From the inner to outer Milky Way: a photometric sample of 2.6 million red clump stars

Madeline Lucey , Yuan-Sen Ting, Nesar S. Ramachandra and Keith Hawkins

Reporter：Zheng Yu

Date：2025/04/25

Background

Why Study Red Clump (RC) Stars?

- **Precise Distance Measurement:**
  RC stars are standard candles, offering precise distances (<6% uncertainty) up to ~10 kpc.
  - ❖Critical for studying the **structure** and **evolution** of the Milky Way.

- **Overcoming Gaia's Limitations:**
  - ❖Gaia parallax uncertainties increase **quadratically with distance**.
  - ❖RC stars complement Gaia data, especially in the **distant Galaxy** (>3 kpc).

[1] $\sigma_d = \sqrt{\left(\frac{\partial d}{\partial \varpi}\right)^2 \sigma_\varpi^2} = \frac{\sigma_\varpi}{\varpi^2} = \sigma_\varpi d^2$, where d is the distance, ϖ is the parallax and $d = 1/\varpi$.

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

Background

Research Challenges:

- Distinguishing RC stars from Red Giant Branch (RGB) stars is difficult because they have similar properties (Teff, log g).
- Asteroseismic methods (using ΔP and $\Delta \nu$) are accurate but limited to small local samples.

Objective of This Study:

Use photometric data and machine learning to identify RC stars with:

- High accuracy (low contamination).
- Broad Galactic coverage (inner bulge to distant halo).

Data

Key Approach:

Create **Spectral Energy Distributions (SEDs)** from 13 photometric bands and Gaia parallaxes.

Gaia DR2: Parallaxes and photometry (G, BP, RP bands).
Pan-STARRS1 (PS1): Photometry (g, r, i, z, y bands).
2MASS: Infrared photometry (J, H, K bands).
AllWISE: Mid-infrared photometry (W1, W2 bands).

Use **machine learning** (neural networks) to infer stellar parameters:

- Effective temperature (**Teff**).
- Surface gravity (**log g**).
- Asteroseismic parameters (**ΔP**, Δν).

Data

$T_{\rm eff}$ = 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

Spectral and photometric differences between RC and RGB stars.

Top Panel: RGB and RC synthetic spectra show differences in CN and CO molecular bands.

Middle Panel: Magnitude differences in key photometric bands due to [C/N] variations.

Bottom Panel: Magnitude differences are of the same order as photometric uncertainties, making the RC signal challenging to detect.

