

# From the Inner to Outer Milky Way: A Photometric Sample of 2.6 Million Red Clump Stars

Madeline Lucey<sup>1\*</sup>, Yuan-Sen Ting<sup>2,3,4,5†</sup>, Nesar S. Ramachandra<sup>6,7</sup>, Keith Hawkins<sup>1</sup>

<sup>1</sup>Department of Astronomy, the University of Texas at Austin, 2515 Speedway Boulevard, Austin, TX 78712, USA

<sup>2</sup>Institute for Advanced Study, Princeton, NJ 08540, USA

<sup>3</sup>Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA

<sup>4</sup>Observatories of the Carnegie Institution of Washington, 813 Santa Barbara Street, Pasadena, CA 91101, USA

<sup>5</sup>Research School of Astronomy & Astrophysics, Australian National University, Cotter Rd., Weston, ACT 2611, Australia

<sup>6</sup>High Energy Physics Division, Argonne National Laboratory, Lemont, IL 60439, USA

<sup>7</sup>Kavli Institute for Cosmological Physics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA

Accepted .... Received ...; in original form ...

## ABSTRACT

Large pristine samples of red clump stars are highly sought after given that they are standard candles and give precise distances even at large distances. However, it is difficult to cleanly select red clumps stars because they can have the same  $T_{\text{eff}}$  and  $\log g$  as red giant branch stars. Recently, it was shown that the asteroseismic parameters,  $\Delta P$  and  $\Delta \nu$ , which are used to accurately select red clump stars, can be derived from spectra using the change in the surface carbon to nitrogen ratio ( $[C/N]$ ) caused by mixing during the red giant branch. This change in  $[C/N]$  can also impact the spectral energy distribution. In this study, we predict the  $\Delta P$ ,  $\Delta \nu$ ,  $T_{\text{eff}}$  and  $\log g$  using 2MASS, AllWISE, *Gaia*, and Pan-STARRS data in order to select a clean sample of red clump stars. We achieve a contamination rate of  $\sim 20\%$ , equivalent to what is achieved when selecting from  $T_{\text{eff}}$  and  $\log g$  derived from low resolution spectra. Finally, we present two red clump samples. One sample has a contamination rate of  $\sim 20\%$  and  $\sim 405,000$  red clump stars. The other has a contamination of  $\sim 33\%$  and  $\sim 2.6$  million red clump stars which includes  $\sim 75,000$  stars at distances  $> 10$  kpc. For  $|b| > 30$  degrees we find  $\sim 15,000$  stars with contamination rate of  $\sim 9\%$ . The scientific potential of this catalog for studying the structure and formation history of the Galaxy is vast given that it includes millions of precise distances to stars in the inner bulge and distant halo where astrometric distances are imprecise.

**Key words:** stars: distances, techniques: photometric, stars: evolution

## 1 INTRODUCTION

Distances are one of the most important and yet hardest measurements to make in astronomy. Distances to stars are especially important for studying the formation and the chemodynamical evolution of the Galaxy. The *Gaia* mission aims to provide parallax measurements for billions of stars in order to produce a 3D map of the Galaxy (Gaia Collaboration et al. 2016). However, the *Gaia* mission will be plagued by large uncertainties in the distant Galaxy. Through simple uncertainty propagation, the uncertainty on distance derived from parallax scales as distance squared<sup>1</sup>. However,

standard candles, such as red clump (RC) stars, where the distance is derived using the distance modulus have uncertainties that are linear with distance<sup>2</sup>. With high precision photometry, RC stars can provide distances with uncertainties  $< 6\%$  at distances up to  $\sim 10$  kpc (Bovy et al. 2014; Hawkins et al. 2017; Ting et al. 2018a). Therefore, standard candles can provide more precise distances for objects in the distant Galaxy than parallaxes and are an excellent complement to the *Gaia* catalog.

In this context, RC stars are low mass stars ( $< 2.5 M_{\odot}$ ) that have undergone the helium flash to ignite helium burning in their cores. The electron-degenerate cores prior to he-

\* E-mail: m\_lucey@utexas.edu

† NASA Hubble Fellow

<sup>1</sup>  $\sigma_d = \sqrt{\left(\frac{\partial d}{\partial \varpi}\right)^2 \sigma_{\varpi}^2} = \frac{\sigma_{\varpi}}{\varpi^2} = \sigma_{\varpi} d^2$  where  $d$  is the distance,  $\varpi$  is the parallax and  $d = 1/\varpi$

<sup>2</sup>  $\sigma_d = \sqrt{\left(\frac{\partial d}{\partial m}\right)^2 \sigma_m^2} = \sigma_m \ln(10) 10^{(m-M+5)/5} = \sigma_m \ln(10) d$  where  $m$  is apparent magnitude,  $M$  is the absolute magnitude and  $d = 10^{(m-M+5)/5}$

lium ignition causes RC stars to have roughly the same core mass. Therefore, the luminosity of RC stars is only weakly dependent on mass and metallicity (Girardi 2016, and references therein). Beginning in the late 1990s, the nearly constant luminosity of RC stars has been used to determine distances, making RC stars a commonly used standard candle. Selecting from color-magnitude diagrams, Stanek et al. (1997) used RC stars to constrain the shape and orientation of the Galactic bar. Stanek et al. (1998) used a similar technique to determine the distance to M31 using RC stars. However, there are many sciences cases when a more pristine sample of RC stars is desired which can be difficult to select. Bovy et al. (2014) achieved a contamination rate of  $\sim 10\%$  using effective temperature ( $T_{\text{eff}}$ ) and surface gravity ( $\log g$ ) from the high-resolution spectra of the Apache Point Observatory Galactic Evolution Experiment (APOGEE) survey (Majewski et al. 2017). This contamination is from red giant branch (RGB) stars which have inert helium cores with a hydrogen burning shell but can have similar  $T_{\text{eff}}$  and  $\log g$  as RC stars (Iben 1968).

Asteroseismology has proven to be the most accurate method for selecting RC stars (Montalbán et al. 2010; Bedding et al. 2011; Mosser et al. 2011, 2012; Stello et al. 2013; Mosser et al. 2014; Pinsonneault et al. 2014; Vrad et al. 2016; Elsworth et al. 2017). The average large frequency spacing ( $\Delta\nu$ ) goes as the square root of the mean stellar density (Chaplin & Miglio 2013). As such, it has been shown that the distribution of  $\Delta\nu$  values for RGB and RC stars are distinct (Miglio et al. 2009; Mosser et al. 2010). RC stars are more constrained in this space with  $\Delta\nu < 5 \mu\text{Hz}$ . While, RGB stars can have  $\Delta\nu > 20 \mu\text{Hz}$ . Furthermore, evolved stars have been shown to have coupling between the gravity waves in the dense radiative core (g-modes) and the acoustic waves in the envelope, p-modes (Beck et al. 2011). RC stars have a lower core density than RGB stars which causes stronger coupling between g- and p-modes. Therefore, RC and RGB clearly separate in the period spacing ( $\Delta P$ ) of these mixed modes with RC stars typically having  $\Delta P > 200$  s and RGB stars having  $\Delta P < 100$  s (Bedding et al. 2011; Mosser et al. 2011; Stello et al. 2013; Mosser et al. 2014). However,  $\Delta P$  is a difficult measurement and has only been measured for a fraction of the *Kepler* and *CoRoT* samples (Girardi 2016).

Recently, Hawkins et al. (2018) showed the  $\Delta P$  and the  $\Delta\nu$  spacing can be inferred from a stellar spectrum. It has long been thought that the carbon to nitrogen ratio,  $[C/N]$ , would be different in RC stars relative to RGB stars because of mixing that occurs along the upper RGB phase (Martell et al. 2008; Masseron & Gilmore 2015; Masseron & Hawkins 2017; Masseron et al. 2017). Hawkins et al. (2018) showed the carbon and nitrogen bands in spectra from the APOGEE survey can be used to infer the  $\Delta P$  and  $\Delta\nu$  and therefore, used to select RC stars. Building off of this work, Ting et al. (2018a) presented a catalog of  $\sim 100,000$  RC stars from the APOGEE and LAMOST catalogs. However,  $> 70\%$  of that sample is within 3 kpc of the Sun where *Gaia* gives more precise distances (Ting et al. 2018a). This represents one of the largest limitations to recent spectroscopic surveys. Spectra require more flux than photometry and therefore high resolution ( $R = \lambda/\Delta\lambda \sim 22,500$ ) and high signal-to-noise (S/N) data can only be achieved for nearby bright stars as. For example, the magnitude limit of the APOGEE survey is  $H=$

13.8 mag (Majewski et al. 2017). Given that the RC absolute magnitude in the H-band is -1.46 mag (Hawkins et al. 2017) the APOGEE survey is limited to RC stars within 12 kpc, assuming not extinction. In the case of the LAMOST survey, which has much lower resolution ( $R \sim 1800$ ), the magnitude limit is  $r=18.5$  mag which corresponds to a RC distance of  $\sim 40$  kpc given the RC absolute magnitude in the r-band of 0.55 mag (Ruiz-Dern et al. 2018). However, these observation require long exposure times (5400 s) and therefore are number limited. Photometric surveys are able to achieve fainter magnitudes for much greater number of stars. For example, the RC catalog derived using LAMOST spectra has  $\sim 2,000$  RC stars with distances  $> 10$  kpc (Ting et al. 2018a) while we find  $\sim 75,000$  RC stars with distances  $> 10$  kpc using photometric surveys.

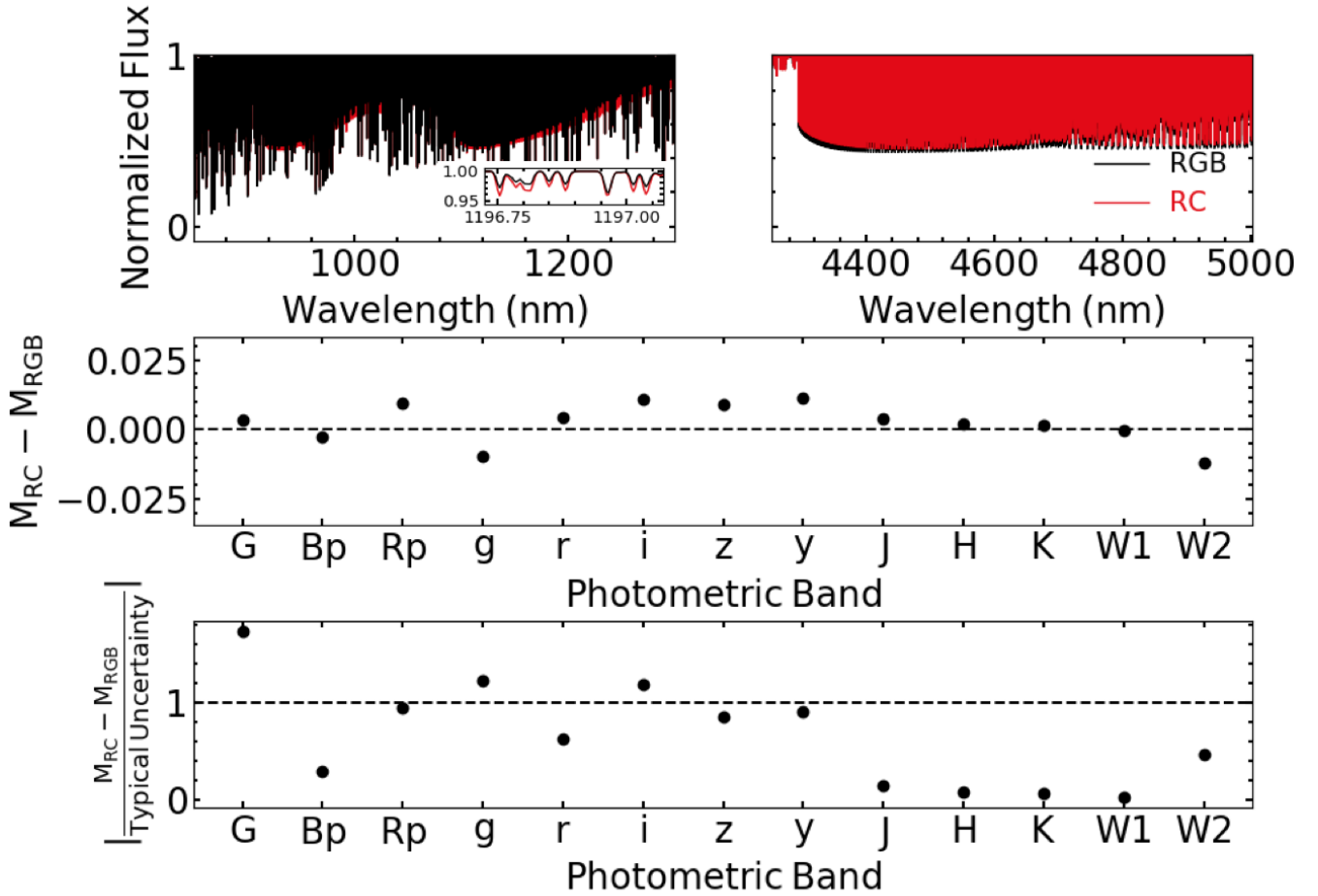
In this work we aim to make use of the vast amount of available photometry to obtain the largest and most distant sample of RC stars yet. In Section 2 we describe the photometric selection we use to create the spectral energy distribution (SED). In Section 3, we describe the method we develop to infer the  $T_{\text{eff}}$ ,  $\log g$ ,  $\Delta P$  and  $\Delta\nu$  from the SEDs using a neural network based method and how we use those parameters to select a sample of RC stars. In Section 4, we present the catalog of RC stars. In Section 5, we demonstrate the interpretability of the neural network and try to confirm the physical mechanism used for the inference. Finally, we summarize the results in Section 6.

