochastic, and node_periodic of both bands are well calibrated (except for one bin of the node_periodic in the *r* band, but this bin contains only a few samples). For the case of the node_transient, the PF is smaller than the average probability, which implies that the probabilities overestimate the reliability of the classifications. These results are expected, though, as the quality of the ZTF DR light curves is not adequate for the study of transients. For the node_variable, the PF is larger than the average probability in

¹¹ The code used to generate these plots corresponds to a modified version of the code presented in <https://github.com/hollance/reliability-diagrams>

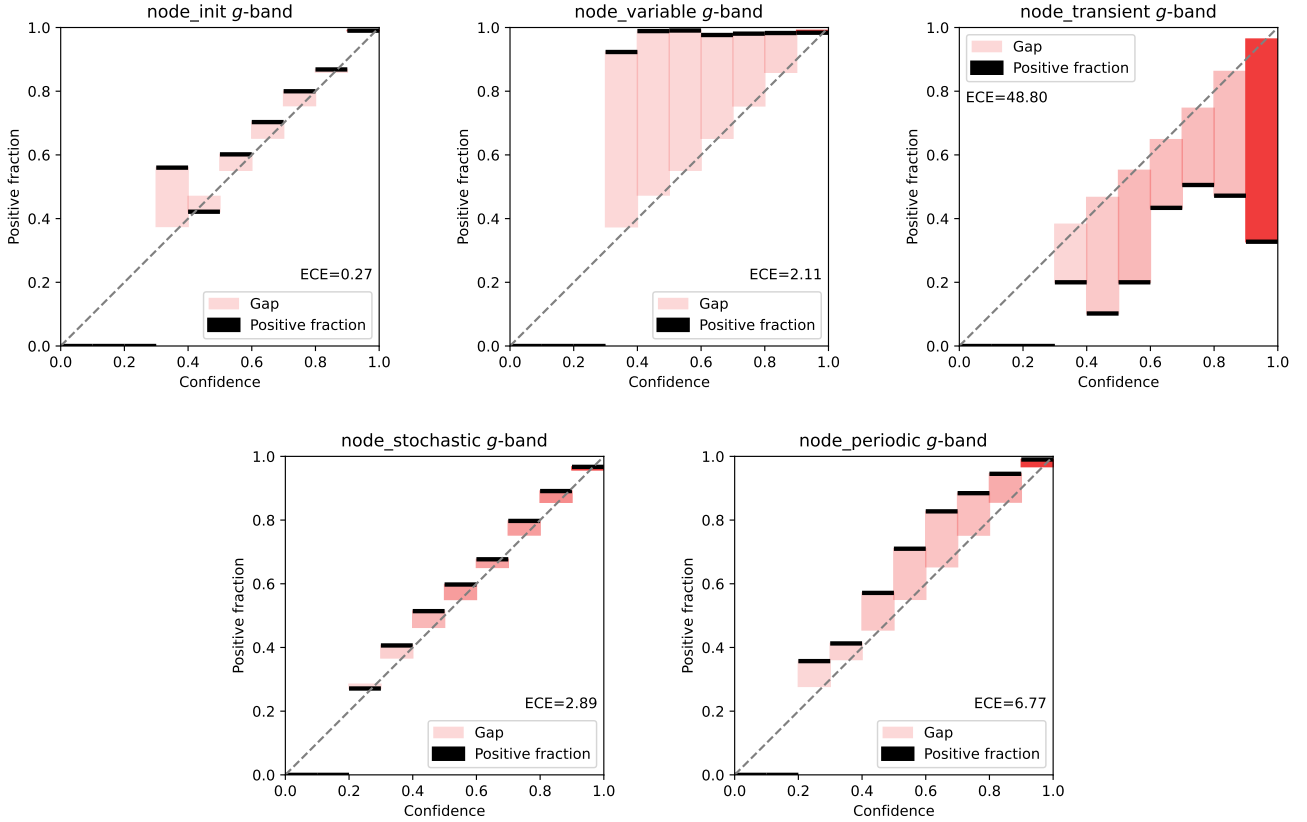


Fig. 5. Reliability diagrams for each node of the g -band HBRF classifier. We plot the PF (shown in black) and the gap (in red) versus the confidence. The transparency of the gap represents how much each bin contributes to the ECE score (less transparent for bins with more samples). The identity function is shown as a reference. The ECE score is shown next to the legend for each node.

most of the probability bins; however, only a small fraction of the predictions have a probability smaller than 0.9, and thus the ECE score for the node_variable is smaller than the score obtained for the node_transient, node_stochastic, and node_periodic. From this we can conclude that when the predicted probabilities of the node_variable are lower than 0.9, there is an underestimated reliability of the classifications. Since this is a hierarchical model, this result favors the quality of the classifications, because we can expect that the sources that go to each of the following nodes (node_transient, node_stochastic, and node_periodic) are properly classified. We suggest that users of our classifications take these results into account when filtering the candidates from each node by probability.

6. Results for the ZTF/4MOST sky

We applied the g -band and r -band HBRF models to the sources in the ZTF/4MOST sky described in Sect. 4, which includes 86 576 577 and 140 409 824 individual light curves in the g band and the r band, respectively. Figure 7 shows the number of candidates per class for these sources. The fraction of sources classified as transient or persistently variable is 3.84% in the g band, and 19.20% in the r band. This is a surprising difference and is not reflected at all in the results obtained for the testing set (which corresponds to 20% of the LS), presented in Sect. 5.2. Figure 8 shows the number of sources per class, when only the sources with a classification probability in the node_init equal to or greater than 0.9 are considered ($P_{\text{init}} \geq 0.9$). The fraction of sources classified as transient or persistently variable drops to 0.73 and 2.62% in the g and r bands, respectively. In this case,

the differences between the results of the g and r bands have narrowed considerably, but still remain much higher when using the r -band light curves. From Figs. 7 and 8 we see that the main difference in the results, when going from the g to the r band, occurs in the cases of the SN-other, CV/Nova, highz-AGN, and YSO classes.

In Figs. 9 and 10, we show the probability distributions obtained for all the sources in the ZTF/4MOST sky, in the g and r bands. Figure 9 shows the probability obtained by the node_init, P_{init} , that separates the sources into nonvar-star, nonvar-galaxy, and variable, while Fig. 10 shows the final prediction of the model, P , which corresponds to the final leaves of each node (i.e., the probability of the node_init for nonvar-star and nonvar-galaxy, and the probability of the node_transient, node_stochastic or node_periodic, for sources classified as variable in the node_init). From these figures, it can be noticed that in the node_init the classes with the lowest probabilities are the transients, and the ones with the highest probabilities are the non-variable and AGN classes. We can also notice that the distribution of P_{init} for most classes in both bands presents a bimodal distribution, with several sources having $P_{\text{init}} \sim 1$, and others having lower values. The final predicted probabilities P vary strongly between classes, as was also observed in SS21b, and is explained by the confusion between some similar classes. The classes with the highest P in both bands are CV/Nova, midz-AGN, LPV, and the non-variable classes.

The coordinates of the objects are not included as features in the HBRF model, and thus we use them to inspect the quality of the classifier by exploring the sky densities of each class in the Galactic coordinate space. Figures 11 and 12 show the sky

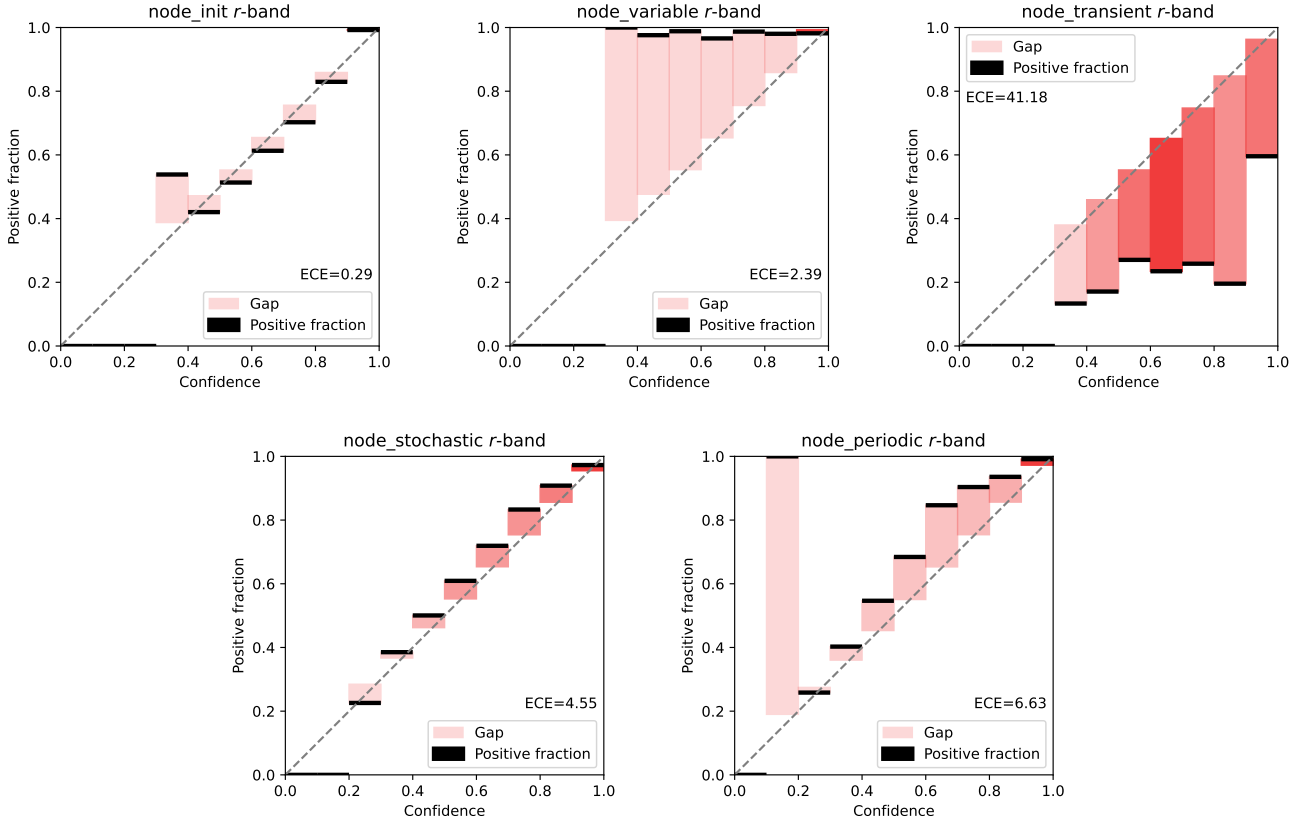


Fig. 6. As in Fig. 5, but for the *r*-band HBRF classifier.

density of midz-AGN and highz-AGN candidates, respectively, in Galactic coordinates. The density maps of the remaining 15 classes are shown in Appendix A. From these maps, we can notice that for some extragalactic classes, like highz-AGN (and SNIa, SN-other, or Blazar), there is a high overdensity of candidates in the Galactic plane ($gal_b \sim 0$), while for others, like midz-AGN (and lowz-AGN) the densities drop at lower Galactic latitudes. The overdensities of extragalactic classes in the Galactic plane are unexpected, as we should see a decrease in the number of sources belonging to these classes in this region due to Galactic extinction. On the other hand, the enormously larger numbers of stars in the Galactic plane provide a natural source of false positives in this region. In the case of the *g* band, these overdensities disappear when only sources with a high probability in the node_init ($P_{\text{init}} \geq 0.9$) are selected. For classes that are expected to be more prominent in the Galactic plane, like YSO, CEP, and LPV, we see a much lower density at larger Galactic latitudes. For all classes, we obtain more reliable results in the *g* band, when $P_{\text{init}} \geq 0.9$. We, therefore, recommend giving priority to candidates selected from the *g*-band light curves, when using ZTF DR light curves to identify variable and transient objects.

7. Discussion

7.1. Differences in the number of variables and transient candidates selected using the *g* and *r* ZTF bands

For some classes, like SNIa, SN-other, CV/Nova, highz-AGN, and YSO, there is a clear difference in the number of persistent variables and transients selected using the *g* and *r* ZTF bands. Moreover, as previously mentioned, some of the extragalactic

classes (SNIa, SN-other, highz-AGN, and Blazar) show large overdensities in the Galactic plane. Our hypothesis is that these results can be explained either by the very intense observing seasons carried out in the *r* band by the ZTF project, as can be seen in the ZTF DR11 documentation¹², or by source crowding.

In order to disentangle what is producing these differences in the classification of the *g*-band and *r*-band light curves, we first visually inspected the science images of the ZTF *r* band in the zones with a high density of extragalactic candidates (around $gal_l \sim 220$ and $gal_b \sim 0$), but we did not find evidence of large numbers of source blends, or other photometric issues (like ghosts). To perform a more quantitative analysis, we selected the ZTF field 360, which is centered at $gal_l \sim 225$ and $gal_b \sim 2.6$, and measured the number of sources with unique ZTF ID in the *r* band that have a neighboring source within a radius of 3". Of the 3 356 426 sources in this field in the *r* band, only 61 314 (2%) have a neighboring source within this radius. This corresponds to a low number of sources with close neighbors, and thus source crowding cannot explain the high density of extragalactic candidates in the galactic plane when using the *r* band.