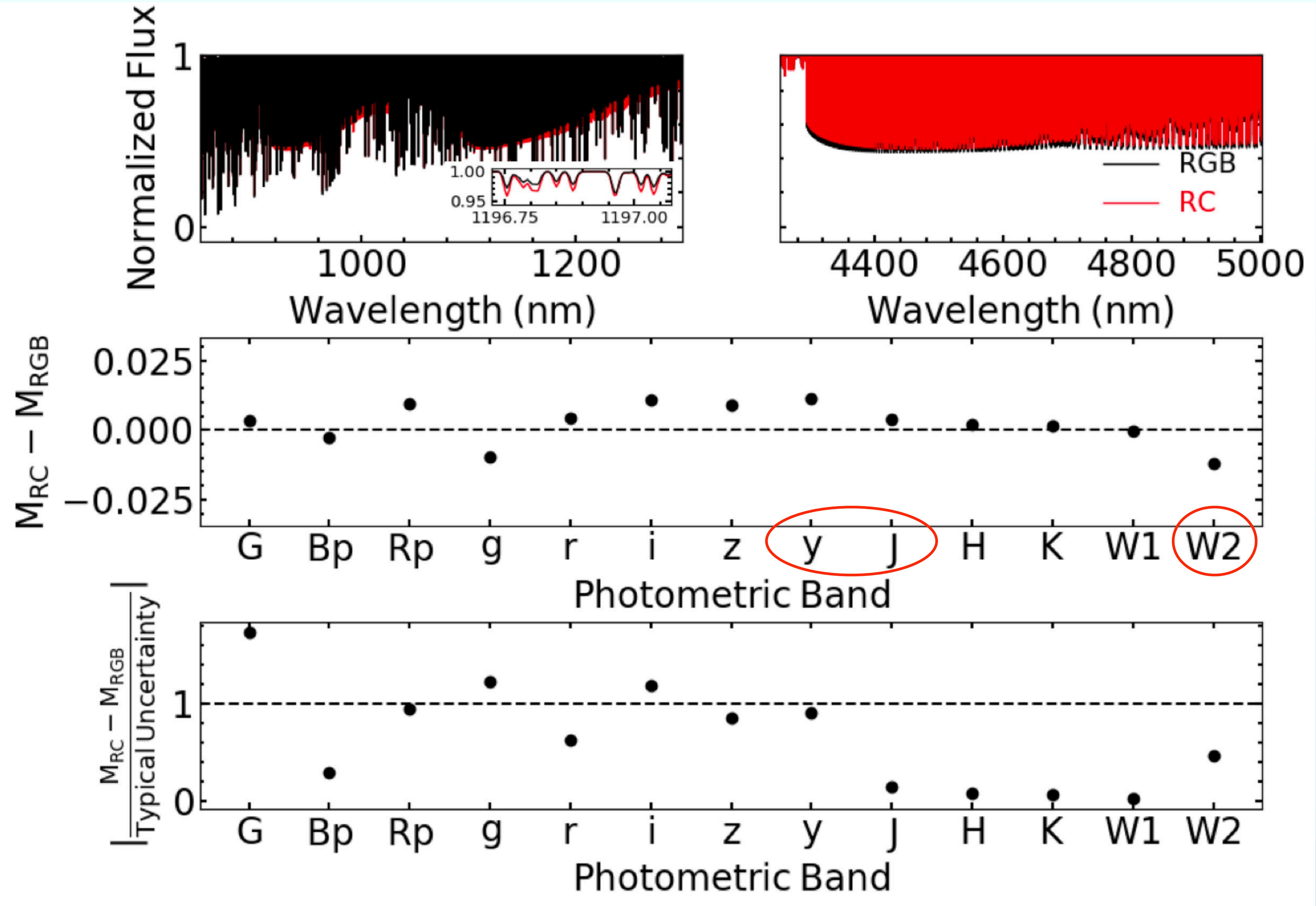

Figure 1. In the top panel, we compare the synthetic spectra of an RGB and RC star. Specifically, we show two CN molecular bands that impact the y and J photometry (top left) and a CO band which impacts the W2 photometry (top right). In the top left-hand panel we also show a zoomed-in spectrum. Both stars are synthesized with $T_{\rm eff}$ = 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

METHOD

1. <u>Mixture Density Network (MDN)</u>:
A neural network that predicts a Gaussian Mixture Model (GMM) for each output parameter.

2. <u>Inferring Teff and log g, and Selecting Giant Stars</u>

Step 1: Use MDN to infer Teff and log g for all stars in the dataset.

Step 2: Train a second MDN specifically on spectroscopic data of giant stars for improved accuracy.

T_{eff}: ±153 K ; log g: ±0.32 dex.

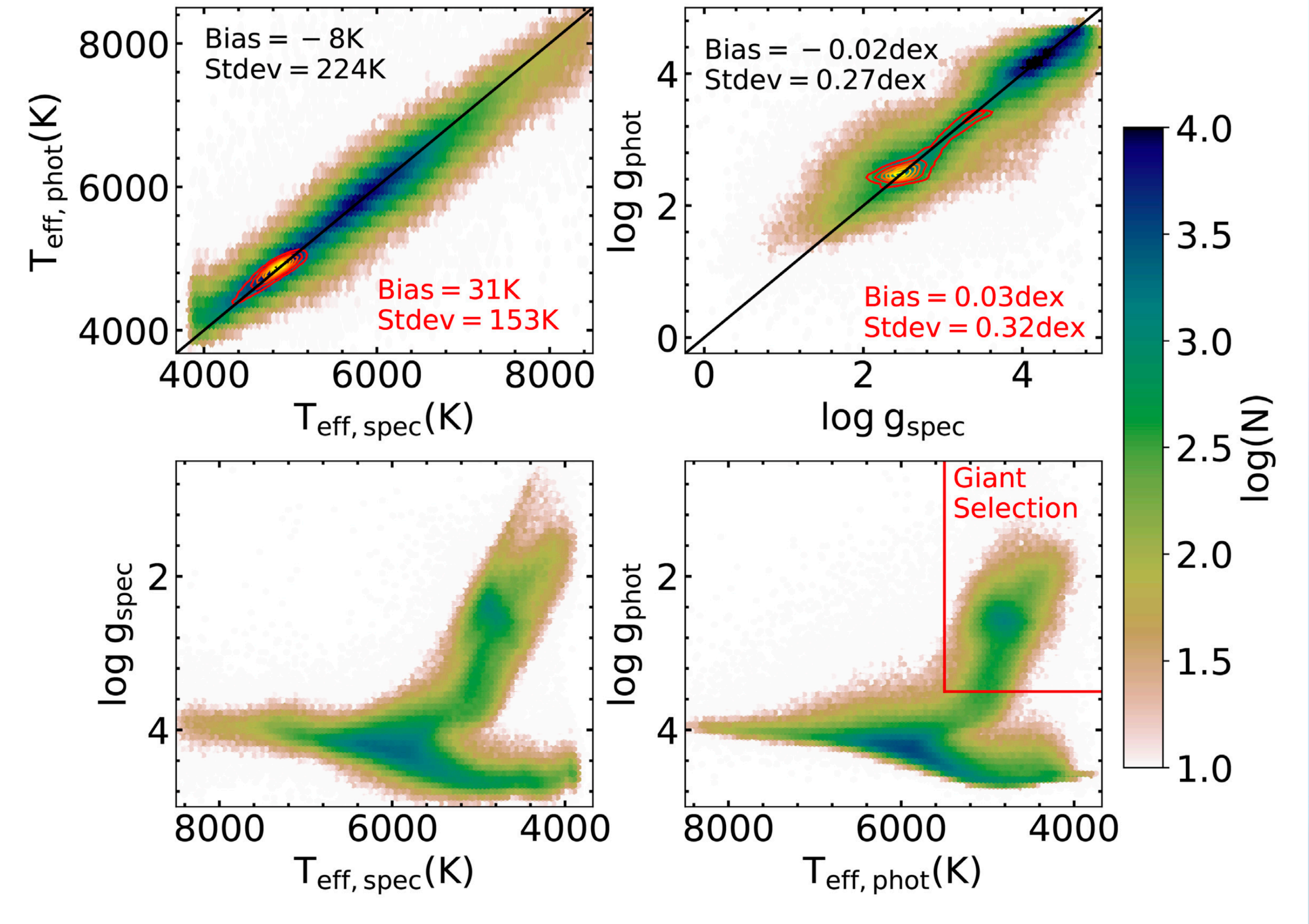

Figure 2. Comparing the stellar parameters (T_{eff} and log *g*), we derive from the photometry compared to the spectroscopically derived values from LAMOST using the validation set which was not used to train the network. The top panels show directly how the inferred values compare to the spectroscopic values for

METHOD

3. <u>Inferring Asteroseismic Parameters (ΔP, $\Delta\nu$)</u>

ΔP: ±80 s.
$\Delta\nu$: ±1.64 μHz.