## 2 DATA

It is reasonable to assume that if the information to accurately select RC stars is in the spectra, then the same information is the SED just with a much weaker signal. In Figure 1, we show the theoretical differences in magnitudes between an RGB and RC star. These values are calculated using PYPHOT<sup>3</sup> and the synthetic spectra shown in the upper panel of Figure 1. The spectra are synthesized in 1D local thermodynamic equilibrium (LTE) using ATLAS12 and SYNTHE maintained by R. Kurucz (Kurucz 1970; Kurucz & Avrett 1981; Kurucz 1993, 2005, 2013, 2017, and references therein). We use a resolution of  $R \sim 300,000$  and the line list provided by R. Kurucz<sup>4</sup>. Both spectra are calculated at  $T_{\text{eff}} = 4875$  K,  $\log g = 2.32$  dex and  $Z = 0.004$ . The RGB spectra has  $[C/N] = 0$  while the RC spectra has  $[C/N] = -0.79$ . These  $T_{\text{eff}}$ ,  $\log g$  and  $[C/N]$  values are the theoretical values of a RC star with mass of  $1.00 M_{\odot}$  and  $Z = 0.004$  (Lagarde et al. 2012). The change in the  $[C/N]$  ratio is caused by an increase in the nitrogen abundance as well as a decrease in the carbon abundance. In addition to this change, the CNO cycle causes a slight decrease in the oxygen abundance. This decrease in the oxygen abundance will decrease the abundance of CO molecules which releases carbon atoms. Combined with the increase in nitrogen, this leads to an increase in the abundance of CN molecules (Ting et al. 2018b). Therefore, specific photometric bands will experience higher magnitudes due to more absorption from CN molecules. Additionally, the magnitude for other bands will

<sup>3</sup> <https://mfouesneau.github.io/docs/pyphot>

<sup>4</sup> <http://kurucz.harvard.edu>



**Figure 1.** In the top panel we compare the synthetic spectra of an RGB and RC star. Specifically, we show two CN molecular bands which impact the y and J photometry (top left) and a CO band which impacts the W2 photometry (top right). In the top left panel we also show a zoomed-in spectrum. Both stars are synthesized with  $T_{\text{eff}} = 4875$  K,  $\log g = 2.32$  dex and  $Z=0.004$ . The RGB star has  $[C/N] = 0$  while the RC star has  $[C/N]=-0.79$ . The increase in the N abundance causes more absorption in the CN bands which increase the magnitudes for the z, y and J passbands. On the other hand, the decrease in C causes less absorption by CO molecules which decreases the W2 magnitude. In the middle panel, we show the impacts of a change in  $[C/N]$  of -0.79 on each photometric passbands used in this work. These values are calculated using the synthetic spectra shown above. In the lower panel we show how the change in magnitudes shown in the middle panel relate to the typical uncertainty in that passband. Most bands have the  $[C/N]$  signal on the order of the uncertainties making it marginally detectable.

increase due to less absorption from CO molecules. The decrease in oxygen is a much smaller effect and is not taken into account in Figure 1. Therefore, our estimations in the change of magnitudes between RGB and RC stars are likely slightly underestimated. As shown in the lower panel of Figure 1, the difference in magnitude is typically on the order of the uncertainties (0.002, 0.010, 0.010, 0.008, 0.007, 0.009, 0.011, 0.012, 0.026, 0.027, 0.026, 0.026, 0.026 mag for G, BP, RP, g, r, i, z, y, J, H, K, W1 and W2, respectively). This makes it challenging to pick up this signal. Therefore, we use a neural network which is flexible and sensitive enough to detect the photometric signal of a  $[C/N]$  change in RC stars. We confirm the network harnesses this information to perform the inference in Section 5.2.

In order to train the network and evaluate the method we require a set of data with known spectroscopic and asteroseismic parameters. We describe the data set we use to

train and validate the network in Section 2.1. The validation set is used to evaluate the accuracy of the neural network. It is comprised of data with known spectroscopic and asteroseismic parameters that were never exposed to the network for the training. Therefore, it can be used to determine the network's precision on the data we wish to infer parameters for. This set is also described in Section 2.1. In Section 2.2 we describe the photometric data that for which we infer  $T_{\text{eff}}$ ,  $\log g$ ,  $\Delta P$  and  $\Delta \nu$  in order to select RC stars.

## 2.1 Training and Validation Data

The accuracy of neural networks highly depends on the size of the training set. Specifically, neural networks typically outperform other machine learning algorithms when the training data set is large. Therefore, we choose to use the large catalog of spectroscopically derived  $\Delta P$  and  $\Delta \nu$  pa-

rameters from Ting et al. (2018a) rather than smaller catalogs of asteroseismically derived parameters. The catalog from Ting et al. (2018a) is derived using a data-driven neural network which finds a mapping between the normalized spectral fluxes to the asteroseismic parameters  $\Delta P$  and  $\Delta \nu$ . It then infers on LAMOST DR3 (Xiang et al. 2017) spectra that are within a 3D convex hull<sup>5</sup> of  $T_{\text{eff}}$ ,  $\log g$  and  $[\text{Fe}/\text{H}]$  values that are within the parameter range of the training set used (see Figure 3 in Ting et al. 2018a). For our training set we include only high quality  $\Delta P$  and  $\Delta \nu$  measurements from this catalog, requiring the spectra to have signal-to-noise per pixel  $S/N_{\text{pix}} > 75 \text{ pixel}^{-1}$ . With this quality cut, the training sample has a contamination rate of  $\sim 3\%$  (Ting et al. 2018a). We train the  $\Delta P$  and  $\Delta \nu$  network on a subset of 30,000 stars from this sample. The rest serves as a validation set. We include data with  $S/N_{\text{pix}} < 75 \text{ pixel}^{-1}$  in the validation set of 100,000 of stars to determine the accuracy of the inference on lower S/N data. Since the LAMOST sample has a magnitude limit that corresponds to a RC distance that is greater than that given by the AllWISE magnitude limit, this test sample spans the entire distance range of our data set for inference. Therefore, the validation sample is representative of our final sample.

In addition to this training set we require another training set to determine photometric  $T_{\text{eff}}$  and  $\log g$  outside of the convex hull in order to separate the dwarf stars from the giant stars in the data. We adopt the  $T_{\text{eff}}$  and  $\log g$  values from the LAMOST DR3 stellar catalog as our training and validation sets for these separate networks (see Section 3.2). Here, we require the training set of 200,000 to have LAMOST g-band  $S/N > 50$ . Again, our validation set of 900,000 stars does not have a S/N cut in order to cover the same magnitude range as our final sample.

With both training sets we perform a sky crossmatch with *Gaia* DR2. We then use the provided *Gaia* DR2 crossmatches (Marrese et al. 2019) to obtain the photometry from AllWISE, 2MASS, and Pan-STARRS1.

## 2.2 SEDs & Parallaxes for Inference

For the inference, we only require photometry and parallax. Therefore, our sample is much larger than spectroscopic samples given that spectra require longer exposures because of the need for more flux than imaging. We can also take advantage of the hundreds of millions of stars that already have photometry that spans from the near-UV to the mid-infrared from large surveys.

In this work we make use of data from *Gaia* DR2 (Gaia Collaboration et al. 2018), Pan-STARRS1 (Chambers et al. 2016), 2MASS (Skrutskie et al. 2006) and AllWISE (Wright et al. 2010; Mainzer et al. 2011) photometric catalogs, along with *Gaia* DR2 parallaxes. We include all the pass bands from *Gaia* (G, BP, RP), Pan-STARRS (g,r,i,z,y) and 2MASS (J, H, K). We use only the two bluest AllWISE bands, W1 and W2. The two other bands, W3 and W4, are shallower, have lower spatial resolution and do not contain much additional information about the SED of RC stars. We also

make use of the provided *Gaia* DR2 crossmatches with Pan-STARRS1, 2MASS, and AllWISE which take into account the motions of the targets and the varying epochs of the different surveys (Marrese et al. 2019).

We perform multiple quality cuts to ensure we only use accurate photometry and parallaxes:

- (i)  $1.0 + 0.015(\text{BP} - \text{RP})^2 < \text{phot\_bp\_rp\_excess\_factor} < 1.3 + 0.060(\text{BP} - \text{RP})^2$  (Arenou et al. 2018; Evans et al. 2018)
- (ii) RUWE (renormalized unit weight error)  $< 1.4$  (Lindgren 2018)
- (iii) “A” quality photometry from 2MASS and AllWISE
- (iv)  $(\sigma_G^6 - 0.0018)/G < 0.0002$  to remove nearby stars that are near saturation in *Gaia* DR2

We do not apply a quality cut to the Pan-STARRS (PS1) catalog because the Image Processing Pipeline (IPP) (Magnier et al. 2016) achieves relative photometric calibration better than 1% for much of the sky (Schlafly et al. 2012).

Since neural networks perform poorly on data that is dissimilar from what it has been trained on –i.e., the network is not portable and cannot extrapolate, we create a convex hull in the photometric and parallax space based on the validation set and do not infer on any data that does not fall within that convex hull. To create the convex hull we randomly select 10,000 stars from our training set and use the BP, g, K, W1 and W2 bands along with the parallax. As the convex hull algorithm can only handle up to 6 dimensions we choose a subset of the photometric bands to use. Bp, g, K, W1 and W2 are selected because we want to span the entire wavelength range while also prioritizing bands important for the  $\Delta P$  inference (see Figure 1) and infrared bands that will be least impacted by extinction. Although, we find that changes in the band selection has an insignificant impact on the final red clump sample.

## 3 METHOD

From synthetic photometry, we know that the signal of the evolved  $[\text{C}/\text{N}]$  ratio that we hope to use to infer the  $\Delta P$  is much smaller than the effects of  $T_{\text{eff}}$  and extinction on the photometry. It is also on the order of the noise in the photometry (see Figure 1). Given these constraints and access to a large training set, we find a neural network with probabilistic inference to be an effective tool to infer the parameters from the photometry. In this work, we infer the  $\Delta P$ ,  $\Delta \nu$ ,  $T_{\text{eff}}$ , and  $\log g$  of stars from 13 photometric bands and *Gaia* DR2 parallaxes using Mixture Density Networks (MDNs; Bishop 1994). We use 6 separate networks, 2 each for  $T_{\text{eff}}$  and  $\log g$  for all stars and for just giant stars and 2 for  $\Delta P$  and  $\Delta \nu$  of giant stars. We defer the simultaneous inference of multiple parameters with a single MDN which will better account for covariances to a future study.

<sup>5</sup> The convex hull is the smallest polygon that contains all of the data. For more information on convex hulls see Ting et al. (2016)

<sup>6</sup>  $\sigma_G = \sqrt{(\frac{\sigma_I}{I})^2 + \sigma_{ZP}^2}$  where  $\frac{\sigma_I}{I}$  is the mean G band error over the flux and  $\sigma_{ZP}$  is the zero-point error which for *Gaia* DR2 is 0.0018 mag



### 3.1 Mixture Density Network

A MDN is a neural network where the outputs parametrize a Gaussian mixture model (GMM):

$$p(\theta|x) = \sum_{j=1}^n \varpi_j(x) \mathcal{N}(\mu_j(x), \sigma_j(x)) \quad (1)$$

where one can choose the number of components ( $n$ ). Unlike a typical GMM, the weights ( $\varpi(x)$ ), and widths ( $\sigma(x)$ ) along with the means ( $\mu(x)$ ) of each component are a function of the input parameters ( $x$ ) which in our case is the 13-band SED plus parallax. In other words, given an SED and parallax ( $x$ ), we expect the output parameter ( $\theta$ ), in this case  $T_{\text{eff}}$ ,  $\log g$ ,  $\Delta P$ , or  $\Delta \nu$ , can span a distribution, and we try to model that distribution using the training data, essentially doing a conditional density estimation of the training set. Having estimated such density during the training, when the network is exposed to new input data for the inference, we can then read off the probability density of the output parameter ( $T_{\text{eff}}$ ,  $\log g$ ,  $\Delta P$ , or  $\Delta \nu$ ). Thus the outputs of our networks ( $\varpi(x)$ ,  $\sigma(x)$ , and  $\mu(x)$ ) parameterize a GMM which acts as a probability distribution function (PDF) and we use the negative log likelihood as the loss function. In other words, we train the network to maximize the sum of the log likelihoods of the output PDFs given the training data. This method differs from [Ting et al. \(2018a\)](#) in that we are essentially doing a density estimation of the training set with a neural network while [Ting et al. \(2018a\)](#) performs non-linear regression with their neural network.

The advantage of using a MDN over a standard neural network is that the GMM output is probabilistic instead of single best-estimate and therefore gives us the ability to estimate uncertainties for our inferred parameters. However, these uncertainty estimates assume there is no uncertainty in the input parameters (SED and parallax). Hence, we do not perform any input uncertainty propagation. In practice, we can also integrate the input parameter uncertainties to get the full posterior. For the purpose of this study, in which we are more focused on classifying RC stars, our reported uncertainties do not impact the classification and therefore we do not calculate the full posterior. Therefore, our uncertainties are a reflection of the training set density distribution. However, since the training set density distribution has noise, the uncertainty estimate does capture some of the input noise (see Section 5.1). An additional benefit to the MDN is that we can use a two-component GMM output for the  $\Delta P$  network in order to encapsulate the bimodality of the posterior distribution of  $\Delta P$ . Otherwise we use a one-component network for the  $T_{\text{eff}}$ ,  $\log g$ , and  $\Delta \nu$  inference. When we use a one component model, the mean ( $\mu(x)$ ) and width ( $\sigma(x)$ ) maps directly to the inferred values and uncertainty. However for the  $\Delta P$  network when we use two-component, the inferred value is the weighted mean of the means ( $\sum_{j=1}^2 \omega_j(x) \mu_j(x)$ ) and the adopted uncertainty is the weighted mean of the widths ( $\sum_{j=1}^2 \omega_j(x) \sigma_j(x)$ ). Besides the number of components of the output GMMs, each of the MDNs use the same architecture. We use four fully connected layers with 32, 24, 16 and 8 nodes to map from the 14 input parameters (SED + parallax) to the  $3 \times n$  outputs that parametrize the GMM ( $\omega(x)$ ,  $\sigma(x)$  and  $\mu(x)$ ). We apply

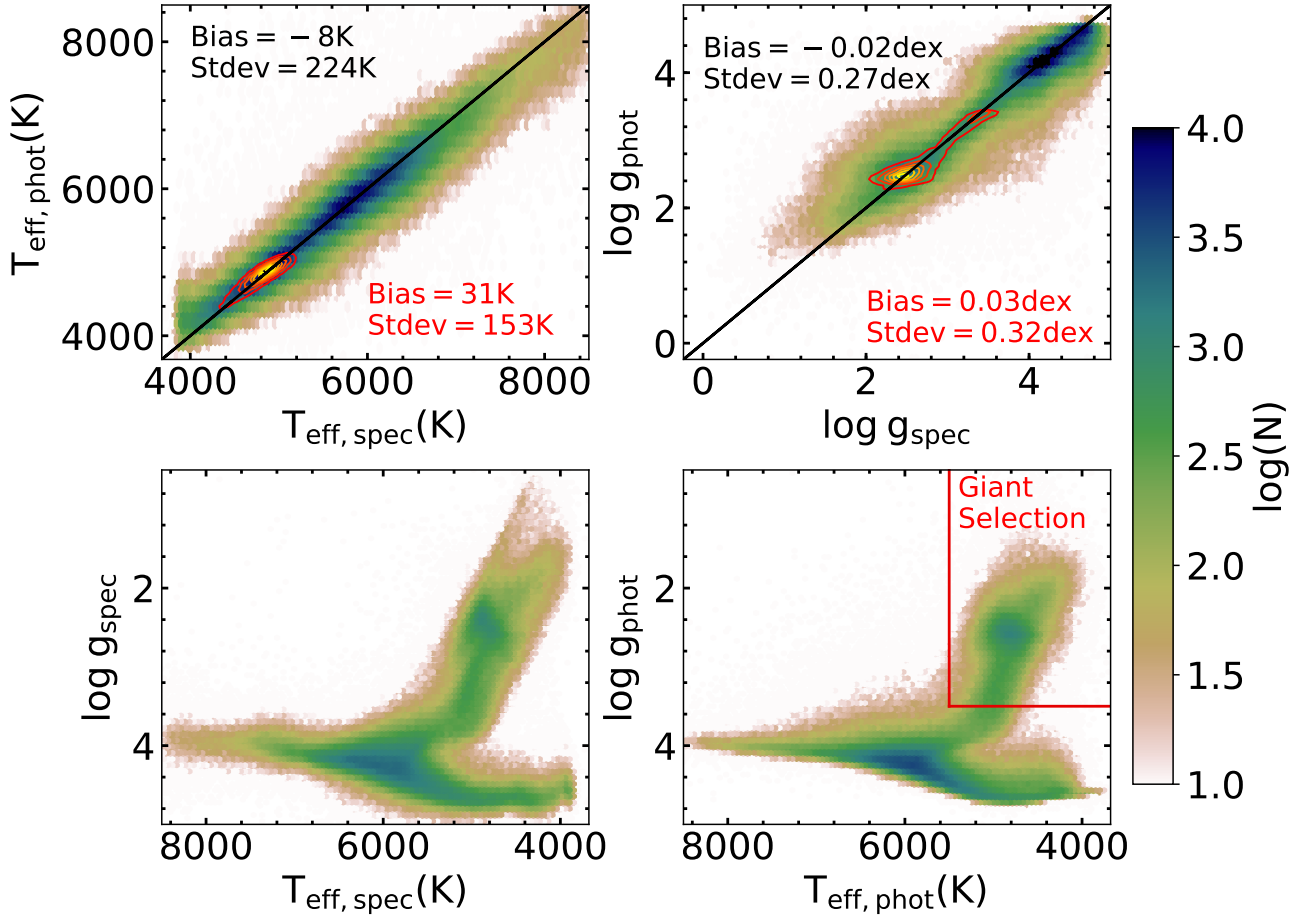
a sigmoid activation function<sup>7</sup> to each node to capture the non-linear mapping from the inputs to the outputs.