We then inspected whether the regions showing the largest number of extragalactic candidates are consistent with regions of very rapid cadence in the ZTF light curves. Figure 13 shows the number of epochs and the MaxSlope (maximum absolute magnitude slope between two consecutive observations) versus the Galactic coordinates for highz-AGN candidates. We can see that the sky regions with a higher number of candidates coincide with regions with a larger number of observations and extreme values

¹² https://irsa.ipac.caltech.edu/data/ZTF/docs/releases/dr11/src/TspanVsDT_fid2.png

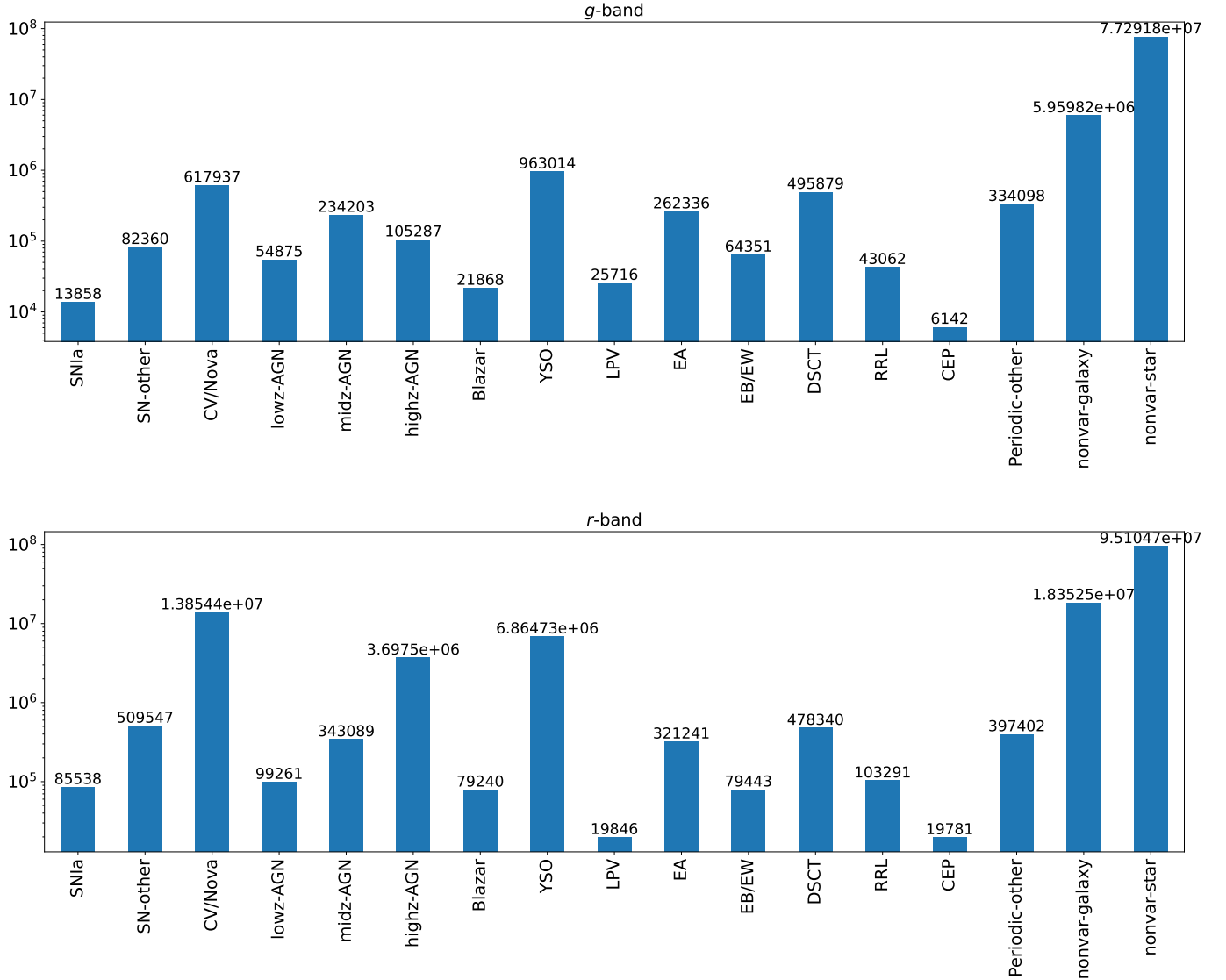


Fig. 7. Number of candidates per class for all the sources in the ZTF/4MOST sky with more than 20 epochs. The results for the *g* band are shown in the top panel, and for the *r* band in the bottom panel. The number of sources per band is specified on top of each bar.

of *MaxSlope*. When *MaxSlope* is large, it could mean that there are two data points very close in time, but with different magnitude values. From the figure, we can see that around the Galactic plane, there is a high concentration of sources with more than 400 epochs and that *MaxSlope* has extreme values in this region. A similar behavior is observed in the *r* band for the other extragalactic classes (SNIa, SN-other, lowz-AGN, midz-AGN, and Blazar). These large values of *MaxSlope* can be produced by the very rapid cadence seasons in the *r* band, where in some cases the separation between two consecutive observations can even be less than a minute. This can be seen in Fig. 14, which shows the light curve in the *r* band of a highz-AGN candidate with ZTF ID 360206100033262, located in the field 360, and the histogram of the time difference between consecutive observations for its light curve. This light curve has a total of 445 observations. During the day with MJD = 58 491, a total of 189 observations were obtained (42% of the total observations), and the day with MJD = 58 867, there were 124 observations (28%). In the histogram of the time difference between consecutive observations,

we can notice that most of the epochs are separated by a few minutes.

In order to test the effect of the rapid cadence in some ZTF fields, we recomputed the features for the field 360 in the *r* band, but keeping only one observation per night (the first observation), and we use the model trained for the *r* band (without any modifications) to classify the new set of features. There were 2 107 241 light curves with more than 20 epochs (after filtering), and of these, 12 961 were classified as one of the AGN classes (lowz-, midz-, highz-AGN, or Blazar), with 1024 (8%) of them having $P_{\text{init}} \geq 0.9$. In the same field, before the modification of the features, there were 54 981 sources classified as AGN (with 21% having $P_{\text{init}} \geq 0.9$), and with the modification of the light curves, only 5% of them are classified as AGN.

From these results, we recommend giving priority to the classifications obtained in the *g* band. When using the *r*-band classifications, we suggest filtering the catalogs by the classification probability in the node_init ($P_{\text{init}} \geq 0.9$), and/or to avoid the use of our model in regions with rapid cadence observations

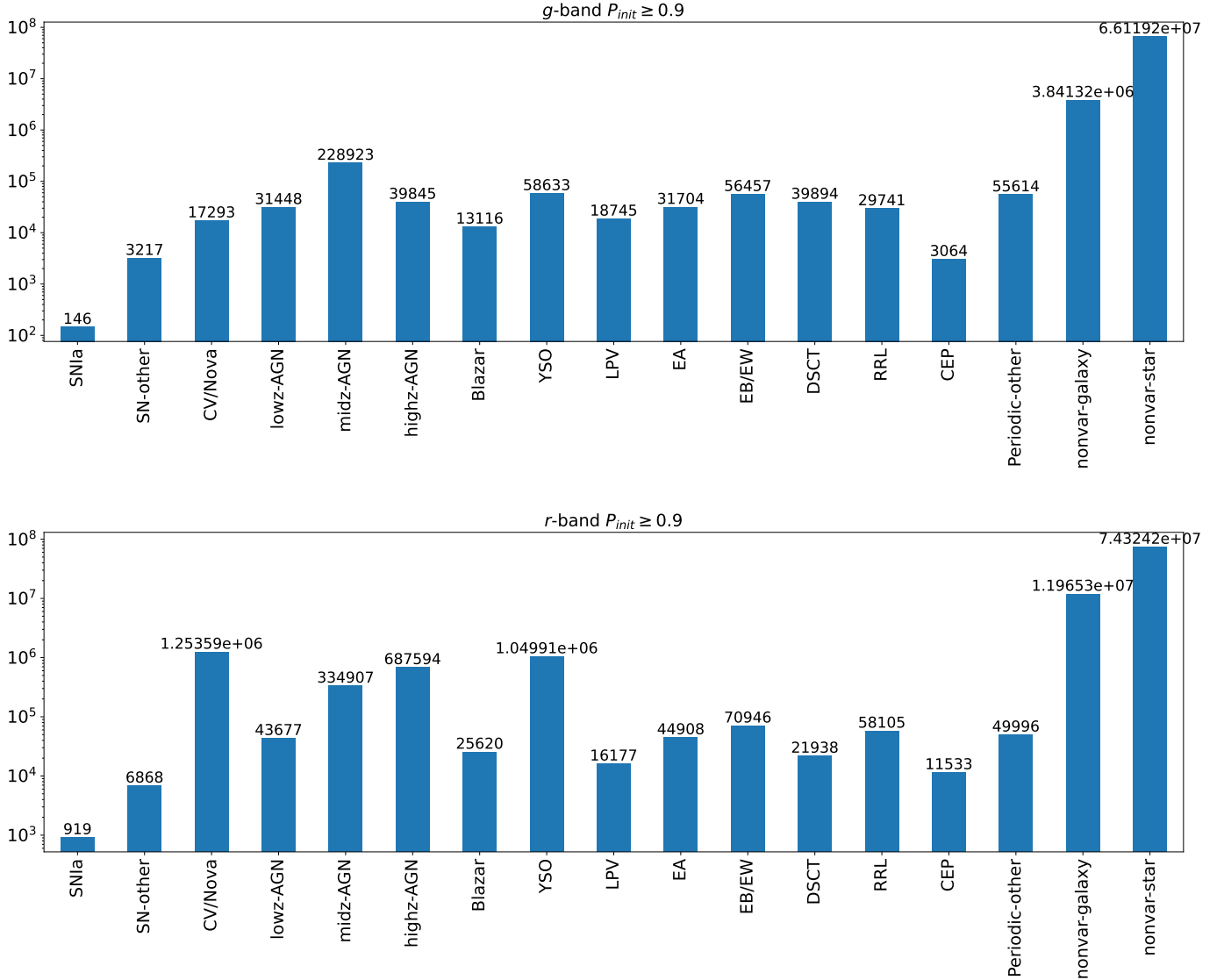


Fig. 8. Number of candidates per class for all the sources in the ZTF/4MOST sky with more than 20 epochs, and with a classification probability in the node_init greater or equal than 0.9. The results for the *g* band are shown in the top panel, and for the *r* band in the bottom panel. The number of sources per band is specified on top of each bar.

(which normally have more than 400 epochs in the *r*-band light curves, according to the ZTF DR11 documentation). We also suggest taking into account the probability distribution shown in Figs. 9 and 10, and the reliability diagrams presented in Figs. 5 and 6 when filtering our classifications by probability.

7.2. AGN candidates

As mentioned above, the main motivation for the development of this classifier was to identify AGN and Blazar candidates at different redshifts that will be followed up by the ChANGES survey. There are 416 233 (*g*-band) and 4 219 094 (*r*-band) AGN and Blazar candidates in total. When only sources with $P_{init} \geq 0.9$ are considered, we have 313 332 candidates in the *g* band and 1 091 798 in the *r* band. This difference in the number of candidates can be explained by the larger number of light curves available in the *r* band, but also by the issues described in Sect. 7.1; there is a large overdensity in the Galactic plane of highz-AGN candidates detected in the *r* band.

Since a source can have more than one light curve per band in the ZTF DRs, we did an internal match of the candidates in each band, using a radius of $1.54''$, and kept only one light curve per source per band. After this, we ended up with 384 242 candidates in the *g* band, with 287 156 having $P_{init} \geq 0.9$, and 4 048 299 candidates in the *r* band, with 1 020 327 having $P_{init} \geq 0.9$. We cross-matched these two samples using a radius of $1.5''$, and found that there are 356 631 candidates that are classified as AGN or Blazar in both bands. This corresponds to 92.8% of the *g*-band candidates and only 8.8% of the *r*-band candidates.

To have an idea of the quality of the AGN selection, we cross-matched the set of AGN candidates in both bands, and the AGN sample from the LS, with the *Gaia* DR3 catalog of proper motions (Gaia Collaboration 2022). From this, for each sample we measured the signal-to-noise of the proper motion (PMsig), by dividing the proper motion (PM) of each source by its error in the following way: $PMsig = pm / \sqrt{e_{pmRA}^2 + e_{pmDE}^2}$, where e_{pmRA} and e_{pmDE} are the PM errors in RA and Dec, respectively. We would expect to measure low PMsig values for

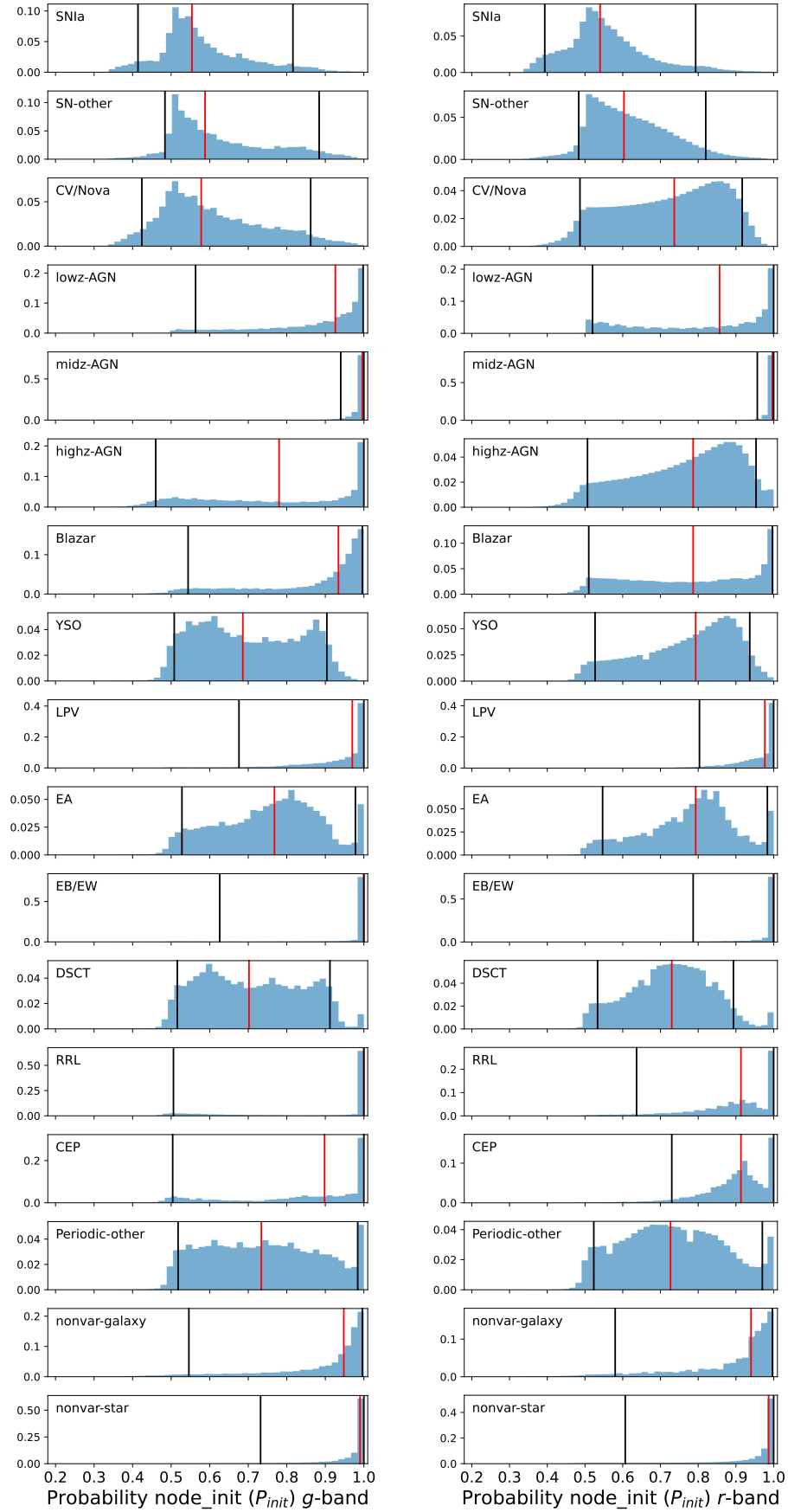


Fig. 9. Normalized probability distributions of the node_init (P_{init}) for the 17 classes considered in this work. The results for the g band are shown in the left panel, and for the r band in the right panel. The red lines show the median probability for each class. The black lines show the 5th and 95th percentiles of the probabilities.