Key Feature: RC stars are distinct from RGB stars in ΔP and $\Delta\nu$ space.

- RC stars: **$\Delta P > 200$ s** and **$\Delta\nu < 5$ μHz**.
- RGB stars: **$\Delta P < 100$ s** and typically higher $\Delta\nu$.

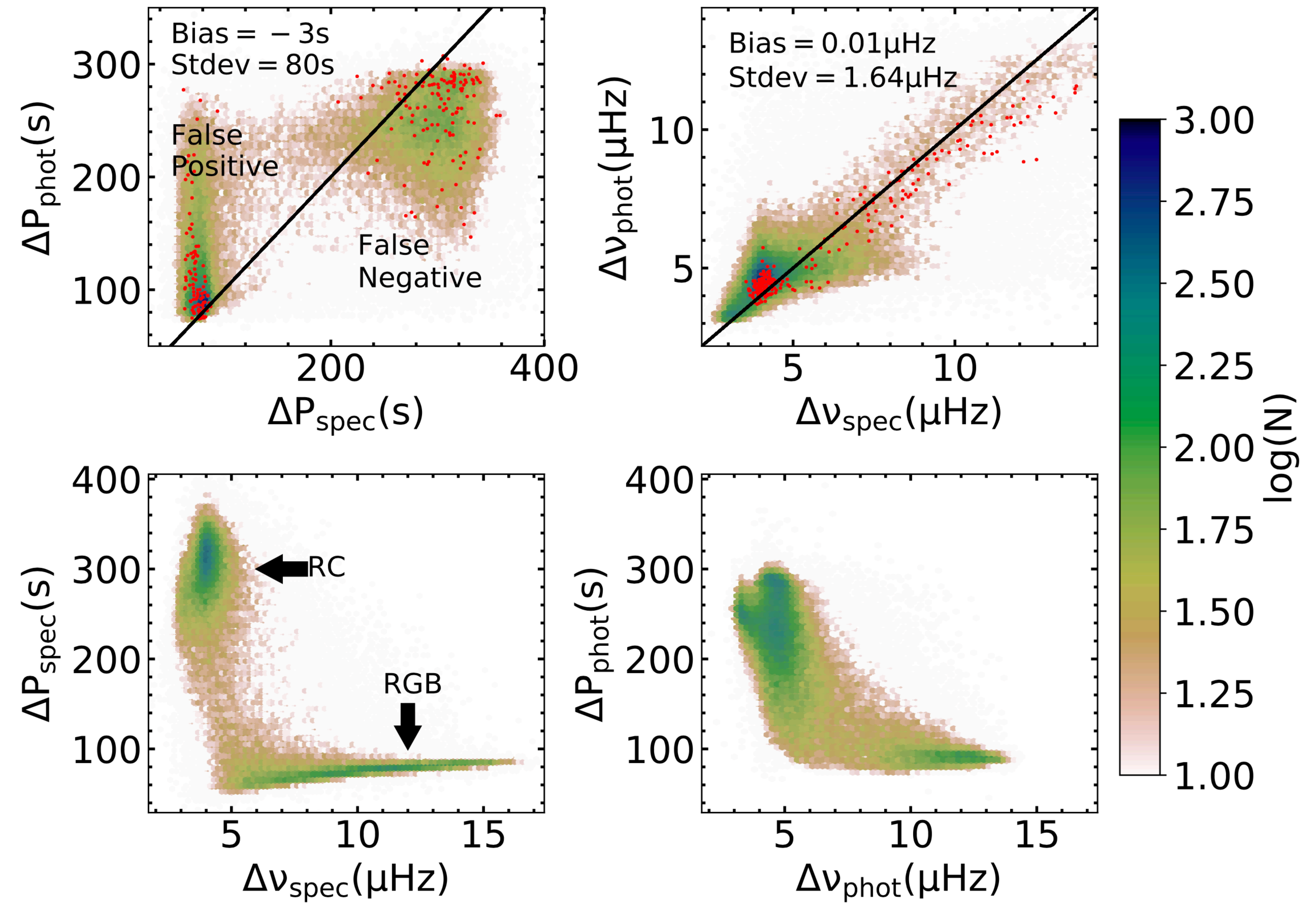

Figure 3. The asteroseismic parameters (ΔP and $\Delta\nu$) we derive for giant stars compared to the spectroscopically derived values from LAMOST spectra (Ting et al. 2018a). The red points in the top panels compare our derived photometric values with asteroseismc parameters from Vrard et al. (2016). The ΔP is the most effective parameter for selecting RC stars, since they have ΔP values more than 100s larger than RGB ΔP values. From the bottom left-hand panel, we

METHOD

4. Deriving Distances

Extinction Correction:

- Establish a **linear relationship** between photometric **Teff** and **G – W1** color for low-extinction stars.
- Use $\mathbf{A_G / A_{W1} = 16}$ (Hawkins et al. 2017) to calculate and correct for extinction in the **W1 band**.