### 3.2 Inferring $T_{\text{eff}}$ and $\log g$ and Selecting Giant Stars

We first infer the  $\log g$  and  $T_{\text{eff}}$  of all of the stars in the sample in order to select the giant stars. To do this, we train two MDNs using a random sample of 200,000 stars from the LAMOST DR3 catalog which pass the quality cuts (see Section 2.1). Both of these networks have one mixture component as the output so the inferred value is the mean of the Gaussian and the uncertainty is the width. We show a comparison of the derived values with the LAMOST results for 900,000 validation stars in Figure 2. In order to quantitatively describe the accuracy and precision of the inferred parameters, we report the bias ( $\langle \text{LAMOST} - \mu(x) \rangle$ ), the standard deviation ( $\langle (\text{LAMOST} - \mu(x))^2 \rangle$ ) and the typical uncertainty ( $\langle \sigma(x) \rangle$ ). However, as we noted in Section 3.1, the uncertainty estimate does not fully take into account the uncertainty on the input data. Therefore, it is likely that our uncertainties are underestimated, especially given that the network is trained on higher S/N data. The standard deviation from external validation is a more accurate estimate of the precision. Nonetheless, as we will see, the standard deviation is typically on the order of the uncertainty estimate  $\langle \sigma(x) \rangle \approx \langle (\text{LAMOST} - \mu(x))^2 \rangle$ . This shows that, since the training input is noisy, the adopted uncertainty does capture some of the input uncertainty. The bias for photometric  $T_{\text{eff}}$  values is -8 K with a standard deviation of 224 K. The typical uncertainty on  $T_{\text{eff}}$  is 200 K which is reasonable given that the typical uncertainty for  $T_{\text{eff}}$  in the LAMOST catalog is 168 K. We also find a bias for photometric  $\log g$  values of -0.02 dex with a standard deviation of 0.27 dex. The typical  $\log g$  uncertainty is 0.23 dex which is again consistent with the typical uncertainty for  $\log g$  in the LAMOST catalog is 0.24 dex. The lack of bias in the  $T_{\text{eff}}$  and  $\log g$  is expected given that we are using a data-driven method and the parameters are tied to the scale of the training set. After inferring on all of the data that passes the quality cuts (see Section 2.2), we select stars with inferred  $\log g < 3.5$  dex and  $2500 \text{ K} < T_{\text{eff}} < 5500 \text{ K}$  as giant stars (see Figure 2).

Once the giant sample is created we refine the  $T_{\text{eff}}$  and  $\log g$  inference by using a network trained on the spectroscopic parameters of just giant stars. This network is trained on a subset of 200,000 giant stars selected using the method described above. We use the initial photometrically inferred parameters to make the giant selection for the training set to ensure our training set will be similar to the one we will use for the inference, which will only have photometrically inferred parameters. Again, we use one mixture component as the output so the inferred value is the mean of the Gaussian and the uncertainty is the width. The red contours in Figure 2, demonstrate how the inference has been refined. These new photometric  $T_{\text{eff}}$  values have a bias of 31 K with a standard deviation of 153 K. The typical  $T_{\text{eff}}$  uncertainty is 97 K. The bias of photometric  $\log g$  values is 0.03 dex with

<sup>7</sup> The choice in activation function is rather arbitrary here given the low-dimensionality of the inputs (14) and outputs ( $3 \times n$  where  $n$  is 1 or 2)



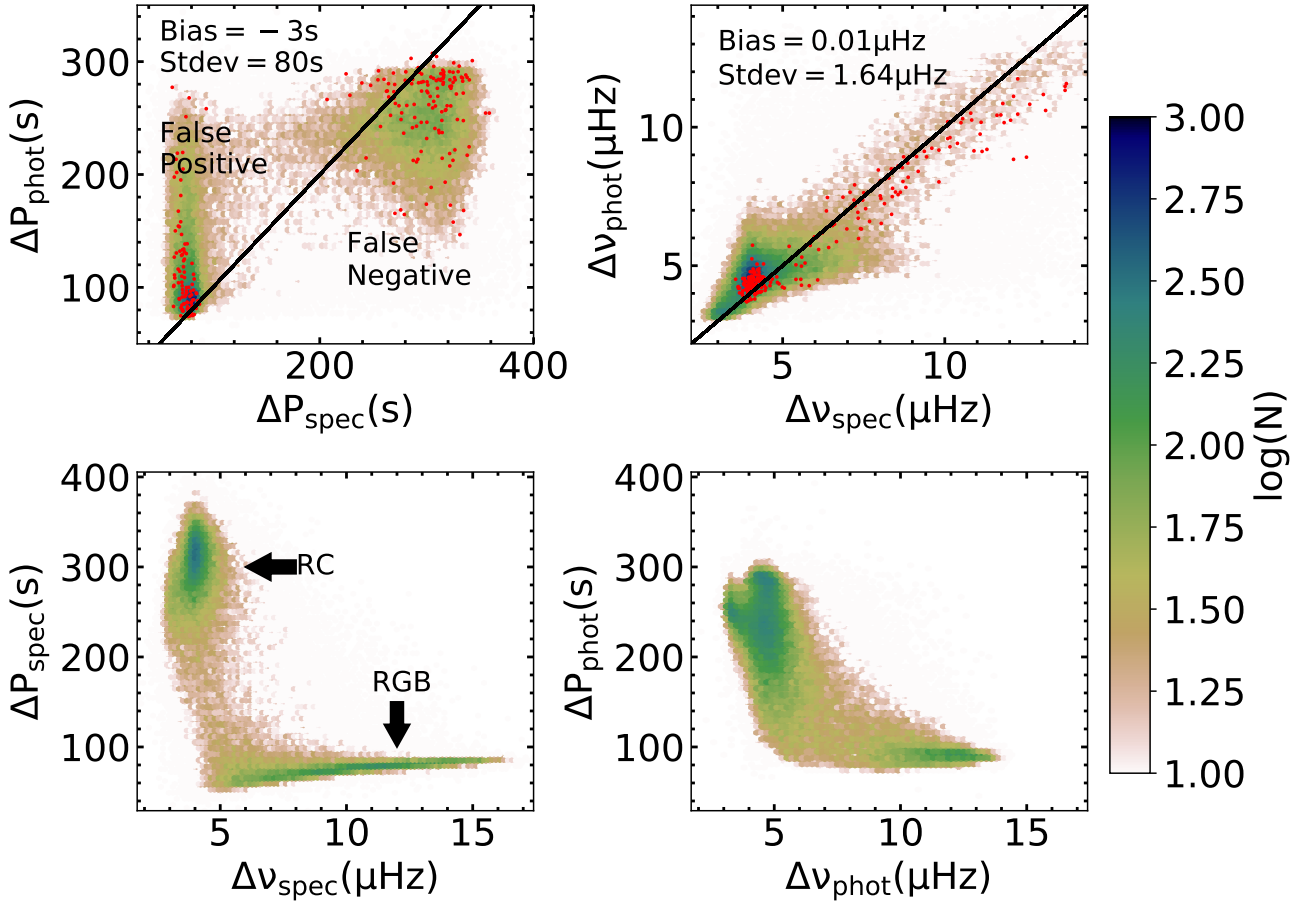
**Figure 2.** Comparing the stellar parameters ( $T_{\text{eff}}$  and  $\log g$ ) we derive from the photometry compared to the spectroscopically derived values from LAMOST using the validation set which was not used to train the network. The top panels show directly how the inferred values compare to the spectroscopic values for the entire validation set. The results from the first  $T_{\text{eff}}$  and  $\log g$  networks which all of the data go through are shown in the beige to blue density plot. The bias and standard deviation for this inference are shown in black. In the red to yellow contours we show our refined inference which uses networks trained on only spectroscopic giant stars and applied to photometrically selected giant stars. The inference improves when we retrain the network with a smaller parameter range. We show the statistics for this inference in red. In the bottom panel, we show the LAMOST Kiel diagram ( $T_{\text{eff}}$ - $\log g$  “Hertzsprung-Russell” diagram) for the validation sample (left) and the inferred photometric Kiel diagram (right). The main sequence and giant branch are preserved in the photometric diagram. Therefore we can confidently select giant stars using our inferred  $T_{\text{eff}}$  and  $\log g$ . The red box in the photometric Kiel diagram (right) shows the selection of giant stars which we determine  $T_{\text{eff}}$  and  $\log g$  using the refined networks.

a standard deviation of 0.32 dex. The typical uncertainty is 0.22 dex. The typical uncertainties for  $T_{\text{eff}}$  and  $\log g$  reported here (97 K and 0.22 dex) are smaller than the typical uncertainty in the LAMOST catalog (168 K and 0.24 dex). As explained, this is likely because our uncertainties are underestimated by not taking into account the uncertainty on the input data. In addition, the training set has higher S/N than the validation set. The external validation gives us a more robust estimate of the precision of the inference.

### 3.3 Inferring Asteroseismic Parameters and Selecting Red Clump Stars

We make the selection of RC stars using inferred  $\log g$ ,  $T_{\text{eff}}$ ,  $\Delta\nu$  and  $\Delta P$ . For each of these parameters we train another MDN. For  $\Delta\nu$  and  $\Delta P$ , we train on a subset of 30,000 stars from the sample presented in [Ting et al. \(2018a\)](#). For the

$\Delta\nu$  the network also has one mixture component as the output and we infer using the same method as  $T_{\text{eff}}$  and  $\log g$ . The distribution of  $\Delta P$  is bi-modal (see Figure 3). Using two mixture components helps the network to learn to reproduce this bi-modality, especially when the input is noisy and the posterior is likely bimodal. Therefore, the output of the  $\Delta P$  network is a two component Gaussian mixture model. We define the inferred  $\Delta P$  to be the weighted mean of the means of the two components and the uncertainty is the weighted mean of the widths of the two components. As we are trying to infer the  $\Delta P$  based on the signal of a change  $[C/N]$  in the photometry, which is roughly on the order of the uncertainty (see Figure 1), we know that the network will not be perfect. However, the inference result is encouraging (see Figure 3). The bias for the photometric  $\Delta\nu$  is only 0.01  $\mu\text{Hz}$  with a standard deviation of 1.64  $\mu\text{Hz}$ . The typical uncertainty is 1.51  $\mu\text{Hz}$ . For  $\Delta P$ , the bias is -4 s with a standard deviation