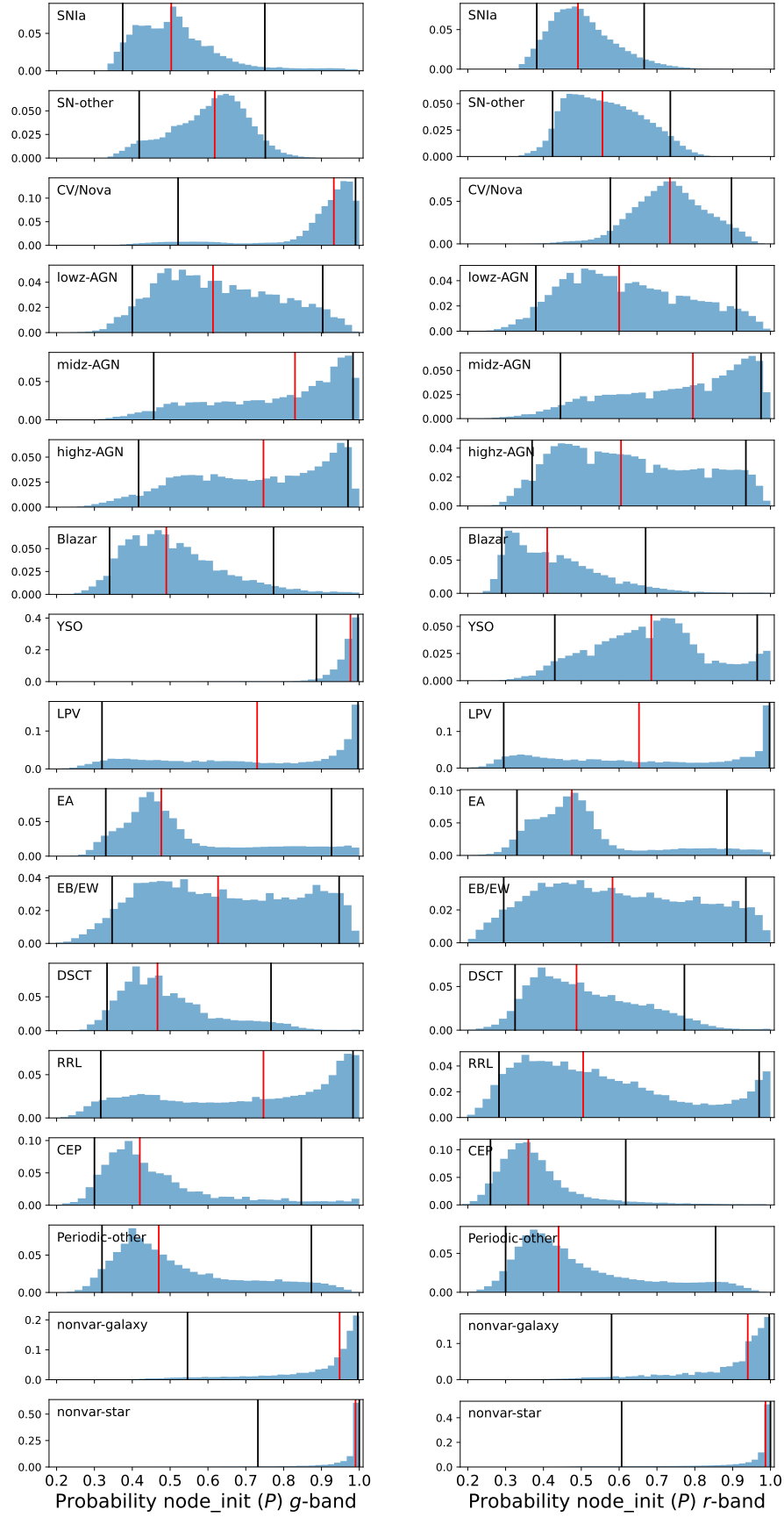


Fig. 10. As in Fig. 9, but for the final leaves (final prediction) of the model (P).

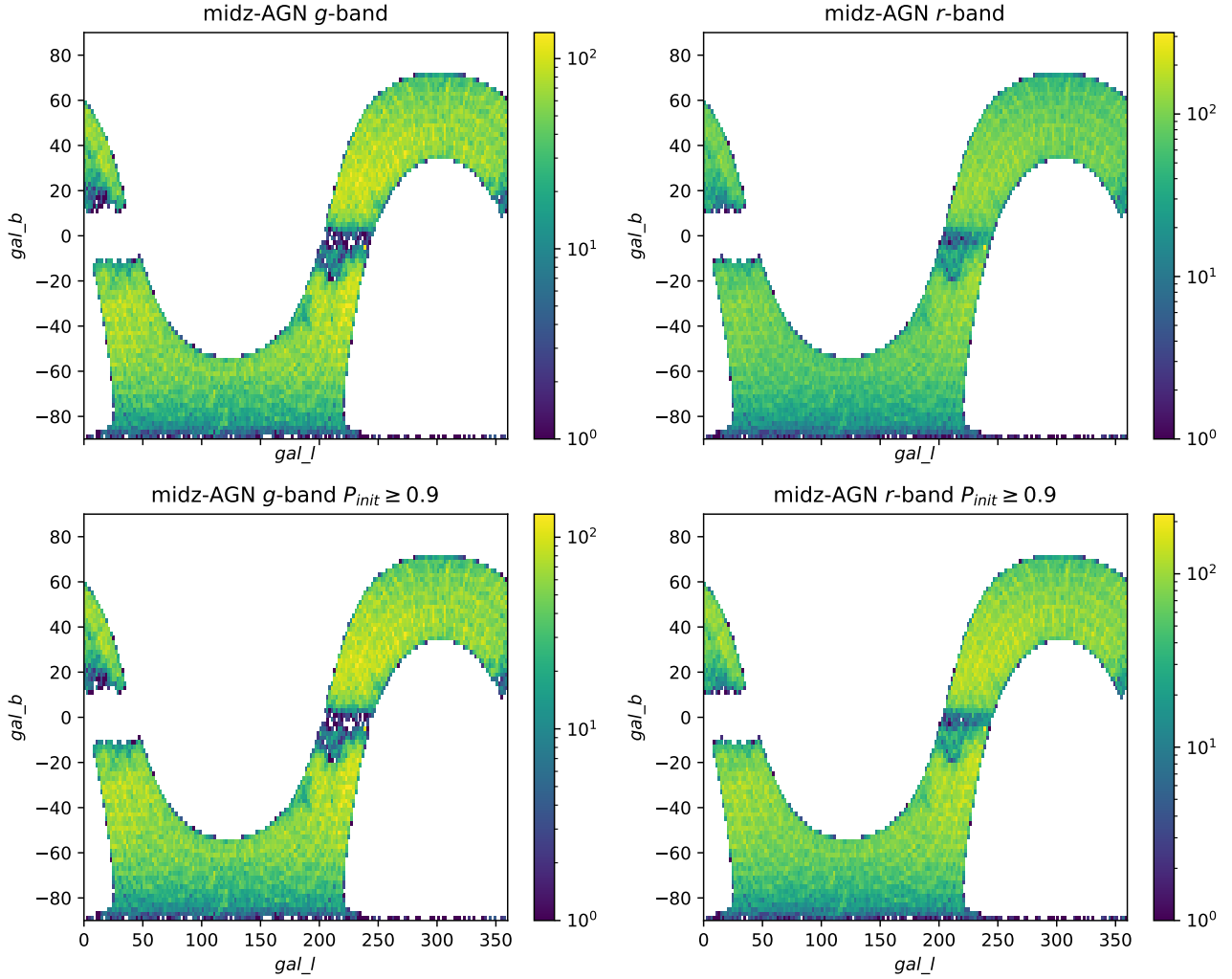


Fig. 11. Number of targets (see the color bars) per 4 deg^2 for the midz-AGN class, in Galactic coordinates. The upper panels show the densities for all the selected candidates in the g band (upper left) and the r band (upper right). The bottom panels show the densities for candidates with large probability in the node_init ($P_{\text{init}} \geq 0.9$) for the g band (bottom left) and the r band (bottom right). Note that the range of the number of targets covered by each color bar can differ in each panel.

AGN, since they are distant sources, but a fraction of them will present large PMSig values when they look like extended sources in the *Gaia* DR3 images. We noted that 84% of the AGN in the LS have $\text{PMSig} \leq 3$, the same fraction is observed for the candidates selected in the g band, while only 39% of the candidates selected from the r band have $\text{PMSig} \leq 3$. Visual inspection indicates that the vast majority of the significant PM sources appear to have a stellar origin, although a slim minority seem to be legitimate AGN located in the centers of resolved, host-dominated nearby galaxies.

Moreover, we compared the optical and mid-infrared color distributions of these three samples. These are shown in the top panels of Fig. 15. The left panel shows the $g - r$ distribution for AGN from the LS, and AGN candidates in both bands, the central panel shows the $r - W1$ color, and the right panel shows their $W1 - W2$ distributions. We can note that the AGN candidates selected from the g band have very similar colors to those from the LS, while the candidates selected from the r band have different color distributions. We also compared the feature distributions in each band. Some of these are shown in the lower panels of Fig. 15. The bottom panel shows the feature Amplitude (left; half of the difference between the median magnitudes obtained with the 5% brighter and 5%

fainter measurements), GP_DRW_tau (center; relaxation time, or timescale, from damped random walk –DRW– modeling), and the mean magnitude (right), in the g band, for AGN in the LS (blue) and AGN candidates (green), and in the r band, for AGN in the LS (yellow) and AGN candidates (pink). From these figures we can see that in general the g -band candidates have feature distributions closer to the ones of the AGN in the LS, compared to the r band, although they are not exactly the same. This is expected, as the LS is dominated by blue and bright AGN.

From these results, we can conclude that the AGN sample selected with the g band is consistent with the AGN sample from the LS, and we can suspect that the number of false positives selected from the r band are non-negligible. Therefore, for the selection of candidates to be observed by ChANGES, we decided to give priority to the candidates selected using the g band, and for objects without g -band classification, the classification from the r band was used, but including only objects with high probability in the node_init ($P_{\text{init}} \geq 0.9$).

7.3. Variable star candidates

The variable star classes considered in this work can be roughly distinguished in a period versus amplitude diagram. In Fig. 16

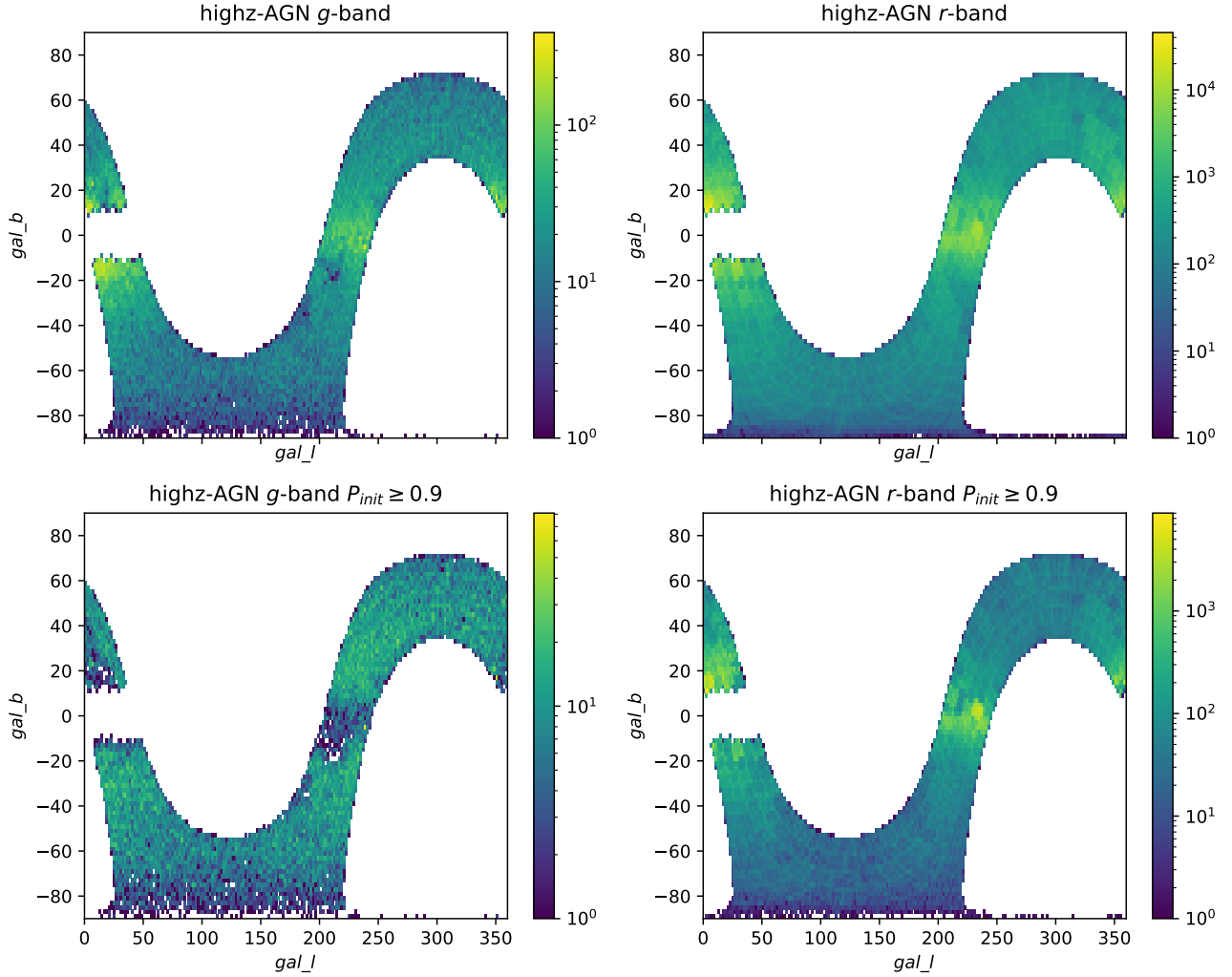


Fig. 12. As in Fig. 11 but for the highz-AGN class.

we show the feature **Period** versus the feature **Amplitude** for the sources in the LS (left panel), and for candidates in the ZTF/4MOST sky filtered by their probability in the `node_init` and their final predicted probability ($P_{\text{init}} \geq 0.9$ and $P \geq 0.5$; right panel). We can see in the figure that the periods of some classes have distributions that are different from expectations (e.g., Catelan & Smith 2015). This is produced by aliasing, when the periodogram finds a multiple of the true period, or spurious periods related to the light curve cadence (most prominently, 0.5 day, 1 day, and a month). For instance, some LPVs show periods of one day, which can affect their proper classification. Fortunately, the model uses other features that are related to the timescale of the variations, which can compensate for these miscalculated periods, such as `IAR_phi` (level of autocorrelation using a discrete-time representation of a damped random walk or DRW model) and `GP_DRW_tau` (relaxation time from a DRW model), which are among the most relevant features in the `node_periodic` (see Table 4). We can also see that there is a large fraction of DSCT candidates with very low amplitudes and short periods: these sources drop out when only sources with $P \geq 0.9$ are selected. In general, the distribution of the high probability candidates (right panel in Fig. 16) is consistent with previous findings (e.g., Catelan & Smith 2015; Chen et al. 2020). Therefore, in order to obtain reliable candidates from this model, we suggest applying probability cuts, taking into account

the probability distribution per class shown in Figs. 9 and 10, and the reliability diagrams presented in Figs. 5 and 6.

7.4. Comparison with other works

The ALerCE broker alert light curve classifier (SS21b) introduced a taxonomy similar to the one used in this work, but their model does not include non-variable classes, as they were dealing with alerts and not with DR light curves. We find that our precision, recall, and F1-score are quite similar to those obtained by them. Excluding the non-variable classes, the confusion matrices presented in Fig. 4 are comparable to the one presented in Fig. 7 of SS21b. However, for some classes, like SNIa, lowz-AGN, or YSO, the recall is lower when using DR light curves; this result is expected, since the noisy DR light curves hinder the detection of the low-amplitude variations from AGN and YSOs, while the alert light curves are dominated by the most variable sources belonging to these classes. Moreover, the DR light curves of transients have more noise compared to their alert light curves, since the photometry of the former is centered in the host galaxy, and the contribution from the host is removed in the alert light curves.