Intrinsic Absolute Magnitude:

- Calibrate M_{W1} for RC stars to the fixed value of -1.68 ± 0.02 mag (Hawkins et al. 2017; Ruiz-Dern et al. 2018).

Comparing to *Gaia* parallaxes with uncertainties <5 per cent and assuming the *Gaia* distance as ground truth, they find their distances have uncertainties ~9 percent.

Results

THE RED CLUMP SAMPLE

Provide two RC star samples with varying contamination and completeness rates:

- **Tier 1:**

$\Delta P > 265$ s, $\Delta \nu < 5.25$ μHz and $T_{\rm eff} > 4700$ K

405,000 stars, ~20% contamination and ~25% completeness.

- **Tier 2:**

$\Delta P > 225$ s and $\Delta \nu < 5.5$ μHz

2.6 million stars, ~33% contamination and ~94% completeness.

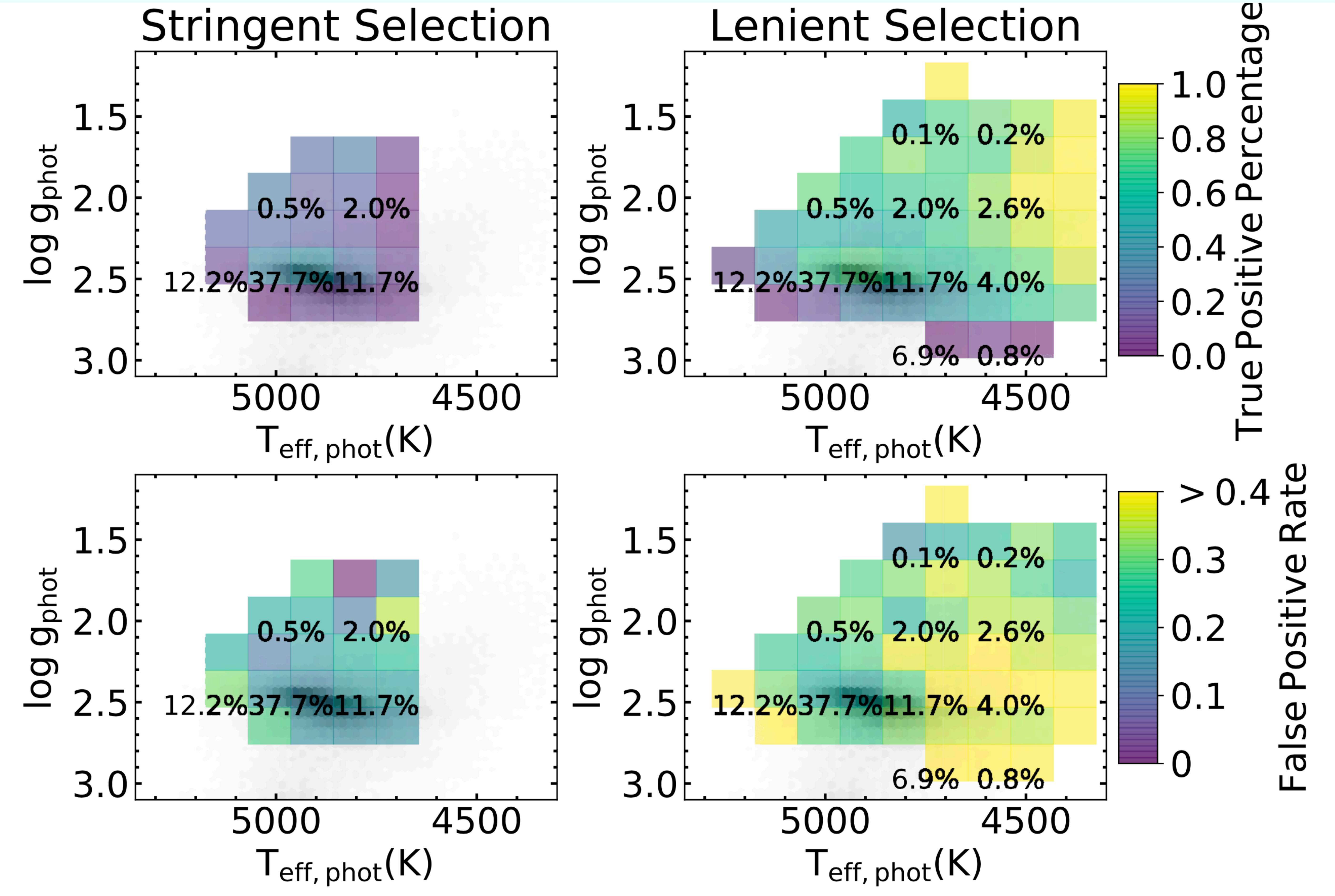

Figure 4. The true positive percentage (upper panels) and the contamination rate (lower panels) as a function of the photometrically derived $T_{\rm eff}$ and log g for both of our recommended samples. These are evaluated using the spectroscopic sample for which we know the 'ground truth'. The black text shows the

Results

Contamination rate

1. Contamination is higher at low Galactic latitudes (|b| < 10°) due to extinction, while high-latitude regions (|b| > 30°) have contamination as low as ~9%.
2. Stars with larger photometric errors, such significantly rates.
3. Selecting and with lo greatly imp RC star sam

