



# Robust Measurement of Stellar Streams around the Milky Way: Correcting Spatially Variable Observational Selection Effects in Optical Imaging Surveys

K.K. Boone<sup>1,2</sup> , P.S. Ferguson<sup>2,3</sup> , M. Tabbutt<sup>2</sup> , K. Bechtol<sup>2,4</sup> , T.-Y. Cheng<sup>5</sup> , A. Drlica-Wagner<sup>6,7,8</sup> ,  
 C. E. Martínez-Vázquez<sup>9</sup> , B. Mutlu-Pakdil<sup>10</sup> , T. M. C. Abbott<sup>9</sup> , M. Aguena<sup>11,12</sup> , O. Alves<sup>13</sup> , F. Andrade-Oliveira<sup>14</sup> ,  
 D. Bacon<sup>15</sup> , S. Bocquet<sup>16</sup> , D. Brooks<sup>17</sup> , R. Camilleri<sup>18</sup> , A. Carnero Rosell<sup>12,19,20</sup> , L. N. da Costa<sup>12</sup> ,  
 M. E. da Silva Pereira<sup>21</sup> , T. M. Davis<sup>18</sup> , J. De Vicente<sup>22</sup> , S. Desai<sup>23</sup> , P. Doel<sup>17</sup> , S. Everett<sup>24</sup> , B. Flaugher<sup>6</sup> ,  
 J. Frieman<sup>6,7,8</sup> , J. García-Bellido<sup>25</sup> , D. Gruen<sup>16</sup> , G. Gutierrez<sup>6</sup> , S. R. Hinton<sup>18</sup> , D. L. Hollowood<sup>26</sup> ,  
 K. Honscheid<sup>27,28</sup> , D. J. James<sup>29</sup> , K. Kuehn<sup>30,31</sup> , J. L. Marshall<sup>32</sup> , J. Mena-Fernández<sup>33</sup> , F. Menanteau<sup>34,35</sup> ,  
 R. Miquel<sup>36,37</sup> , J. Myles<sup>38</sup> , R. L. C. Ogando<sup>39</sup> , A. A. Plazas Malagón<sup>40,41</sup> , A. Porredon<sup>22,42</sup> ,  
 M. Rodríguez-Monroy<sup>25,43</sup> , E. Sanchez<sup>22</sup> , D. Sanchez Cid<sup>14,22</sup> , I. Sevilla-Noarbe<sup>22</sup> , M. Smith<sup>44</sup> , E. Suchyta<sup>45</sup> ,  
 M. E. C. Swanson<sup>34</sup> , V. Vikram<sup>46</sup> , and N. Weaverdyck<sup>47,48</sup>

(DES Collaboration)

<sup>1</sup> Department of Physics, Harvard University, Cambridge, MA 02138, USA; [kboone@g.harvard.edu](mailto:kboone@g.harvard.edu)<sup>2</sup> Physics Department, University of Wisconsin-Madison, 1150 University Avenue Madison, WI 53706, USA; [pferguso@uw.edu](mailto:pferguso@uw.edu)<sup>3</sup> DiRAC Institute and the Department of Astronomy, University of Washington, Seattle, WA, USA<sup>4</sup> Legacy Survey of Space and Time, 933 North Cherry Avenue, Tucson, AZ 85721, USA<sup>5</sup> Kapteyn Astronomical Institute, University of Groningen, Landleven 12 (Kapteynborg, 5419), 9747 AD Groningen, The Netherlands<sup>6</sup> Fermi National Accelerator Laboratory, PO Box 500, Batavia, IL 60510, USA<sup>7</sup> Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA<sup>8</sup> Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA<sup>9</sup> Cerro Tololo Inter-American Observatory, NSF's National Optical-Infrared Astronomy Research Laboratory, Casilla 603, La Serena, Chile<sup>10</sup> Department of Physics and Astronomy, Dartmouth College, Hanover, NH 03755, USA<sup>11</sup> INAF-Osservatorio Astronomico di Trieste, via G. B. Tiepolo 11, I-34143 Trieste, Italy<sup>12</sup> Laboratório Interinstitucional de e-Astronomia—LInEA, Av. Pastor Martin Luther King Jr., 126 Del Castilho, Nova América Offices, Torre 3000/sala 817 CEP: 20765-000, Brazil<sup>13</sup> Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA<sup>14</sup> Physik-Institut, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland<sup>15</sup> Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth, PO1 3FX, UK<sup>16</sup> University Observatory, LMU Faculty of Physics, Scheinerstr. 1, 81679 Munich, Germany<sup>17</sup> Department of Physics & Astronomy, University College London, Gower Street, London, WC1E 6BT, UK<sup>18</sup> School of Mathematics and Physics, University of Queensland, Brisbane, QLD 4072, Australia<sup>19</sup> Instituto de Astrofísica de Canarias, E-38205 La Laguna, Tenerife, Spain<sup>20</sup> Universidad de La Laguna, Dpto. Astrofísica, E-38206 La Laguna, Tenerife, Spain<sup>21</sup> Hamburger Sternwarte, Universität Hamburg, Gojenbergsweg 112, 21029 Hamburg, Germany<sup>22</sup> Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain<sup>23</sup> Department of Physics, IIT Hyderabad, Kandi, Telangana 502285, India<sup>24</sup> California Institute of Technology, 1200 East California Boulevard, MC 249-17, Pasadena, CA 91125, USA<sup>25</sup> Instituto de Física Teórica UAM/CSIC, Universidad Autónoma de Madrid, 28049 Madrid, Spain<sup>26</sup> Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA<sup>27</sup> Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA<sup>28</sup> Department of Physics, The Ohio State University, Columbus, OH 43210, USA<sup>29</sup> Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA<sup>30</sup> Australian Astronomical Optics, Macquarie University, North Ryde, NSW 2113, Australia<sup>31</sup> Lowell Observatory, 1400 Mars Hill Road, Flagstaff, AZ 86001, USA<sup>32</sup> George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, and Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA<sup>33</sup> Université Grenoble Alpes, CNRS, LPSC-IN2P3, 38000 Grenoble, France<sup>34</sup> Center for Astrophysical Surveys, National Center for Supercomputing Applications, 1205 West Clark Street, Urbana, IL 61801, USA<sup>35</sup> Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA<sup>36</sup> Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain<sup>37</sup> Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193, Bellaterra (Barcelona), Spain<sup>38</sup> Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA<sup>39</sup> Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil<sup>40</sup> Kavli Institute for Particle Astrophysics & Cosmology, PO Box 2450, Stanford University, Stanford, CA 94305, USA<sup>41</sup> SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA<sup>42</sup> Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute, German Centre for Cosmological Lensing, 44780 Bochum, Germany<sup>43</sup> Laboratoire de physique des 2 Infinis Irène Joliot-Curie, CNRS Université Paris-Saclay, Bât. 100, F-91405 Orsay Cedex, France<sup>44</sup> Physics Department, Lancaster University, Lancaster, LA1 4YB, UK<sup>45</sup> Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA<sup>46</sup> Department of Physics, Central University of Kerala, 93VR+RWF, CUK Rd, Kerala 671316, India<sup>47</sup> Berkeley Center for Cosmological Physics, Department of Physics, University of California, Berkeley, CA 94720, USA

<sup>48</sup>Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA  
 Received 2025 October 20; revised 2026 March 26; accepted 2026 March 27; published 2026 April 17

## Abstract

Observations of density variations in stellar streams are a promising probe of low-mass dark matter substructure in the Milky Way. However, survey systematics such as variations in seeing and sky brightness can also induce artificial fluctuations in the observed densities of known stellar streams. These variations arise because survey conditions affect both object detection and star–galaxy misclassification rates. To mitigate these effects, we use `Balrog` synthetic source injections in the Dark Energy Survey (DES) Y3 data to calculate detection rate variations and classification rates as functions of survey properties. We show that these rates are nearly separable with respect to survey properties and can be estimated with sufficient statistics from the synthetic catalogs. Applying these corrections reduces the standard deviation of relative detection rates across the DES footprint by a factor of 5, and our corrections significantly change the inferred linear density of the Phoenix stream when including faint objects. Additionally, for artificial streams with DES-like survey properties we are able to recover density power spectra with reduced bias. We also find that uncorrected power-spectrum results for Legacy Survey of Space and Time (LSST)-like data can be around 5 times more biased, highlighting the need for such corrections in future ground-based surveys.

*Unified Astronomy Thesaurus concepts:* [Stellar streams \(2166\)](#); [Cosmology \(343\)](#); [Sky surveys \(1464\)](#); [Milky Way dark matter halo \(1049\)](#); [Dark matter \(353\)](#)

## 1. Introduction

The fundamental nature of dark matter is an outstanding question in physics and astronomy. The standard model of dark energy plus cold dark matter ( $\Lambda$ CDM) fits data at large scales and predicts the existence of smaller dark matter subhalos in our Galaxy (e.g., J. S. Bullock & M. Boylan-Kolchin 2017; M. R. Buckley & A. H. G. Peter 2018; S. Chabanier et al. 2019; K. Bechtol et al. 2022). At these smaller scales the population of Milky Way satellites provides insight into the nature of dark matter down to the threshold of star formation ( $M \sim 10^8 M_\odot$ ), helping to constrain the microphysics of dark matter models and abundance of low mass subhalos. Currently, these observations are consistent with the standard model (e.g., P. Jethwa et al. 2018; E. O. Nadler et al. 2020; O. Newton et al. 2021; A. Dekker et al. 2022). To continue to stress test  $\Lambda$ CDM and detect the impact of fully dark subhalos, we must look to additional probes. In the far field, galaxy-scale gravitational lenses are expected to be sensitive to these dark subhalos down to a mass of  $\sim 10^7 M_\odot$ ; either through flux-ratio anomalies, image positions, or time delays (T. Treu 2010; S. Vegetti et al. 2024, and references therein). Additionally, a promising near-field probe of dark subhalos comes from stellar streams around the Milky Way.

Stellar streams are the tidally disrupting remnants of star clusters and satellite galaxies (H. J. Newberg & J. L. Carlin 2016). The discovery and characterization of these systems around the Milky Way has been enabled by wide-field astronomical surveys, initially through matched filter searches of photometric data (e.g., SDSS; M. Odenkirchen et al. 2001, Pan-STARRs; E. J. Bernard et al. 2014; C. J. Grillmair 2017, DES; N. Shipp et al. 2018), and subsequently using combinations of photometric, astrometric, and spectroscopic data (e.g., Gaia; K. Malhan et al. 2018, S5; T. S. Li et al. 2019). Currently, there are more than 120 identified stellar streams around our Galaxy (C. Mateu 2023; A. Bonaca & A. M. Price-Whelan 2025).

Within an individual stream, the stellar positions and velocities are probes of the local acceleration field experienced

by the stream stars as they orbit around the Milky Way’s gravitational potential (J. Bovy et al. 2017). Therefore, in a Milky Way potential absent of substructure, it is expected that streams would show less substructure; although some will be present due to variations in the stripping rate and the formation of epicyclic overdensities (A. H. W. Küpper et al. 2010). But the presence of dark subhalos passing by a stream will induce perturbations in the on-sky positions, density, and velocity distribution of stream stars. This can result in the formation of gaps and other small-scale deviations in the stream track (e.g., D. Erkal et al. 2016; A. Bonaca et al. 2019; N. Banik et al. 2021a, 2021b; M. S. Delos & F. Schmidt 2022).

The population of kinematically cold stellar streams (i.e., ones with globular cluster progenitors) comprise the best probe of this effect due to their small intrinsic velocity dispersions and stream widths (J. Lu et al. 2026). One such kinematically cold stellar stream is the Phoenix stream. This substructure, originally discovered in early DES data (E. Balbinot et al. 2016), has a length of 4.6 kpc ( $\sim 15^\circ$ ), a width of  $0.14$ , and a measured velocity dispersion of  $\sigma_v = 2.66 \text{ km s}^{-1}$  (N. Shipp et al. 2018; Z. Wan et al. 2020). Interestingly, this stream also shows small-scale density fluctuations, making it a promising candidate for probing the low mass regime of the subhalo mass function (K. Tavangar et al. 2022). Due to these characteristics, we use this stream as an example application in our analysis. For thin streams in general, there has been a large effort to follow up on discoveries using deeper imaging surveys to access fainter stars and provide the best constraints on the track and density variations (e.g., D. Erkal et al. 2017; S. E. Koposov et al. 2019; A. Bonaca et al. 2020; T. S. Li et al. 2021; J. M. Patrick et al. 2022; K. Tavangar et al. 2022), while proper motion measurements (N. Shipp et al. 2019) and targeted spectroscopy of brighter members (A. Bonaca et al. 2021; T. S. Li et al. 2022) can be used to refine the stellar samples, constrain dynamics and further characterize the interaction history of these substructures.

One common approach to quantify the observable effects of gravitational interactions with dark matter is through power-spectrum-type analyses of density variations (stars/degree) along a stream (N. Banik et al. 2021a; M. S. Delos & F. Schmidt 2022). Such inferences rely on accurate measurements of the density and track along the stream. At small spatial scales, shot noise from the limited number of stars



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

**Table 1**

These Are the Survey Properties Used in the Analysis, Along With Their Units and the Different Statistics Used

| Quantity                  | Units                                  | Statistics          |
|---------------------------|--|---------------------|
| <i>airmass</i>            | ...                                    | WMEAN<br>MIN<br>MAX |
| <i>fwhm</i>               | arcsec                                 | WMEAN<br>MIN<br>MAX |
| <i>fwhm_fluxrad</i>       | arcsec                                 | WMEAN<br>MIN<br>MAX |
| <i>exptime</i>            | seconds                                | SUM                 |
| <i>t_eff</i>              | ...                                    | WMEAN<br>MIN<br>MAX |
| <i>t_eff_exptime</i>      | seconds                                | SUM                 |
| <i>skybrite</i>           | e <sup>-</sup> /CCD pix                | WMEAN               |
| <i>skyvar</i>             | (e <sup>-</sup> /CCD pix) <sup>2</sup> | WMEAN<br>MIN<br>MAX |
| <i>skyvar_sqrt</i>        | e <sup>-</sup> /CCD pix                | WMEAN               |
| <i>skyvar_uncertainty</i> | e <sup>-</sup> /s-coadd pix            |                     |
| <i>sigma_mag_zero</i>     | mag                                    | QSUM                |
| <i>fgcm_gry</i>           | mag                                    | WMEAN<br>MIN        |
| <i>stellar_dens</i>       | stars/deg <sup>2</sup>                 | ...                 |

**Note.** Each property has maps in *griz* bands with the exception of *stellar\_dens* and *skyvar\_sqrt* (which was lacking an *r*-band map). Adding up all the statistics and bands across each quantity gives a total of 92 maps used.

dominates the stream density power spectrum, thus it is most important to reduce bias on measurements at larger scales (tens of degrees; J. Bovy et al. 2017).