In Fig. 17 we compare the classifications provided by the ALerCE light curve classifier with our results. The top panels show the comparison for all the sources in common in both bands

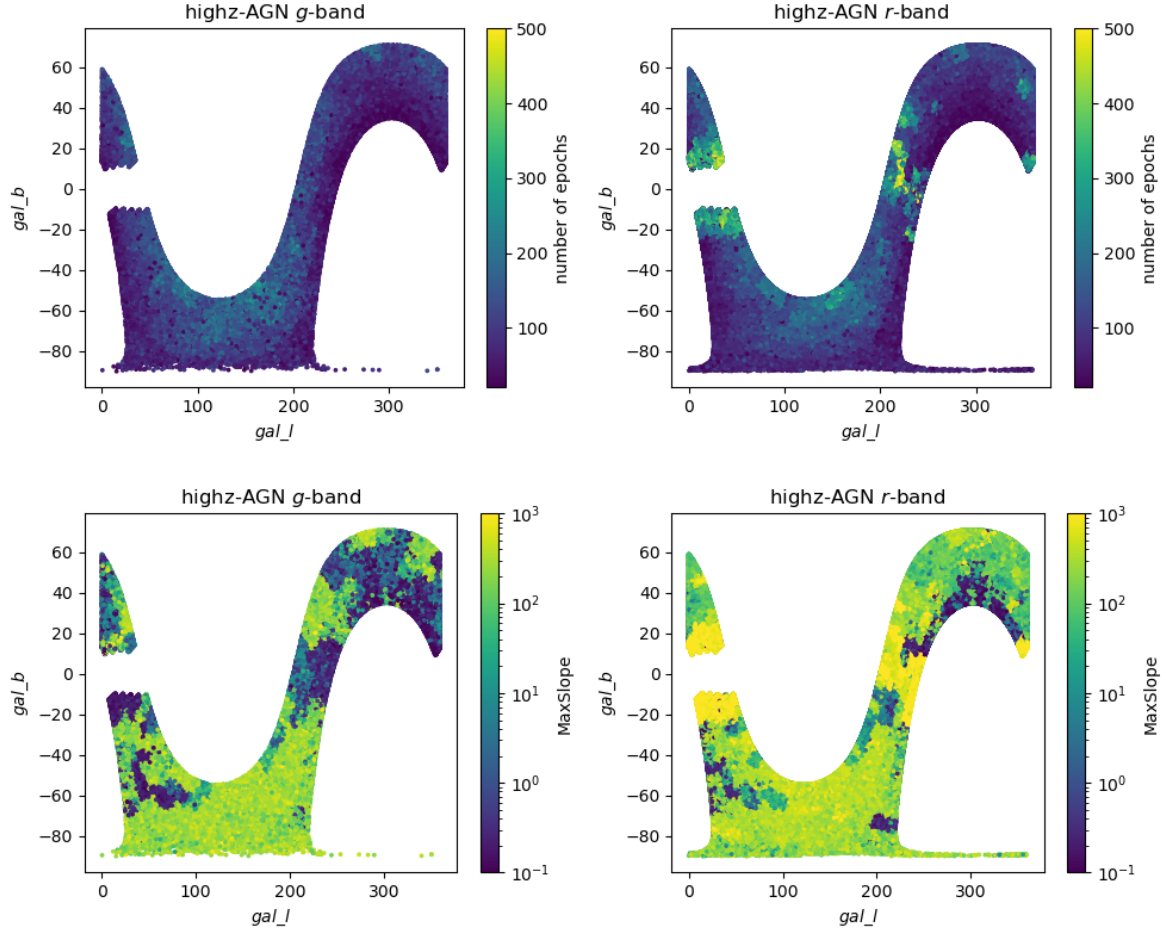


Fig. 13. Distribution of light curve properties as a function of the Galactic coordinates for highz-AGN candidates. The top panels show the number of epochs, and the bottom panels the MaxSlope, for the *g* (left) and *r* (right) bands.

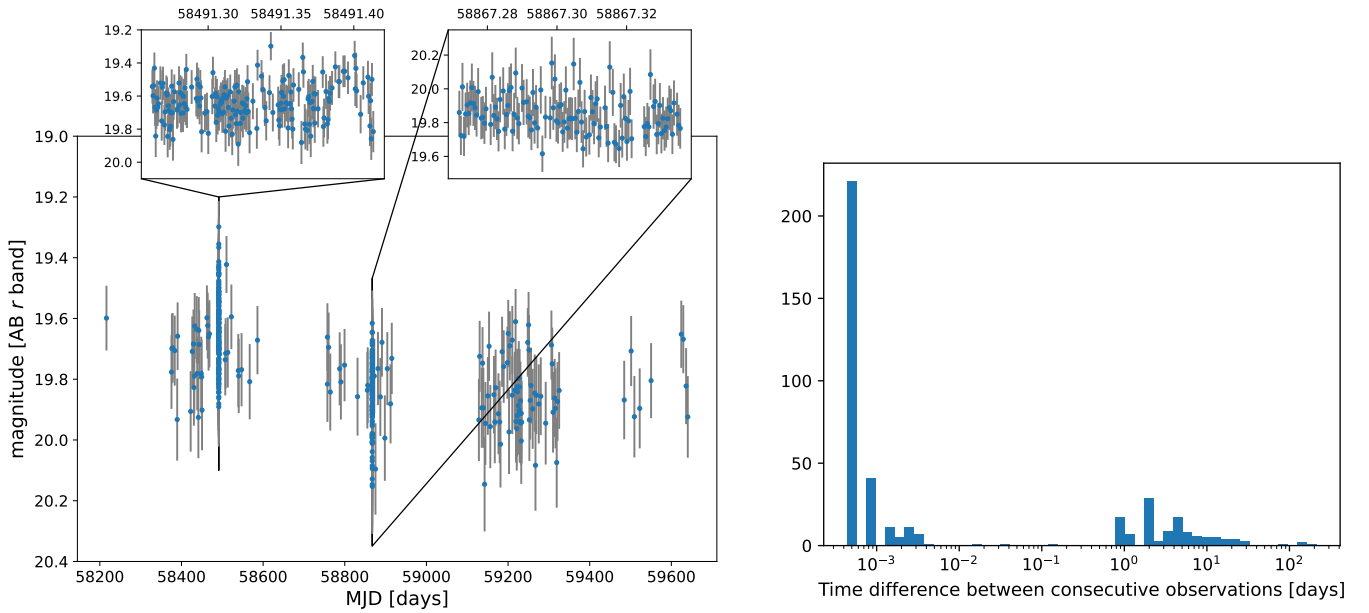


Fig. 14. Highz-AGN candidate with ZTF DR11 ID 360206100033262. The left panel shows its *r*-band light curve. Zoom-in at MJD = 58 491 and MJD = 58 867 are shown. The right panel shows the histogram of the time difference between consecutive observations for the *r*-band light curve.

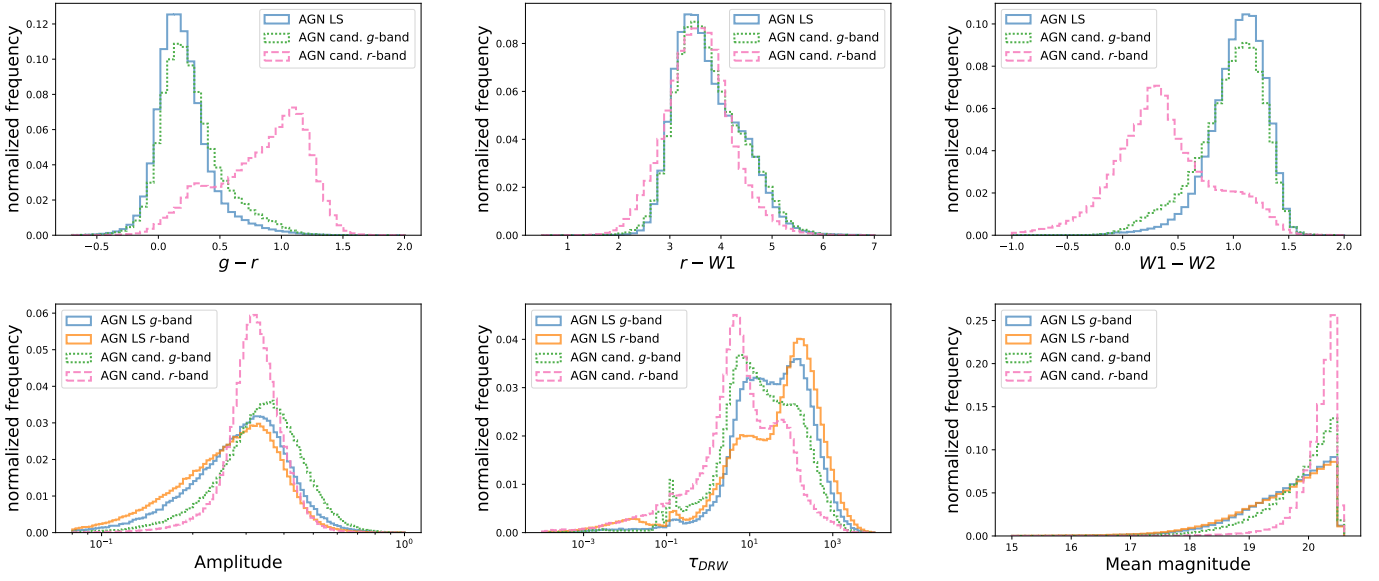


Fig. 15. Feature distributions of AGN candidates and AGN from the LS. Top panels: $g-r$ (left), $r-W1$ (center), and $W1-W2$ (right) color distributions of AGN from the LS (blue), AGN candidates selected from the g band (green), and AGN candidates selected from the r band (pink). Bottom panels: Amplitude (left), GP_DRW_tau (timescale of the DRW model, center), and mean magnitude (right), for AGN from the LS in the g band (blue), AGN from the LS in the r band (yellow), AGN candidates selected from the g band (green), and AGN candidates selected from the r band (pink).

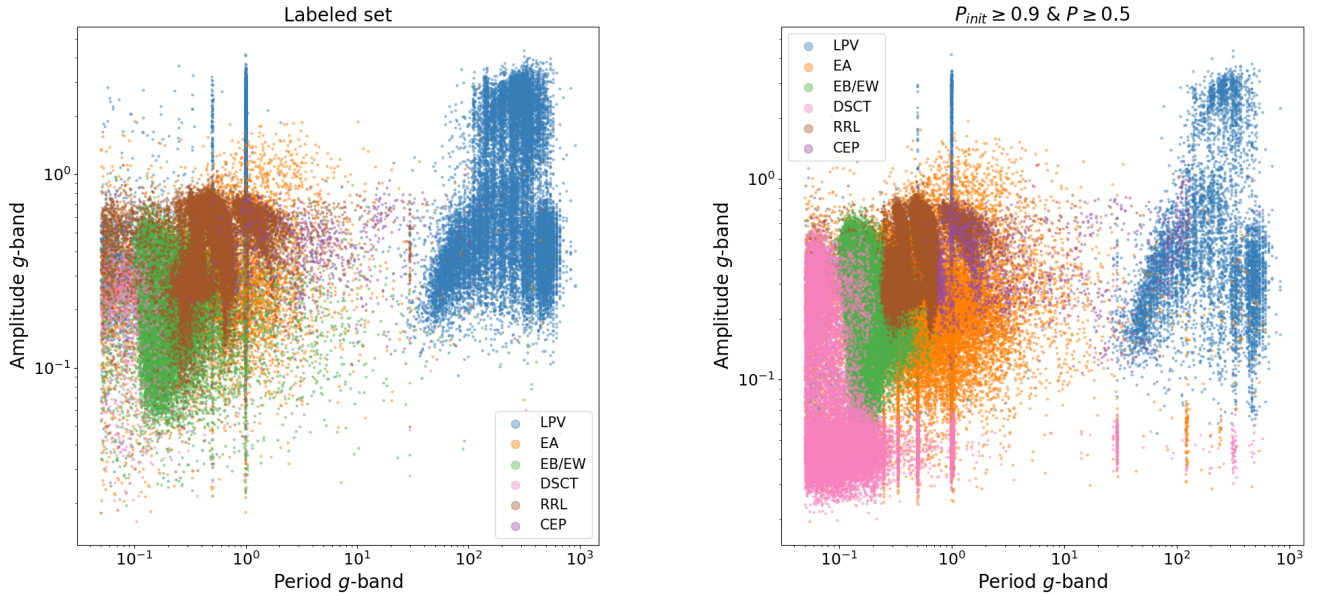


Fig. 16. Period versus Amplitude in the g band for different periodic variable star classes. The left panel shows sources from the LS, and the right panel shows periodic variable star candidates, filtered by the probability of the node_init $P_{init} \geq 0.9$ and by the final probability $P \geq 0.5$.

(372 125 in g and 360 322 in r), while the bottom panels show the comparison when only sources with a final predicted probability $P \geq 0.5$; in both ZTF DR11 and ZTF alert models; are considered (123 829 in g and 116 827 in r). Clearly, there is more confusion when all sources in common are included in the analysis, but for those with probabilities larger than 0.5, the results agree well. We can see that a high fraction of the SNe candidates from the ZTF alert stream are classified as non-variable galaxies in the ZTF DR11; this is expected since the DR light curves do not necessarily contain the flux from the transient, as the PSF photometry is centered in the position of the host galaxy and not at the position of the transient itself. Moreover, other classes

from the alert classification are classified as non-variable in the ZTF DR model. This can be due to problems in the ZTF image subtraction that produces alerts for non-variable objects (like the bogus class from Carrasco-Davis et al. 2021), or by the difficulty in detecting low-amplitude variations when DR light curves are used.