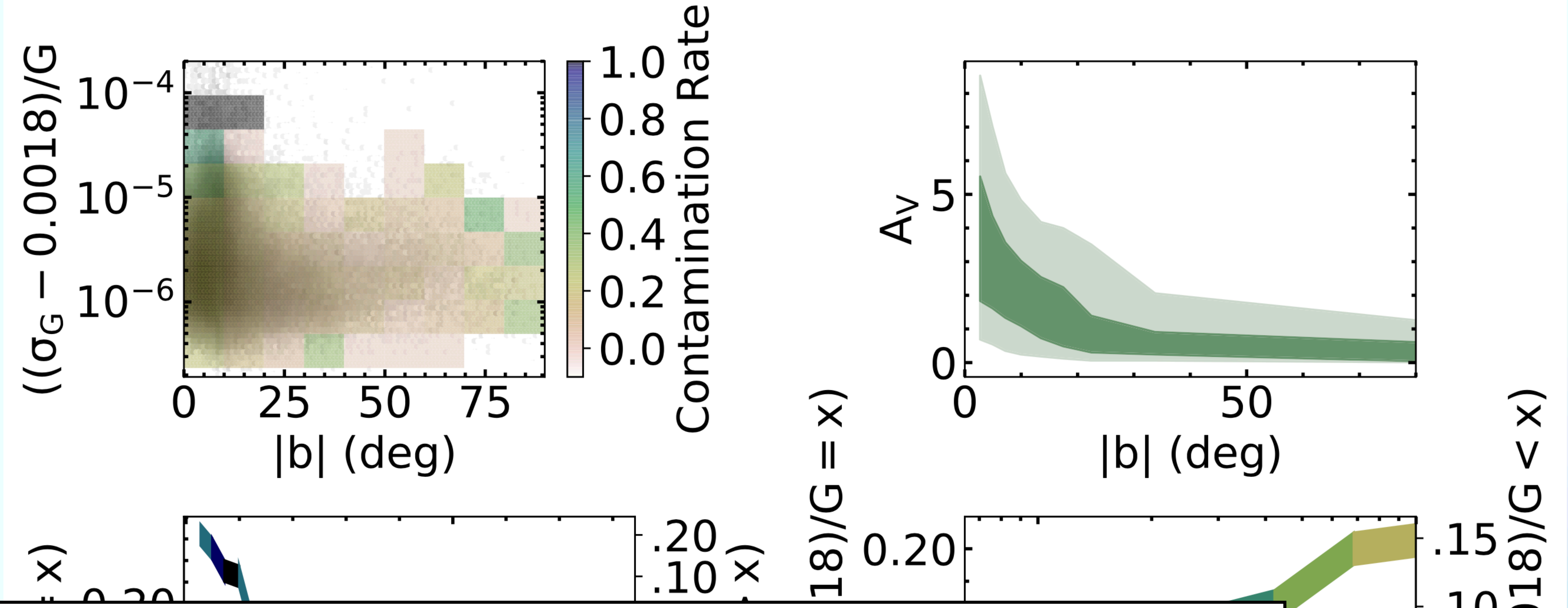

Table 1. The cuts, contamination rate, true positive percentage, and total number of stars for each of our provided samples.

Selection	ΔP cut (s)	$\Delta\nu$ cut (μHz)	T_{eff} cut	Contamination (per cent)	Completeness (per cent)	Number of stars
Tier 1	>265	<5.25	> 4700 K	20	25	405 000
Tier 2	>225	<5.5	None	33	94	2.6 million
$\|b\| > 30$ deg	>265	<5.25	>4700 K	9	12	15 000

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

er Density

Galactic latitude (b). In the
he contamination rate in the

Results

Galactic distribution of red clump sample

Distance Precision:

RC stars provide more precise distances than Gaia DR2 beyond 3 kpc, with uncertainties scaling linearly, unlike Gaia's quadratic scaling.

Sample Size and Reach:

The photometric sample is larger and extends farther into the Galactic halo and bulge, with over 1.8 million stars having more precise distances than Gaia DR2.

Scientific Potential:

Reaching up to ~20 kpc, the sample enables detailed studies of the distant Galaxy and its structure, surpassing Gaia's reliability at large distances.