In photometric imaging data, systematic biases can be produced by unaccounted differences between intrinsic and observed source populations, i.e., the observational selection function. The observational selection function depends on intrinsic source properties (e.g., flux in multiple photometric bands, surface brightness) together with external survey properties (see Table 1) that include both astrophysical effects (e.g., projected source density, interstellar extinction) and observational details (e.g., integrated exposure time) that generally vary over the survey footprint. Streams covering tens of degrees on the sky will span multiple telescope fields of view and are likely to be observed with a variety of survey properties.

In addition, for most optical surveys, the observations are taken over multiple epochs under different observing conditions. This produces variations in detection rates for both stars and galaxies in imaging data. It also causes variations in how often stars and galaxies are incorrectly classified. Already with DES Year 3 data, we find that spatially variable observational selection effects across the survey footprint induce statistically significant effects (Section 4). As our data and analyses become more sophisticated, percent-level variations in the

survey-transfer function can induce systematic errors that are much larger than the statistical errors and therefore limit the power of analyses (S. Everett et al. 2022).

Traditionally, to avoid these systematic biases, a selection for high S/N objects is applied to minimize variations in observational selection effects. For example, in the DES DR1 stream search, the stellar selection was limited to  $g < 23.5$  (N. Shipp et al. 2018). However, at the magnitudes that will be accessible to future surveys, such as the NSF-DOE Vera C. Rubin Observatory’s Legacy Survey of Space and Time (LSST;  $5\sigma$  *g*-band depth  $\sim 27.4$  corresponding to around 2.3 mag deeper than DES DR1; Z. Ivezić et al. 2019), galaxies are much more common, and even small variations in galaxy misclassification rates could produce large effects on the stellar sample (K. Tsiane et al. 2025). Therefore, new tools are needed to fully leverage the potential LSST-like data.

In the context of large-scale structure analyses for cosmology, galaxy weight maps have been derived to account for spatially variable observational selection effects that would otherwise produce systematic errors in galaxy clustering measurements. These weight maps are derived using the ansatz that galaxies are isotropically distributed on large angular scales and that empirical relations between observed galaxy densities and survey properties (e.g., seeing, integrated exposure time) can be used to learn the observational selection function for galaxies across the survey footprint (see Section 5.3 of M. Rodríguez-Monroy et al. 2022). This approach is not possible for stellar samples because the intrinsic distribution of stars across the survey footprint is highly nonisotropic.

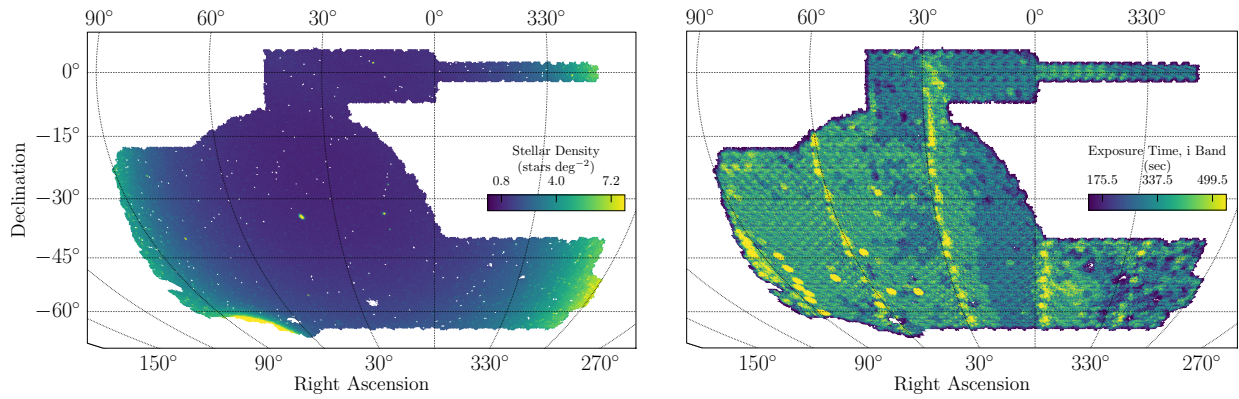
In this work, we explore the use of synthetic source injection (SSI) in combination with survey metadata (i.e., survey property maps) to correct stellar samples for spatially variable observational selection effects. The SSI pipeline inserts realistic artificial stars and galaxies directly into pixel-level image data and reruns source detection and measurement algorithms to effectively sample the observational selection function at locations across the survey footprint. Specifically, we use the `Balrog` implementation of SSI for DES Y3 as a testing ground to develop the methodology (S. Everett et al. 2022).

The paper is organized as follows: in Section 2 we discuss the data from DES Y3 and `Balrog` that will be used in our corrections. In Section 3 we discuss the calculations involved in our corrections. In Section 4 we apply our corrections to the full DES Y3 footprint and compare these results to precorrected data. In Section 5 we test the overall corrective power of our algorithm, how this power changes with larger training sets, and how repeatable our algorithm is. In Section 6 we apply our corrections to observations of the Phoenix stream and investigate the resulting changes in linear density. We then apply our corrections to simulated stellar streams and compare density power spectra. Finally, we conclude and motivate potential future works in Section 7. All code used in this project is available on github<sup>49</sup> and Zenodo (K. Boone 2026).

## 2. Data

In this section we describe the observations, DES Y3 Gold catalog (Section 2.1), and synthetic source injection runs, `Balrog` (Section 2.2), used in our analysis.

<sup>49</sup> [https://github.com/Kyle-Boone/ssi\\_corrections\\_des\\_y3\\_balrog](https://github.com/Kyle-Boone/ssi_corrections_des_y3_balrog)



**Figure 1.** Distributions of two of the survey properties used, stellar density (left) and exposure time sum in the  $i$  band (right).

### 2.1. DES DR1 and Y3 GOLD

We use data products from the first data release of DES (DR1; T. M. C. Abbott et al. 2018) based on 3 yr of observations using the Dark Energy Camera (B. Flaugher et al. 2015) mounted on the Blanco 4 m telescope at the Cerro Tololo Inter-American Observatory (CTIO). The survey data covers an area of  $\sim 5000 \text{ deg}^2$  in five broadband filters,  $grizY$ . With 38,850 total exposures in DES DR1, each location in the footprint typically contains four to six exposures in each band (H. T. Diehl et al. 2016), corresponding to a median depth of  $i \sim 23.3$  at signal-to-noise ratio ( $S/N$ ) = 10 for unresolved sources.

We also use object classifications, quality flags, and survey property maps from the value-added DES Y3 GOLD release (I. Sevilla-Noarbe et al. 2021). The survey property maps consist of spatial survey property values stored as HEALPix<sup>50</sup> (K. M. Górski et al. 2005) pixel (hereafter healpixel) maps. These survey property maps track spatial variations of observing conditions at a high resolution ( $N_{\text{SIDE}} = 4096$ ) and are described in more detail in Section 7.3 and Appendix E of I. Sevilla-Noarbe et al. (2021). Table 1 lists the survey properties that are used in this analysis. Examples of survey property maps for stellar density and exposure time in the  $i$  band are shown in Figure 1.

For the flags, to obtain a high quality sample of objects we base our selection on S. Everett et al. (2022; see their Section 4) and place the following cuts on all Y3 GOLD objects:

$$\begin{aligned} \text{FLAGS\_FOREGROUND} &= 0 \\ \text{AND } \text{FLAGS\_BADREGIONS} &< 2 \\ \text{AND } \text{FLAGS\_FOOTPRINT} &= 1. \end{aligned}$$

For the above cuts, the footprint cut is used to select regions that had good coverage in multiple observing bands, the foreground cut is used to remove regions near bright objects, and the bad region cut is used to remove high densities of anomalous colors and tiles where the multiobject fitting pipeline failed to finish. More details on the above cuts can be found in Sections 7.1 and 7.2 of I. Sevilla-Noarbe et al. (2021).

Object classification is done using single-object fitting (SOF) classification with `EXTENDED_CLASS_SOF`, which is the only classification provided within our synthetic source catalog. Stars are defined as objects with  $0 \leq \text{EXTENDED\_CLASS\_SOF} < 1$ , and galaxies are objects with  $2 \leq \text{EXTENDED\_CLASS\_SOF} \leq 3$ .

Object magnitudes are also incorporated in our methodology; in all cases we use the SOF magnitudes, which are

single-object multiepoch measurements described in I. Sevilla-Noarbe et al. (2021). However, in this work point-spread function (`SOE_PSF_MAG`) magnitudes are used for objects classified as stars, while composite model (`SOE_CM_MAG`) magnitudes (K. Abazajian et al. 2004) are used for objects classified as galaxies.

### 2.2. Balrog Synthetic Sources

To compliment the Y3 GOLD data, the synthetic source catalog generated by Balrog (E. Suchyta et al. 2016; S. Everett et al. 2022) is used. Balrog is a software package that synthetically injects sources into individual DES images before coaddition and processes them with the same DESDM pipelines as observations. Balrog is the name used to refer to both the software and synthetic object catalog, but for the remainder of this work we will use this name only to refer to the catalog. The injected sources were chosen to be empirical populations of artificial galaxies and stars taken from the deeper DES Deep Field observations (W. G. Hartley et al. 2021). Additionally, delta-function stars (delta functions convolved with a local PSF; Section 2.2.4 of S. Everett et al. 2022) are injected, which we use for our synthetic stellar sample to avoid any potential galaxy contamination from incorrect Deep Field classifications. These synthetic objects are subject to the same systematic impacts from survey properties as physical objects would be. For this work we use all Balrog objects that pass our quality cuts. This consists of 7.4M total objects (90% Deep Field objects, 10% delta stars). These objects were injected on a uniform grid on 2041 randomly chosen tiles out of the 10,338 total Y3 tiles (for a coverage map, see Figure 6 of S. Everett et al. 2022, and more information on the injected objects can be found in their Section 3).

In this work, Balrog objects are subject to the same flag cuts as described previously for Y3 GOLD objects. Classification cutoffs in terms of `EXTENDED_CLASS_SOF` are also the same. Likewise, PSF magnitudes are used for objects classified as stars while CM magnitudes are used for objects classified as galaxies. To be consistent with Y3 GOLD objects, measured magnitudes are used for objects instead of true magnitudes. Finally, a cut of `MATCH_FLAG_1.5_ASEC` < 2 is applied, which reduces the risk of ambiguous matching to Y3 GOLD objects within  $1''.5$ . More details on this flag cut can be found in Section 3.5 of S. Everett et al. (2022). With this synthetic source injector, we are able to perform corrections on DES data, which we turn to next.

<sup>50</sup> <http://healpix.sourceforge.net>

**Table 2**

This Is a Description of Notation and Terminology That Will Be Used Throughout This Paper

| Notation        | Terminology              | Definition  |
|-----------------|--------------------------|---|
| $-s$            |                          | Stars.  |
| $-G$            |                          | Galaxies.   |
| $I_-$           |                          | Injected objects.   |
| $T_S$           | True Stars               | True number of stars in a given area.   |
| $O_S$           | Observed Stars           | Count of detected true stars in a given area, regardless of given classification. |
| $C_S$           | Classified Stars         | Count of objects classified as stars in a given area.                             |
| $P(C_S O_S)$    |                          | Probability of giving an observed star a classification of star.                  |
| $RO_S$          | Recovered Observed Stars | Recovered count for observed stars after applying maximum likelihood separation.  |
| $D_R(C_S, O_S)$ |                          | Relative detection rate of observed stars classified as stars.                    |
| $FC_S$          | (Final) Corrected Stars  | Final star counts after applying corrections.                                     |

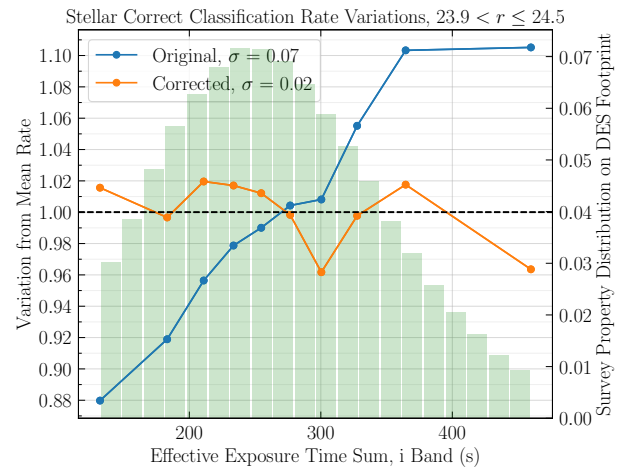
**Note.** Lesser used notation will be defined when used.

### 3. Methods

#### 3.1. Overview and Notation

Spatial variations in survey properties lead to correlated variations in the stellar selection function in two ways: (i) variations in the correct classification rate of detected objects and (ii) variations in the object detection rate. The notation used in this work is summarized in Table 2. In particular,  $T_S$  is the true number of stars, detected and undetected, in any area of the sky.  $O_S$  is the number of these true stars that are detected, and  $C_S$  is the number of objects that are classified as stars (this includes misclassified galaxies).

Our algorithm uses SSI objects and survey property maps to derive a relation between the survey properties and the likelihood that a given object will be detected and classified correctly. This will allow us to take catalog level data ( $C_S$  and  $C_G$ ) and correct it based on the survey properties at a given on-sky location. Our correction will be based on the assumption that the synthetic objects will respond in the same way to survey property variations as real data. For Balrog objects, this assumption is appropriate based on Figure 11 of S. Everett et al. (2022). Since we know whether Balrog objects are truly stars or galaxies, we can use them to measure the rates at which objects are classified correctly. This measurement is critical for getting high accuracy counts of stars at faint magnitudes since misclassified galaxies are the dominant contributor to uncorrected faint star counts.



**Figure 2.** This figure shows Balrog delta star correct classification rates before (blue) and after (orange) corrections as a function of the effective exposure time sum in the  $i$  band. The green histogram shows the distribution of effective exposure time in the  $i$  band across the DES footprint. Training was performed on an 80% subsample of the Balrog objects with the remaining sources used for testing. As well as magnitude cuts, a color cut to the Phoenix isochrone is made (see Section 4.1). A drop in standard deviation after corrections shows a mitigation of some of the variations caused by this survey property.