When comparing the classifications provided by the ALerCE light curve classifier and those obtained using the ZTF DR11, we notice that there are 49 sources classified as SNIa and 113 as SN-other, in both g and r bands, that are classified as AGN, quasi-stellar object, or Blazar based on their alerts. After inspecting some examples of these objects, we noticed that most

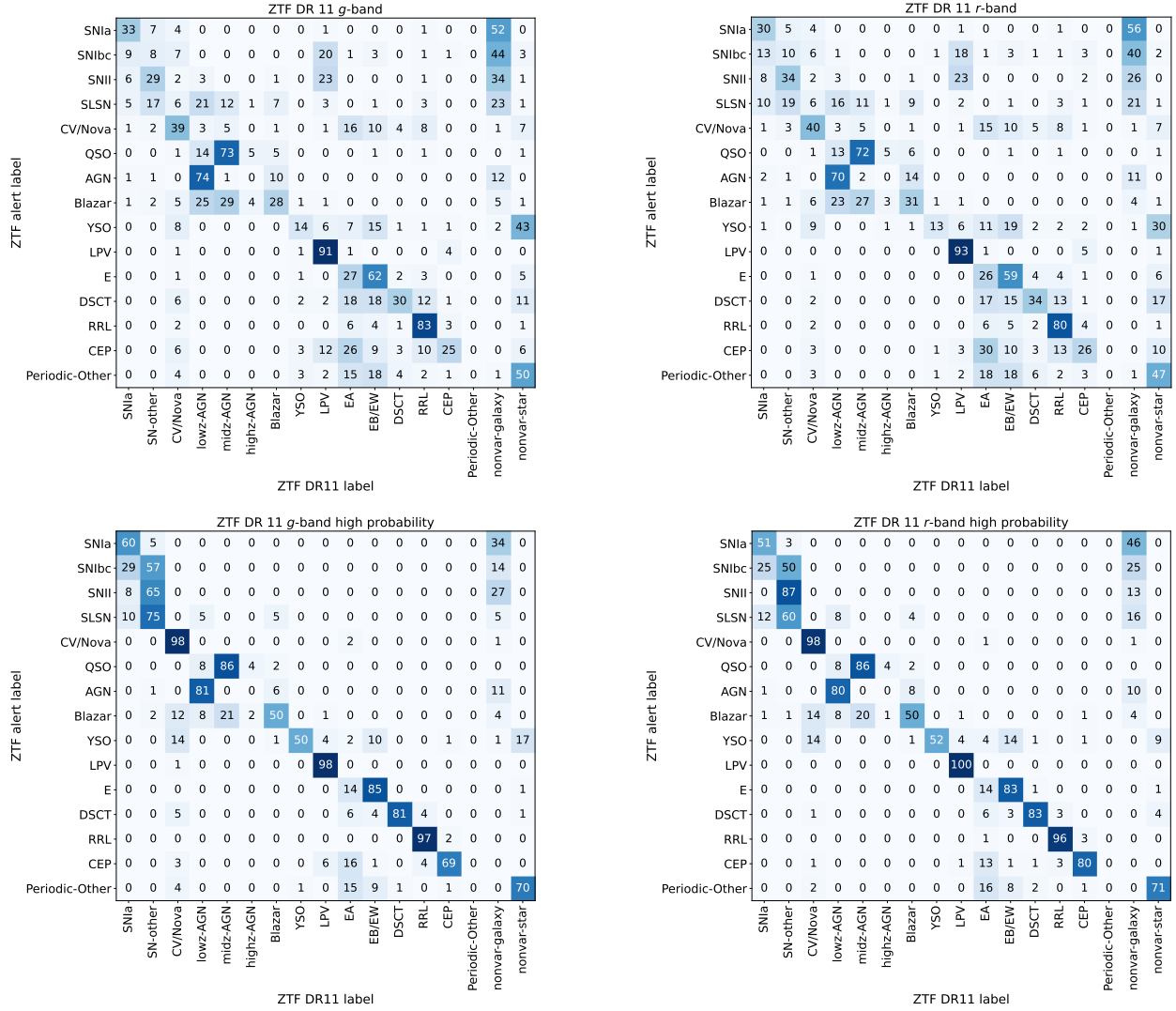


Fig. 17. Comparison of the classifications obtained by the ALerCE broker ZTF alert light curve classifier, and the classifications obtained by our model for the ZTF DR11 light curves. The left panel shows the results when the g -band features are used for the DR classifier, the right panel shows the results when the r -band features are used for the DR classifier. The bottom panels show the same results, but keeping only those sources with high classification probabilities ($P \geq 0.5$) in both the alert and DR samples. We divided each row by the total number of objects per class with alert labels, and we rounded these percentages to integer values.

of them were transients present in ZTF that produced alerts when the transient event ended. An example of this is presented in Fig. 18. The figure shows the ZTF DR11 and the alert light curve of the source ZTF19abcezc. This source was reported to TNS by the ATLAS team (Tonry et al. 2018), and it is classified as SNIa in both bands from its ZTF DR11 light curves. From the figure, we can see that the alerts are generated after the SN event finished. The source was classified by the ALerCE broker as AGN. This confusion can be explained by the lack of a transient event in the alerts, and WISE and optical colors similar to those of AGN (the color of the host galaxy). These results demonstrate the advantage of combining alert and DR light curves when dealing with ZTF data.

Chen et al. (2020) classified ZTF DR2 light curves into 11 classes of periodic variable stars, and provided a catalog with classifications for 781 602 objects. There are 57 128 and 57 086 sources in their catalog located in the ZTF/4MOST sky that also have classifications from our model in the g and r bands,

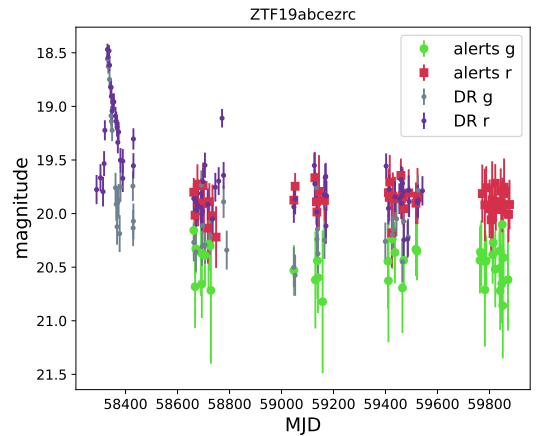


Fig. 18. ZTF alert and ZTF DR11 light curve for the source ZTF19-abcezc.

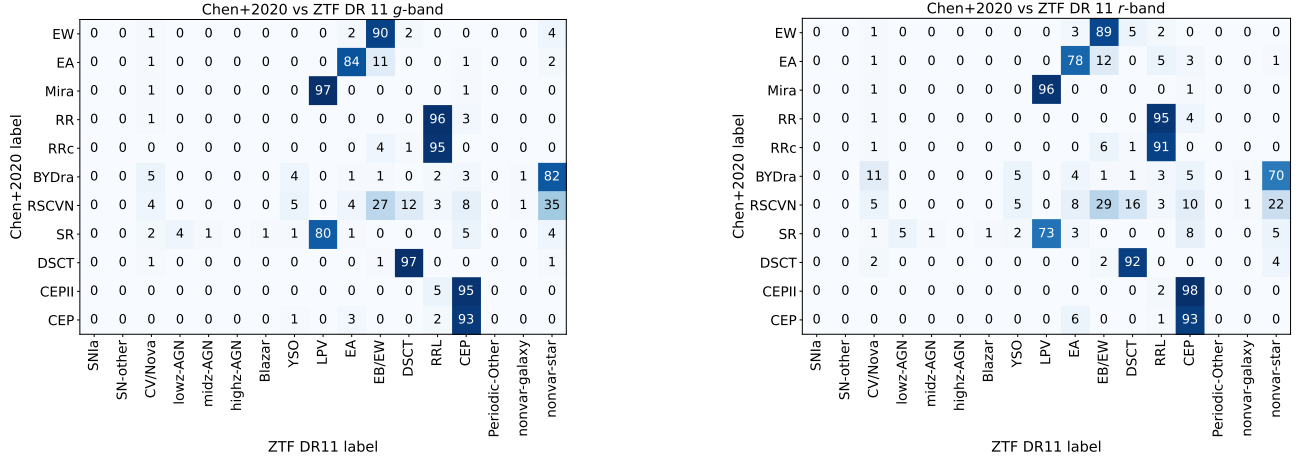


Fig. 19. Comparison of the classifications obtained by [Chen et al. \(2020\)](#), using ZTF DR2, and the classifications obtained by our model for the ZTF DR11 light curves. The left panel shows the results when the g -band features are used for the DR classifier, the right panel shows the results when the r -band features are used for the DR classifier. We divided each row by the total number of objects per class with [Chen et al. \(2020\)](#) labels, and we rounded these percentages to integer values.

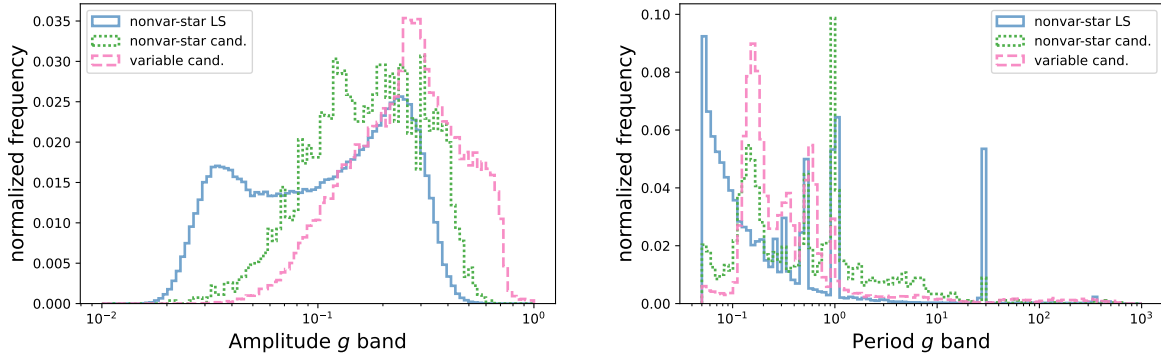


Fig. 20. Variability properties of variable star candidates from [Chen et al. \(2020\)](#), classified as Variable (pink) and non-variable (green) by our model. The left panel shows the **Amplitude** distribution, and the right panel the **Period** distribution. Sources from the LS classified as nonvar-star are included for reference (blue).

respectively. In Fig. 19 we compare the results obtained by [Chen et al. \(2020\)](#) with ours. We can see that there is broad agreement in the classification of the sources classified as variable by our model, when the classes considered are present in both models (i.e., EA, EB/EW, LPV, RRL, DSCT, and CEP). We can also see that very few of their objects are classified as transient or stochastic by our model, as expected since their catalog does not include this kind of variability. Although around 6% of their semi-regular (SR) variables are classified as AGN by our model (65 of them are known AGN from the literature). The largest discrepancy by far, however, is the number of objects that are classified as variable stars by them but non-variable stars by us (including several of their BY Dra and RS CVn variables in this field). In order to understand the origin of this discrepancy, we show in Fig. 20 the **Amplitude** and the **Period** of the common sources in the g band, separating them according to our classification into nonvar-star and variable. Sources from the LS, classified as nonvar-star are also included for reference. A similar distribution is observed in the r band. We can see that sources classified as nonvar-star have lower **Amplitude** values compared to the ones classified as variable, which could explain why we do not classify them as variables with our model. We can also see that we are missing sources with **Amplitude** larger than 0.1, probably due to the

noise of the ZTF DR11 light curves, since known nonvar-stars from the LS also show amplitudes larger than 0.1. The **Period** distribution of the nonvar-star and the variable candidates is also different, with many of the sources classified as non-variable by our model, having periods of around one day. [Chen et al. \(2020\)](#) did a very good job in ensuring that their periods were correct, including masking out obvious aliases (0.5 day, 1 day, 29.5 days, etc.), so their classification of low-amplitude sources should be better than ours.

Finally, we decided to exclude from this comparison the works presented by [van Roestel et al. \(2021\)](#) and [Aleo et al. \(2022\)](#), as they do not provide full sky catalogs. We also decided to exclude from the analysis a comparison with works that used other surveys (such as *Gaia* or CRTS), since several of their candidates were included in our training set (see Sect. 3.2.3).

8. Conclusions

In this work we present a hierarchical classifier designed to process ZTF DR11 PSF light curves (measured over the science images for objects detected in the reference images) of both extended and point sources. To the best of our knowledge, this is the first attempt to classify light curves of extended sources using ZTF DR light curves, which are not optimal for the study

of the variability of extended objects. In this work, however, we demonstrate that these light curves can still be used to identify variable extended objects when a suitable set of features is available.

The main purpose of our classifier is to select AGN and Blazars at different redshifts from their variability, which will be spectroscopically followed up on by the 4MOST survey ChANGES. The main goal of ChANGES is to target an unrivaled legacy sample of AGN with 4MOST, selected from several existing surveys (including ZTF). Although this was the main motivation for developing this classifier, inspired by our previous experience with the ALERCE broker light curve classifier (SS21b), we decided to make a more general classifier that considers different classes of transients, persistently variable objects, and non-variable objects. The model considers a total of 17 classes, with two non-variable, three transient, five stochastic, and seven periodic classes.

The model was designed in a hierarchical fashion, following the nature of the classes considered. We used a local classifier per parent node approach, where each node is composed of a BRF classifier. For this reason, we call our model the HBRF classifier. We trained two versions of this model, one that uses 61 variability features computed using the ZTF g band and another that uses the same number of features but measured from the r -band light curves. We also included ten additional features, which correspond to nine CatWISE and PS1 colors and one PS1 morphology score. We decided not to include as features the coordinates of the sources, to avoid the position-dependent biases present in the LS. The 61 variability features used in this work are taken from the detection features used by SS21b, with multiband features excluded. We also modified some of the features to properly deal with DR light curves. The LS was constructed using catalogs taken from SS21b, but additional catalogs were included, notably for the non-variable classes.

We notice that, in general, better results are obtained when using the ZTF g band. We propose that this is due to the intensive cadence campaigns conducted by ZTF in the r band that skew the feature distributions and produce a large number of candidates in some classes, particularly CV/Nova, high- z -AGN, and YSO. Therefore, we recommend that any user of this model give priority to the results obtained using the g band. When using the classifications obtained in the r band, we recommend filtering the catalogs by the classification probability in the node_init ($P_{\text{init}} \geq 0.9$) and/or avoiding the use of the model in regions with rapid cadence (using sources with a number of epochs lower than 400 in DR11 light curves). Finally, when filtering the classifications by probability, we suggest taking into account the probability distributions per class presented in Sect. 6 and the reliability diagrams discussed in Sect. 5.4.

Using this model, we were able to identify 384 242 unique AGN and Blazar candidates (within a radius of $1.5''$) in the g band and 4048 299 candidates in the r band, with 356 631 of these candidates being classified as AGN or Blazar in both bands. Considering the issues in the selection done with the r band, for the selection of sources to be observed by ChANGES, we decided to give priority to the candidates selected in the g band.

We expect to include more classes in a future version of this classifier. In particular, with the growing sample of known tidal disruption events (e.g., van Velzen et al. 2021; Hammerstein et al. 2023), we can start testing new models that include them as a class in the transients node. We also plan to expand the periodic node to new classes and subclasses of periodic variable stars, including, for instance, rotational classes. Furthermore, future

versions of this classifier can take advantage of new models that use light curves directly, without the need of a feature extraction procedure (e.g., Becker et al. 2020; Donoso-Oliva et al. 2023; Pimentel et al. 2023; Astorga et al. 2023).

Acknowledgements. The authors acknowledge support from the National Agency for Research and Development (ANID) grants: Millennium Science Initiative Program ICN12_12009 (PSS,FEB,GCV,MC,PAE,FF,PH,LHG,RK,AMMA,GP), and NCN19_058 (PA, PL); BASAL Center of Mathematical Modelling Grant PAI AFB-170001 (FF,AMMA); Basal CATA FB210003 (FEB,MC); FONDECYT Regular 1190818 (FEB), 1200495 (FEB), 1220829 (PAE), 1200710 (FF), 1211374 (PH), 1201748 (PL); FONDECYT Initiation 11191130 (GCV); FONDECYT Postdoctorado 3200250 (PSS); and from the Max-Planck Society through a Partner Group grant (PA). This work has been possible thanks to the use of AWS-U.Chile-NLHPC credits. Powered@NLHPC: This research was partially supported by the supercomputing infrastructure of the NLHPC (ECM-02). Based on observations obtained with the Samuel Oschin 48-inch Telescope at the Palomar Observatory as part of the Zwicky Transient Facility project. ZTF 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 Elektronen-Synchrotron and Humboldt University, Los Alamos National Laboratories, the TANGO Consortium of Taiwan, the University of Wisconsin at Milwaukee, and Lawrence Berkeley National Laboratories. Operations are conducted by COO, IPAC, and UW. Based on observations obtained with the Samuel Oschin Telescope 48-inch and the 60-inch Telescope at the Palomar Observatory as part of the Zwicky Transient Facility project. Z.T.F. is supported by the National Science Foundation under Grants No. AST-1440341 and AST-2034437 and a collaboration including current partners Caltech, IPAC, the Weizmann Institute for Science, the Oskar Klein Center at Stockholm University, the University of Maryland, Deutsches Elektronen-Synchrotron and Humboldt University, the TANGO Consortium of Taiwan, the University of Wisconsin at Milwaukee, Trinity College Dublin, Lawrence Livermore National Laboratories, IN2P3, University of Warwick, Ruhr University Bochum, Northwestern University and former partners the University of Washington, Los Alamos National Laboratories, and Lawrence Berkeley National Laboratories. Operations are conducted by COO, IPAC, and UW. 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. We acknowledge with thanks the variable star observations from the AAVSO International Database contributed by observers worldwide and used in this research.