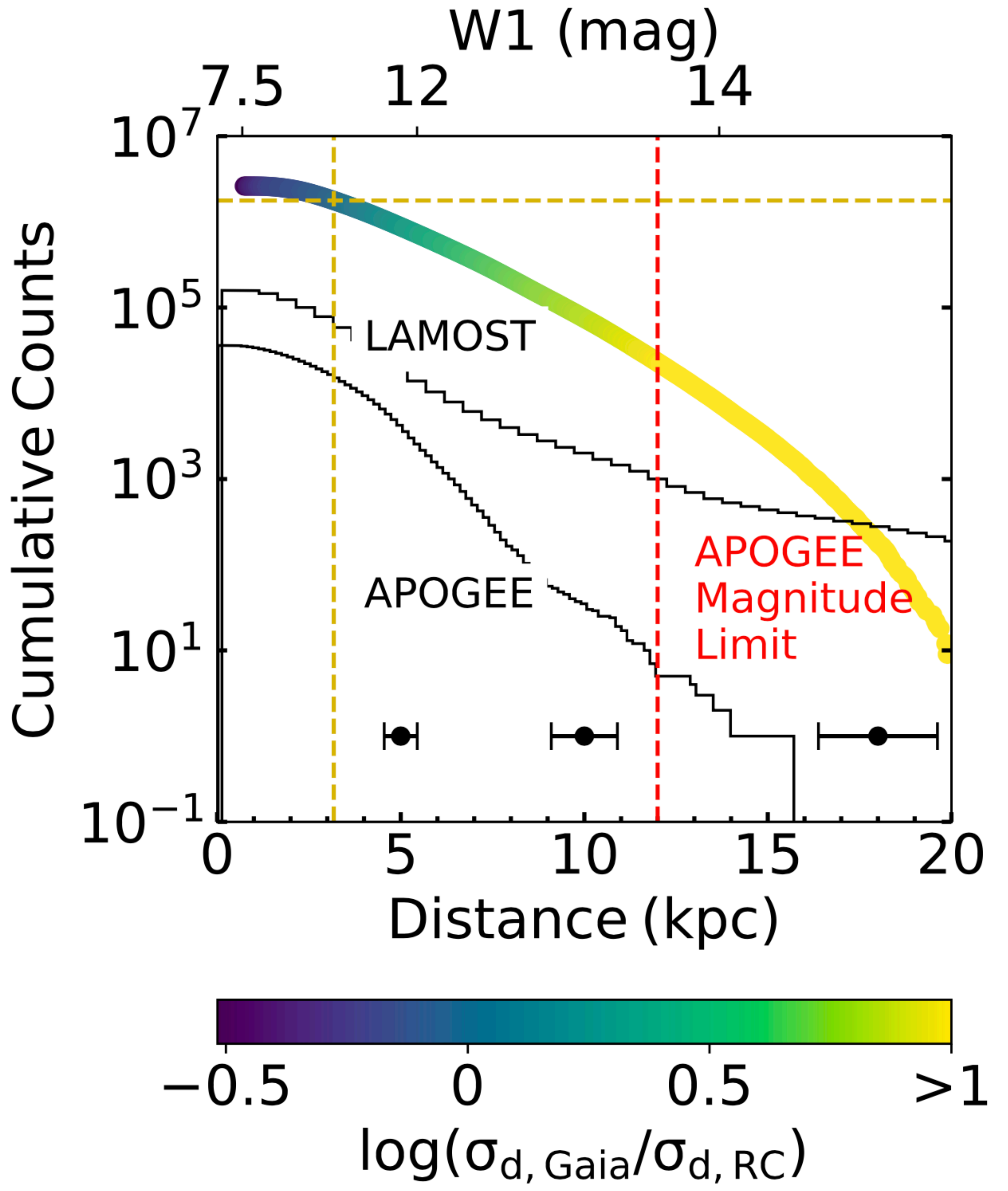

Figure 7. Cumulative distribution of distances for our sample of 2.6 million RC stars. We show the corresponding RC *W1* magnitude on the top *x*-axis

Results

Galactic distribution of red clump sample

Correlation with Input Uncertainties:
The network captures rising input noise (e.g., G−W2 for Teff, parallax for log g) and reflects it in the inferred uncertainties.

Trends by Parameter Range:
Cooler stars and low log g giants show higher uncertainties due to greater distances and larger fractional parallax errors.

Validation of Uncertainty Estimates:
The network effectively models input noise and parameter specific uncertainties, ensuring reliable inferences.