To start our correction, we therefore first use SSI objects to obtain the probability that a detected SSI star/galaxy will be classified correctly ( $P(IC_S|IO_S)$  and  $P(IC_G|IO_G)$ ). We assume that the probability that a detected star (galaxy) will be classified correctly is the same as that probability for an SSI star (galaxy):

$$P(C_S|O_S) = P(IC_S|IO_S). \quad (1)$$

These probabilities allow us to estimate the number of true stars (galaxies) that were detected  $RO_S$  ( $RO_G$ ) at each position. After obtaining  $RO_S$  and  $RO_G$ , we use the synthetic sources to estimate relative detection rates (for correctly classified stars,  $D_R(C_S, O_S)$  is the rate at which true stars  $T_S$  enter the stellar sample  $C_S$  relative to the average rate across the footprint). We find that correctly classified stars are subject to *distinct* variations in the relative detection rate compared to misclassified galaxies (see Appendix C), which necessitates the calculation of both for a full correction. Therefore, we calculate four relative detection rates:  $D_R(IC_S, IO_S)$ ,  $D_R(IC_G, IO_S)$ ,  $D_R(IC_S, IO_G)$ , and  $D_R(IC_G, IO_G)$ . As in Equation (1), we assume the relative detection rates for real objects in these groups are the same as for SSI objects.

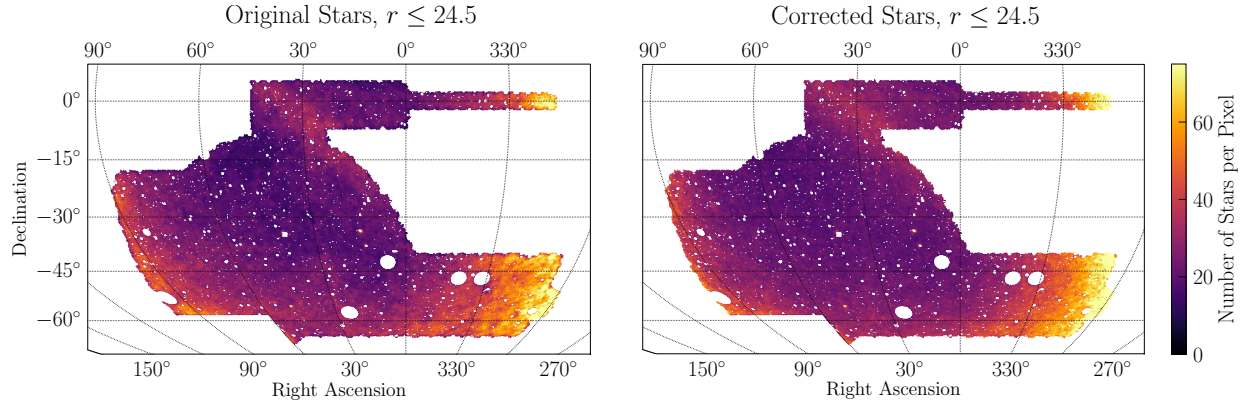
Classification probabilities and relative detection rates are used to obtain final corrected counts at each position ( $FC_S$  and  $FC_G$ ). We present two different types of corrections: one estimates what  $C_S$  would have been with uniform survey properties, and one estimates what  $O_S$  would have been with uniform survey properties. More details are given in Section 3.3.

For large surveys, the stellar magnitude distribution is not uniform over the full footprint. To account for this, we bin all sources based on magnitudes. All probabilities, detection rates, and corrections are calculated for each magnitude bin.

#### 3.2. Probability Calculations

This section outlines the calculation of  $P(C_S|O_S)$  and  $D_R(C_S, O_S)$ . A more detailed description can be found in Appendices A and B.

As a specific example to illustrate our method, we present  $P(IC_S|IO_S)$  in our faintest magnitude bin as a function of a



**Figure 3.** Star counts before and after corrections are applied. The matched filter described in Section 4.1 is used to crop for color and magnitude. Masked regions are primarily masked due to bright foreground objects, which is also true for future plots. For more physical units, at the above  $N_{\text{SIDE}} = 512$  each pixel has an area of  $\sim 47.2$  arcmin<sup>2</sup>. Notable improvements are the suppression of the depth feature at R.A.  $\sim 30^\circ$  and the better continuity at the edge of the footprint, both of which are more obvious in the galaxy plot below.

single survey property in Figure 2. The green histogram shows the distribution of effective exposure time in the  $i$  band for DES Y3. The blue (orange) line shows the ratio before (after) corrections of correctly classified SSI stars to all detected SSI stars as a function of the survey property value (Table 1). Since we only care about the variations in this rate, we divide each line by its average to center them at 1. For corrections, this ratio of correctly classified SSI stars to all detected SSI stars is computed for all survey properties. We then take the survey property with the largest variation in the ratio and apply an empirical correction. This process is repeated iteratively until one of our termination criteria is met: either  $\sigma < 0.01$  for all survey properties or, to avoid overfitting and/or long runtimes, after 150 iterations (see Section 5.3). To ensure that we are not just removing noise, we take 80% of our SSI stars to be a training sample and the remaining 20% to be a testing sample. All corrections were calculated on the training sample but then applied to the testing sample to make the plot. In this way, the drop in standard deviation from  $\sigma = 0.07$  to  $\sigma = 0.02$  indicates that our correction pipeline has reduced the impact variations in this survey property have on the stellar classification rate. We see  $\sigma > 0.01$  after corrections since the test set is independent from the training set.

Relative detection rates are calculated in a similar way as classification probabilities, and details are left in Appendix A. We find that variations in relative detection rates are larger than variations in classification probabilities. Before corrections, Figure 2 shows that among all bins  $P(C_S|O_S)$  varies from the average by a maximum of  $\sim 12\%$ . For these same survey property bins,  $D_R(C_S, O_S)$  varies from the average by a maximum of  $\sim 39\%$ , so we allow training to continue for 300 cycles before stopping if convergence had not been reached.

### 3.3. Algorithm Design

With classification probabilities and relative detection rates established, we now detail the correction algorithms introduced in Section 3.1, which share identical steps except for their final stage.

The first step uses  $P(C_S|O_S)$ ,  $P(C_G|O_G)$ ,  $C_S$ , and  $C_G$  to estimate the number of true stars (galaxies) that were detected,  $RO_S$  ( $RO_G$ ). To start, we note that the expected number of classified stars is:

$$\langle C_S \rangle = O_S P(C_S|O_S) + O_G [P(C_G|O_G) - 1], \quad (2)$$

with an analogous equation for  $\langle C_G \rangle$ . For the above equation, we note that  $[P(C_G|O_G) - 1] = P(C_S|O_G)$ .

These equations can be inverted to solve for the expected number of observed stars and observed galaxies ( $RO_S$  and  $RO_G$ ). This gives the maximum likelihood estimate for  $RO_S$  and  $RO_G$  of:

$$RO_S = \frac{C_S P(C_G|O_G) + C_G [P(C_G|O_G) - 1]}{P(C_S|O_S) + P(C_G|O_G) - 1}, \quad (3)$$

This value can be negative or larger than  $C_S + C_G$ , both of which are not physical. We crop  $RO_S$  to be between zero and  $C_S + C_G$ . Using conservation of counts, we get  $RO_G$  by demanding that  $RO_S + RO_G = C_S + C_G$ .

All relative detection rates behave distinctly, so each one must be used to make corrections to specific subsets of the objects. For more details, refer to Appendix C. Due to this, our next step is getting estimates on the number of objects in the groups  $C_S \cap RO_S$ ,  $C_S \cap RO_G$ ,  $C_G \cap RO_S$ , and  $C_G \cap RO_G$  (this is just the classification distribution for  $RO_S$  and  $RO_G$ ), given by:

$$C_S \cap RO_S = RO_S P(C_S|O_S). \quad (4)$$

Other combinations are calculated analogously. Correcting any of these four sets of objects for variable detection rates is done by dividing by their respective relative detection rate, for example:

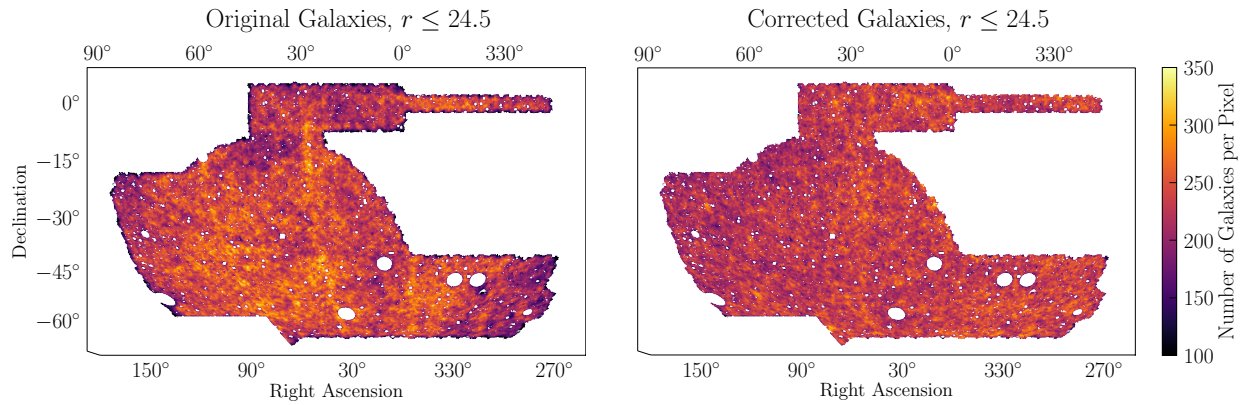
$$(C_S \cap RO_S)_{\text{Corr}} = \frac{C_S \cap RO_S}{D_R(C_S, O_S)}. \quad (5)$$

Obtaining  $FC_S$  (and  $FC_G$ ) is now just a matter of algorithmic design. We could take  $FC_S$  to be the sum of the corrected counts for  $C_S \cap RO_S$  and  $C_S \cap RO_G$  as in Equation (6). This design tries to remove variations induced by survey properties in  $C_S$ . Alternatively we could take  $FC_S$  to be the sum of the corrected counts for  $C_S \cap RO_S$  and  $C_G \cap RO_S$  as in Equation (7). This design tries to remove variations induced by survey properties in  $O_S$ .

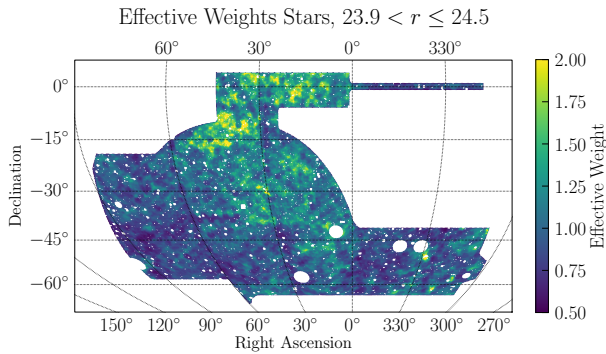
$$FC_S = (C_S \cap RO_S)_{\text{Corr}} + (C_S \cap RO_G)_{\text{Corr}} \quad (6)$$

$$FC_S = (C_S \cap RO_S)_{\text{Corr}} + (C_G \cap RO_S)_{\text{Corr}}. \quad (7)$$

In this work, we use the first of these algorithms, Equation (6), which is justified in Appendix D.



**Figure 4.** Galaxy counts before and after corrections are applied.



**Figure 5.** The effective weight map (corrected/uncorrected), smoothed with a  $0.15$  kernel, for stellar objects in our faintest magnitude bin. This map is used for the validation tests in Section 5.

#### 4. Correction of the DES Data

In this section, we present the results of applying our correction algorithm to DES data. To select a realistic dataset for this pipeline that will be applicable to a stream, we follow the study of the Phoenix stream by K. Tavangar et al. (2022). Phoenix is a  $15^\circ$  long,  $0.16^\circ$  wide, and dynamically cold stellar stream in the southern hemisphere (E. Balbinot et al. 2016; N. Shipp et al. 2018; K. Tavangar et al. 2022). Phoenix shows density fluctuations on small scales, making it a well-suited candidate for studying potential perturbations. In addition to its intrinsic properties, the stream lies in the middle of the DES footprint, near a prominent survey depth feature at a R.A. of  $\sim 30^\circ$  (Figure 1). This feature has increased depth and is one of the most readily apparent survey property features in the entire DES footprint. Therefore, we can use observations of the Phoenix stream to demonstrate the application of our pipeline and its effect on the generation of density maps.

Initially, we define a color–magnitude-based matched filter data selection used to obtain a realistic set of catalog objects. Then, we discuss the adjustments needed to apply our method to this dataset and generate corrections for both stars and galaxies.

##### 4.1. Matched Filter Data Selection

To derive corrections for a given stellar or galactic density map we want to place the same observational selection criteria on the injected objects as we would place on observations. In the case of the Phoenix stream, we use the matched filter from K. Tavangar et al. (2022). This filter is described in detail in

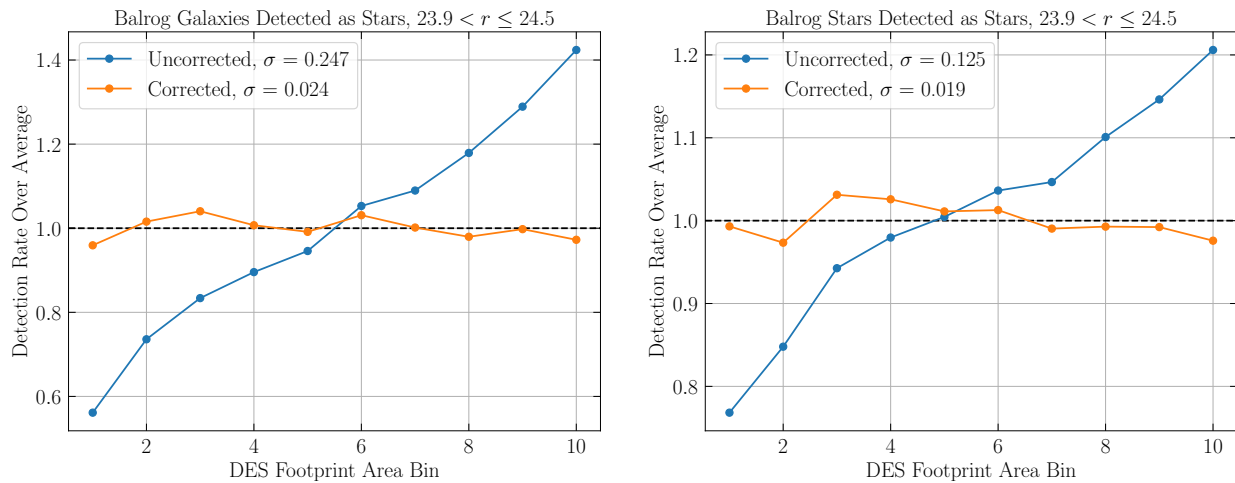
their analysis, but generally it uses a synthetic isochrone to generate a selection region in color–magnitude space. The size of this region is defined by the expected width of the stellar population convolved with observational uncertainties. We use an isochrone from A. Bressan et al. (2012) as implemented in *ugali* (K. Bechtol et al. 2015; A. Drlica-Wagner et al. 2020) with an age of  $a = 12.8$  Gyr, metallicity of  $[\text{Fe}/\text{H}] = -2.5$ , and distance of  $d = 17.4$  kpc (K. Tavangar et al. 2022). In practice, extinction corrected data should be used to apply the isochrone cut and with apparent magnitudes then being used to apply corrections. This could be achieved by adding an extinction map as an additional survey property, but more *Balrog* objects would be needed to cover the full range of apparent magnitude space necessary for corrections, and binning would likely have to be done by color as well as magnitude. Since this paper is focused on proof of concept corrective power rather than legitimate DM constraints, we simplify our pipeline to only work with apparent magnitudes.