References

- Abril, J., Schmidtbreick, L., Ederoclite, A., & López-Sanjuan, C. 2020, *MNRAS*, **492**, L40
- Aleo, P., Malanchev, K., Pruzhinskaya, M., et al. 2022, *New Astron.*, **96**, 101846
- Ansdell, M., Oelkers, R. J., Rodriguez, J. E., et al. 2018, *MNRAS*, **473**, 1231
- Astorga, N., Reyes-Jainaga, I., Cabrera-Vives, M., et al. 2023, *Nat. Astron.*, submitted
- Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, *PASP*, **93**, 5
- Becker, I., Pichara, K., Catelan, M., et al. 2020, *MNRAS*, **493**, 2981
- Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2019, *PASP*, **131**, 018002
- Bentley, J. L. 1975, *Commun. ACM*, **18**, 509
- Bhardwaj, A., Panwar, N., Herczeg, G. J., Chen, W. P., & Singh, H. P. 2019, *A&A*, **627**, A135
- Brahem, M., Yeh, L., & Zeitouni, K. 2018, in Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, SIGSPATIAL '18 (New York: Association for Computing Machinery), 229
- Carrasco-Davis, R., Reyes, E., Valenzuela, C., et al. 2021, *AJ*, **162**, 231
- Catelan, M., & Smith, H. A. 2015, *Pulsating Stars* (Weinheim: Wiley-VCH)
- Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv e-prints [arXiv:1612.05560]
- Chen, C., Liaw, A., Breiman, L., et al. 2004, *University of California, Berkeley*, **110**, 24
- Chen, X., Wang, S., Deng, L., et al. 2020, *ApJS*, **249**, 18
- Chen, Z. H., Yan, L., Kangas, T., et al. 2023, *ApJ*, **943**, 41
- de Jong, R. S., Agertz, O., Berbel, A. A., et al. 2019, *The Messenger*, **175**, 3
- Donoso-Oliva, C., Becker, I., Protopapas, P., et al. 2023, *A&A*, **670**, A54
- Drake, A. J., Graham, M. J., Djorgovski, S. G., et al. 2014, *ApJS*, **213**, 9

- Drake, A. J., Djorgovski, S. G., Catelan, M., et al. 2017, *MNRAS*, **469**, 3688
- Eisenhardt, P. R. M., Marocco, F., Fowler, J. W., et al. 2020, *ApJS*, **247**, 69
- Elliott, P., Bayo, A., Melo, C. H. F., et al. 2016, *A&A*, **590**, A13
- Flesch, E. W. 2019, arXiv e-prints [arXiv:1912.05614]
- Förster, F., Cabrera-Vives, G., Castillo-Navarrete, E., et al. 2021, *AJ*, **161**, 242
- Gaia Collaboration (Prusti, T., et al.) 2016, *A&A*, **595**, A1
- Gaia Collaboration (Brown, A. G. A., et al.) 2021, *A&A*, **649**, A1
- Gaia Collaboration (Vallenari, A., et al.) 2023, *A&A*, **674**, A1
- Günther, H. M., Cody, A. M., Covey, K. R., et al. 2014, *AJ*, **148**, 122
- Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. 2017, in Proceedings of the 34th International Conference on Machine Learning, ICML'17 (JMLR.org), 70, 1321
- Hammerstein, E., van Velzen, S., Gezari, S., et al. 2023, *ApJ*, **942**, 9
- Huijse, P., Estévez, P. A., Förster, F., et al. 2018, *ApJS*, **236**, 12
- Ivezić, Ž., Smith, J. A., Miknaitis, G., et al. 2007, *AJ*, **134**, 973
- Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, **873**, 111
- Jayasinghe, T., Kochanek, C. S., Stanek, K. Z., et al. 2018, *MNRAS*, **477**, 3145
- Jayasinghe, T., Stanek, K. Z., Kochanek, C. S., et al. 2019a, *MNRAS*, **485**, 961
- Jayasinghe, T., Stanek, K. Z., Kochanek, C. S., et al. 2019b, *MNRAS*, **486**, 1907
- Jayasinghe, T., Stanek, K. Z., Kochanek, C. S., et al. 2020, *MNRAS*, **491**, 13
- Kuhn, M. A., de Souza, R. S., Krone-Martins, A., et al. 2021, *ApJS*, **254**, 33
- Lakeland, B. S., & Naylor, T. 2022, *MNRAS*, **514**, 2736
- Lemaître, G., Nogueira, F., & Aridas, C. K. 2017, *J. Mach. Learn. Res.*, **18**, 1
- Leoni, M., Ishida, E. E. O., Peloton, J., & Möller, A. 2022, *A&A*, **663**, A13
- Lomb, N. R. 1976, *Ap&SS*, **39**, 447
- López-Navas, E., Martínez-Aldama, M. L., Bernal, S., et al. 2022, *MNRAS*, **513**, L57
- López-Navas, E., Arévalo, P., Bernal, S., et al. 2023, *MNRAS*, **518**, 1531
- Mahabal, A., Sheth, K., Gieseke, F., et al. 2017, in 2017 IEEE Symposium Series on Computational Intelligence (SSCI), 1
- Malanchev, K. L., Pruzhinskaya, M. V., Korolev, V. S., et al. 2021, *MNRAS*, **502**, 5147
- Marocco, F., Eisenhardt, P. R. M., Fowler, J. W., et al. 2021, *ApJS*, **253**, 8
- Masci, F. J., Laher, R. R., Rusholme, B., et al. 2019, *PASP*, **131**, 018003
- Massaro, E., Maselli, A., Leto, C., et al. 2015, *Ap&SS*, **357**, 75
- Miranda, N., Freytag, J. C., Nordin, J., et al. 2022, *A&A*, **665**, A99
- Möller, A., Peloton, J., Ishida, E. E. O., et al. 2021, *MNRAS*, **501**, 3272
- Mowlavi, N., Lecoeur-Taïbi, I., Lebzelter, T., et al. 2018, *A&A*, **618**, A58
- Mowlavi, N., Rimoldini, L., Evans, D. W., et al. 2021, *A&A*, **648**, A44
- Naeini, M. P., Cooper, G. F., & Hauskrecht, M. 2015, in Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI'15 (AAAI Press), 2901
- Naik, A., & Rangwala, H. 2018, Large-Scale Hierarchical Classification with Feature Selection, 61
- Narayan, G., Zaidi, T., Soraisam, M. D., et al. 2018, *ApJS*, **236**, 9
- Niculescu-Mizil, A., & Caruana, R. 2005, in Proceedings of the 22nd International Conference on Machine Learning, ICML '05 (New York, NY, USA: Association for Computing Machinery), 625
- Nordin, J., Brinnel, V., van Santen, J., et al. 2019, *A&A*, **631**, A147
- Oh, K., Yi, S. K., Schawinski, K., et al. 2015, *ApJS*, **219**, 1
- Palaversa, L., Ivezić, Ž., Eyer, L., et al. 2013, *AJ*, **146**, 101
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *J. Mach. Learn. Res.*, **12**, 2825
- Pimentel, Ó., Estévez, P. A., & Förster, F. 2023, *AJ*, **165**, 18
- Platt, J. 2000, in *Advances in Large Margin Classifiers* (USA: MIT Press)
- Poppenhaeger, K., Cody, A. M., Covey, K. R., et al. 2015, *AJ*, **150**, 118
- Rebull, L. M., Stauffer, J. R., Cody, A. M., et al. 2015, *AJ*, **150**, 175
- Rebull, L. M., Stauffer, J. R., Hillenbrand, L. A., et al. 2017, *ApJ*, **839**, 92
- Rebull, L. M., Stauffer, J. R., Cody, A. M., et al. 2018, *AJ*, **155**, 196
- Rebull, L. M., Stauffer, J. R., Cody, A. M., et al. 2020, *AJ*, **159**, 273
- Riello, M., De Angeli, F., Evans, D. W., et al. 2021, *A&A*, **649**, A3
- Rimoldini, L., Holl, B., Audard, M., et al. 2019, *A&A*, **625**, A97
- Ritter, H., & Kolb, U. 2003, *A&A*, **404**, 301
- Rodríguez, J. E., Ansdell, M., Oelkers, R. J., et al. 2017, *ApJ*, **848**, 97
- Roquette, J., Alencar, S. H. P., Bouvier, J., Guarcello, M. G., & Reipurth, B. 2020, *A&A*, **640**, A128
- Sánchez-Sáez, P., Lira, H., Martí, L., et al. 2021a, *AJ*, **162**, 206
- Sánchez-Sáez, P., Reyes, I., Valenzuela, C., et al. 2021b, *AJ*, **161**, 141
- Scargle, J. D. 1982, *ApJ*, **263**, 835
- Sergison, D. J., Naylor, T., Littlefair, S. P., Bell, C. P. M., & Williams, C. D. H. 2020, *MNRAS*, **491**, 5035
- Silla, C. N., & Freitas, A. A. 2011, *Data Mining Knowledge Discovery*, **22**, 31
- Smith, K. W., Williams, R. D., Young, D. R., et al. 2019, *Res. Notes Am. Astron. Soc.*, **3**, 26
- Tachibana, Y., & Miller, A. A. 2018, *PASP*, **130**, 128001
- Taylor, M. B. 2005, *ASP Conf. Ser.*, **347**, 29
- Thanjavur, K., Ivezić, Ž., Allam, S. S., et al. 2021, *MNRAS*, **505**, 5941
- Thomas, D., Steele, O., Maraston, C., et al. 2013, *MNRAS*, **431**, 1383
- Tonry, J., Stalder, B., Denneau, L., et al. 2018, *Transient Name Server Discovery Report*, 2018–1123, 1
- Torres, C. A. O., Quast, G. R., da Silva, L., et al. 2006, *A&A*, **460**, 695
- van Roestel, J., Duev, D. A., Mahabal, A. A., et al. 2021, *AJ*, **161**, 267
- van Velzen, S., Gezari, S., Hammerstein, E., et al. 2021, *ApJ*, **908**, 4
- Villar, V. A., Berger, E., Miller, G., et al. 2019, *ApJ*, **884**, 83
- Vioque, M., Oudmaijer, R. D., Schreiner, M., et al. 2020, *A&A*, **638**, A21
- Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, *A&AS*, **143**, 9
- Wolk, S. J., Günther, H. M., Poppenhaeger, K., et al. 2015, *AJ*, **150**, 145
- York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, *AJ*, **120**, 1579
- Zadrozny, B., & Elkan, C. 2001, in *Proceedings of the Eighteenth International Conference on Machine Learning, ICML '01* (San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.), 609
- Zadrozny, B., & Elkan, C. 2002, in *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '02* (New York, NY, USA: Association for Computing Machinery), 694
- Zaharia, M., Xin, R. S., Wendell, P., et al. 2016, *Commun. ACM*, **59**, 56
- Zúñiga-Fernández, S., Bayo, A., Elliott, P., et al. 2021, *A&A*, **645**, A30

Appendix A: Sky densities per class

The following figures (from A.1 to A.15) show the target sky densities per class, in the Galactic coordinate space.

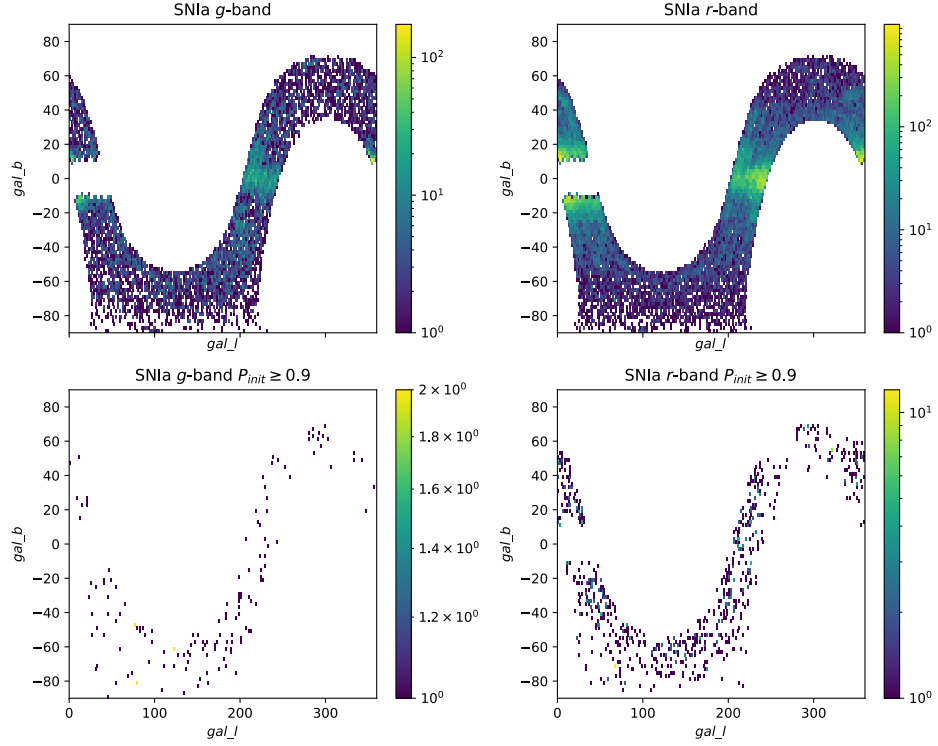


Fig. A.1. As in Fig. 11 but for the SNIa class.

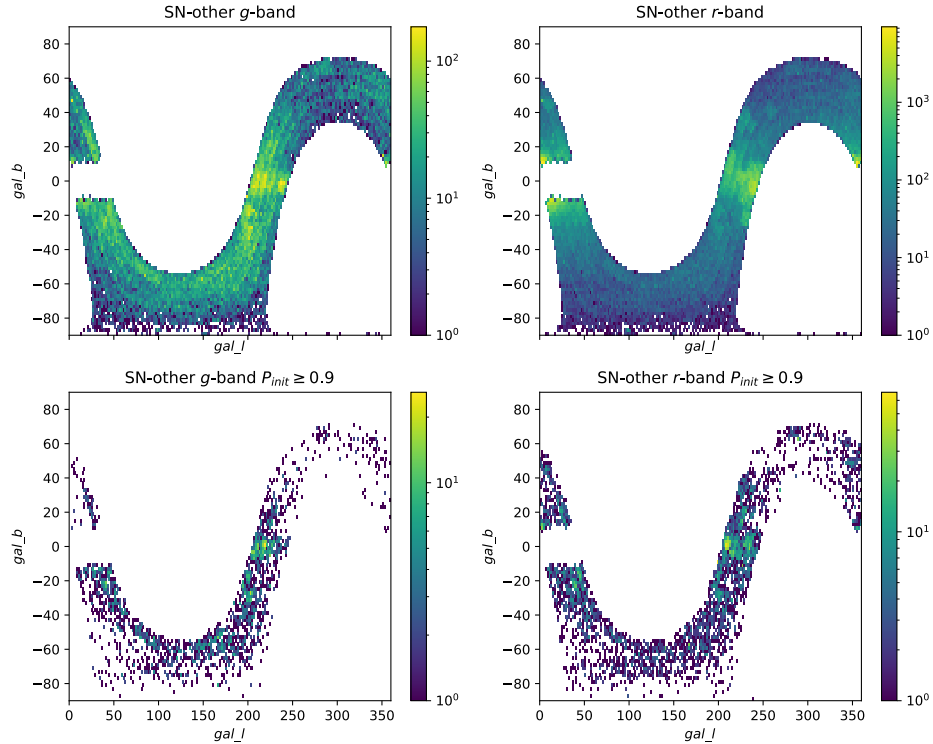


Fig. A.2. As in Fig. 11 but for the SN-other class.