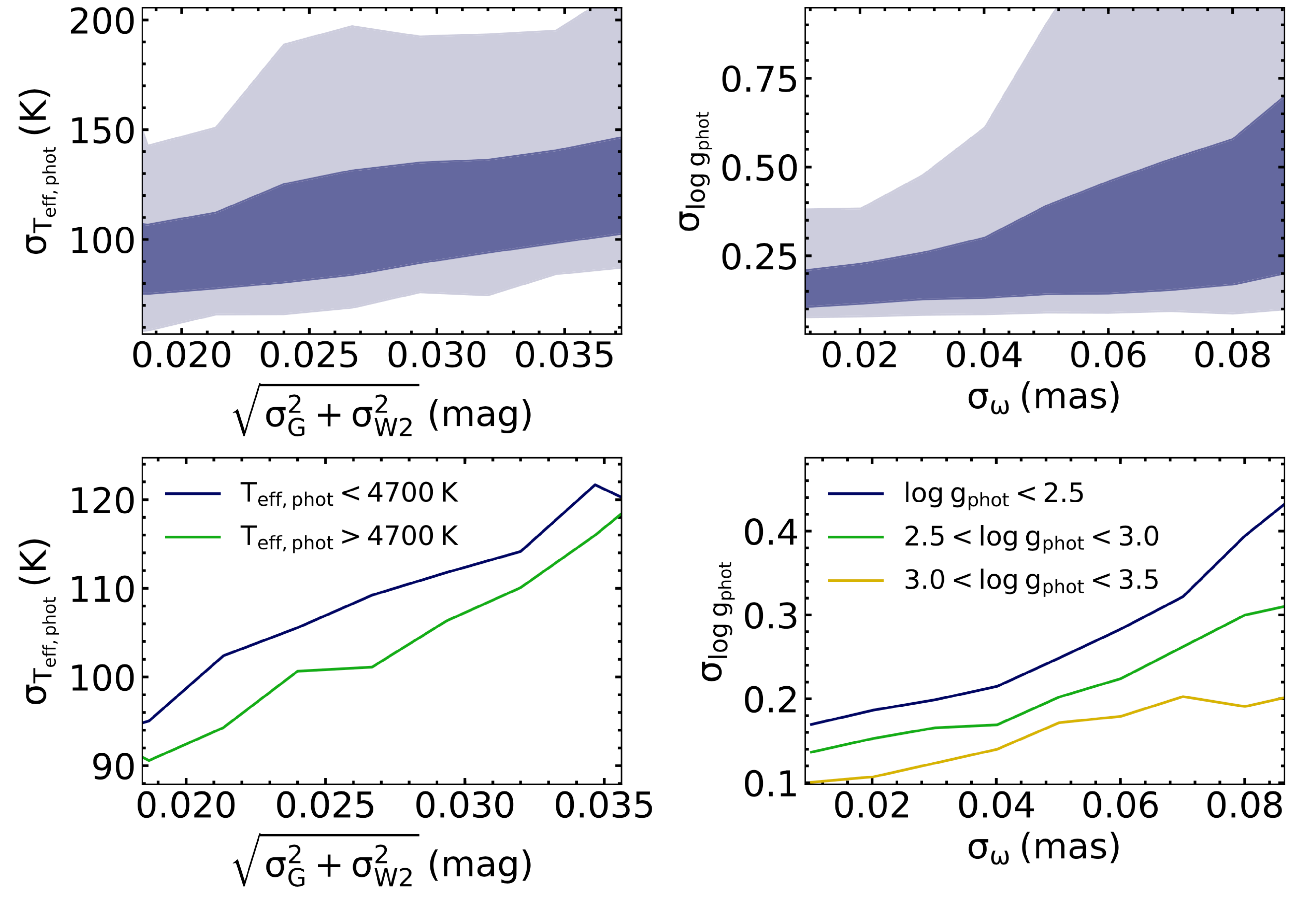

Figure 9. The uncertainties of inferred parameters $T_{\rm eff}$ and $\log g$ as a function of the uncertainties in the input data. Specifically, we compare them to

Results

The derivative of ΔP with respect to the H band peaks in RC region, showing the network uses [C/N] related photometric signals to infer ΔP. Distinct gradient patterns for Teff, log g, and ΔP confirm that ΔP inference is not solely based on correlations with T_{eff} and log g. These patterns align with theoretical expectations, validating the network's use of physical features like [C/N] differences.

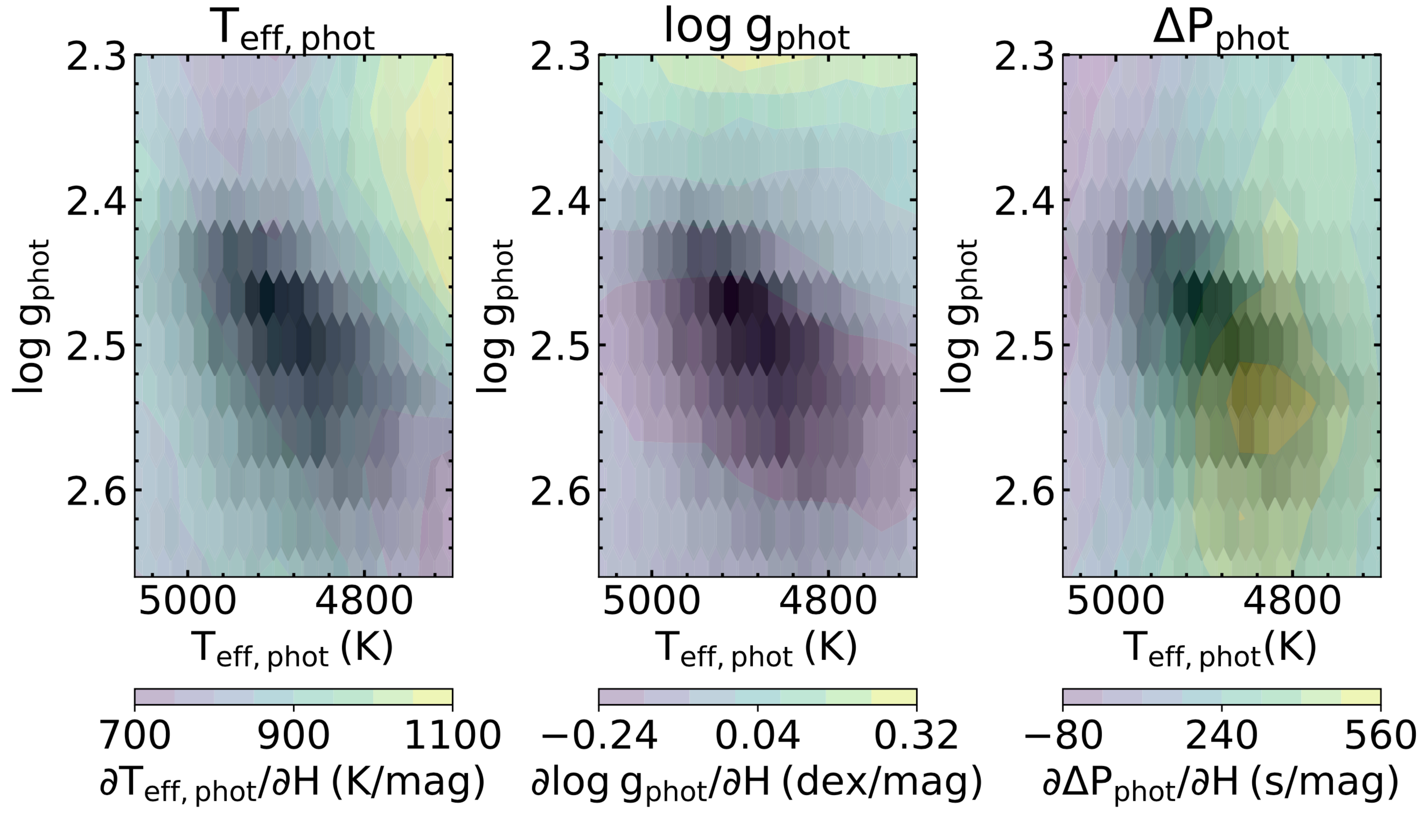

Figure 10. The empirical derivatives of $T_{\rm eff}$, log g, and ΔP with respect to the H band as a function of $T_{\rm eff}$ and log g. In the background, we show the counts

Results

Stellar evolutionary models from Lagarde et al. (2012)

Solid red lines: Represent stars with constant mass.