We also use a magnitude limit of  $r < 24.5$ , where completeness falls to  $\sim 20\%$  (see S. Everett et al. 2022 Figure 9). This can be compared to N. Shipp et al. (2018), who use the same observations (DES Y3) but place a limit around one magnitude brighter ( $g < 23.5$ ). It is also 0.3 mag fainter than the  $r < 24.2$  limit used in K. Tavangar et al. (2022), who use DES Y6 data (roughly twice as many observations as DES Y3).

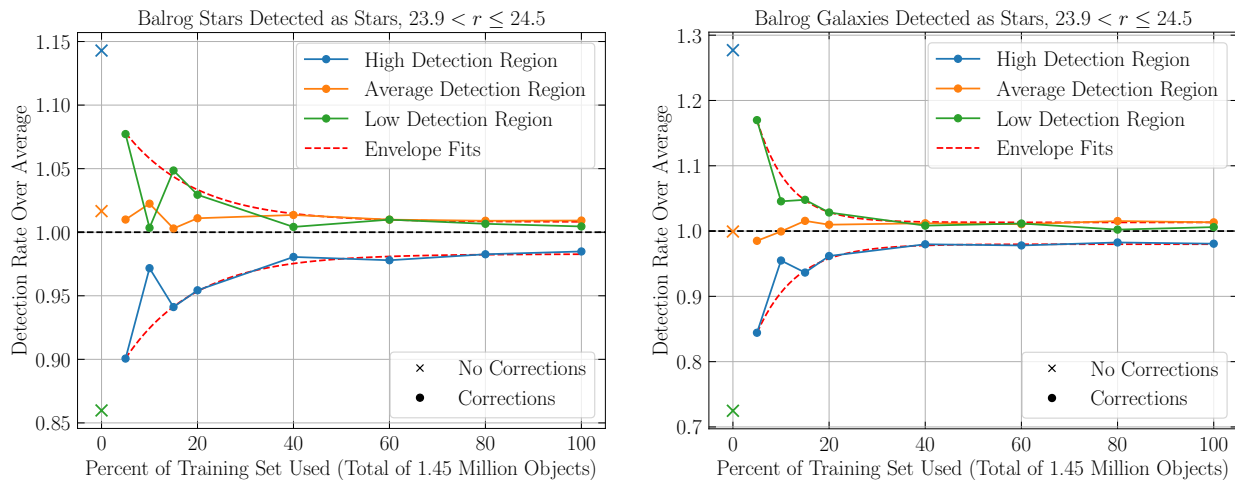
As mentioned in Section 2.1, in our correction algorithm we apply corrections to different magnitude bins. Our magnitude bins are as follows:  $r \leq 22.9$ ,  $22.9 < r \leq 23.9$ , and  $23.9 < r \leq 24.5$ . These magnitude bins were chosen as they had similar relative detection rate variations in testing sets when these variations were extrapolated out to expected values for a more current *Balrog* run. For more details, refer to Appendix E. These selections were placed in addition to the ones described in Section 2. Throughout this section we show figures for these magnitude bins only, as this best illustrates the effectiveness of our method in extending the usable magnitude range for stream analyses; at brighter magnitudes the catalog is significantly more uniform.

##### 4.2. Algorithm Adjustments for Real Data

Due to noise concerns, when applying to observational data, we perform the corrections at a healpixel resolution of  $\text{NSIDE} = 512$ . A *FRACDET* map (I. Sevilla-Noarbe et al. 2021) is used for a weighted resolution degradation and for a



**Figure 6.** Detection rates relative to the average for a 20% testing subset of the `Balrog` objects before and after corrections by effective weight bin number. Before corrections are shown in blue and after corrections in orange. The plot titles describe which relative detection rates are being shown. A drop in variance like observed shows that selection effects are being mitigated.



**Figure 7.** The convergence of corrections as more objects are used for training. A constant 20% subset of `Balrog` objects are withheld for testing. The remaining 80% is then cropped to different levels. The 0% level indicates that no corrections had been performed. Three area bins are tested. Blue, orange, and green lines represent area bins with above average detection rates, average detection rates, and below average detection rates, respectively. Red dotted lines show an exponent fit to an outer envelope. The plot titles describe which relative detection rates are being shown.

first-order correction to counts. We crop to healpixels that have a `FRACDET` value  $> 0.5$ .

In this work, Deep Field object classifications are taken as the truth. By performing spatial matching between the DES Deep Fields and Wide Fields and looking at the classifications given in both cases, we are able to get true correct classification rates. When comparing these rates against the classification rates in that area based on our training we found systematic errors. To correct this we multiplied our classification probabilities by scalars to match the probabilities shown by the deep fields.

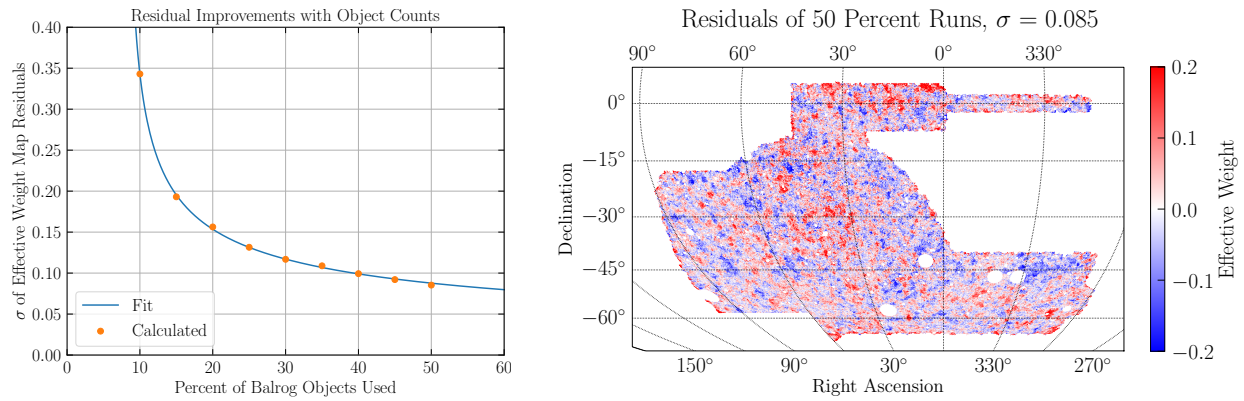
### 4.3. DES Corrections

In Figure 3 we show the observed stellar map on the left and our corrected map on the right. For all plots in this section we applied a Gaussian smoothing with a kernel of  $0.15^\circ$ . The Phoenix isochrone was used to select the objects in all plots in this section. The stellar density map is nonuniform on large scales, with counts increasing near the Galactic plane (along the left and right edges of the footprint)

and near the Sagittarius stream located at R.A.  $\sim 30^\circ$  and decl.  $\sim -5^\circ$ . As well as these physical variations, features correlated with survey properties (such as the RA  $\approx 30^\circ$  vertical stripe and the artifacts at the edge of the footprint) are present in the initial map and mitigated in the corrected map.

Galaxy counts can provide a probe as to how well the algorithm is working since they should be uniform on large scales across the DES footprint. Figure 4 shows initial counts on the left and the corrected map on the right. As with stars, in the original galaxy map a number of features correlated with survey properties can be seen. These include large-scale variations in the average number of objects (e.g., lower right of the map), vertical stripes at constant R.A. (e.g., R.A.  $\approx 30^\circ$ ) due to increased depth, and artifacts at the edge of the footprint. The corrected map does not exhibit these features and is more uniform, indicating the observational selection function has been mitigated.

Before applying these corrections to the Phoenix stellar stream specifically, we validate our methodology (Section 5).



**Figure 8.** Left: standard deviation of the effective weight map residuals as a function of percentage of training objects used. Right: residuals of the total effective weight map for two runs using 50% of the `Balrog` objects.

For our tests, we construct an effective weight map for stellar objects by taking corrected star counts divided by original star counts in Figure 3. The smoothed effective weight map is shown in Figure 5 with the boundary removed as it suffers from additional systematics. Bins of this map are used to sample a large range of impacts on stars from survey properties.

## 5. Validation of Methodology

This section tests the accuracy of the probabilities calculated in Section 3. We test the total corrective power of our relative detection rates (Section 5.1), how this changes with the number of objects used to train (Section 5.2), and how consistent our final corrections are as a function of the number of objects used for training (Section 5.3).

### 5.1. Overall Corrective Power

Initially we test the calculations of  $D_R(C_S, O_S)$  and  $D_R(C_G, O_G)$  in our faintest magnitude bin since these are the two relative detection rates necessary for a stellar correction. We focus on relative detection rates as they have larger variations than classification probabilities. Assuming uniform injections, the relative detection rate of a type of object (such as true stars classified as stars) is calculated as the number of detected objects of this type per healpixel divided by the overall average. If the corrections are effective we would expect this ratio to become independent of the amount of correction applied (effective weight).

Subsets of the data are created by binning the healpixels based on their effective weight from Figure 5, to obtain 10 bins of equal sized samples. From each area bin, we select 80% of the `Balrog` objects to be in our training sample, which is used to calculate corrections applied to the remaining test set. Figure 6 shows relative detection rates on the test set across the 10 effective weight bins, which are referred to in the figure as the DES footprint area bins. Comparing rates before (blue) and after (orange) corrections shows that variations in relative detection rates for both stars and galaxies drop by a factor of  $\sim 5$ .

### 5.2. Convergence from Object Counts

Characterizing how corrective power scales with more SSI objects will help inform SSI strategies for future analyses. To investigate this, we perform the test from Section 5.1 while

varying the size of the training sample. Due to noise concerns, we reduce the number of effective weight bins to 3. Our results are shown in Figure 7 where we plot the detection rates for correctly classified stars (left) and incorrectly classified galaxies (right) as a function of the size of the training dataset. Blue (orange, green) points represent an effective weight bin that started with above (approximately, below) average relative detection rates. To illustrate the general behavior we include an exponential decay envelope in red. This line shows convergence is reached with  $\sim 40\%$  of the training set in both cases. Beyond this point we are limited by the Poisson noise of the test dataset and cannot probe further convergence.

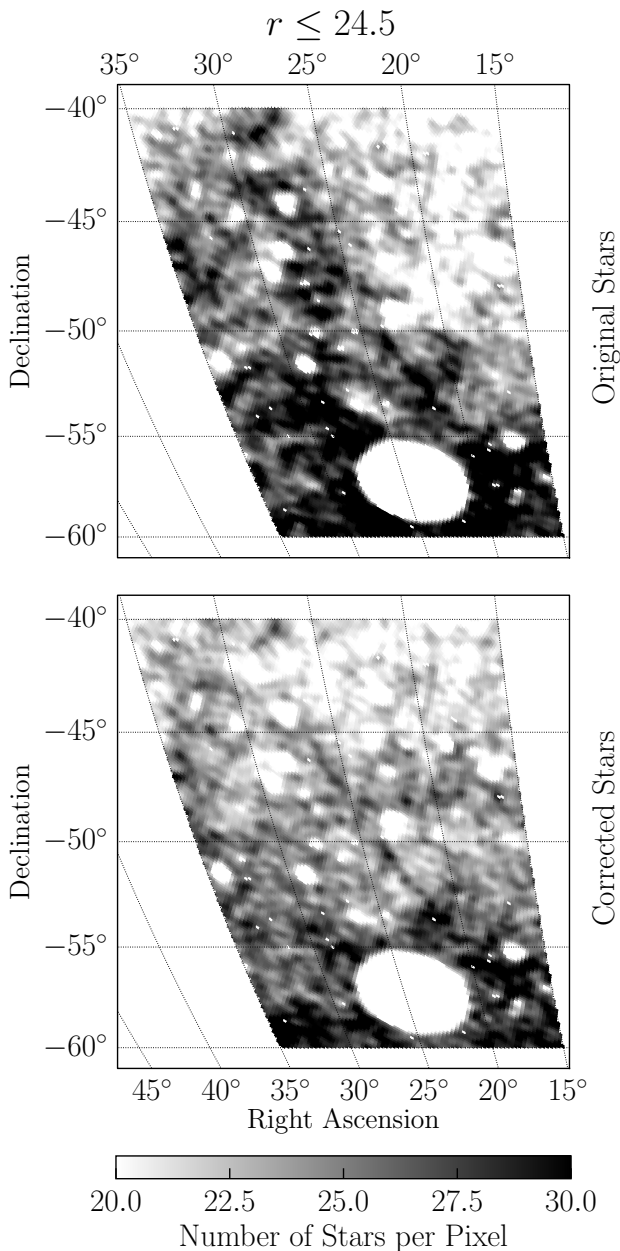
### 5.3. Repeatability of Results

Given a training set, the derived corrections are completely deterministic, but it is interesting to see how these derived corrections change as the training sample is varied. This is important because it shows the inherent noise in our corrections as a function of training set size.

To test this, we vary the objects used to derive corrections. At a fixed training set size, we create two disjoint groups of training objects. For each subset, we derive corrections and effective weight maps. We then compute the difference in effective weight, and the variance of this residual gives an estimate of our algorithm's repeatability. This process is repeated for nine training set sizes ranging from 10% to 50% of the total sample ( $\sim 26$  million total objects). Residuals for the two 50% runs are shown in the right plot of Figure 8. The left plot of Figure 8 shows standard deviations of the residuals as a function of the training set size. We found our results could be well fit by a power law  $y \sim x^{-0.45}$ .

With `Balrog` Y6 covering the entire DES footprint instead of just 20% (D. Anbajagane et al. 2025a), we can use our fit to predict the consistency of results when  $\sim 5$  times more objects are injected. In this case, we expect two full runs of `Balrog` Y6 to have stellar effective weight map residuals with a standard deviation of  $\sim 0.03$ , compared to  $\sim 0.06$  for `Balrog` Y3. Other differences in `Balrog` Y6 mentioned in S. Everett et al. (2022) and M. Tabbutt (2023) such as no longer using delta stars will impact the accuracy of this prediction.