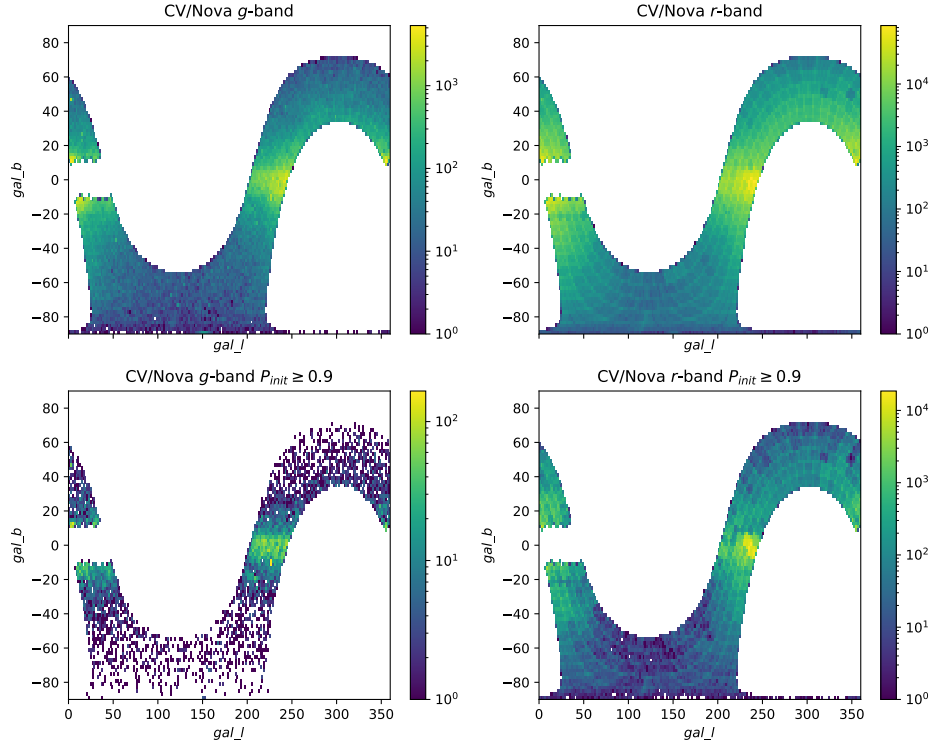


Fig. A.3. As in Fig. 11 but for the CV/Nova class.

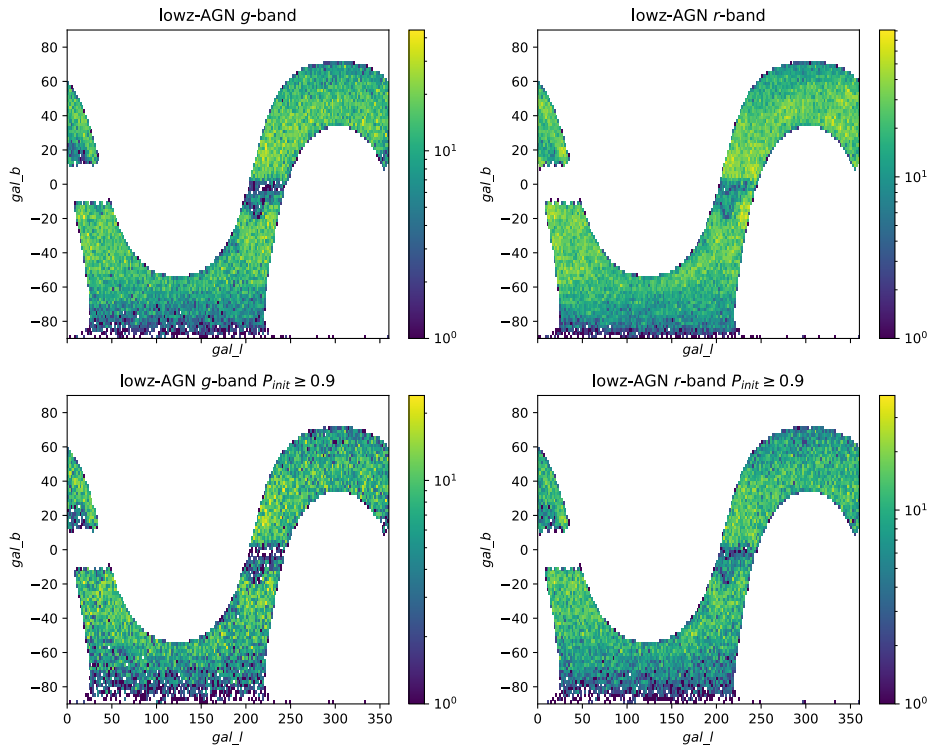


Fig. A.4. As in Fig. 11 but for the lowz-AGN class.

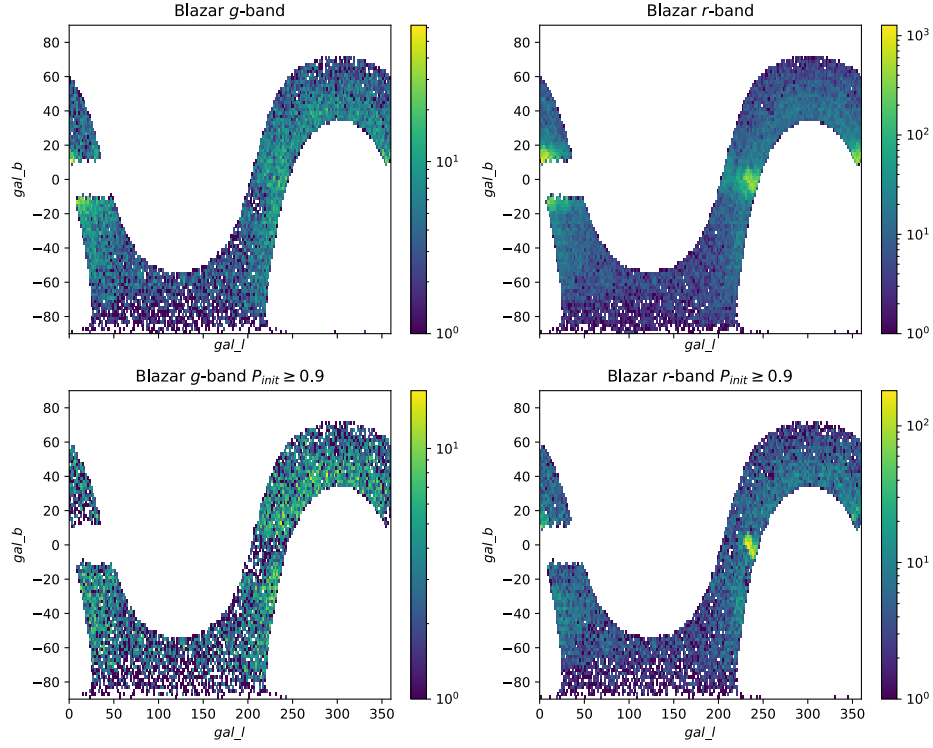


Fig. A.5. As in Fig. 11 but for the Blazar class.

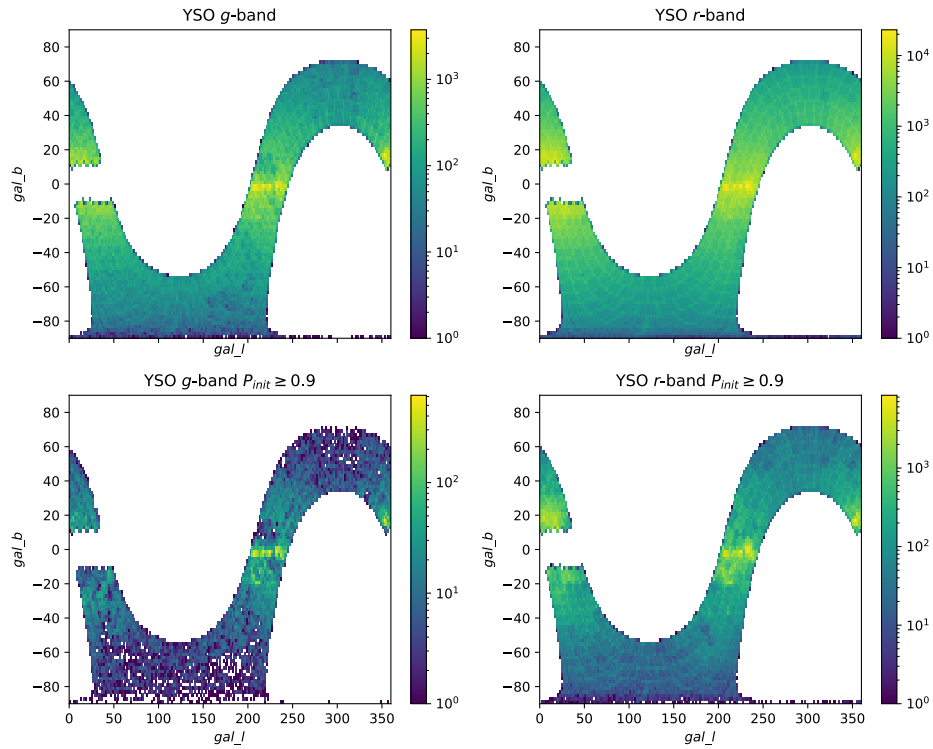


Fig. A.6. As in Fig. 11 but for the YSO class.

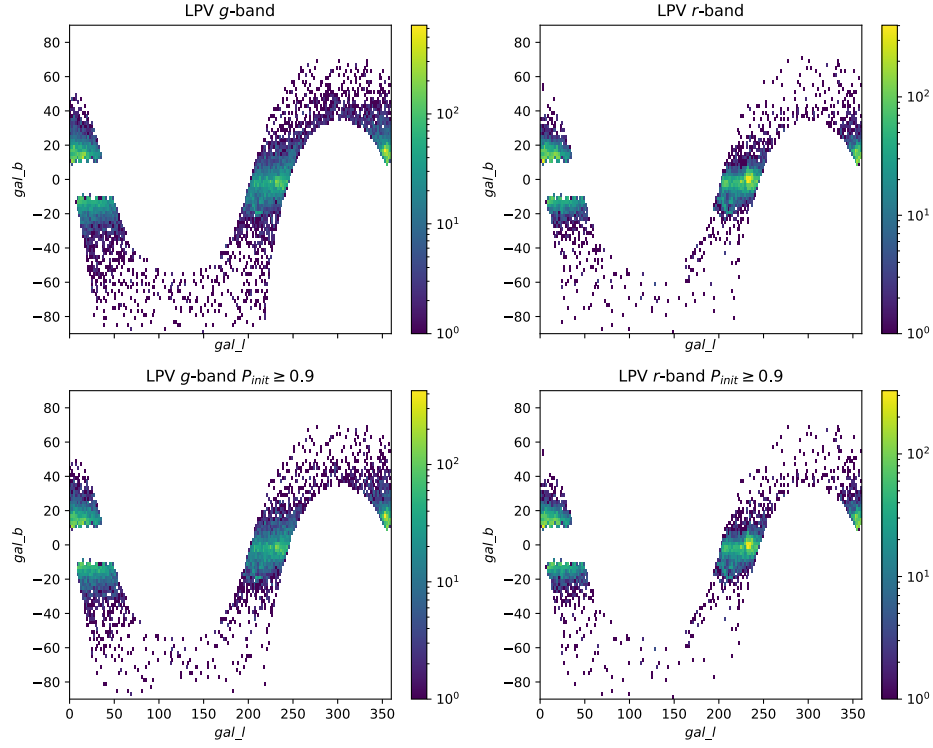


Fig. A.7. As in Fig. 11 but for the LPV class.

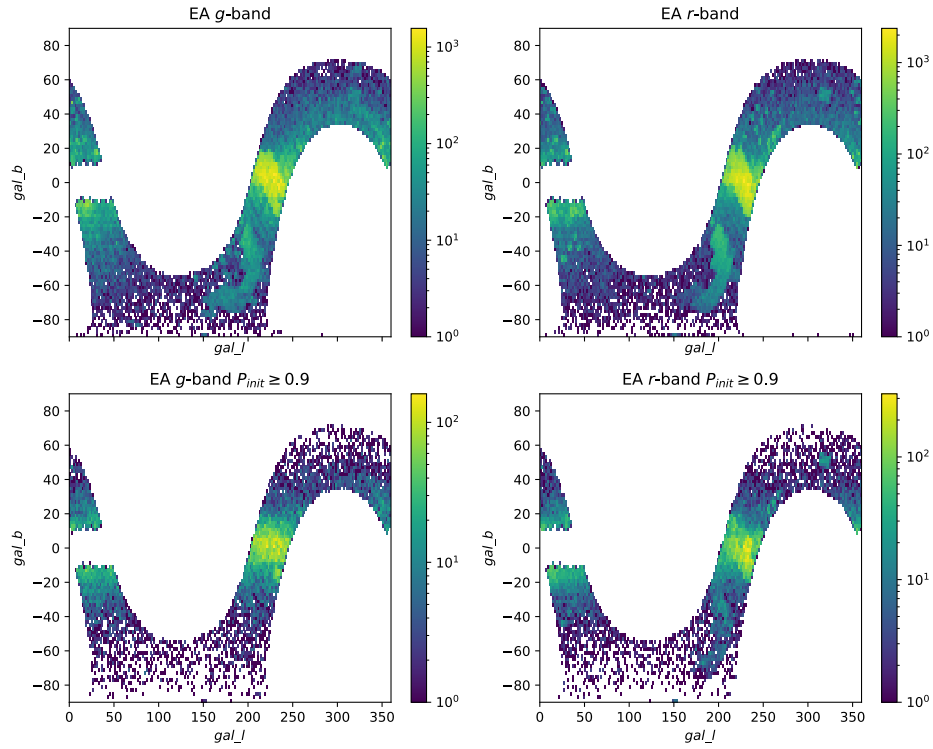


Fig. A.8. As in Fig. 11 but for the EA class.