Dashed red lines: Represent stars with constant metallicity.

1. **Low metallicity stars** show stronger changes in [C/N], especially at the **low-mass end**.
2. At **high masses**, metallicity has minimal effect on [C/N].

These trends are caused by differences in the depth of the convective zone:

- Lower metallicity and lower mass stars have deeper convective zones, leading to more extensive mixing.
- Greater mixing increases the transport of processed material (e.g., nitrogen) to the surface.

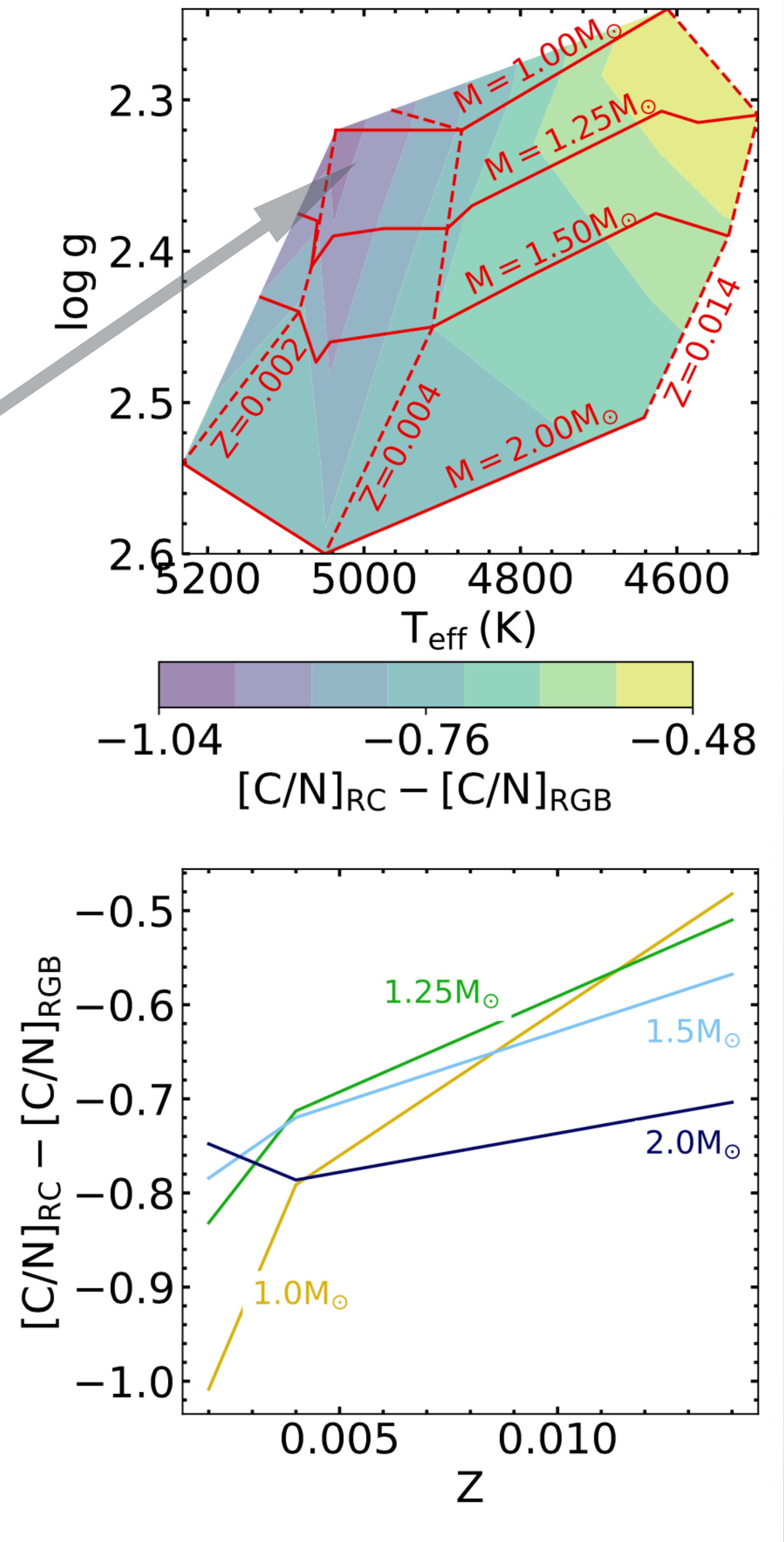

Figure 12. In the upper panel, we show how the expected difference in [C/N] changes as a function of T_{eff} and log g. The models used are from

Summary

- A photometric catalogue of 2.6 million red clump (RC) stars was created using machine learning and stellar models.
- RC stars were identified through [C/N] variations caused by mixing during the red giant branch phase.
- Catalogue combines data from 2MASS, AllWISE, Gaia, and Pan-STARRS with precise distances (~9% uncertainty).
- Two samples:
    - Tier 1: ~405,000 stars with 20% contamination.
    - Tier 2: ~2.6 million stars with 33% contamination.
- Extends to >10 kpc, enabling studies of the Milky Way's structure and evolution.