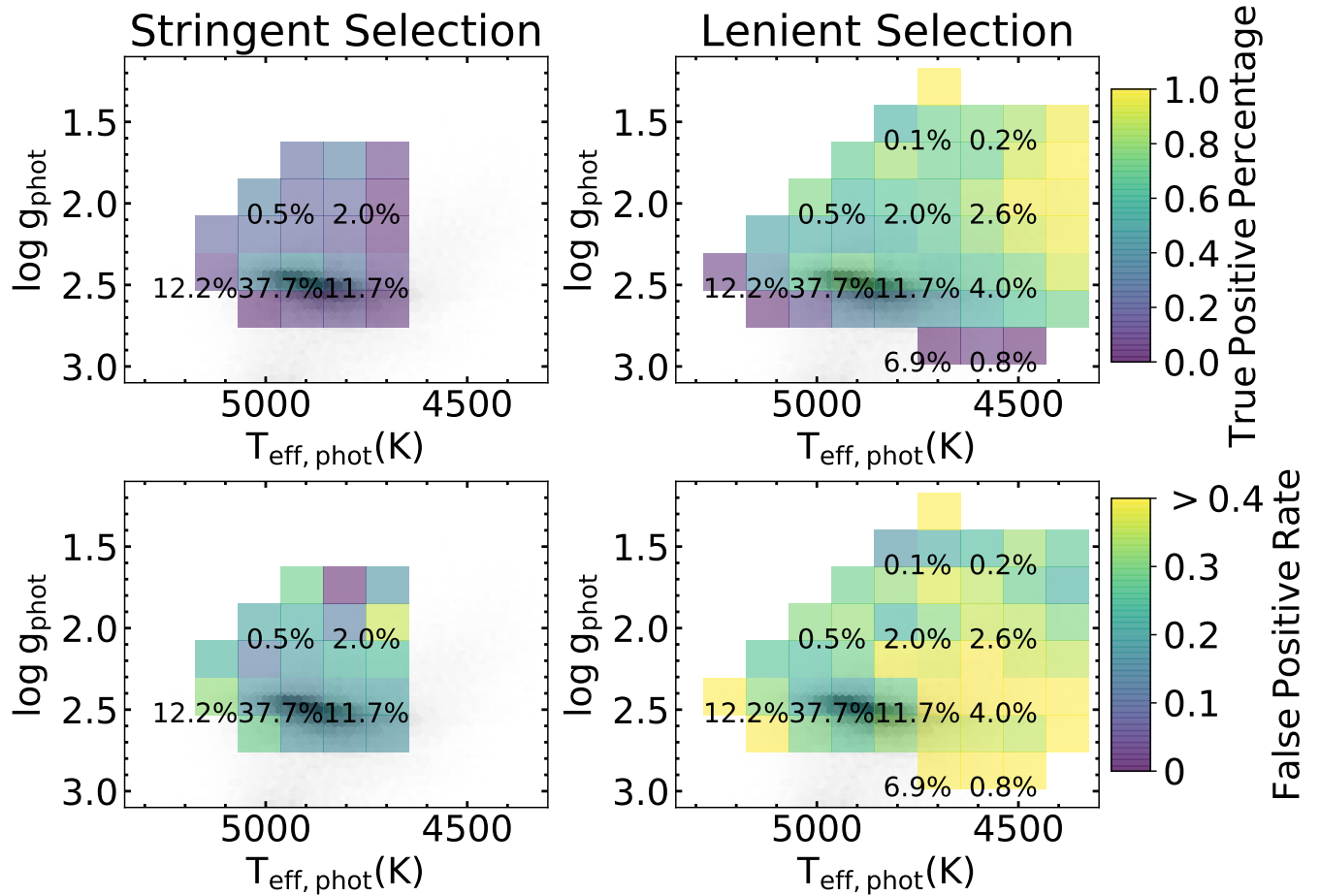


**Figure 3.** The asteroseismic parameters ( $\Delta P$  and  $\Delta \nu$ ) we derive for giant stars compared to the spectroscopically derived values from LAMOST spectra (Ting et al. 2018a). The red points in the top panels compare our derived photometric values with asteroseismic parameters from Vrad et al. (2016). The  $\Delta P$  is the most effective parameter for selecting RC stars since they have  $\Delta P$  values more than 100s larger than RGB  $\Delta P$  values. From the bottom left panel, we can see that RC stars have  $\Delta \nu \lesssim 5 \mu\text{Hz}$  and  $\Delta P \gtrsim 200\text{s}$ . While, RGB stars have  $\Delta \nu \gtrsim 5 \mu\text{Hz}$  and  $\Delta P \lesssim 100\text{s}$ . From the top right panel, we can see we are effective at picking up all of the RC stars as there are few false negatives. However, we do suffer from contamination with false positives. Combining the  $\Delta P$  with  $T_{\text{eff}}$ ,  $\log g$  and  $\Delta \nu$  can help limit this contamination. On the bottom left we show the asteroseismic parameters ( $\Delta P$  and  $\Delta \nu$ ) derived from LAMOST spectra in Ting et al. (2018a). On the bottom right we show the derived asteroseismic parameters from photometry. We can pick out the RC with  $\Delta P > 200 \text{ s}$  from the photometrically inferred asteroseismic parameters.

of 80 s. The typical uncertainty is 37 s. Although this uncertainty is not negligible, it is much smaller than the expected difference in  $\Delta P$  between RGB and RC stars ( $\sim 200 \text{ s}$ ). We also compare our photometrically derived values to asteroseismically derived values from Vrad et al. (2016). We find a bias of  $0.27 \mu\text{Hz}$  and a standard deviation of  $1.28 \mu\text{Hz}$  for  $\Delta \nu$  and a bias of 6 s and standard deviation 61 s for  $\Delta P$ . It is interesting to note that the biases are slightly higher for the asteroseismic comparison while the standard deviations are slightly smaller. The bias is likely higher because it is compounding with the spectroscopically derived bias. The standard deviation likely shrinks because the asteroseismic sample is more nearby and therefore has less noisy photometry, which improves the performance of the network.

To find the ideal RC selection criteria, we have to balance completeness with the contamination. In general, we want to maximize the completeness while minimizing the contamination. However, there may be science cases where

one may wish to sacrifice completeness for a more pristine sample or vice versa. In this work, we quantify completeness by the true positive rate, which is defined as the percentage of true spectroscopic RC stars in the validation sample that are successfully identified as RC stars (1-false negative rate, see Figure 3). We quantify the contamination rate as the percentage of stars which are selected as RC stars that are not true RC stars (false positive rate, see Figure 3). We expect the contamination rate and true positive percentage to be a function of  $T_{\text{eff}}$  and  $\log g$  along with  $\Delta \nu$  and  $\Delta P$  because the density of true RC versus RGB stars varies as a function of  $T_{\text{eff}}$  and  $\log g$ . Therefore, we decide to bin the data in  $T_{\text{eff}}$ ,  $\log g$ ,  $\Delta \nu$  and  $\Delta P$  in order to find the ideal selection criteria. Using our validation set, we bin the data using the photometrically inferred parameters and calculate the contamination and true positive rate within each bin. Although this is done in four dimensions ( $T_{\text{eff}}$ ,  $\log g$ ,  $\Delta \nu$  and  $\Delta P$ ), we show flattened two dimensional examples in Figure 4. It is



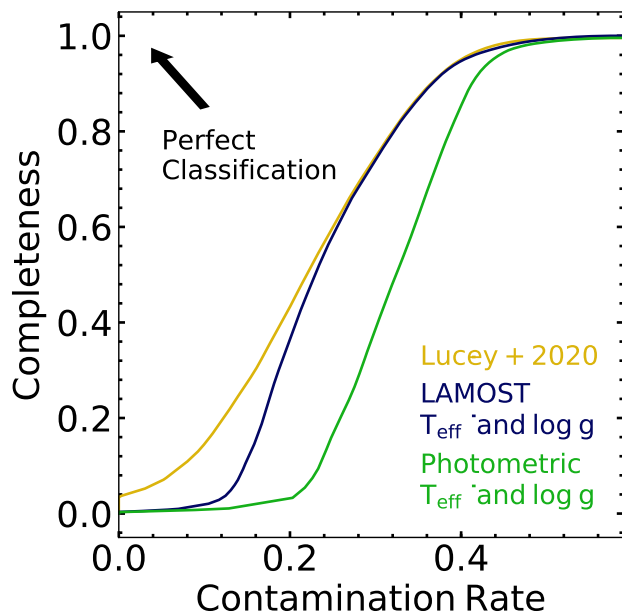
**Figure 4.** The true positive percentage (upper panels) and the contamination rate (lower panels) as a function of the photometrically derived  $T_{\text{eff}}$  and  $\log g$  for both of our recommended samples. These are evaluated using the spectroscopic sample for which we know the “ground truth”. The black text shows the percentage of true spectroscopic RC stars in each sample that are in the adjacent four bins. We define the contamination rate to be the number of false positives divided by the total number of the stars selected per box. The true positive percentage is defined as the number of stars successfully identified as RC divided by the number of true RC stars in the selected box. The panels on the left correspond to our stringent one selection which is  $\Delta P > 265$  s,  $\Delta \nu < 5.25$   $\mu\text{Hz}$  and  $T_{\text{eff}} > 4700$  K. The panels on the right correspond to a more lenient selection of  $\Delta P > 225$  s and  $\Delta \nu < 5.5$   $\mu\text{Hz}$ . When we soften our selection criteria the true positive percentage increases. However, so does the contamination rate. Integrating over all of the bins, the tier 1 (stringent) selection leads to an overall contamination rate of  $\sim 20\%$  and overall true positive percentage of  $\sim 25\%$ . While the tier 2 (lenient) selection leads to an overall contamination rate of  $\sim 33\%$  and overall true positive percentage of  $\sim 94\%$ .

interesting to note that the contamination rate for the tier 2 (lenient) sample (top right panel) may be artificially low at the lowest  $T_{\text{eff}}$  and  $\log g$  values. This is likely because the tip of the red giant branch stars that are located there would have a  $[\text{C}/\text{N}]$  most similar to the RC values. However, the fraction of stars in the region is limited. We bin each parameter into 20 bins so that in total we have 160,000 bins. This results in bin sizes of roughly 50K for  $T_{\text{eff}}$ , 0.10 dex for  $\log g$ , 1  $\mu\text{Hz}$  for  $\Delta \nu$  and 20 s for  $\Delta P$ . It is important to note that we bin using the photometrically derived parameters so that the results are applicable to our final sample which do not always have spectroscopically derived values. We include bins with a low contamination while also having a significant percentage of the true RC stars. Further information about the final sample can be found in Section 4.

### 3.4 Deriving Distances

Once we have the RC sample, we infer the heliocentric distances using the AllWISE W1 band similar to [Ting & Rix \(2019\)](#). First, we perform an extinction correction by finding a linear relationship between the photometric  $T_{\text{eff}}$  and the  $G-W1$  color for less extinguished stars. We then use this along with  $A_G/A_{W1}=16$  ([Hawkins et al. 2017](#)) to find the extinction in the W1 magnitude and correct for it. Next, we again use the less extinguished stars with *Gaia* parallax uncertainties  $< 5\%$  to derive a linear relationship between the inferred  $T_{\text{eff}}$  and the intrinsic  $M_{W1}$ . In order for our absolute W1 magnitudes to be consistent with the calibrated absolute W1 magnitude for RC stars ( $M_{W1} = -1.68 \pm 0.02$  mag) from [Hawkins et al. \(2017\)](#), which is consistent with [Ruiz-Dern et al. \(2018\)](#), we use a parallax offset of 0.04 mas. This parallax offset is roughly consistent with those found





**Figure 5.** The receiver operating characteristic (ROC) curve of the selection method compared to other methods. The ideal classification method gives 100% completion with no contamination. Therefore, the closer the ROC curves gets to the top left corner of the plot, the more effective the classification method. The lines are calculated by cumulatively summing bins in the order of lowest contamination rate (see Figure 4). The result for our method where we bin in all four photometrically derived parameters ( $T_{\text{eff}}$ ,  $\log g$ ,  $\Delta\nu$ , and  $\Delta P$ ) and select from them is shown in yellow. The result of only selecting in the photometric  $T_{\text{eff}}$  and  $\log g$  space is shown in green. The results of making the same selection but using LAMOST spectroscopically ( $R \sim 1800$ ) derived  $T_{\text{eff}}$  and  $\log g$  is shown in blue. This shows our method can obtain a lower contamination rate than the other two standard methods.

using other methods (Lindgren et al. 2018; Schönrich et al. 2019). If we do not include the parallax offset we find an absolute W1 magnitude of -1.99 mag. Finally, we derive the distance using the distance modulus with the inferred  $M_{W1}$  and the extinction corrected W1 magnitudes. In principal, a Bayesian approach would be more appropriate for determining the distances (e.g., Hawkins et al. 2017; Anders et al. 2019). However, the main focus of this work is to derive a clean sample of RC stars and we will refine the distances in future work. Comparing to *Gaia* parallaxes with uncertainties  $< 5\%$  and assuming the *Gaia* distance as ground truth, we find our distances have uncertainties  $\sim 9\%$ . This is consistent with reported theoretical precision estimates given the intrinsic dispersion in the absolute magnitude of RC stars (Bovy et al. 2014; Hawkins et al. 2017).

#### 4 THE RED CLUMP SAMPLE

Having inferred four parameters ( $T_{\text{eff}}$ ,  $\log g$ ,  $\Delta\nu$  and  $\Delta P$ ) for each giant stars in our sample, we must now determine how we will use these parameters to determine our RC sample. Different selection criteria will result in varying completeness and contamination rates, given that we know the true

positive percentage and contamination rate vary as a function of  $T_{\text{eff}}$ ,  $\log g$ ,  $\Delta\nu$  and  $\Delta P$  (see Figure 4). Ideally, we would like to find a selection that maximizes completeness while minimizing contamination. In this work, we prioritize a low contamination over a complete sample. However, different science cases may require different selection criteria. Various selections can make retracing the selection function more difficult. Therefore, along with the recommended samples, we provide all of the inferred parameters for the entire giants catalog and encourage the reader to choose their own selection criteria.

To choose which bins from Figure 4 we will use in the final selection, we perform a cumulative summation sorted by the contamination rate. The results of the cumulative summation are shown in Figure 5. Our method is able to achieve a lower contamination rate than when selecting just using spectroscopically (at low resolution,  $R \sim 2,000$ ) or photometrically  $T_{\text{eff}}$  and  $\log g$ . We note that we do not include using  $[C/N]$  information from spectra in our comparison. Using the  $[C/N]$  from spectra would result in a lower contamination rate for the spectroscopic sample (Hawkins et al. 2018; Ting et al. 2018a). However, in Figure 5, we are comparing to more “classical” methods which are still frequently used today (e.g., Wegg et al. 2019; Chan & Bovy 2019). It is also important to note that completeness is evaluated based on the same sample which is confined by the spectroscopic sample to brighter magnitudes. Therefore, a given completeness fraction corresponds to a much greater number of stars for the photometric sample than the spectroscopic sample given that the photometric sample contains much more data. However, these results are also slightly idealized. The selection method used to create Figure 5 selects parameter bins in order of the lowest contamination rate rather than selecting adjacent bins as we do for our final selection.

For ease of selection and to better model the selection function in the future, for our recommended catalogs we choose to only use adjacent bins with simple parameter boundaries. We provide two RC samples with different contamination rates and true positive percentages. Again, we provide the inferred parameters for all of the giant stars and encourage the readers to make their own selection. However, we will mark the stars in our two RC samples as “tier 1” and “tier 2” in the catalog. For our smaller, less contaminated sample (tier 1), we achieve a contamination rate of  $\sim 20\%$  (see Section 4.1 for how these contamination rates are calculated) and a true positive percentage of  $\sim 25\%$  which results in a final sample of  $\sim 405,000$  RC stars. These results are achieved by selecting giant stars with  $T_{\text{eff}} > 4700$  K,  $\Delta\nu < 5.25 \mu\text{Hz}$ , and  $\Delta P > 265$  s. We also select a larger sample (tier 2) ( $\sim 2.6$  million RC stars) with a contamination rate of  $\sim 33\%$  and a true positive percentage of  $\sim 94\%$ . This sample is chosen by selecting giant stars with  $\Delta\nu < 5.5 \mu\text{Hz}$ , and  $\Delta P > 225$  s. Therefore, The “tier 1” sample is subset of the “tier 2” sample. As such, everything labeled as “tier 1” in the catalog is also included in the “tier 2” sample.

##### 4.1 Contamination Rate

In order to accurately evaluate the contamination rate and completeness of our final sample we need to determine how the contamination and completeness of our validation set behaves as a function of extinction and uncertainty in the

**Table 1.** The cuts, contamination rate, true positive percentage and total number of stars for each of our provided samples

Selection	$\Delta P$ cut	$\Delta\nu$ cut	$T_{\text{eff}}$ cut	Contamination	Completeness	# of Stars
Tier 1	>265 s	<5.25 $\mu\text{Hz}$	> 4700 K	20%	25%	405,000
Tier 2	>225 s	<5.5 $\mu\text{Hz}$	None	33%	94%	2.6 million
$ b  > 30$ degrees	>265 s	<5.25 $\mu\text{Hz}$	>4700 K	9%	12%	15,000

Description of the final provided samples of RC stars. We also provide the inferred  $T_{\text{eff}}$ ,  $\log g$ ,  $\Delta\nu$  and  $\Delta P$  in the final catalog in order for the reader to be able to choose their own selection criteria for their science case.

**Table 2.** Photometrically derived  $T_{\text{eff}}$ ,  $\log g$ ,  $\Delta P$  and  $\Delta\nu$  for all giant stars along with distances for selected RC stars.