Residuals can also be used to validate our cycle limit for convergence as mentioned in Section 3. We compare our calculated effective weight map to an effective weight map where we put no limit on the number of cycles in the iterative

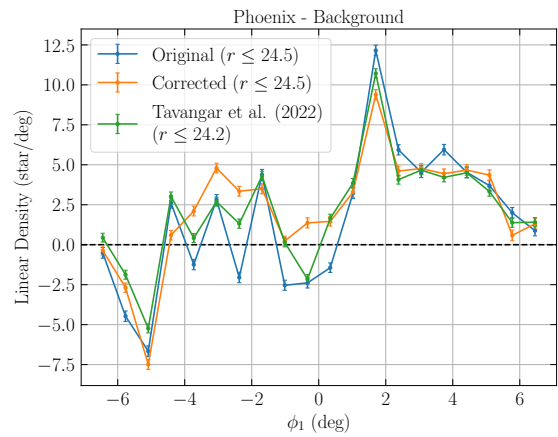


**Figure 9.** Phoenix stellar stream before and after corrections are applied. A Gaussian smoothing with a kernel of  $0.15$  is used to smooth these maps. The (R. A., decl.) endpoints in degrees of this stream are  $(20.1, -55.3)$  and  $(27.9, -42.7)$  (N. Shipp et al. 2018).

training. The standard deviation for the residuals of these maps was  $0.003$ , showing that the cycle limit we have chosen has allowed results to converge.

### 6. Application to Stellar Streams

This section applies the corrections from Section 4 to observations of stellar streams. First we look at the Phoenix stellar stream and show that our corrections cause statistically significant linear density shifts (Section 6.1). We then look at synthetic streams and show that our correction algorithm properly suppresses artificial power at length scales larger than the DECam field of view (FOV; Section 6.2) of  $3 \text{ deg}^2$ . This suppression of power is detectable on individual stream



**Figure 10.** Linear densities of stars per degree of Phoenix minus background. Results are shown for the uncorrected stream, the corrected stream, and the uncorrected stream with the more conservative magnitude cut used in K. Tavangar et al. (2022). Statistical uncertainties on densities are given by error bars.  $\phi_1$  is the angular stream track coordinate. Whether or not these changes are beneficial will be addressed later, for now we just note that the changes are statistically significant.

realizations at length scales relevant for subhalo interactions (N. Banik et al. 2021b).

#### 6.1. Phoenix Stellar Stream

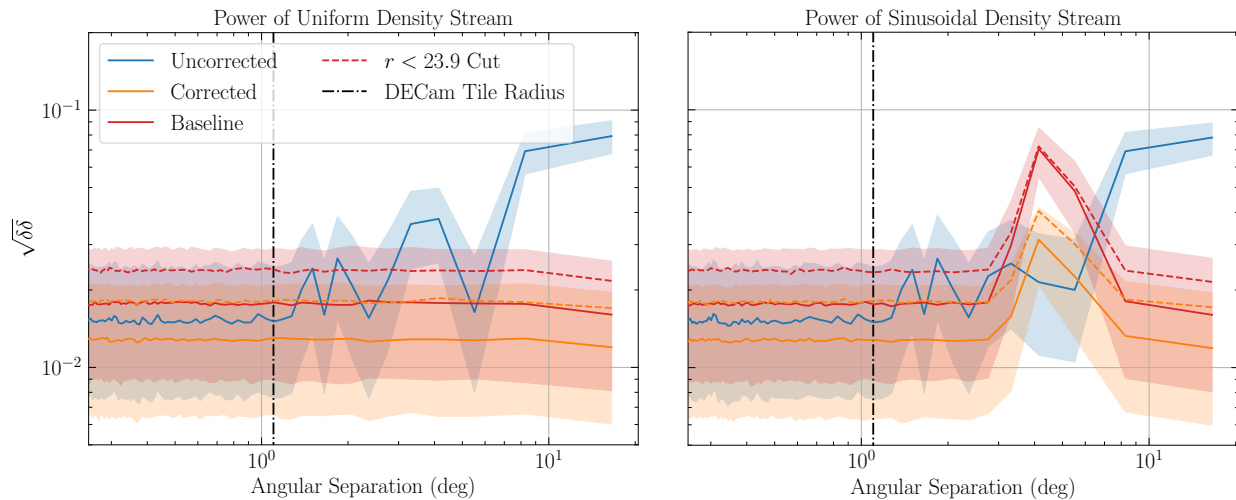
Using the color and magnitude cuts from Section 4.1, we show the original (top) and corrected (bottom) counts for the on-sky region around the Phoenix stream in Figure 9. It is immediately obvious that the depth feature along R.A.  $\sim 27^\circ$  is removed in the corrected map, enhancing the signal of the stellar stream. For a quantitative measurement of this effect, we compute the number of excess stars along the stream relative to the background.

We select a background region along the stream but offset by  $1^\circ$ . Then we define the linear density as the star counts on stream minus the background. The results of this are shown in Figure 10. The blue (orange) line signifies the stream before (after) corrections, and the green line signifies the stream with a more conservative ( $r \leq 24.2$ ) magnitude cut from K. Tavangar et al. (2022). Error bars represent the statistical uncertainty of each point, showing that the corrections made to the Phoenix stellar stream lead to statistically significant changes in the linear density of the stream. We note that this is not meant as a direct comparison to K. Tavangar et al. (2022) as they use DES Y6 data, but instead is meant to show the dependence of linear density on magnitude cutoffs.

The two most prominent changes as a result of these corrections are the reduced “gap” centered at  $\phi_1 \sim -0.5$  deg and the fluctuations seen at the local peak centered around  $\phi_1 \sim -3$  deg. Conversely, the density peak at  $\phi_1 \sim 1.5$  deg, hump at  $2 < \phi_1 < 6$  deg, and underdensity at  $\phi_1 \sim -5$  remain generally the same in all cases. This analysis also suggests that it would be interesting to apply this correction technique to the deeper Y6 data analyzed in K. Tavangar et al. (2022). In the next section, we turn to tests on the power spectra of simulated streams to assess the impact of these corrections.

#### 6.2. Synthetic Stellar Streams

To probe the effect of our corrections on the 1D density autocorrelation function, we compute this power spectra for



**Figure 11.** Power spectra for synthetic streams. Shown are the baseline for the true stream (red), the observed stream before corrections (blue), and the observed stream after corrections (orange). We also show the baseline and corrected streams with a more conservative  $r < 23.9$  cut, denoted with dashed lines. The true stream has perfect classification and no variations in detection rates. Solid lines represent the median power spectra at each point for 5000 realizations. Shaded regions represent the 16th and 84th percentiles among the 5000 simulations for each point along the power spectra. Power spectra are shown for a uniform density stream on the left and a sinusoidal density stream, with a period of  $4.4$  (two times the DECAM FOV), on the right.

5000 simulated streams before and after corrections. These simulated streams are injected into an arbitrary region of the sky at high Galactic latitude with large variations in effective weight (the area is fixed between realizations). For each realization, we assume a uniform distribution for the true number of stars and galaxies plus a stream population of stars. Since we do not know the absolute detection rate, we fix the number of detected objects for each realization. We then use relative detection rates to create realistic spatial distributions for all of our mocks. Our stream is defined to, on average, have a 25% excess of stars compared to the background stellar population. Corrections are then performed on the data based on the relative detection rates and correct classification probabilities as prescribed in Section 3. For more details, refer to Appendix D.

To compare these uncorrected and corrected results to a baseline, we independently do a run with perfect classification and no variations in detection rates. Two runs were performed, one with a uniform density stream, and one with a sinusoidal density stream with period  $4.4$ , 2 times the DECAM FOV. In Figure 11, we show the results of this exercise. The lines are the median power spectra for baseline (red), corrected (orange), and uncorrected (blue) results. Shaded regions represent the 16th and 84th percentiles among the 5000 simulations for each point along the power spectra. For the uniform density stream (left), the corrected stream obtains a near constant power spectra while the uncorrected stream obtains extra power at scales larger than the DECAM FOV. For the sinusoidal density stream (right), the corrected stream does gain signal power at the expected scales, although its amplitude is lower due to the contamination of galaxies. The differences from the uncorrected stream can be seen even on individual runs, and as is shown in N. Banik et al. (2021b), these are the same angular scales expected to be sensitive to dark matter subhalo interactions. Dashed lines in Figure 11 represent results with a more conservative magnitude cut of  $r < 23.9$ . Noise levels increase as expected, and in the sinusoidal case the signal-to-noise ratio of the peak is reduced by a factor of 10% compared to the full  $r < 24.5$  data, showing the advantage of using fainter objects.

We can estimate the importance of proper corrections at even fainter magnitudes where the ratio of galaxies to stars becomes even larger. This is particularly relevant for future surveys such as the Vera C. Rubin Observatory LSST (Z. Ivezić et al. 2019). To test this we use the DES deep fields to obtain a realistic distribution of object counts at fainter magnitudes ( $24.5 < r \leq 25$ ). Assuming survey properties will have similar impacts as in DES, we can get a lower bound for detection and classification variations by setting them equal to the variations from our previous faintest bin ( $23.9 < r \leq 24.5$ ). Running the same power spectrum test as before, the multiplicative difference between the baseline and uncorrected power spectra increased by a factor of 5 at the largest tested scales, showing that corrections will be even more necessary to fully leverage LSST-like data.

## 7. Conclusion

Accurate stellar stream density measurements offer a promising probe into dark matter substructure in the Milky Way. While variable selection effects from survey properties can introduce artificial density variations, synthetic source injection can be used to correct for these effects. In this work we describe and offer a method to correct for artificial density variations that will allow for a more accurate characterization of stellar streams and their parameters.

Section 4 shows our corrections applied to DES DR1 data. After these corrections were applied, galaxy counts became more uniform on large scales. Features correlated with survey properties such as the depth feature at R.A.  $\sim 30^\circ$  and edge artifacts are also visually diminished after corrections.

Validation of our algorithm is performed in Section 5. For overall corrective power, we find that key relative detection rates in our faintest magnitude bin had their standard deviations drop by a factor of at least 5 after corrections. Running this same test with a variable training set size showed that relative detection rates improved in uniformity when more training objects were used, but only up to a limit. Finally, we find that the repeatability of our results improved as more

training objects were used. These improvements persisted over the entire range of training set sizes that we were able to test.

We apply our corrections to stellar streams in Section 6. Here we find that the changes in the linear density of Phoenix due to our corrections are statistically significant, which calls into question some of the apparent density variations reported in the literature. For uniform density simulated streams, our corrections mitigated all signals from survey property variations. For sinusoidal density simulated streams, our corrections saw a signal at the proper frequency, although the amplitude of the corrected signal was reduced.

One clear next step is to derive corrections for the DES Y6 run of Balrog (D. Anbajagane et al. 2025a). This SSI run has injections across the entire DES footprint compared to the 20% coverage for Balrog Y3. One large advantage of this will come in the form of testing set size for the tests performed in Section 5. The number of sources in a 20% testing set of the Balrog Y6 data would be on the order of the entire Balrog Y3 catalog. This will make Balrog Y6 less susceptible to potential bias in estimates of relative detection rates that are more likely to arise with smaller sample sizes. This could lead to the relative detection rates in Figure 7 moving closer to one in the future if the current limits on accuracy were due to biased testing sets. Additionally, this framework could easily be applied to data from the DECam Local Volume Exploration survey (DELVE; A. Drlica-Wagner et al. 2021) processed through the DECam All Data Everywhere (DECADE; D. Anbajagane et al. 2025b) campaign. This dataset uses the same instrument with the same data reduction pipeline, but contains much more heterogeneous survey properties.

An interesting problem to investigate is how well these corrections transfer to other color–magnitude regions. Our color–magnitude cuts were specific to the isochrone of Phoenix. If possible, less restrictive color cuts would allow one training session to potentially provide corrections for multiple stellar streams with different isochrones, and allow for more objects to be used in training. The potential cost would come from how well the Balrog sources would actually model the stars in the stellar stream if they’re in a different area of color–magnitude space.

This work demonstrates that the use of synthetic sources to correct for variable selection effects is a viable approach. It also provides a groundwork methodology for applying these corrections in future surveys such as the LSST, which will cover the DES footprint (and extend beyond) and use a similar photometric system. Rudimentary tests show that we can expect the impacts of survey properties to increase by a factor of at least 5 when fainter magnitude bins are used in LSST. Therefore, accurate corrections and a deep understanding of the observational selection function will be even more important in making unbiased measurements of density variations in stellar streams.

### Acknowledgments

P.S.F. acknowledges support from the DIRAC Institute in the Department of Astronomy at the University of Washington. The DIRAC Institute is supported through generous gifts from the Charles and Lisa Simonyi Fund for Arts and Sciences, and the Washington Research Foundation.

Contribution statement: K.B. conducted the analysis, created all plots and tables presented here, and led the writing. P.S.F. provided direct supervision of the research, guidance on

the analysis, and helped with writing. M.T. provided technical expertise in accessing and correctly using the SSI data products. K. Bechtol guided major analysis and content decisions as well as cosupervising the research. A.D.W. and C.M.V. internally reviewed the paper. T.Y.C. and B.M.P. provided valuable comments that improved the paper’s clarity and quality. The remaining authors contributed to this work through the construction of DECam and other aspects of data collection; data processing and calibration; developing widely used methods, codes, and simulations; running pipelines and validation tests; and promoting the science analysis.

Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft, and the Collaborating Institutions in the Dark Energy Survey.

The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l’Espai (IEEC/CSIC), the Institut de Física d’Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, NSF NOIRLab, the University of Nottingham, The Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, Texas A&M University, and the OzDES Membership Consortium.

Based in part on observations at NSF Cerro Tololo Inter-American Observatory at NSF NOIRLab (NOIRLab Prop. ID 2012B-0001; PI: J. Frieman), which is managed by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

The DES data management system is supported by the National Science Foundation under Grant Numbers AST-1138766 and AST-1536171. The DES participants from Spanish institutions are partially supported by MICINN under grants PID2021-123012, PID2021-128989, PID2022-141079, SEV-2016-0588, CEX2020-001058-M, and CEX2020-001007-S, some of which include ERDF funds from the European Union. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya.

We acknowledge support from the Brazilian Instituto Nacional de Ciência e Tecnologia (INCT) do e-Universo (CNPq grant 465376/2014-2).

This document was prepared by the DES Collaboration using the resources of the Fermi National Accelerator Laboratory (Fermilab), a U.S. Department of Energy, Office of Science, Office of High Energy Physics HEP User Facility. Fermilab is managed by Fermi Forward Discovery Group, LLC, acting under Contract No. 89243024CSC000002

Some of the results in this paper have been derived using the healpy and HEALPix package

*Facility:* Blanco (DECam).

*Software:* `astropy` (Astropy Collaboration et al. 2013, 2018, 2022), `fitsio`,<sup>51</sup> `HEALPix` (K. M. Górski et al. 2005), `healpy` (A. Zonca et al. 2019), `healsparse`,<sup>52</sup> `matplotlib` (J. D. Hunter 2007), `numpy` (C. R. Harris et al. 2020), `scipy` (P. Virtanen et al. 2020), `skyproj`,<sup>53</sup> `ugali` (K. Bechtol et al. 2015; A. Drlica-Wagner et al. 2020).