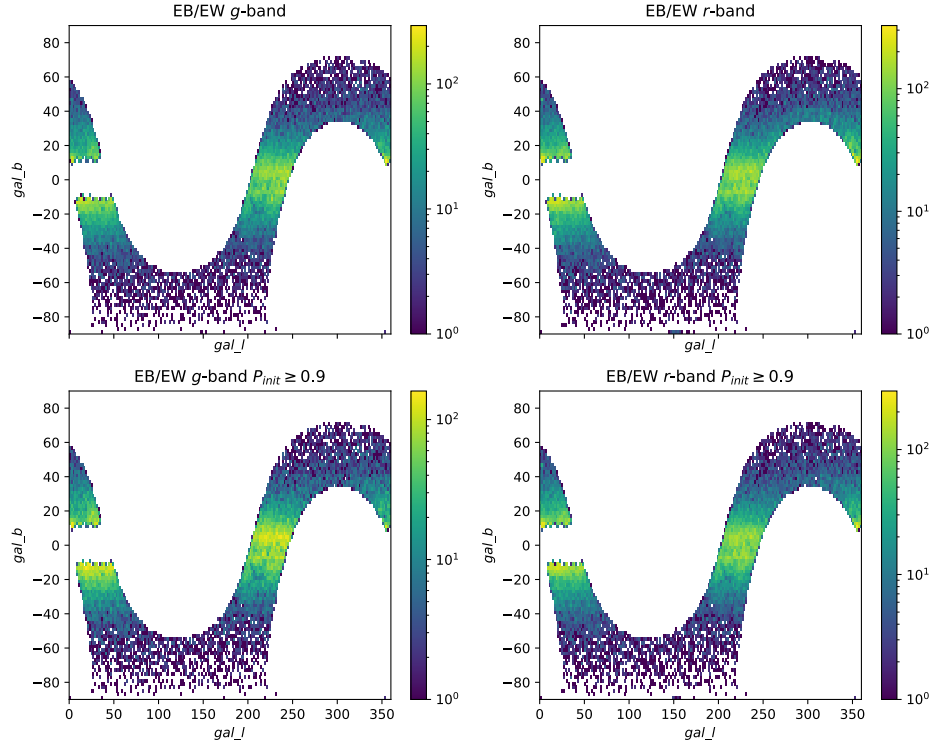


Fig. A.9. As in Fig. 11 but for the EB/EW class.

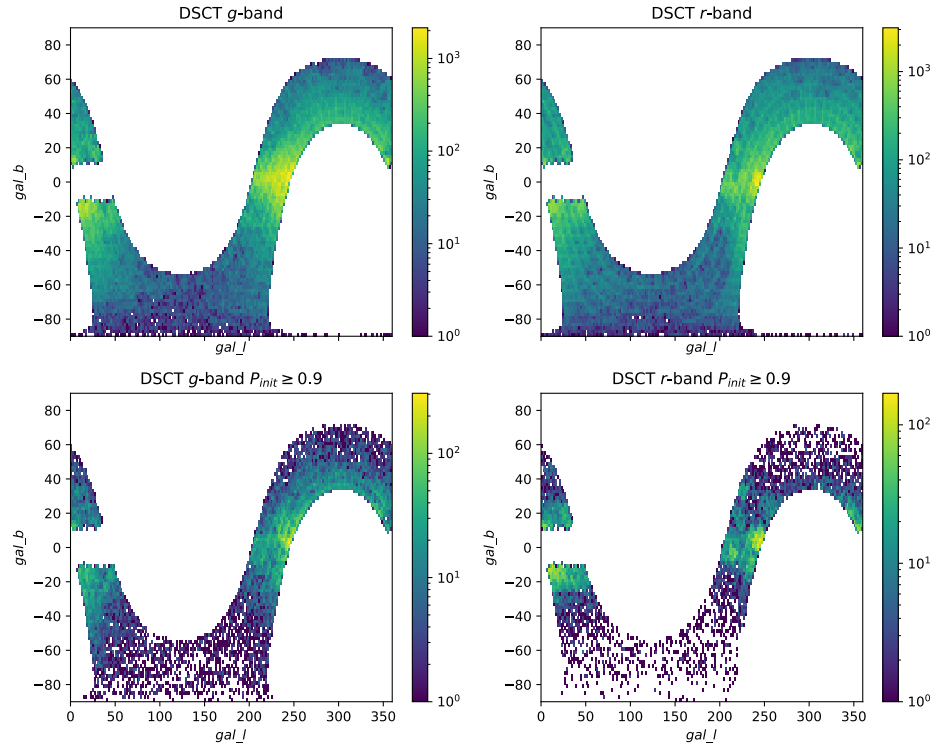


Fig. A.10. As in Fig. 11 but for the DSCT class.

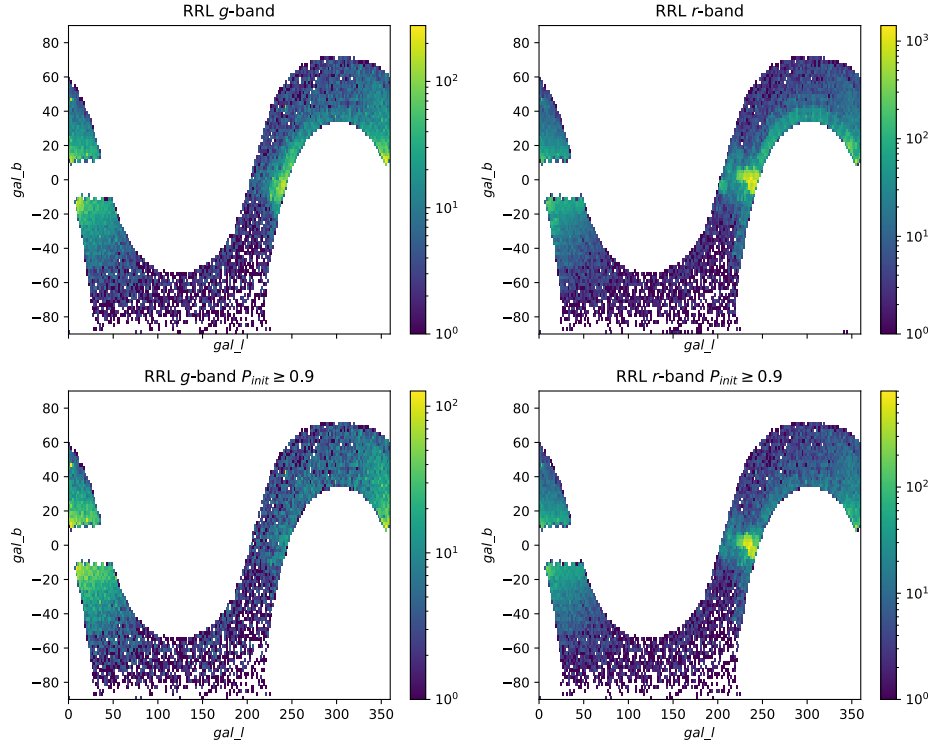


Fig. A.11. As in Fig. 11 but for the RRL class.

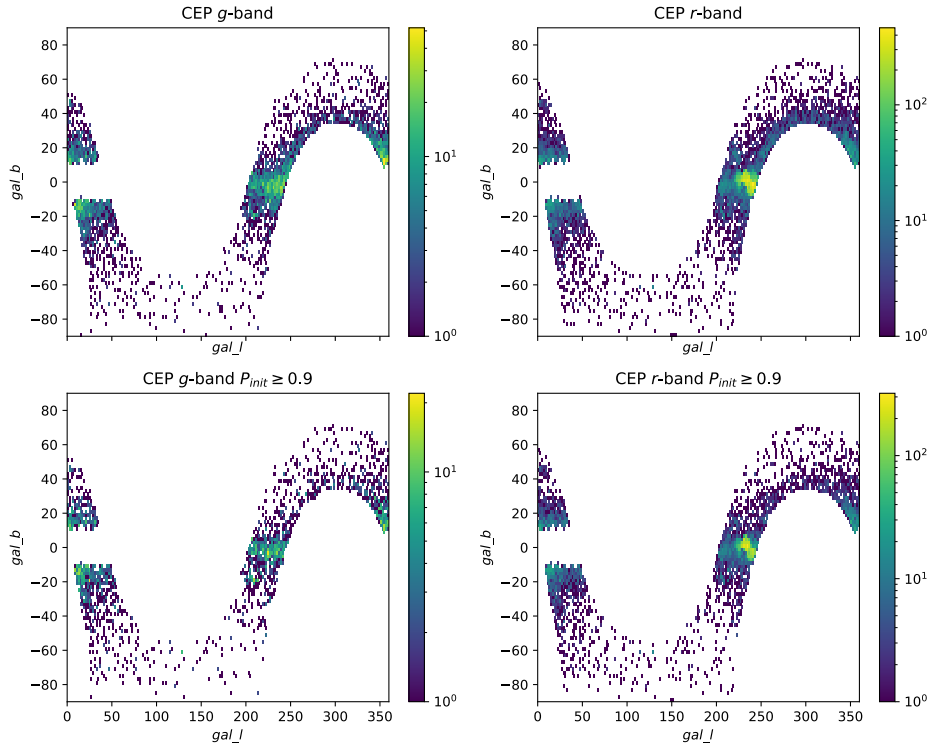


Fig. A.12. As in Fig. 11 but for the CEP class.

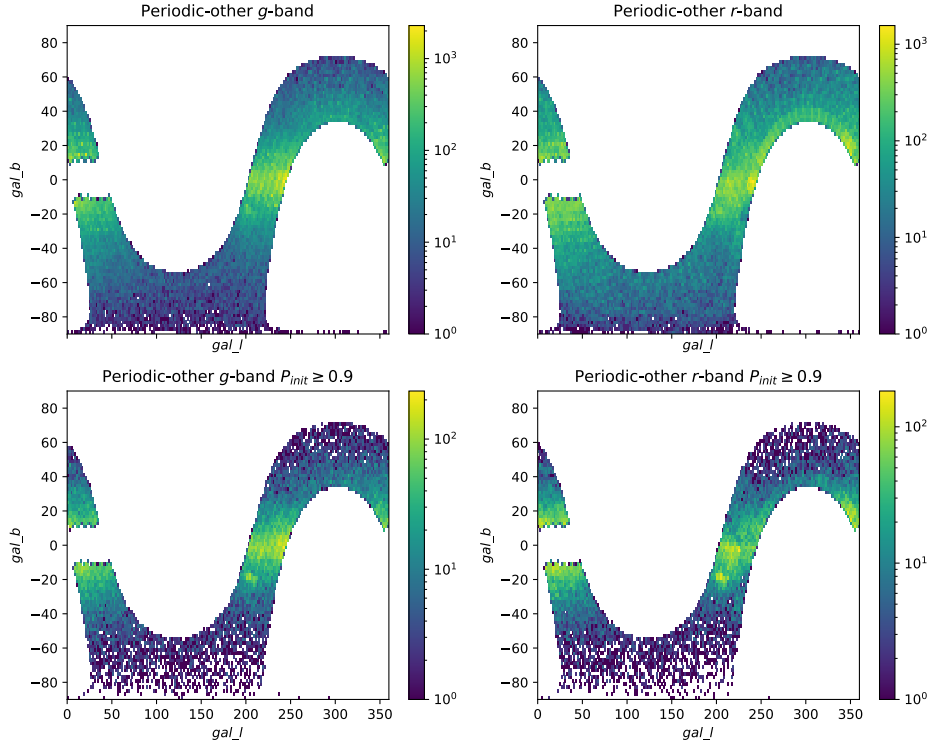


Fig. A.13. As in Fig. 11 but for the Periodic-other class.

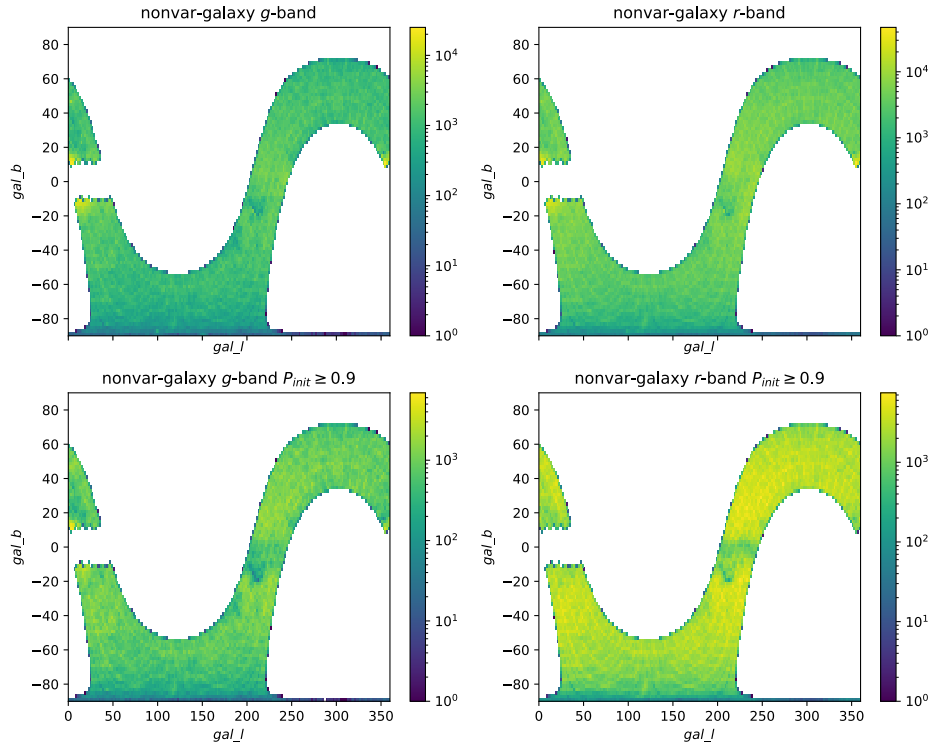


Fig. A.14. As in Fig. 11 but for the nonvar-galaxy class.

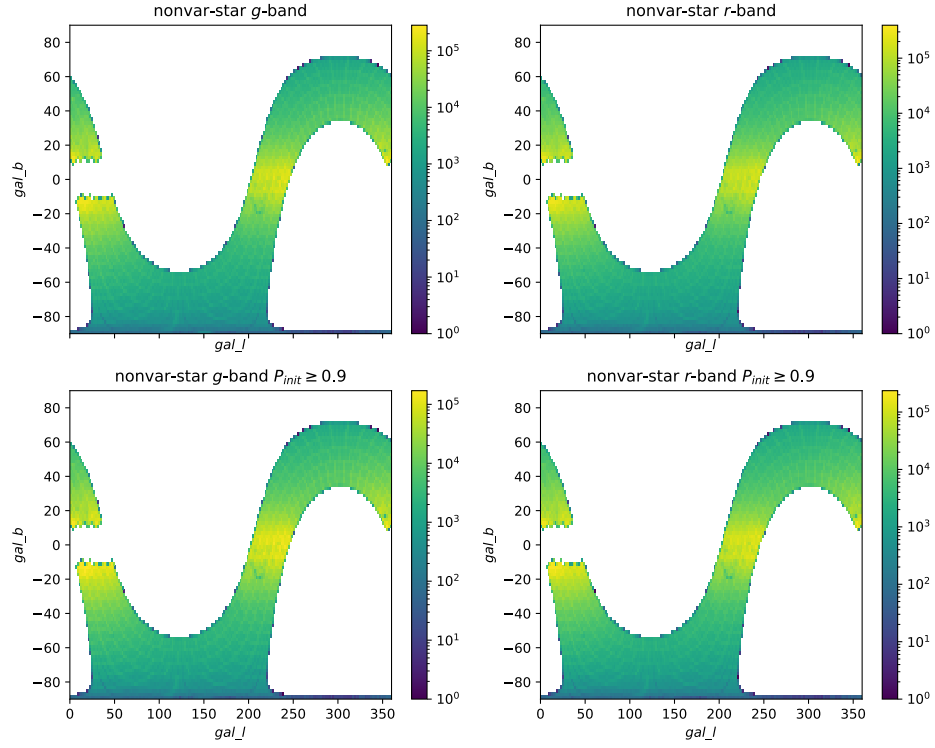


Fig. A.15. As in Fig. 11 but for the nonvar-star class.