<i>Gaia</i> Source ID	RA	DEC	$T_{\text{eff}}$	$\log g$	$\Delta P$	$\Delta\nu$	Selection	d
	(deg)	(deg)	(K)	(dex)	(s)	( $\mu\text{Hz}$ )		(kpc)
4130510288119401088	252.313	-19.93	5021 $\pm$ 119	2.68 $\pm$ 0.39	185 $\pm$ 40	7.65 $\pm$ 2.95		
4114024073354437376	255.807	-22.352	4574 $\pm$ 89	2.64 $\pm$ 0.16	247 $\pm$ 44	4.9 $\pm$ 1.08	tier 2	4.72
4109961240517001856	261.658	-25.704	3741 $\pm$ 31	2.49 $\pm$ 0.2	185 $\pm$ 42	0.13 $\pm$ 0.05		
4114024550075584768	255.861	-22.332	4948 $\pm$ 120	2.28 $\pm$ 0.64	185 $\pm$ 53	7.12 $\pm$ 10.15		
4114024623110255232	255.855	-22.32	4102 $\pm$ 93	1.69 $\pm$ 0.24	255 $\pm$ 34	2.6 $\pm$ 0.12	tier 2	2.94
4114027539372846848	255.743	-22.316	4956 $\pm$ 119	2.99 $\pm$ 0.72	274 $\pm$ 37	10.71 $\pm$ 3.1		
4114027917328986240	255.807	-22.281	5037 $\pm$ 134	3.23 $\pm$ 0.18	103 $\pm$ 29	12.57 $\pm$ 1.75		
4126036886406973952	255.633	-22.265	4035 $\pm$ 84	2.45 $\pm$ 0.29	229 $\pm$ 33	2.11 $\pm$ 0.16	tier 2	4.13
6734964158462988160	277.756	-33.984	4920 $\pm$ 144	2.39 $\pm$ 0.58	315 $\pm$ 26	4.59 $\pm$ 1.15	tier 1	7.86
6734962754045102848	277.679	-34.036	4849 $\pm$ 223	2.86 $\pm$ 0.05	76 $\pm$ 4	9.59 $\pm$ 2.82		
...	...	...	...	...	...	...	...	...

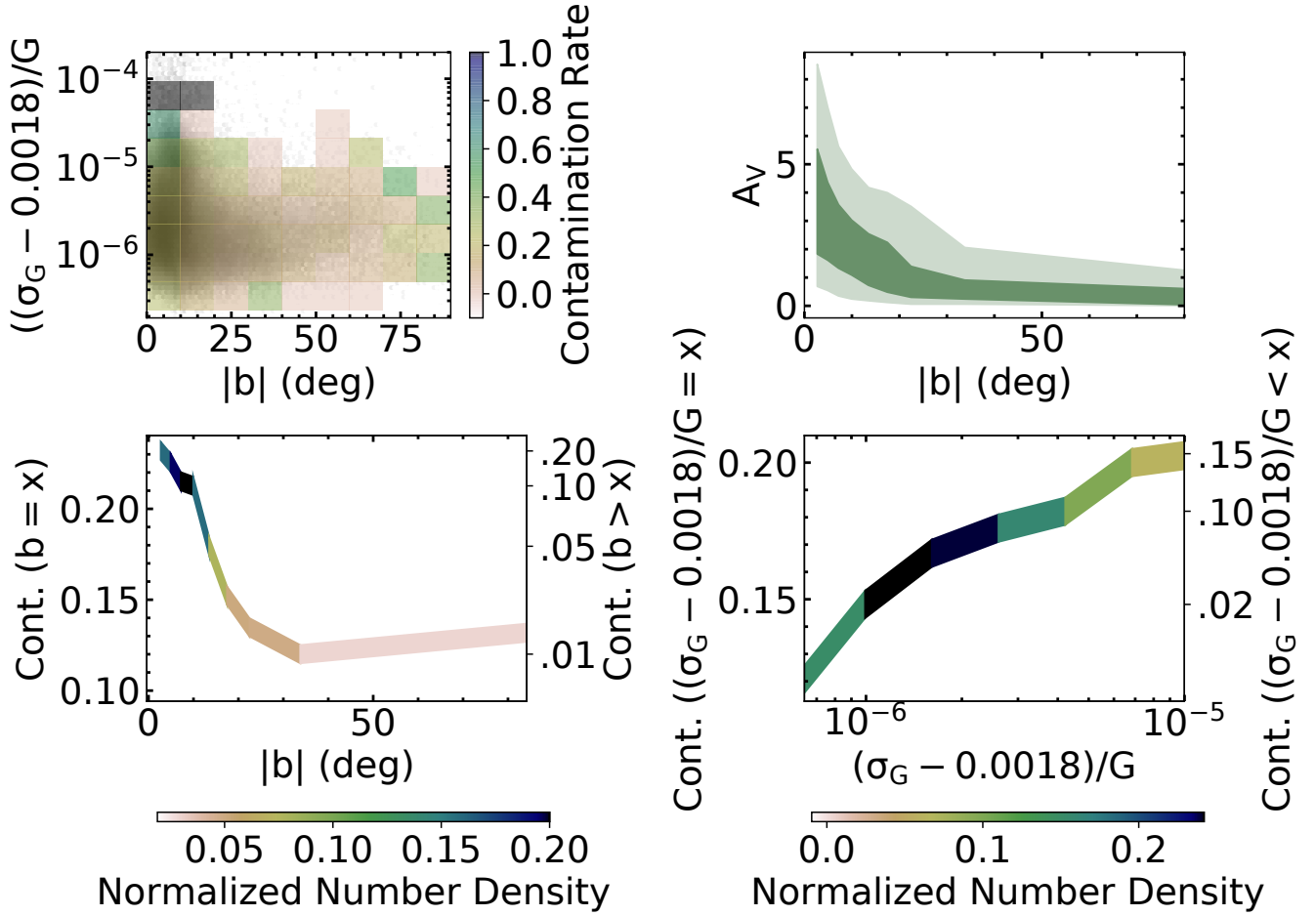
A sample of the provided catalog with  $T_{\text{eff}}$ ,  $\log g$ ,  $\Delta P$  and  $\Delta\nu$  for all giant stars along with uncertainties. For a cautionary note about the uncertainties, see Section 5.1. We also give the “tier 1” and “tier 2” RC selection as described in Section 4. The “tier 1” is a less contaminated subset of the “tier 2” sample. However, we only label it as “tier 1” in the catalog for simplicity. We provide distance estimates for stars selected as RC.

SED. Therefore, we can account for differences between our final samples distribution in these parameters compared to our validation set.

In Figure 6, we show the contamination rate as a function of Galactic latitude ( $b$ ) and the fractional *Gaia* G band magnitude uncertainty  $((\sigma_G - 0.0018)/G)$ . We evaluate the fractional G band uncertainty as  $(\sigma_G - 0.0018)/G$  because we want to subtract off the contribution of the uncertainty in the zero-point which converts the measured flux to a magnitude using the Vega magnitude system. This conversion accounts for the majority of the uncertainty for a large fraction of the data (see bottom right panel of Figure 6), so that we have a more dynamic range of uncertainties with which to evaluate the contamination. We also test calculating the contamination rate as a function of just  $\sigma_G$  and the results are similar to  $(\sigma_G - 0.0018)/G$ . In the top right panel of Figure 6, we show the extinction ( $A_V$ ) as a function of Galactic latitude for the RC sample. We calculate the extinction ( $A_V$ ) using the Python package DUSTMAPS (Green 2018) and the map provided by Green et al. (2019). Ideally, we would calculate the contamination as a function of  $A_V$ . However, For accurate estimates of  $A_V$ , we require accurate distances to our objects. Given that we do not have accurate distance for the RGB stars in our validation sample, we cannot calculate  $A_V$  for these stars and therefore cannot calculate the contamination rate as a function of  $A_V$ . However, because the extinction is highly correlated to the Galactic latitude (see Figure 6), we can use the Galactic latitude as a proxy for extinction. This correlation is expected given that in the disk, at  $|b| < 10$  degrees, there are high levels of dust, which causes extinction. However, at  $|b| > 10$  degrees, we are looking out

of the plane of the Galaxy, where dust density and therefore extinction is much lower. As extinction can greatly impact the shape of an SED, we expect it to negatively impact our ability to infer the  $T_{\text{eff}}$ ,  $\log g$ ,  $\Delta\nu$  and  $\Delta P$ . Therefore, we expect higher contamination rates at lower Galactic latitude. We also expect our inference of the  $\Delta P$  to be highly sensitive to the error of the input SED given that the signal of an RC star is on the order of the typical uncertainty (see Figure 1). Evaluating the contamination rates with our validation sample, in Figure 6 we show that indeed the contamination rate steeply increases with larger G band uncertainties and lower Galactic latitude.

Given that our data has a slightly different distribution of Galactic latitudes and G band errors than our validation sample, the contamination rate and completeness for a given sample may be different than that calculated by the validation sample. To account for this, we calculate our final sample contamination rate by integrating the contamination rate as a function of Galactic latitude and fractional G band uncertainty multiplied by the normalized number density of our final sample. This leads to a contamination rate of  $\sim 20\%$  for the “tier 1” sample and  $\sim 35\%$  for the “tier 2” sample. This is an increase of  $\sim 3\text{--}5\%$  compared to the validation sample. However, as shown in Figure 6, a cut in either the fractional G band magnitude errors or Galactic latitude can lead to much lower contamination rates. For example, applying a cut of  $|b| > 30$  degrees to the “tier 2” sample leads to a sample of 15,000 stars with an integrated contamination rate of  $\sim 9\%$ .



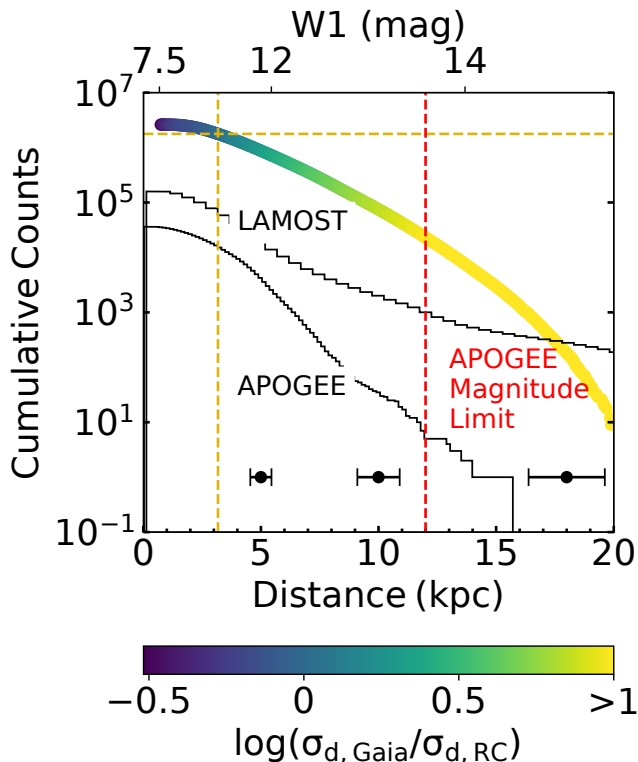
**Figure 6.** The contamination rate as a function of the fractional *Gaia* G band magnitude uncertainty  $((\sigma_G - 0.0018)/G)$ , and the Galactic latitude ( $b$ ). In the top left plot, we show the distribution of our “tier 1” sample in G band fractional uncertainty and Galactic latitude in grey with the contamination rate in the colored boxes. In the top right plot we show the extinction ( $A_V$ ) as a function of Galactic latitude and we can therefore use Galactic latitude as a proxy for extinction when evaluating the contamination rate. In the bottom two plots, we show the 1D versions of the top left plot. For these plots, the left y-axis shows the contamination rate at  $x$ . The right y-axis shows the cumulative contamination rate for all data  $>x$  for Galactic latitude and  $<x$  for the fractional G band uncertainty. These values are calculated by summing the contamination rate multiplied by the normalized number density of the “tier 1” sample. The lines are colored by the value of the normalized number density at that point.

#### 4.2 Galactic Distribution of Red Clump Sample

Figure 7 shows the reverse cumulative distribution of the derived heliocentric distances for the “complete” sample (“tier 2”). For comparison, we show the LAMOST and APOGEE spectroscopic RC samples from Ting et al. (2018a) in black. Even though the LAMOST sample is the largest and most distant sample of spectroscopically derived RC stars, our photometrically derived sample is significantly larger even at large distances. Therefore, our sample provides more precise distances where they matter, e.g., in the distant Galaxy where there are many open questions about its structure and where the distances are most difficult to derive. For comparison, we also derive uncertainties for distances derived using *Gaia* DR2 parallaxes. For this we use the equation<sup>1</sup>  $\sigma_d = \sigma_\pi d^2$ , where  $\sigma_\pi$  is the individual reported *Gaia* DR2 parallax uncertainty. In Figure 7, we color our distribution of stars by the ratio of the *Gaia* distance uncertainties to the typical uncertainty for RC stars ( $\sim 9\%$ ) at the given

distance. The dark blue dashed lines indicate where distances measured using RC stars become more precise than distances provided by *Gaia* DR2 parallaxes. *In total, our sample contains over 1.8 million stars with distances more precise than the distance given by Gaia DR2 parallaxes (see Figure 7).*

We show the Galactic distribution of the sample of  $\sim 2.6$  million RC stars in Figure 8. Our sample reaches farther into the Galactic halo and Galactic center than previous pristine RC samples (Bovy et al. 2014; Ting et al. 2018a; Wu et al. 2019) with our most distant RC star at a heliocentric distance of  $\sim 20$  kpc. This maximum distance is consistent with the W1 limiting magnitude ( $W1 = 17.1$  mag) and the absolute magnitude of RC stars in W1 ( $M_{W1} = -1.68$  mag) which corresponds to a distance of 30.06 kpc. As shown in Figure 8, extinction prevents us from observing the Galactic center. Assuming an  $A_V = 30$  mag at the Galactic center, in order to observe a RC star at that distance (8.3 kpc) we



**Figure 7.** Cumulative distribution of distances for our sample of 2.6 million RC stars. We show the corresponding RC W1 magnitude on the top x-axis for reference. Our sample is colored by the ratio of *Gaia* DR2 distance uncertainty to the distance uncertainties of RC stars. We also show the spectroscopic LAMOST and APOGEE RC samples from [Ting et al. \(2018a\)](#) in black for comparison. The red dashed line corresponds to the RC distance for the APOGEE magnitude limit ( $H=14$  mag). The yellow dashed lines show where the *Gaia* DR2 uncertainties become larger than the RC distance uncertainties. Our RC distances are more precise at distance  $> 3$  kpc. The *Gaia* uncertainties degrade faster because they scale as distance squared rather while the RC distance uncertainties are linear with distance. The black error bars show the typical RC distance uncertainty for distances of 5, 10 and 18 kpc.

would need to use the infrared. A RC star at that distance and level of extinction is just on the edge of the magnitude limit of the 2MASS survey for the K band ( $K=14.3$  mag). The only bands in which RC stars at the Galactic center are observable in current large surveys are the K, W1 and W2 bands. However, given the typical uncertainties and CN information content, these bands would not be sufficient to effectively select RC stars (see Figure 1). Future infrared photometric surveys (e.g., WFIRST) may provide sufficient depth and precision to select an RC sample in the Galactic center.