## Appendix A

### Calculating Classification Probabilities and Relative Detection Rates

This section will discuss the details in calculating classification probabilities and relative detection rates. The calculations are nearly identical, so more focus will be given to calculating classification probabilities. The notation used will be the same as that given in Table 2. As additional notation,  $SP$ , will refer to a general survey property.

#### A.1. Classification Probabilities

This section will focus on calculating  $P(C_S|O_S)$  for some magnitude bin using Balrog data. The calculation of  $P(C_G|O_G)$  is completely analogous after switching out the Balrog delta star sample for the Balrog galaxies. For this subsection, valid healpixels will refer to healpixels with valid values for each survey property that also has a detected Balrog delta star within the magnitude bin of interest.

After applying measured magnitude cuts, two numbers are stored for each valid healpixel. First is the number of detected Balrog delta stars in the healpixel,  $IO_S$ . Second is the number of correctly classified Balrog delta stars in the healpixel,  $IC_S \cap IO_S$ . Both of these counts are subjected to quality and magnitude cuts. The sum over all valid healpixels of  $IC_S \cap IO_S$  divided by the sum of  $IO_S$  can be thought of as the average correct classification probability:

$$\langle P(IC_S|IO_S) \rangle = \frac{\sum IC_S \cap IO_S}{\sum IO_S}. \quad (A1)$$

Next we compute the relation between each survey property and the correct classification probability. For each survey property, the valid healpixels are binned according to the survey property. For this analysis, we choose to use 10 bins. In each bin the average survey property value is calculated in each bin as well as the relative correct classification rate compared to the full DES footprint. These are calculated using Equations (A2) and (A3), respectively. Here, summing over a

bin means taking the sum over the healpixels within the bin.

$$\langle SP \rangle_{\text{Bin}} = \frac{\sum_{\text{Bin}} SP}{\sum_{\text{Bin}} 1} \quad (A2)$$

$$\begin{aligned} & \frac{\langle P(IC_S|IO_S) \rangle_{\text{Bin}}}{\langle P(IC_S|IO_S) \rangle} \\ &= \frac{(\sum_{\text{Bin}} IC_S \cap IO_S) / (\sum_{\text{Bin}} IO_S)}{\langle P(IC_S|IO_S) \rangle}. \end{aligned} \quad (A3)$$

These 10 ordered pairs show the impact the survey property has on classification probabilities. With this dependency in mind for each survey property individually, we have to construct the classification probability as a function of every survey property. This is done in an iterative process. First we select whichever survey property causes the most variance in  $P(IC_S|IO_S)$  among its bins. For the  $n$ th iteration, we will refer to this survey property as  $SP_{\text{Max},n}$ . Once this survey property is chosen, we correct for the dependency it causes. This is repeated iteratively until a termination condition is met. Notably, there is nothing preventing a correction for one survey property introducing a dependency in another survey property that has already been corrected for. Due to this, the same survey property could be corrected for multiple times during the training procedure. For example, we could very well have  $SP_{\text{Max},1} = SP_{\text{Max},50}$ .

To begin training, we want to figure out which survey property to correct for first. Equation (A3) is calculated for each bin of each survey property and we use a least squares deviation from one to determine which survey property has the largest impact on classification rates. This is shown in expression (A4), which is calculated for each survey property.

$$\sum_{\text{Bin}=1}^{10} \left( \frac{\langle P(IC_S|IO_S) \rangle_{\text{Bin}}}{\langle P(IC_S|IO_S) \rangle} - 1 \right)^2. \quad (A4)$$

Using the notation defined earlier, we will let  $SP_{\text{Max},1}$  be whichever survey property has the largest value for expression (A4). With this survey property chosen, we can use the 10 ordered pairs defined in Equations (A2) and (A3) to construct a function from  $SP_{\text{Max},1}$  values to a relative classification rate. To achieve this we use a linear interpolation function  $f_1$  defined such that the following holds for each bin:

$$f_1(\langle SP_{\text{Max},1} \rangle_{\text{Bin}}) = \frac{\langle P(IC_S|IO_S) \rangle_{\text{Bin}}}{\langle P(IC_S|IO_S) \rangle}. \quad (A5)$$

We note that a constant value is used for extrapolation.

To correct the Balrog data for the dependency on  $SP_{\text{Max},1}$  we analyze  $f_1(SP_{\text{Max},1})$  for each valid healpixel with inputs of the  $SP_{\text{Max},1}$  values on each healpixel. The output of  $f_1$  can be interpreted as how likely a star is to be classified correctly, for a given healpixel relative to the average correct classification rate. Therefore to remove the  $SP_{\text{Max},1}$  dependency, at each valid healpixel the number of correct classifications is divided by the output of the interpolation function. This is shown explicitly in Equation (A6), where each value shown is assumed to be the value on one specific

<sup>51</sup> <https://github.com/esheldon/fitsio>

<sup>52</sup> <https://github.com/LSSTDESC/healsparse>

<sup>53</sup> <https://github.com/LSSTDESC/skyproj>

healpixel.

$$(IC_S \cap IO_S)_{\text{New}} = \frac{(IC_S \cap IO_S)_{\text{Old}}}{f_1(SP_{\text{Max},1})}. \quad (\text{A6})$$

Substituting these new values on each valid healpixel to Equation (A1), we can start this process over to make another correction for a survey property dependency. This process is repeated until values from Equation (A3) fall between 0.99 and 1.01 for each bin for each survey property. For notation purposes, let the survey property and interpolation function used on the  $n$ th iteration be denoted as  $SP_{\text{Max},n}$  and  $f_n$  respectively. As mentioned briefly in Section 2.2, training is stopped after 150 cycles even if convergence has not been reached to avoid overfitting data.

With training completed,  $P(IC_S|IO_S)$  must now be calculated for all healpixels with valid survey properties so it can be applied to data. This is done in an iterative process. To start, we assign each of these healpixels the average classification rate, shown in Equation (A7). We note that  $\langle P(IC_S|IO_S) \rangle$  in this case will refer to the initial value before any corrections were made to  $IC_S \cap IO_S$ , which could change the value slightly. In practice, the specific starting point is not important as the average is recalibrated (see Section 4.2).

$$P_0(IC_S|IO_S) = \langle P(IC_S|IO_S) \rangle. \quad (\text{A7})$$

After this, for each correction to  $IC_S \cap IO_S$  applied during training in Equation (A6), we must apply a corresponding correction to our classification rate. For concreteness, suppose there were a total of  $N$  cycles used in training. Then, for  $n$  between 1 and  $N$  we define the recursive relation given in Equation (A8), which is calculated on each healpixel. Once all iterations have been completed, we have our overall correct classification probability. Combining that with the relation in Equation (1) gives us Equation (A9).

$$P_n(IC_S|IO_S) = P_{n-1}(IC_S|IO_S)f_n(SP_{\text{Max},n}) \quad (\text{A8})$$

$$P(C_S|O_S) = P_0(IC_S|IO_S)\prod_{n=1}^N f_n(SP_{\text{Max},n}). \quad (\text{A9})$$

To avoid a nonphysical probability, this is cropped to be less than 1 (negative values will not appear by construction).

### A.2. Relative Detection Rates

In this section we will calculate four relative detection rates as functions of survey properties for some particular magnitude bin: the rates at which Balrog delta stars are classified as stars  $D_R(IC_S, IO_S)$ , Balrog delta stars are classified as galaxies  $D_R(IC_G, IO_S)$ , Balrog galaxies are classified as stars  $D_R(IC_S, IO_G)$ , and Balrog galaxies are classified as galaxies  $D_R(IC_G, IO_G)$ . For now we focus specifically on  $D_R(IC_S, IO_S)$  but the same process is repeated for the other relative detection rates. Valid healpixels in this section will refer to healpixels with valid values for each survey property that also had a Balrog delta star injected that passed quality cuts.

Compared to the classification rates of the previous subsection, with relative detection rates we are unable to get an average detection rate for our objects. This is due to the fact that undetected objects do not have measured magnitudes to bin based on. Therefore, we are restricted to calculating relative detection rates instead of actual detection probabilities.

The calculation of these rates is performed in a similar manner as in Section 5.3.1 of M. Rodríguez-Monroy et al. (2022). Within a bin from some survey property, we look at

the average number of detections of interest (Balrog delta stars classified as stars within our magnitude bin) per healpixel in that bin to find a characteristic detection density. This is then compared to the average number of detections of interest per healpixel across all of the valid healpixels. If the distribution of objects that could cause a detection of interest is uniform on large scales, the ratio of these two densities will then represent the rate at which detections of interest occur in the survey property bin relative to the rate at which detections of interest occur overall. Since Balrog objects are injected on a uniformly spaced hex grid and the distribution of objects on this grid is random, this approach is valid for our calculations.

With this framework in mind, we turn now to the details of the actual calculation. Like in classification calculations, for each survey property we bin the valid healpixels according to the survey property and calculate the relative detection rate as the following:

$$\begin{aligned} \langle D_R(IC_S, IO_S) \rangle_{\text{Bin}} &= \frac{(\sum_{\text{Bin}} IC_S \cap IO_S) / (\sum_{\text{Bin}} 1)}{(\sum IC_S \cap IO_S) / (\sum 1)}. \end{aligned} \quad (\text{A10})$$

Note that the numerator is exactly the average number of detections of interest per healpixel in the bin while the denominator is the average number of detections of interest per healpixel across all valid healpixels.

Treating the value calculated in Equation (A10) like the relative classification rates of the previous subsection, training is then performed. As mentioned briefly in Section 2.2, training is stopped after 300 cycles even if convergence has not been reached.

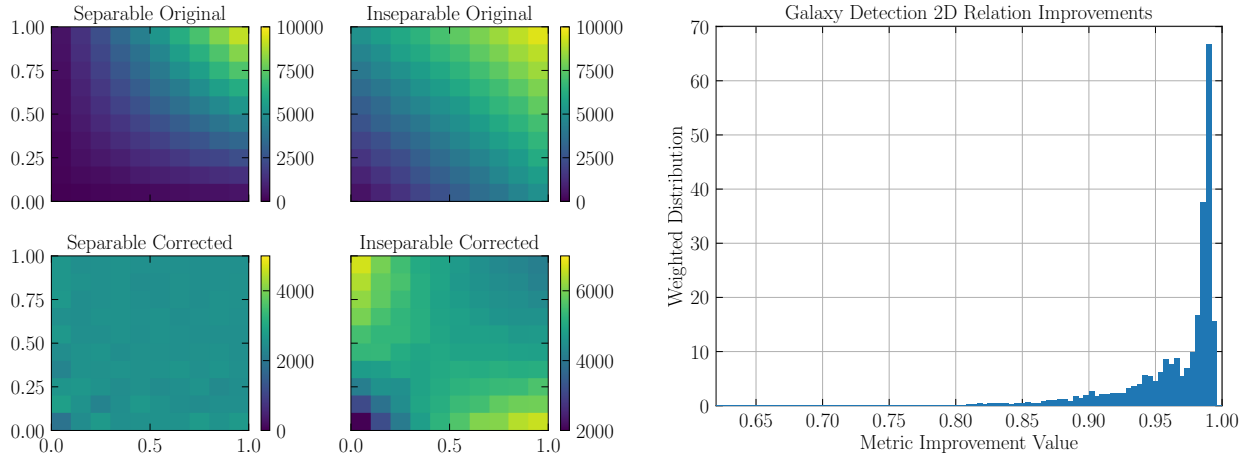
Once training is complete, we must calculate  $D_R(IC_S, IO_S)$  on all healpixels with valid survey properties. This is also performed in a nearly identical way to the calculations in the previous subsection, with the only change being that we start with the assignment:

$$D_{R,0}(IC_S, IO_S) = 1. \quad (\text{A11})$$

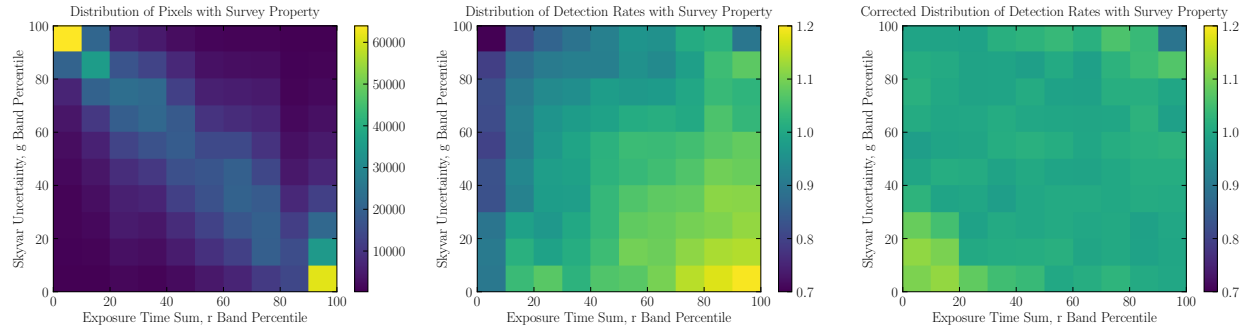
This is done since our relative detection rate should by definition be centered on 1 instead of an average probability like the classification probability. With this assignment done, the recursive relation used to build up to  $D_R(IC_S, IO_S)$  is defined in an identical way as in Equation (A8). As a relative rate can exceed 1, the crop done in the previous section is unnecessary here.

## Appendix B Testing Separability Assumption

When calculating relative detection rate and classification probability variations, our iterative approach means that the selection function we learn is separable. If the selection function is not separable, we can still get convergence in training without properly learning the selection function. This could occur since convergence is defined by looking at the projections of the selection function onto each survey property individually. Higher dimensional dependencies could be hidden from view when these projections occur, which would lead to converging to the wrong selection function. To demonstrate this, simulations were run with injecting objects uniformly over two survey properties,  $x$  and  $y$ . In one simulation, detection rates were given by a separable function of  $x$  and  $y$ , while in another the detection rate function was



**Figure B1.** (Left) Example of 2D distributions of objects in inseparable and separable cases before and after corrections. (Right) A weighted distribution of the metric improvement value, weighted by the precorrection metric. Any value over zero is considered improvement, and a 1 would be perfect removal of any dependency. Every metric improvement value was greater than 0.65, with the vast majority being above 0.95.



**Figure B2.** Above is a sample 2D relation using exposure time sum in the  $r$  band and sky variance uncertainty in the  $g$  band. For the test of 20% of Balrog objects, first we plot out the healpixel distribution in these two dimensions, shown on the left. With this healpixel distribution, we can use original and corrected detection counts to look at detection rate variations before corrections (in the middle) and after corrections (on the right).

inseparable. Results before and after corrections are shown in the left of Figure B1, showing the higher dimensional structure present post corrections in the inseparable case.