## 5 INTERPRETING THE NETWORKS

When using neural networks, it is often useful to understand what information is harnessed by the neural network to make the inference. This can be especially important when there might be confounding variables that the network could be

using to make the inference instead of the desired physical mechanism. After all, neural networks aim to understand the correlations between variables. Domain knowledge is required to determine if the established correlation agrees with our understanding of the causal structure. In this section, we describe our attempts of trying to open the “black box” and determine what correlations the network uses to infer the parameters.

### 5.1 Interpreting the Reported Uncertainties

One of the advantages to using a Mixture Density Network is that the inferred parameters also have uncertainties associated with them. The uncertainties are determined from the widths ( $\sigma(x)$ ) of the output GMMs (see Section 3.1). Although the network is not exposed to the input uncertainties, we find that the uncertainties on the inferred parameters correlate with the input uncertainties.

In the top panels of Figure 9, we show how the uncertainty on the inferred  $T_{\text{eff}}$  and  $\log g$  relate to the uncertainties on the input parameters, which we suspect to be essential for the respective inference. The correlation shown in Figure 9 suggests that the network is able to recognize a noisy SED and assign it a larger width of the GMM. Given that noise acts as a non-physical perturbation to the input, it is possible that the network detects these non-physical perturbations and during the training, learns it is unable to infer the parameter as accurately for SEDs with large perturbations as those with only small perturbations.

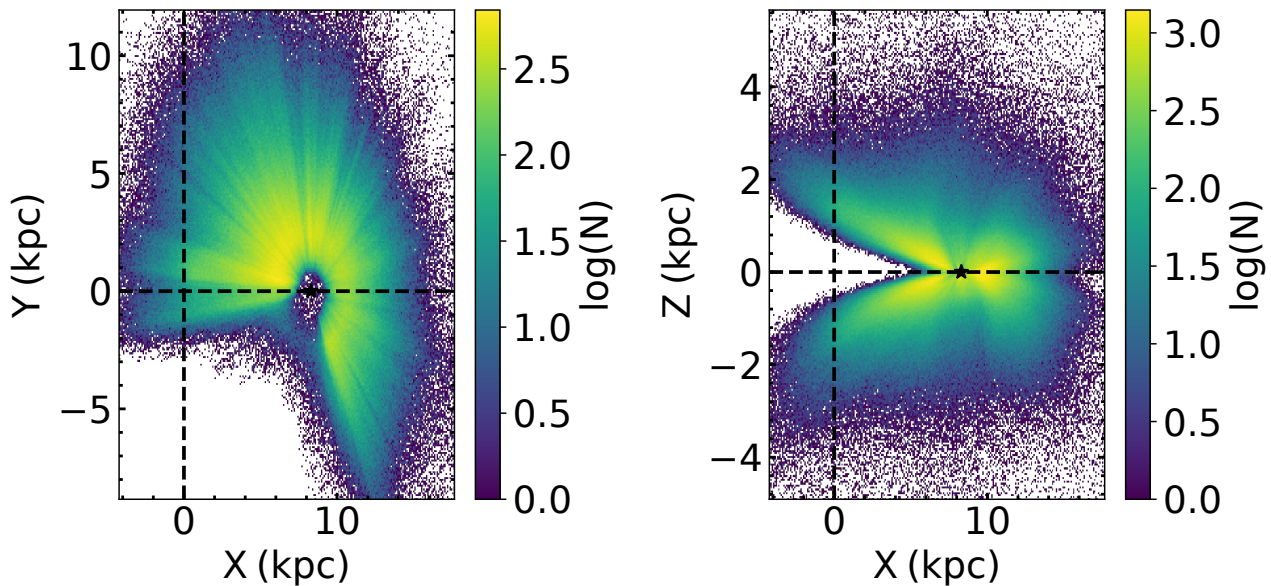
In the lower panels of Figure 9, we show the break down of these trends into different parameter bins. The lower left panels shows that giant stars with  $T_{\text{eff}} < 4700$  K have higher  $T_{\text{eff}}$  uncertainties for a given  $G$  and  $W2$  uncertainty. This is likely because cooler stars are more luminous and due to the brightness limit of the surveys, luminous stars are often not observed when they are close to the Sun. As a consequence, cooler stars are statistically further away from their hotter counterparts. This leads to both a larger fractional uncertainty in photometry as well as parallax, which could explain why cooler stars have slightly larger uncertainties in their temperature estimates. Similarly, giants that are more evolved (lower  $\log g$ ) are preferentially missed when they are very close by; as a consequence, they are biased toward larger distances. Therefore, they typically have a larger fractional uncertainty in parallax which causes are larger uncertainty in  $\log g$ .

### 5.2 Parameter Importance for the $\Delta P$ Inference

We started this work with the conjecture that the network will pick up the difference in the  $[C/N]$  ratios between RC stars and red giant branch stars from the photometry and would use that information to determine the inferred  $\Delta P$  value. Indeed, the results shown in Figure 5 demonstrate that we are more effective at picking RC stars using the  $\Delta P$  compared to just from the photometrically derived  $T_{\text{eff}}$  and  $\log g$ . This indicates that the network is using more information than  $T_{\text{eff}}$  and  $\log g$  to infer the  $\Delta P$  and therefore select RC stars.

In order to better understand the extra information the network is using to infer  $\Delta P$ , we calculate the empirical





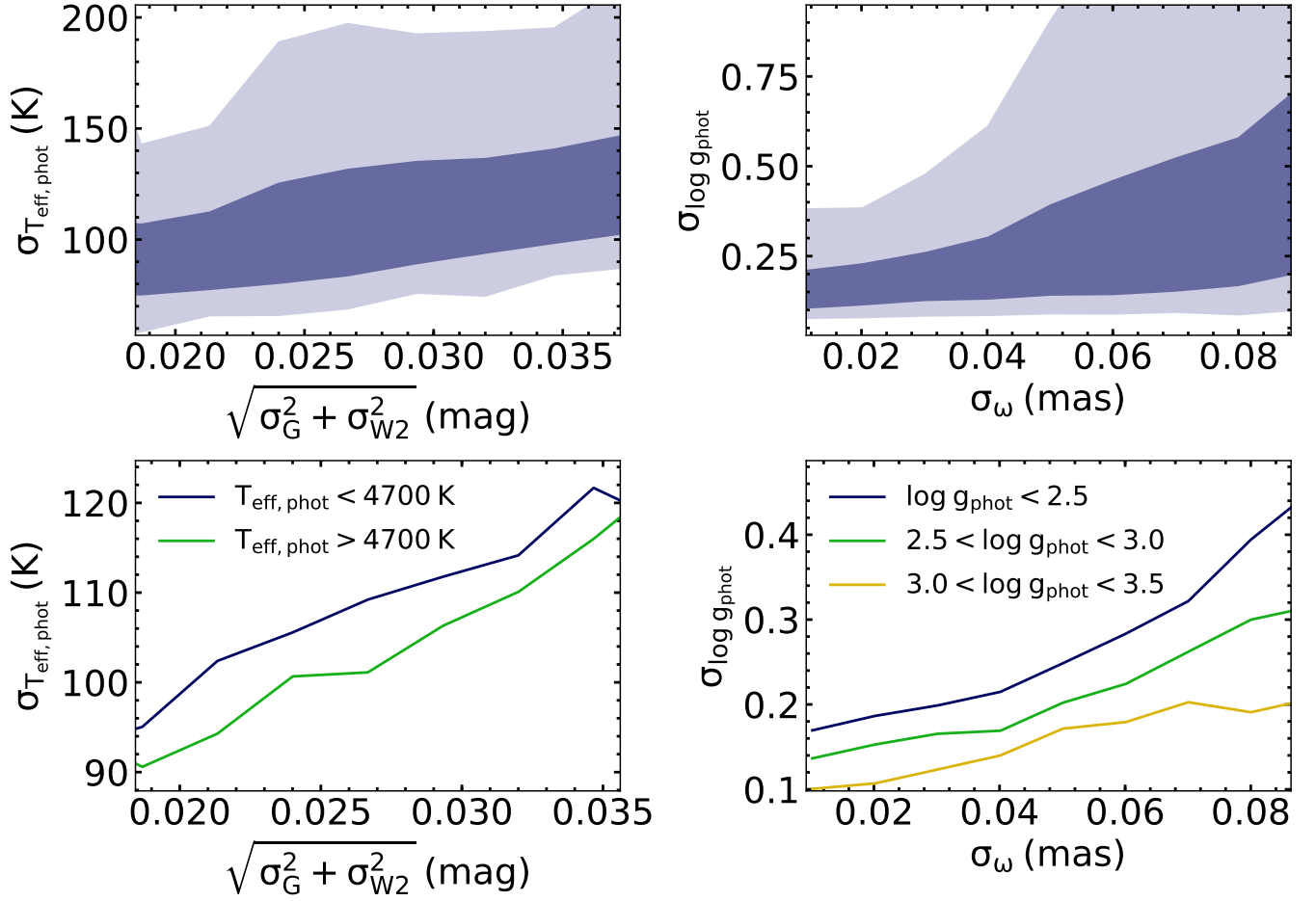
**Figure 8.** Galactic distribution of our sample of 2.6 million RC stars. On the left is the density of stars in Galactic coordinates, X and Y with  $|Z| < 3$  kpc and on the right are the Galactic coordinates X and Z. The Galactic center is located at (0,0,0) in both plots, and the Sun is located at (8.3,0,0) kpc. The missing section of the disk is due to the Pan-STARRS footprint; Pan-STARRS does not reach declinations  $< -30^\circ$ . High levels of extinction prevent current photometric surveys from observing RC stars in bands bluer than the K band at the Galactic center. This observational effect could explain why we do not see the Galactic bar in the X-Y plane.

derivative of the inferred  $\Delta P$  with respect to each of the photometric bands. This is done by adding and subtracting 0.001 mag to the selected band for each SED in our testing set and re-inferring the  $\Delta P$ . Therefore, we have an empirical derivative calculated for giant stars with a range of  $T_{\text{eff}}$  and  $\log g$  values, including both RC and RGB stars. We repeat this process with the  $T_{\text{eff}}$  and  $\log g$  networks so that we have derivatives of  $\Delta P$ ,  $T_{\text{eff}}$ , and  $\log g$  with respect to each band. In Figure 10, we show how the median derivatives of  $T_{\text{eff}}$ ,  $\log g$  and  $\Delta P$  with respect to the H-band vary with  $T_{\text{eff}}$  and  $\log g$ . This is done by calculating the median of the validation data set for 100 ( $10 \times 10$ )  $T_{\text{eff}}$  and  $\log g$  bins. We choose to show the H-band because we know it has CN molecular lines (Hawkins et al. 2018) and therefore would be sensitive to any change in the  $[C/N]$  ratio (see Figure 1).

We expect the network to have learned that if a star is outside the typical RC  $T_{\text{eff}}$  and  $\log g$  parameter range, then it is likely not a RC star and it assigns that star a small and relatively constant  $\Delta P$  ( $\Delta P \approx 80$ s), i.e., the derivative of  $\Delta P$  would be small. However, if the star is in the RC  $T_{\text{eff}}$  and  $\log g$  range, then the network must determine if the SED is consistent with the expected change in  $[C/N]$  in order to infer the value of the  $\Delta P$  to assign. We expect that if the  $\Delta P$  network is using this information it should cause the derivative in the RC region to be large in order to be sensitive to the change in  $[C/N]$ . From Figure 10, it is clear that the derivative of  $\Delta P$  with respect to the H-band has a local maximum near the RC region. Another important thing to note is that the variation of the derivative of  $\Delta P$  as a function of  $T_{\text{eff}}$  and  $\log g$  is not similar to the variation showed for the derivatives of  $T_{\text{eff}}$  and  $\log g$ . For example,

both the derivatives  $T_{\text{eff}}$  and  $\log g$  show local minima in the RC region. Additionally, the surfaces of constant derivative make up very different forms in the  $T_{\text{eff}}$  and  $\log g$  space. This gives further evidence that the  $\Delta P$  inference is based on more than simply the  $T_{\text{eff}}$  and  $\log g$ .

In addition to the derivatives, we calculate synthetic photometry to compare to the empirical results. If the network is using  $[C/N]$  information to infer  $\Delta P$ , we expect the derivative of the  $\Delta P$  to be correlated with the change in magnitude caused by a difference in  $[C/N]$  abundance. In Figure 11, we show how the relative strength of the derivative of  $\Delta P$  relates to the calculated change in magnitude due to a change in  $[C/N]$  of -0.77 dex for each photometric band. The adopted change in  $[C/N]$  is motivated by models from Lagarde et al. (2012) and corresponds to an RC star with  $T_{\text{eff}} 4604$  K and  $\log g = 2.62$  dex. The derivatives shown are calculated by taking the median of the derivatives of  $\Delta P$ ,  $T_{\text{eff}}$ , and  $\log g$  with respect to each band for stars with  $4100 \text{ K} < T_{\text{eff}} < 5100 \text{ K}$  and  $2.3 \text{ dex} < \log g < 2.7 \text{ dex}$ , the  $T_{\text{eff}}$  and  $\log g$  of RC stars. We define the relative strength of the derivative of  $\Delta P$  as the ratio of the derivative of  $\Delta P$  with respect to the magnitude to the derivative of  $T_{\text{eff}}$  and  $\log g$  with respect to the magnitude. We know that the  $\Delta P$  inference is modulated by the  $T_{\text{eff}}$  and  $\log g$  of the star because the inference of  $T_{\text{eff}}$ ,  $\log g$  and  $\Delta P$  are weakly correlated. In addition, the empirical gradients for  $T_{\text{eff}}$ ,  $\log g$ , and  $\Delta P$  are each derived from separate networks, as such the correlation can manifest itself in the gradients. However, for these purposes, we are only interested in the sensitivity of the  $\Delta P$  inference to each band irrespective of a change in  $T_{\text{eff}}$  and  $\log g$ . We take the absolute values of the derivative of  $\Delta P$



**Figure 9.** The uncertainties of inferred parameters  $T_{\text{eff}}$  and  $\log g$  as a function of the uncertainties in the input data. Specifically we compare them to uncertainties in the specific inputs that are deemed crucial for the inference. For  $T_{\text{eff}}$  this is the color of the star which can be described as G-W2. For  $\log g$  it is the parallax ( $\varpi$ ; see Section 5.1). The top plots show the 1 $\sigma$  (dark blue) and 2 $\sigma$  (light blue) intervals while the bottom plots show the median trend in different parameter ranges. The  $\log g$  uncertainties are larger for stars with  $\log g < 2.5$  dex likely because they are typically more luminous than higher  $\log g$  giants and therefore have larger fractional parallax uncertainties for a given parallax uncertainty. Similarly, stars with  $T_{\text{eff}} < 4700$  K likely have higher  $T_{\text{eff}}$  uncertainties because they are at the tip of the RGB and are brighter. Therefore, they are typically farther, leading to higher fractional uncertainties to the magnitude and the parallax.