To test whether this occurred in our pipeline, we used the largest set of Balrog objects possible: all Balrog galaxy injections. To keep the statistical significance as high as possible, we trained and corrected for overall relative detection rates (being classified as either a star or a galaxy). Training was done on an 80% subset of objects, with testing being done on the remaining 20%. For each combination of two survey properties we looked at the distribution of valid healpixels over the two survey properties, and the relative detection rates before and after corrections over the two survey properties. An example of this is shown in Figure B2. Ideal results would be uniform distributions in the corrected plot, especially in regions with high healpixel counts as these are the most common throughout the footprint.

With this example shown, we next designed a metric to concisely summarize results over every combination of survey properties. Overall, both the original and corrected two-dimensional histograms should have values centered about one. Furthermore, the points we care about more are the points with high healpixel densities, as this will impact a greater portion of the DES footprint. To take these factors into account, we use a metric of variance from one weighted by healpixel count. If we let  $i$  index over each

two-dimensional bin in the histograms from Figure B2, let  $PC_i$  be the healpixel counts in bin  $i$ , and let  $D_{R,i}$  be the detection rate variation in bin  $i$  (these are the histogram values of the two rightmost histograms of Figure B2), we write our metric as:

$$\epsilon = \frac{\sum_i PC_i (1 - D_{R,i})^2}{\sum_i PC_i}. \quad (\text{B1})$$

This metric will be nonnegative by construction, and better results will lead to values closer to zero. Using these features we then define an improvement metric. If for any given survey property combination  $\epsilon_{\text{orig}}$  is the metric value on the original detection variations and  $\epsilon_{\text{corr}}$  is the metric value on the corrected detection variations, we define our improvement metric as:

$$\epsilon_{\text{impr}} = 1 - \frac{\epsilon_{\text{corr}}}{\epsilon_{\text{orig}}}. \quad (\text{B2})$$

Any positive value for  $\epsilon_{\text{impr}}$  suggests an improvement, with a value of 1 being perfect. We calculated  $\epsilon_{\text{impr}}$  values for every combination of survey properties using the 20% testing data from earlier. We do not necessarily care about each  $\epsilon_{\text{impr}}$  value equally, so weights of  $\epsilon_{\text{orig}}$  are also stored for each survey property combination. If two survey properties hardly had any detection rate dependencies in their 2D relation to begin with their improvement matters far less than two survey properties

that initially had high dependencies. With these weights in mind, a weighted histogram of  $\epsilon_{\text{impr}}$  was generated, as is shown in the right plot of Figure B1. The overwhelming trend toward 1 shows that our separable assumption is reasonable.

### Appendix C Relative Detection Rate Correlations

In Figure C1 we show the distributions of two different relative detection rate combinations across the DES footprint. For both plots we focus on relative detection rates in our faintest magnitude bins. The plot on the left shows the relative detection rates of stars detected as stars versus galaxies detected as stars. In general this plot shows that the relative detection rate of galaxies classified as stars has a higher variance, and there is a clear positive correlation. This correlation is likely due to the fact that galaxies classified as stars appear to be point sources and so survey properties that affect detection rates of stars will affect these galaxies as well. The plot on the right shows the relative detection rates of stars detected as stars versus stars detected as galaxies. This plot does not show strong correlations between the two relative detection rates. The absence of a correlation implies that the detection rate of stars that are then classified as galaxies is not affected by survey properties in the same manner. One possible explanation for this could be that unrecognized blends are causing brighter stars to be classified as galaxies, and therefore higher changes in survey properties will not affect the detection rate of these objects.

Overall, the scatter in these relations eliminates the possibility of certain simplifications that could have potentially been made to our correction algorithm. Referring back to Section 3, our current correction algorithm, originally given in Equation (6), is given by:

$$FC_S = \frac{C_S \cap RO_S}{D_R(C_S, O_S)} + \frac{C_S \cap RO_G}{D_R(C_S, O_G)}. \quad (C1)$$

If the left plot of Figure C1 showed that the two relative detection rates used above were equal, we could simplify that correction to be:

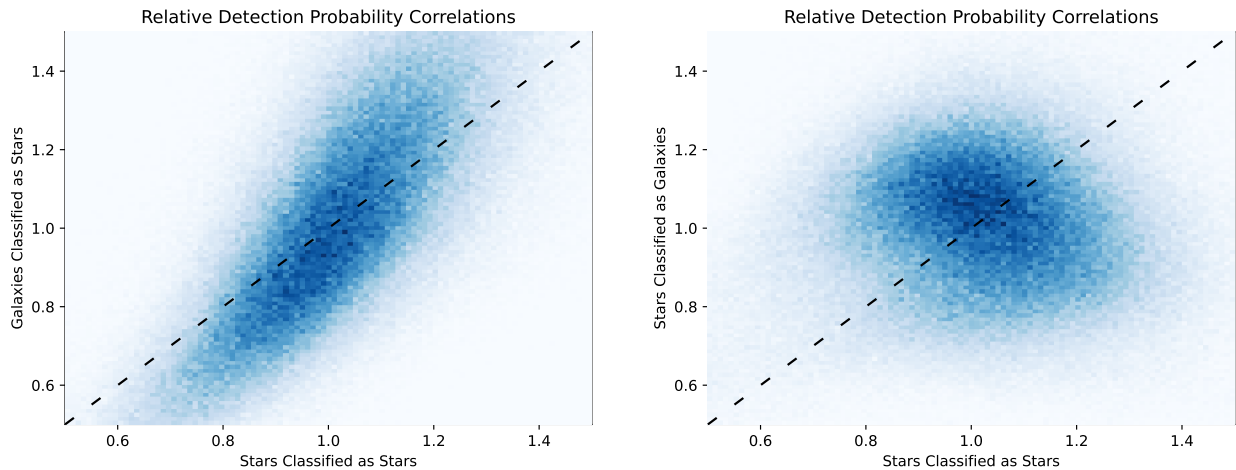
$$FC_S = \frac{C_S}{D_R(C_S, O_S)} = \frac{C_S}{D_R(C_S, O_G)}. \quad (C2)$$

If this were the case, there would be no need to calculate classification probabilities since we would not need  $RO_S$  or  $RO_G$ . As well as this, we would only need to calculate one of  $D_R(C_S, O_S)$  or  $D_R(C_S, O_G)$ .

The correction method that we did not wind up using, originally given in Equation (7), is given by:

$$FC_S = \frac{C_S \cap RO_S}{D_R(C_S, O_S)} + \frac{C_G \cap RO_S}{D_R(C_G, O_S)}. \quad (C3)$$

If the right plot of Figure C1 showed that the two relative detection rates used above were equal, we could simplify this correction in a similar fashion as in Equation (C2). The scatter in both plots shows that neither correction algorithm can be simplified, so all relative detection rates must be calculated.



**Figure C1.** Each plot is using relative detection rates for our faintest magnitude bin. The dotted black line shows where the two rates are equal. (Left) Relative detection rates of stars detected as stars vs. galaxies detected as stars. (Right) Relative detection rates of stars detected as stars vs. stars detected as galaxies.

## Appendix D Algorithm Design Tests

In Section 3, we presented two different correction algorithms: one to estimate what the number of objects classified as a stars would have been with uniform survey properties (we will refer to it as the recovered classified algorithm), and one to estimate what the true number of observed stars would have been with uniform survey properties (recovered observed). The recovered classified algorithm was chosen for use throughout this work. This section will discuss tests performed on the two algorithms that were used in the making of that decision.

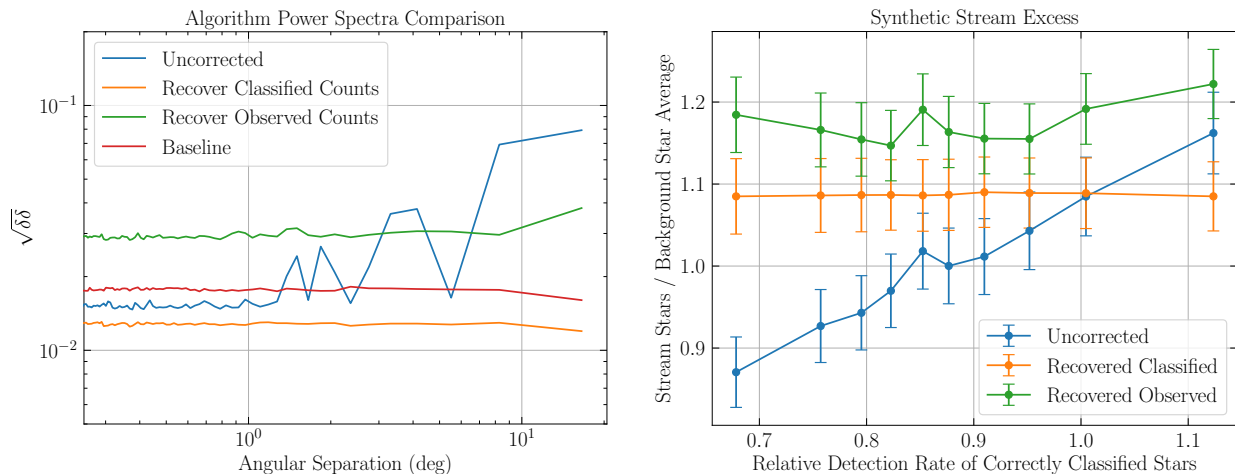
Each test we run uses the same modeling sandbox to generate the counts we analyze. For an area on which to test, we initially chose a  $3 \times 3$  square of 32 resolution healpixels centered at a position of R.A. =  $30^\circ$ , decl. =  $-35^\circ$ . This area was chosen since it contains the depth feature at  $RA \approx 30^\circ$ . This area then has its resolution increased to 512 (giving 2304 total healpixels), since this is the resolution at which we apply our corrections to real data. A subset of these healpixels (which is test dependent) is chosen to hold the simulated stellar stream.

Applying our correction pipeline to this area of sky with real data gives us estimates for the counts in the area for objects of type  $O_S \cap C_S$ ,  $O_G \cap C_S$ ,  $O_S \cap C_G$ , and  $O_G \cap C_G$  for each magnitude bin we use. Injections for each of these four object types occur randomly on healpixels, and detection probabilities for injections are made to be proportional to the relative detection rate for the type of object being injected. This continues until we reach our calculated counts for each type of object. For our simulated stream, an excess percentage of  $O_S \cap C_S$  and  $O_S \cap C_G$  type objects are then injected and then cropped to the stellar stream healpixels. After sampling,  $O_S \cap C_S$  and  $O_G \cap C_S$  counts are summed to get uncorrected star counts, while  $O_S \cap C_G$  and  $O_G \cap C_G$  counts are summed to get uncorrected galaxy counts.

For our first test, the different correction algorithms were tested on their ability to recover the power spectrum of a stellar stream. For this test the overall size of the sandbox was increased to a  $3 \times 9$  rectangle of 32 resolution healpixels

before being increased to a resolution of 512. 432 healpixels were used for the stream ( $3 \times 144$ ) to make a stream that was approximately  $17^\circ$  long, slightly longer than Phoenix. An excess stream percentage over the background of 25% was used for each of the 5000 realizations. As a baseline to compare to, 5000 realizations were done with uniform detection rates and no misclassification. For each of the 5000 realizations, a density power spectrum was calculated. The median of these power spectra for each algorithm is shown in the left plot of Figure D1. The underlying baseline result is shown in red, uncorrected in blue, recovered classified in orange, and recovered observed in green. Both correction algorithms remove most structure from the power spectra, but the recovered classified algorithm results in a median power spectrum that has lower noise than the recovered observed algorithm.

The next test we ran looked at dependencies of stellar stream density as a function of the different relative detection rate variations. For this test, a random 230 of the total 2304 healpixels are chosen to host the stream (the structure of the simulated stream is unimportant for this test). The excess stream percentage was chosen to be 25%. After both correction algorithms are applied, we bin the stellar stream healpixels into 10 bins in terms of each of the 12 relative detection rates used (4 types of relative detection rates in 3 magnitude bins). With this, we check average star counts in each bin and compare it to the average number of stars in the background. An ideal result would be a flat curve (showing no detection rate dependency) at a value of 1.25 (since 25% extra stars were injected). Results of this test are displayed in the right plot of Figure D1. In this plot we look at the stream excess over background as a function of the relative detection rate of correctly classified stars in our faintest magnitude bin. Uncorrected results are shown in blue, recovered classified in orange, and recovered observed in green. For our selected stream healpixels relative detection rates were lower than average, which causes the  $x$ -axis to be centered below 1. The recovered classified curve can be seen to have lower variance while the recovered observed curve is closer to 1.25.



**Figure D1.** (Left) Median power spectra over 5000 realizations for a baseline stream (red), uncorrected data (blue), and corrected data with each correction algorithm (orange and green). The recovered classified counts algorithm shows lower noise levels than the recovered observed counts algorithm. (Right) One-dimensional dependencies of stream density from the relative detection rate of correctly classified stars in our faintest magnitude bin ( $23.9 < r \leq 24.5$ ). 5000 realizations are performed, and the averages are plotted with standard deviation error bars. Both corrected lines show decreased variance.

Both of these tests show the increase in noise when using the recovered observed algorithm. For our application in particular of looking at stellar stream density fluctuations, the power spectrum test is the most useful since power spectra can be used for dark matter constraints. In the power spectrum test in particular, the recovered classified algorithm matched the baseline stream to a higher degree than the recovered observed algorithm. This was our reasoning for using the recovered classified algorithm. It should be noted that for structured streams, if the power of the structure exceeds the noise from the recovered classified algorithm, then the recovered algorithm gives a more accurate value for the power. This could motivate a combination of the two algorithms being used in certain applications.

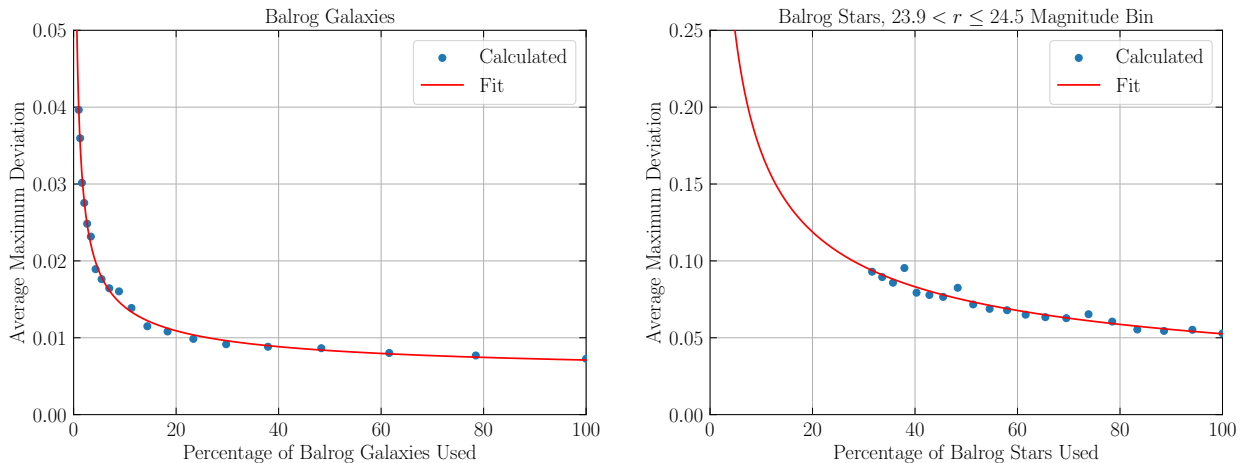
### Appendix E Determining Magnitude Bins

Object magnitude plays a role in detection rate and classification probability variations, so to address this in this work we binned our objects based on magnitudes. As mentioned in Section 4.1, a faint end limit on the  $r$ -band magnitude was placed at 24.5. Figure 9 of S. Everett et al. (2022) claims that detection rates of objects at this magnitude were around 20%, so we decided not to push fainter due to shot noise concerns with such low completion.

When binning based on magnitude, general areas of magnitude space with similar properties made sense to be grouped together. As a general example, bright objects will have nearly perfect detection and classification rates, so that

entire region of magnitude space will have very similar corrections. When detection rates start to fall, this argument no longer applies and finer binning is necessary. However, with a limited number of `Balrog` objects to split among the magnitude bins, more bins leads to less statistical significance for calculated rates. We settled on three bins in an effort to have a magnitude bin for bright objects, a bin for faint objects where detection rates were significantly lower, and a transition bin between those two extremes.

To determine the magnitude bins used, we focused on how effective they would be in `Balrog` Y6, where we assumed we would have access to 5 times as many objects. Since relative detection rates have larger variations than relative classification probabilities, we focused on them for constructing a metric to analyze our magnitude bins. Once a magnitude bin was chosen and we had the appropriate `Balrog` objects, various percentage crops would be applied to simulate different injection counts. We split this smaller sample into an 80% training sample and a 20% testing sample. Training for relative detection rates was performed, and we viewed dependencies from each survey property on relative detection rates within the 20% testing sample. Each survey property would have 1 out of the 10 bins as the furthest away from 1. For this bin we stored the deviation of the relative detection rate from 1 to get an idea of the maximum impact that survey property would have on relative detection rates. We then averaged this value over all survey properties to get a general idea of the deviations one would expect to be caused by survey properties. This average value will be referred to as the average maximum deviation.



**Figure E1.** (Left) Average maximum deviations for relative detection rate variations for `Balrog` galaxies. Higher object counts allowed for probing at lower percentages of the `Balrog` sample. (Right) Average maximum deviations for relative detection rate variations for `Balrog` stars classified as stars in the faintest magnitude bin used.

We found that this average maximum deviation as a function of percentage of `Balrog` objects used could be modeled well by a power-law function, which we would then extrapolate out to 500% to get an idea of accuracy expected from Y6. For relative detection rates, when calculating stellar counts we were only interested in the rates at which stars are classified as stars and galaxies are classified as stars. Galaxies classified as stars are uncommon at bright magnitudes, so our correction for this relative detection rate is not as consistently important across magnitude bins as our correction for stars classified as stars. Due to this, we focused on the relative detection rate of stars classified as stars for our work. Magnitude bins were tested to get similar average maximum deviations for the relative detection rate of stars classified as stars across all three magnitude bins. When using magnitude bins of  $r \leq 22.9$ ,  $22.9 < r \leq 23.9$ , and  $23.9 < r \leq 24.5$ , we found that the average maximum deviation would take values of 2.2%, 2.4%, and 2.5%, respectively, when extrapolated to Y6 object counts. An example of this average maximum deviation as a function of `Balrog` object percentage used is shown in Figure E1. Also shown is a plot generated from detection rates for `Balrog` galaxies classified as either stars or galaxies. This was the largest possible sample of objects that could be used for such a test, which made it useful for determining the form of the fitting function as we could probe lower object percentages.

### ORCID iDs

K.K. Boone <https://orcid.org/0009-0005-7207-3342>  
P.S. Ferguson <https://orcid.org/0000-0001-6957-1627>  
M. Tabbutt <https://orcid.org/0000-0002-0690-1737>  
K. Bechtol <https://orcid.org/0000-0001-8156-0429>  
T.-Y. Cheng <https://orcid.org/0000-0001-8670-4495>  
A. Drlica-Wagner <https://orcid.org/0000-0001-8251-933X>  
C. E. Martínez-Vázquez <https://orcid.org/0000-0002-9144-7726>  
B. Mutlu-Pakdil <https://orcid.org/0000-0001-9649-4815>  
T. M. C. Abbott <https://orcid.org/0000-0003-1587-3931>  
M. Aguena <https://orcid.org/0000-0001-5679-6747>  
O. Alves <https://orcid.org/0000-0002-7394-9466>  
F. Andrade-Oliveira <https://orcid.org/0000-0003-0171-6900>  
D. Bacon <https://orcid.org/0000-0002-2562-8537>  
S. Bocquet <https://orcid.org/0000-0002-4900-805X>  
D. Brooks <https://orcid.org/0000-0002-8458-5047>  
R. Camilleri <https://orcid.org/0000-0002-7436-3950>  
A. Carnero Rosell <https://orcid.org/0000-0003-3044-5150>  
L. N. da Costa <https://orcid.org/0000-0002-7731-277X>  
M. E. da Silva Pereira <https://orcid.org/0000-0002-7131-7684>  
T. M. Davis <https://orcid.org/0000-0002-4213-8783>  
J. De Vicente <https://orcid.org/0000-0001-8318-6813>  
S. Desai <https://orcid.org/0000-0002-0466-3288>  
P. Doel <https://orcid.org/0000-0002-6397-4457>  
S. Everett <https://orcid.org/0000-0002-3745-2882>  
B. Flaugher <https://orcid.org/0000-0002-2367-5049>  
J. Frieman <https://orcid.org/0000-0003-4079-3263>  
J. García-Bellido <https://orcid.org/0000-0002-9370-8360>  
D. Gruen <https://orcid.org/0000-0003-3270-7644>  
G. Gutierrez <https://orcid.org/0000-0003-0825-0517>  
S. R. Hinton <https://orcid.org/0000-0003-2071-9349>  
D. L. Hollowood <https://orcid.org/0000-0002-9369-4157>

K. Honscheid <https://orcid.org/0000-0002-6550-2023>  
D. J. James <https://orcid.org/0000-0001-5160-4486>  
K. Kuehn <https://orcid.org/0000-0003-0120-0808>  
J. L. Marshall <https://orcid.org/0000-0003-0710-9474>  
J. Mena-Fernández <https://orcid.org/0000-0001-9497-7266>  
F. Menanteau <https://orcid.org/0000-0002-1372-2534>  
R. Miquel <https://orcid.org/0000-0002-6610-4836>  
J. Myles <https://orcid.org/0000-0001-6145-5859>  
R. L. C. Ogando <https://orcid.org/0000-0003-2120-1154>  
A. A. Plazas Malagón <https://orcid.org/0000-0002-2598-0514>  
A. Porredon <https://orcid.org/0000-0002-2762-2024>  
M. Rodríguez-Monroy <https://orcid.org/0000-0001-6163-1058>  
E. Sanchez <https://orcid.org/0000-0002-9646-8198>  
D. Sanchez Cid <https://orcid.org/0000-0003-3054-7907>  
I. Sevilla-Noarbe <https://orcid.org/0000-0002-1831-1953>  
M. Smith <https://orcid.org/0000-0002-3321-1432>  
E. Suchyta <https://orcid.org/0000-0002-7047-9358>  
M. E. C. Swanson <https://orcid.org/0000-0002-1488-8552>  
N. Weaverdyck <https://orcid.org/0000-0001-9382-5199>

### References

- Abazajian, K., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2004, *AJ*, 128, 502
- Abbott, T. M. C., Abdalla, F. B., Allam, S., et al. 2018, *ApJS*, 239, 18
- Anbajagane, D., Tabbutt, M., Beas-Gonzalez, J., et al. 2025a, *OJAp*, 8, 65
- Anbajagane, D., Chang, C., Zhang, Z., et al. 2025b, *OJAp*, 8, 46158
- Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, 156, 123
- Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, *ApJ*, 935, 167
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, 558, A33
- Balbinot, E., Yanny, B., Li, T. S., et al. 2016, *ApJ*, 820, 58
- Banik, N., Bovy, J., Bertone, G., Erkal, D., & de Boer, T. J. L. 2021a, *MNRAS*, 502, 2364
- Banik, N., Bovy, J., Bertone, G., Erkal, D., & de Boer, T. J. L. 2021b, *JCAP*, 2021, 043
- Bechtol, K., Birrer, S., Cyr-Racine, F.-Y., et al. 2022, arXiv:2203.07354
- Bechtol, K., Drlica-Wagner, A., Balbinot, E., et al. 2015, *ApJ*, 807, 50
- Bernard, E. J., Ferguson, A. M. N., Schlafly, E. F., et al. 2014, *MNRAS*, 443, L84
- Bonaca, A., Hogg, D. W., Price-Whelan, A. M., & Conroy, C. 2019, *ApJ*, 880, 38
- Bonaca, A., Pearson, S., Price-Whelan, A. M., et al. 2020, *ApJ*, 889, 70
- Bonaca, A., & Price-Whelan, A. M. 2025, *NewAR*, 100, 101713
- Bonaca, A., Naidu, R. P., Conroy, C., et al. 2021, *ApJL*, 909, L26
- Boone, K. 2026, SSI Corrections DES Y3: v1.0.0, Zenodo, doi:10.5281/zenodo.18662142
- Bovy, J., Erkal, D., & Sanders, J. L. 2017, *MNRAS*, 466, 628
- Bressan, A., Marigo, P., Girardi, L., et al. 2012, *MNRAS*, 427, 127
- Buckley, M. R., & Peter, A. H. G. 2018, *PhR*, 761, 1
- Bullock, J. S., & Boylan-Kolchin, M. 2017, *ARA&A*, 55, 343
- Chabanier, S., Millea, M., & Palanque-Delabrouille, N. 2019, *MNRAS*, 489, 2247
- Dekker, A., Ando, S., Correa, C. A., & Ng, K. C. Y. 2022, *PhRvD*, 106, 123026
- Delos, M. S., & Schmidt, F. 2022, *MNRAS*, 513, 3682
- Diehl, H. T., Neilsen, E., Gruendl, R., et al. 2016, *Proc. SPIE*, 9910, 99101D
- Drlica-Wagner, A., Bechtol, K., Mau, S., et al. 2020, *ApJ*, 893, 47
- Drlica-Wagner, A., Carlin, J. L., Nidever, D. L., et al. 2021, *ApJS*, 256, 2
- Erkal, D., Belokurov, V., Bovy, J., & Sanders, J. L. 2016, *MNRAS*, 463, 102
- Erkal, D., Koposov, S. E., & Belokurov, V. 2017, *MNRAS*, 470, 60
- Everett, S., Yanny, B., Kuropatkin, N., et al. 2022, *ApJS*, 258, 15
- Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, *AJ*, 150, 150
- Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, *ApJ*, 622, 759
- Grillmair, C. J. 2017, *ApJ*, 847, 119
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, 585, 357
- Hartley, W. G., Choi, A., Amon, A., et al. 2021, *MNRAS*, 509, 3547

- Hunter, J. D. 2007, *CSE*, **9**, 90
- Ivezić, Z., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, **873**, 111
- Jethwa, P., Erkal, D., & Belokurov, V. 2018, *MNRAS*, **473**, 2060
- Koposov, S. E., Belokurov, V., Li, T. S., et al. 2019, *MNRAS*, **485**, 4726
- Küpper, A. H. W., Kroupa, P., Baumgardt, H., & Heggie, D. C. 2010, *MNRAS*, **401**, 105
- Li, T. S., Koposov, S. E., Zucker, D. B., et al. 2019, *MNRAS*, **490**, 3508
- Li, T. S., Koposov, S. E., Erkal, D., et al. 2021, *ApJ*, **911**, 149
- Li, T. S., Ji, A. P., Pace, A. B., et al. 2022, *ApJ*, **928**, 30
- Lu, J., Lin, T., Sholapurkar, M., & Bonaca, A. 2026, *OJAp*, **9**, 55386
- Malhan, K., Ibata, R. A., & Martin, N. F. 2018, *MNRAS*, **481**, 3442
- Mateu, C. 2023, *MNRAS*, **520**, 5225
- Nadler, E. O., Wechsler, R. H., Bechtol, K., et al. 2020, *ApJ*, **893**, 48
- Newberg, H. J., & Carlin, J. L. 2016, *Tidal Streams in the Local Group and Beyond*, 420 (Springer)
- Newton, O., Leo, M., Cautun, M., et al. 2021, *JCAP*, **2021**, 062
- Odenkirchen, M., Grebel, E. K., Rockosi, C. M., et al. 2001, *ApJL*, **548**, L165
- Patrick, J. M., Koposov, S. E., & Walker, M. G. 2022, *MNRAS*, **514**, 1757
- Rodríguez-Monroy, M., Weaverdyck, N., Elvin-Poole, J., et al. 2022, *MNRAS*, **511**, 2665
- Sevilla-Noarbe, I., Bechtol, K., Carrasco Kind, M., et al. 2021, *ApJS*, **254**, 24
- Shipp, N., Drica-Wagner, A., Balbinot, E., et al. 2018, *ApJ*, **862**, 114
- Shipp, N., Li, T. S., Pace, A. B., et al. 2019, *ApJ*, **885**, 3
- Suchyta, E., Huff, E. M., Aleksić, J., et al. 2016, *MNRAS*, **457**, 786
- Tabbutt, M. 2023, Madison PhD thesis, Univ. Wisconsin
- Tavangar, K., Ferguson, P., Shipp, N., et al. 2022, *ApJ*, **925**, 118
- Treu, T. 2010, *ARA&A*, **48**, 87
- Tsiane, K., Mau, S., Drica-Wagner, A., et al. 2025, *OJAp*, **8**, 89
- Vegetti, S., Birrer, S., Despali, G., et al. 2024, *SSRv*, **220**, 58
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, **17**, 261
- Wan, Z., Lewis, G. F., Li, T. S., et al. 2020, *Natur*, **583**, 768
- Zonca, A., Singer, L., Lenz, D., et al. 2019, *JOSS*, **4**, 1298