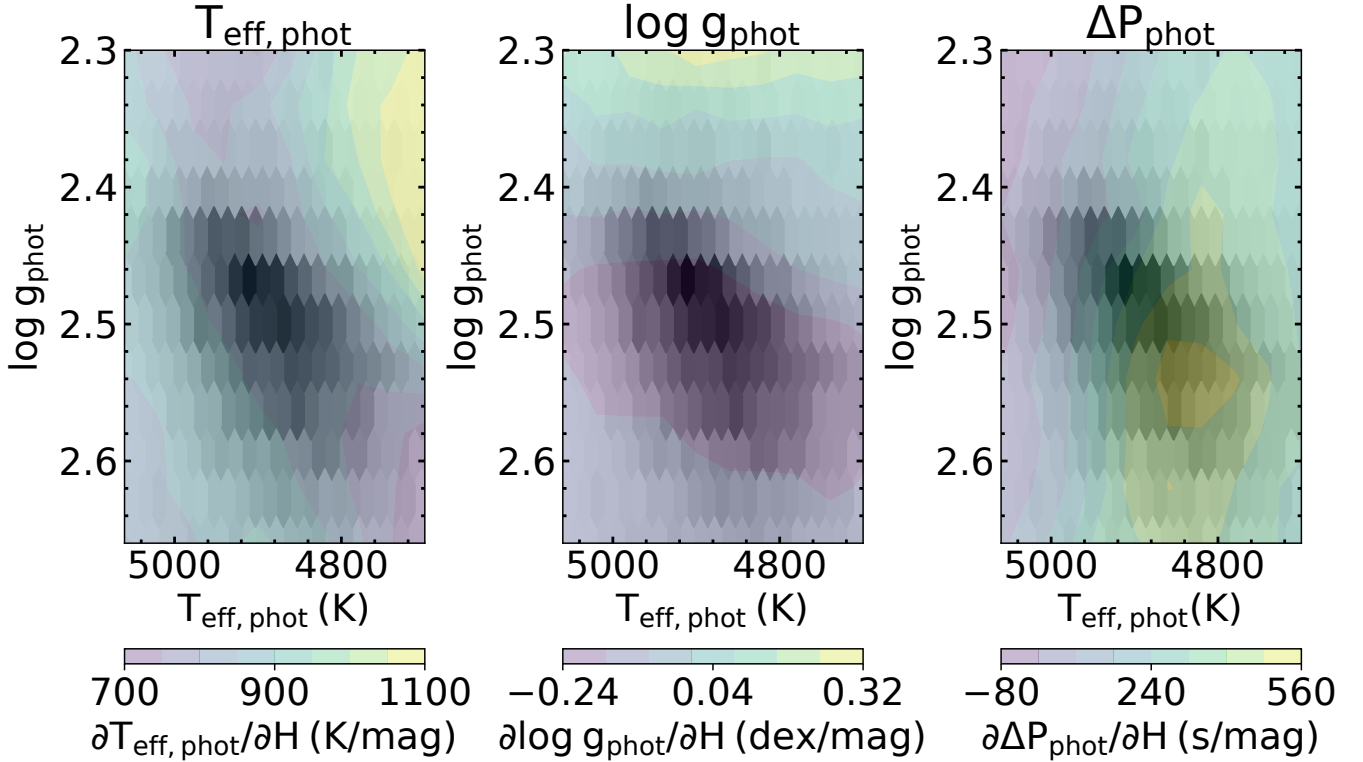
divided by the derivative of  $T_{\text{eff}}$  and  $\log g$  in order to show which bands are most important for the  $\Delta P$  inference beyond their impact on the inferred  $T_{\text{eff}}$  and  $\log g$ . From Figure 11, it is clear that the relative strength of the derivative of the  $\Delta P$  is correlated with the [C/N] information content in the selected photometric band. This indicates that the  $\Delta P$  inference is at least partially based on the change in [C/N] in the parameter range.

We also calculate the expected change in magnitude due to a change in [C/N] as a function of  $T_{\text{eff}}$  and  $\log g$  using synthetic photometry. To do this we first calculate the theoretical change in [C/H], and [N/H] as a star transitions from RGB to RC using models from Lagarde et al. (2012). We calculate these changes for stars with a variety of initial masses (0.85, 1.0, 1.5 or 2.0  $M_\odot$ ) and metallicities ( $Z=0.002, 0.004$  or 0.014) which correspond to a variety of  $T_{\text{eff}}$  and  $\log g$  during the RC phase. Therefore, we have the change in [C/N] as a function of  $T_{\text{eff}}$  and  $\log g$  for RC stars. We use these values to synthesize a grid of stellar spectra (see Section 2.2),

with RGB and RC [C/N] values. We then calculate the corresponding magnitudes for each photometric band (G, BP, RP, g, r, i, z, y, J, H, K, W1 and W2) using PYPHOT<sup>8</sup>. We calculate the expected difference in magnitude between in RGB and RC stars as a function of  $T_{\text{eff}}$  and  $\log g$  for each photometric band.

As shown in Figure 10, the derivative of  $\Delta P$  with respect to the H-band is not constant over the RC region. This is expected given that the change in [C/N] is also not constant for all RC stars. The  $T_{\text{eff}}$ ,  $\log g$  and [C/N] ratio of a RC star are a function of its initial mass and metallicity. The mass and metallicity of a star impact the depth of the convective zone during the RGB phase and therefore the amount of CN mixing. As shown in the lower panel of Figure 12, lower metallicity stars have a larger change in [C/N]. This effect is especially large for lower mass stars while stars with mass

<sup>8</sup> <https://mfouesneau.github.io/docs/pyphot>



**Figure 10.** The empirical derivatives of  $T_{\text{eff}}$ ,  $\log g$ , and  $\Delta P$  with respect to the H-band as a function of  $T_{\text{eff}}$  and  $\log g$ . In the background we show the counts of stars in grey hexagonal bins with the largest counts being the darkest. This is to show where the RC is in  $T_{\text{eff}}$  and  $\log g$  and how it correlates with the derivatives. In the left and middle plot we see the  $T_{\text{eff}}$  and  $\log g$  derivatives are smallest in the RC region while in the right plot the  $\Delta P$  derivatives are the largest in the RC region. The gradient of the derivative of  $\Delta P$  as a function of  $T_{\text{eff}}$  and  $\log g$  is consistent with what is expected for a change in  $[C/N]$  due to mixing (See Figure 13), demonstrating that the network is learning the  $[C/N]$  ratio of RC stars.

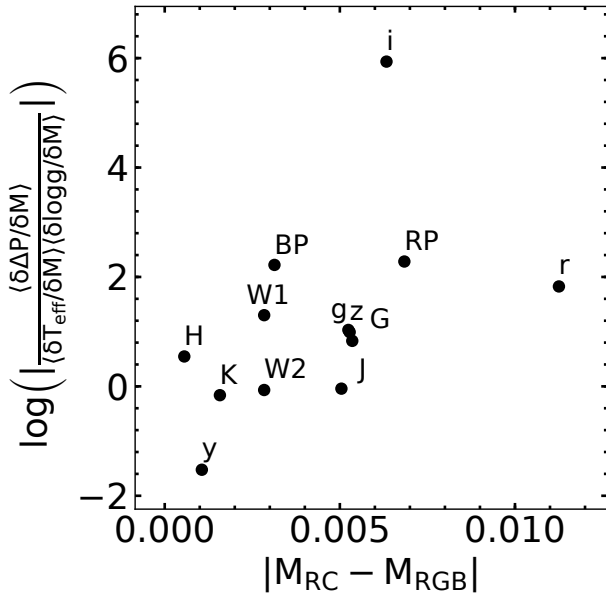
of  $2 M_{\odot}$  have almost no trend in  $[C/N]$  with metallicity. The cumulative effect of this is shown in the upper panel of Figure 12, where the change in  $[C/N]$  is shown as a function of  $T_{\text{eff}}$  and  $\log g$ . To expand on this further, we convert the change in  $[C/N]$  into a change in magnitude for selected bands using synthetic spectra. In Figure 13, we show how the theoretical changes in magnitude relate to the empirical derivatives of  $\Delta P$  with respect to the selected band as a function of  $T_{\text{eff}}$  and  $\log g$ . The shape of the contours are highly correlated. This provide further evidence that the  $\Delta P$  network is learning these small differences in magnitude between RGB and RC stars caused by the change in  $[C/N]$  and using this to infer the  $\Delta P$ .

To summarize, we have found that the inferred output uncertainties are correlated with the input uncertainties which suggests the network can identify noisy data. In addition, the effectiveness of the selection when using the inferred  $\Delta P$  compared to using just  $T_{\text{eff}}$  and  $\log g$  indicates that the  $\Delta P$  inference is derived from more than a correlation with  $T_{\text{eff}}$  and  $\log g$ . Furthermore, we calculated empirical derivatives of the  $\Delta P$  with respect to each photometric band and found that they correlate with the expected difference in magnitude due to a change in  $[C/N]$ . In total, our investigation of the trained network suggests that the  $\Delta P$  inference

uses the  $[C/N]$  information imprinted on the SED and therefore can be used to effectively select RC stars.

## 6 SUMMARY

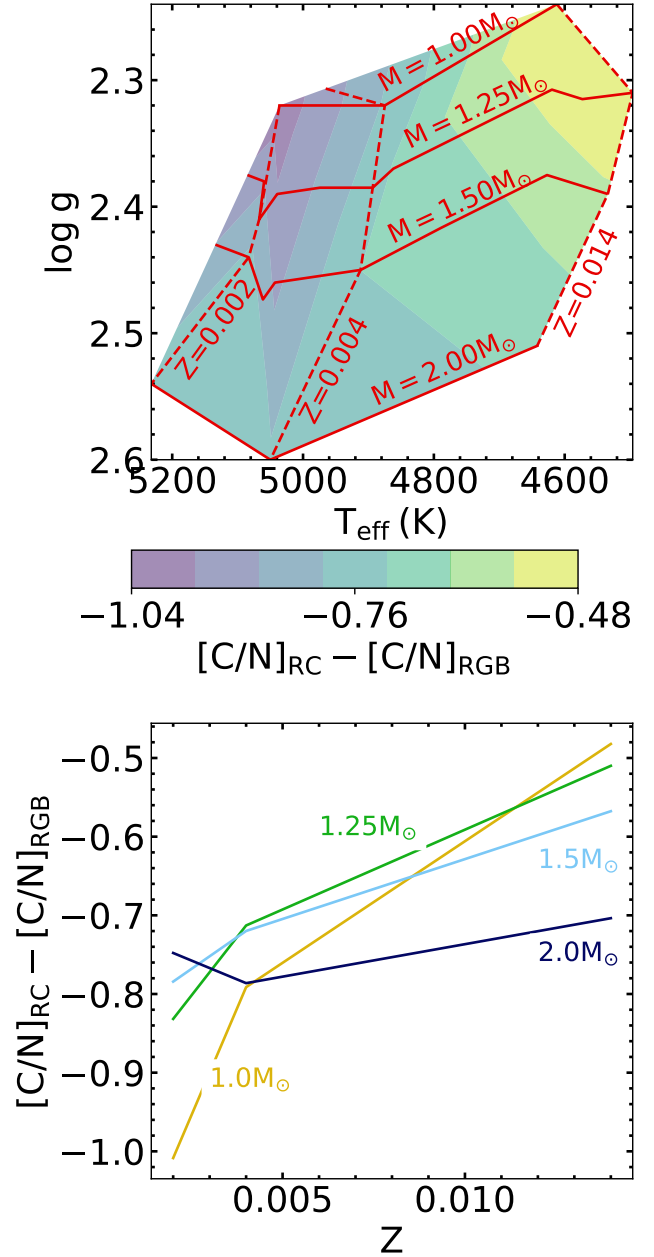
RC stars are standard candles proven to provide more accurate distance measurements than end-of-mission *Gaia* parallaxes at distances  $> 3$  kpc (Ting et al. 2018a). However, identifying large pristine samples of RC stars has historically been difficult. Red giant branch stars can have the same  $T_{\text{eff}}$  and  $\log g$  making it easy to mistake them as RC stars, especially when the  $T_{\text{eff}}$  and  $\log g$  estimates have larger uncertainties, e.g., from low-resolution spectra or SED. The asteroseismic parameters  $\Delta P$  and  $\Delta \nu$  clearly separate helium core-burning RC stars from inert core red giant branch stars. These parameters have only been derived for  $\sim 2,000$  giant stars given the difficulty of the measurement and the amount of time required for light curve observations. Furthermore, the asteroseismic sample is restricted to the Solar neighbourhood due to the magnitude limit. Recently, Hawkins et al. (2018) and Ting et al. (2018a) demonstrated that the  $\Delta P$  and  $\Delta \nu$  can be derived from single-epoch stellar spectra. Specifically, RC stars can be selected from the difference in the carbon to nitrogen ratio due to mixing that occurs at the top of the red giant branch.



**Figure 11.** How the strength of the derivative of  $\Delta P$  with respect to each magnitude relates to the absolute expected change in the magnitude due to the CN mixing between an RGB and RC star. Specifically, on the y-axis, we plot the absolute value of the ratio of the derivative of  $\Delta P$  with respect to the magnitude to the derivative of  $T_{\text{eff}}$  and  $\log g$  with respect to the magnitude. Here we assume the  $T_{\text{eff}}$  and  $\log g$  are the same for both stars ( $T_{\text{eff}} = 4604$  K and  $\log g = 2.62$  dex) but the RGB stars has a  $[C/N] = 0$  while the RC stars has  $[C/N] = -0.77$  dex. The derivatives shown correspond to the median empirical derivative found for stars in the RC region ( $4100 \text{ K} < T_{\text{eff}} < 5100 \text{ K}$  and  $2.3 \text{ dex} < \log g < 2.7 \text{ dex}$ ). The bands with derivatives of  $\Delta P$  that are relatively larger than the derivatives of  $T_{\text{eff}}$  and  $\log g$  generally also have CN molecular lines within them and therefore are more impacted by change in  $[C/N]$ . This indicates that the inference of  $\Delta P$ , and hence our RC selection, is likely aided by the imprint of a change in  $[C/N]$  on the SED.

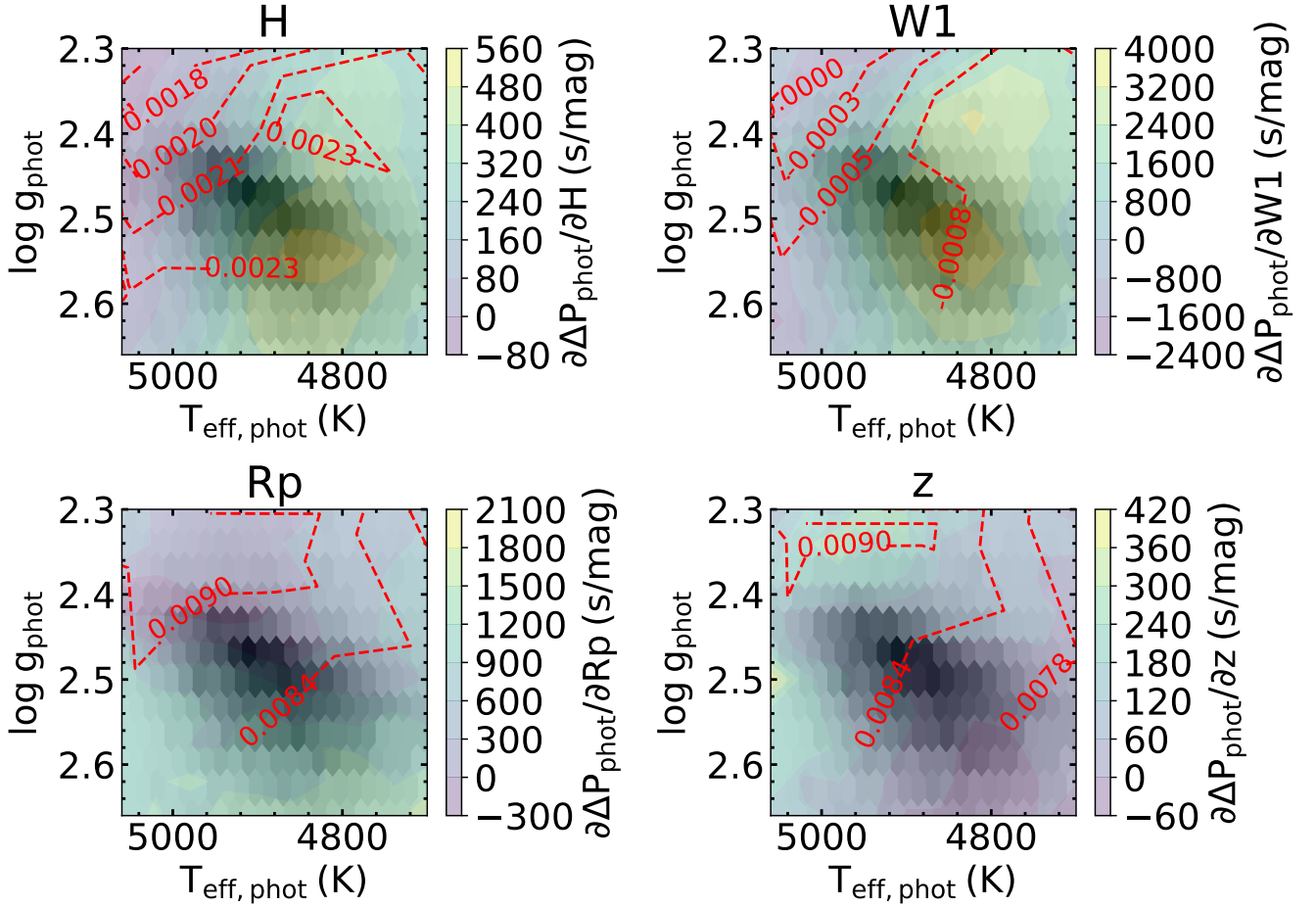
Given that the information to effectively select RC stars are in the spectra, we expect the same information to be in the SED. However, from synthetic photometry, we find that this signal is on the order of the uncertainty in the photometry. Combined with the systematics that limit SED modeling, we resort to using data-driven models and inference with probabilistic neural networks to sort out the RC signal in the photometry. Specifically, we choose to use a Mixture Density Network which provide a probabilistic output rather than a single best-estimate. This allows us to provide uncertainty estimates for our inferred parameters.

We selected RC stars from the  $\sim 200$  million stars which have photometry from 2MASS, AllWISE, *Gaia*, and Pan-STARRS. We derive the  $T_{\text{eff}}$ ,  $\log g$ ,  $\Delta\nu$ , and  $\Delta P$  of these stars from 13 bands of photometry and parallax using a mixture density network. Comparing to external validation samples, we achieve accuracies to the level of 153 K, 0.32 dex, 80s, and  $1.64 \mu\text{Hz}$  for  $T_{\text{eff}}$ ,  $\log g$ ,  $\Delta\nu$  and  $\Delta P$ , respectively. We achieve a contamination rate of  $\sim 20\%$  for a sample of 405,000 RC stars (“tier 1”). This is similar to the contamination rate found when selecting RC stars using  $T_{\text{eff}}$  and



**Figure 12.** In the upper panel we show how the expected difference in  $[C/N]$  changes as a function of  $T_{\text{eff}}$  and  $\log g$ . The models used are from [Lagarde et al. \(2012\)](#). These differences are a result of stars with different initial masses and metallicities have different changes in  $[C/N]$  and different  $T_{\text{eff}}$  and  $\log g$  during the core helium-burning phase. In the upper panel we show lines of constant mass in solid red lines and lines of constant metallicity in dashed red lines. As the mass of the star increases, so does the  $\log g$ . The  $T_{\text{eff}}$  decreases with increasing metallicity. In general, the change in  $[C/N]$  becomes larger with decreasing metallicity. At the low mass end, this effect is very strong while at the high mass end there is almost no trend in  $[C/N]$  with metallicity. This is because the mass and metallicity both impact the depth of the convective zone during the RGB phase and therefore the amount of mixing that occurs.





**Figure 13.** Showing the derivative of  $\Delta P$  with respect to the H, W1, Rp and z band as a function of  $T_{\text{eff}}$  and  $\log g$ . The counts of stars are shown in hexagonal bins with darker bins corresponding to higher counts. The red dashed contours show the theoretical change in magnitude between a RC and RGB star due to the expected change in  $[C/N]$  for the selected band as a function of  $T_{\text{eff}}$  and  $\log g$ . The theoretical change in magnitude contours are correlated with the empirical derivatives of  $\Delta P$ , indicating that the inference of  $\Delta P$  is based on the change in  $[C/N]$  between the RGB and RC.

$\log g$  derived from low-resolution spectra. We also present a catalog of 2.6 million RC stars with a contamination rate of  $\sim 33\%$  (“tier 2”). We find that extinction and the input errors impact the contamination rate. We can obtain a contamination rate of  $\sim 9\%$  for 15,000 stars with  $|b| > 30$  degrees. We derive distances with uncertainties of  $\sim 9\%$  for our sample and show the Galactic distributions. Our sample reaches deep into the Galactic bulge and halo with the most distant stars at a distance of  $\sim 20$  kpc. Therefore, this sample has vast scientific potential for Galactic studies.

We also attempt to interpret the neural networks. We find they are consistent with our expectations from stellar evolution models and synthetic spectral models. Our study is an example showing that combining domain knowledge with the power of inference of machine learning can be a useful combination. This can be used to avoid machine learning algorithms from blindly exploiting unphysical correlations in the data which can cause unknowingly biased results. By calculating empirical derivatives of the  $\Delta P$  and comparing them with model expectations in the change in  $[C/N]$ , including the effects of mass and metallicity, we conclude that

our network harnesses the difference in the  $[C/N]$  ratio to determine the  $\Delta P$ , along with the  $T_{\text{eff}}$  and  $\log g$ .

Increased precision photometric data will help to reduce the contamination. Specifically, WFIRST, Euclid and the Rubin Observatory should all provide improved photometry, especially in the infrared where 2MASS and AllWISE are currently on the order of 0.025 mag uncertainties which is larger than the expected magnitude change in these bands. In addition, these surveys will provide deeper magnitude limits and therefore will allow us to expand our sample to more distant RC stars. This will especially be useful in the Galactic center where extinction currently prevents us from seeing. Last, a larger and less contaminated training set will also improve the results. It would be ideal if we could train our network on asteroseismically derived parameters rather than spectroscopically derived. However, the asteroseismic sample is currently too small. Larger asteroseismic catalogs with derived  $\Delta P$  and  $\Delta \nu$  values, e.g. with Plato and TESS, are needed to improve the training.

## ACKNOWLEDGEMENTS

This project was initiated in the Kavli Summer Program in Astrophysics held at UC Santa Cruz in 2019, funded by the Kavli Foundation and by the National Science Foundation. We thank them for their generous support.

ML acknowledges support from the Wootton Center for Astrophysical Plasma Properties under the United States Department of Energy collaborative agreement DE-NA0003843.

YST is grateful to be supported by the NASA Hubble Fellowship grant HST-HF2-51425.001 awarded by the Space Telescope Science Institute.

NSR is partially supported under the U.S. Department of Energy contract DE-AC02-06CH11357 for the work at Argonne National Laboratory.

KH has been partially supported by a TDA/Scialog grant funded by the Research Corporation and a Scialog grant funded by the Heising-Simons Foundation. KH acknowledges support from the National Science Foundation grant AST-1907417.

The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing high performance computing resources that have contributed to the research results reported within this paper. URL: <http://www.tacc.utexas.edu>

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

This publication makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation.

The Pan-STARRS1 Surveys (PS1) have been made possible through contributions of the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max-Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, Queen's University Belfast, the Harvard-Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the National Central University of Taiwan, the Space Telescope Science Institute, the National Aeronautics and Space Administration under Grant No. NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation under Grant No. AST-1238877, the University of Maryland, and Eotvos Lorand University (ELTE).

## REFERENCES

- Anders F., et al., 2019, *A&A*, **628**, A94  
 Arenou F., et al., 2018, *A&A*, **616**, A17  
 Beck P. G., et al., 2011, *Science*, **332**, 205  
 Bedding T. R., et al., 2011, *Nature*, **471**, 608  
 Bishop C. M., 1994, Technical report, Mixture density networks  
 Bovy J., et al., 2014, *ApJ*, **790**, 127  
 Chambers K. C., et al., 2016, arXiv e-prints, p. [arXiv:1612.05560](https://arxiv.org/abs/1612.05560)  
 Chan V. C., Bovy J., 2019, arXiv e-prints, p. [arXiv:1910.00398](https://arxiv.org/abs/1910.00398)  
 Chaplin W. J., Miglio A., 2013, *ARA&A*, **51**, 353  
 Elsworth Y., Hekker S., Basu S., Davies G. R., 2017, *MNRAS*, **466**, 3344  
 Evans D. W., et al., 2018, *A&A*, **616**, A4  
 Gaia Collaboration et al., 2016, *A&A*, **595**, A1  
 Gaia Collaboration et al., 2018, *A&A*, **616**, A1  
 Girardi L., 2016, *ARA&A*, **54**, 95  
 Green G., 2018, *The Journal of Open Source Software*, **3**, 695  
 Green G. M., Schlafly E., Zucker C., Speagle J. S., Finkbeiner D., 2019, *ApJ*, **887**, 93  
 Hawkins K., Leistedt B., Bovy J., Hogg D. W., 2017, *MNRAS*, **471**, 722  
 Hawkins K., Ting Y.-S., Walter-Rix H., 2018, *ApJ*, **853**, 20  
 Iben Jr. I., 1968, *ApJ*, **154**, 581  
 Kurucz R. L., 1970, SAO Special Report, **309**  
 Kurucz R. L., 1993, A New Opacity-Sampling Model Atmosphere Program for Arbitrary Abundances. p. 87  
 Kurucz R. L., 2005, Memorie della Societa Astronomica Italiana Supplementi, **8**, 14  
 Kurucz R. L., 2013, ATLAS12: Opacity sampling model atmosphere program (ascl:1303.024)  
 Kurucz R. L., 2017, ATLAS9: Model atmosphere program with opacity distribution functions (ascl:1710.017)  
 Kurucz R. L., Avrett E. H., 1981, SAO Special Report, **391**  
 Lagarde N., Decressin T., Charbonnel C., Eggenberger P., Ekström S., Palacios A., 2012, *A&A*, **543**, A108  
 Lindegren L., 2018, Technical report, Re-normalising the astrometric chi-square in Gaia DR2  
 Lindegren L., et al., 2018, *A&A*, **616**, A2  
 Magnier E. A., et al., 2016, arXiv e-prints, p. [arXiv:1612.05240](https://arxiv.org/abs/1612.05240)  
 Mainzer A., et al., 2011, *ApJ*, **743**, 156  
 Majewski S. R., et al., 2017, *AJ*, **154**, 94  
 Marrese P. M., Marinoni S., Fabrizio M., Altavilla G., 2019, *A&A*, **621**, A144  
 Martell S. L., Smith G. H., Briley M. M., 2008, *AJ*, **136**, 2522  
 Masseron T., Gilmore G., 2015, *MNRAS*, **453**, 1855  
 Masseron T., Hawkins K., 2017, *A&A*, **597**, L3  
 Masseron T., Lagarde N., Miglio A., Elsworth Y., Gilmore G., 2017, *MNRAS*, **464**, 3021  
 Miglio A., et al., 2009, *A&A*, **503**, L21  
 Montalbán J., Miglio A., Noels A., Scuflaire R., Ventura P., 2010, *ApJ*, **721**, L182  
 Mosser B., et al., 2010, *A&A*, **517**, A22  
 Mosser B., et al., 2011, *A&A*, **532**, A86  
 Mosser B., et al., 2012, *A&A*, **540**, A143  
 Mosser B., et al., 2014, *A&A*, **572**, L5  
 Pinsonneault M. H., et al., 2014, *ApJS*, **215**, 19  
 Ruiz-Dern L., Babusiaux C., Arenou F., Turon C., Lallement R., 2018, *A&A*, **609**, A116  
 Schlafly E. F., et al., 2012, *ApJ*, **756**, 158  
 Schönrich R., McMillan P., Eyer L., 2019, *MNRAS*, **487**, 3568  
 Skrutskie M. F., et al., 2006, *AJ*, **131**, 1163  
 Stanek K. Z., Udalski A., Szymański M., Kałużny J., Kubiak Z. M., Mateo M., Krzemiński W., 1997, *ApJ*, **477**, 163  
 Stanek K. Z., Zaritsky D., Harris J., 1998, *ApJ*, **500**, L141  
 Stello D., et al., 2013, *ApJ*, **765**, L41  
 Ting Y.-S., Rix H.-W., 2019, *ApJ*, **878**, 21  
 Ting Y.-S., Conroy C., Rix H.-W., 2016, *ApJ*, **826**, 83  
 Ting Y.-S., Hawkins K., Rix H.-W., 2018a, *ApJ*, **858**, L7  
 Ting Y.-S., Conroy C., Rix H.-W., Asplund M., 2018b, *ApJ*, **860**, 159  
 Vrad M., Mosser B., Samadi R., 2016, *A&A*, **588**, A87  
 Wegg C., Rojas-Arriagada A., Schultheis M., Gerhard O., 2019, *A&A*, **632**, A121  
 Wright E. L., et al., 2010, *AJ*, **140**, 1868  
 Wu Y., et al., 2019, *MNRAS*, **484**, 5315  
 Xiang M.-S., et al., 2017, *MNRAS*, **467**, 1890

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.