ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep1452

Evaluation and Comparison of Single-Color and Multi-Parameter Photometric Temperature Calibrations with Gaia DR3 and Hipparcos Catalogue

Sayak Biswas[1]

[1] Barrackpore Rastraguru Surendranath College

Publication Date: 2025/10/03

Abstract: we present a comparative analysis of two photometric calibration methods which is used to calculate the effective temperature and the uncertainties of 109 stars present in both Hipparcos and Gaia Dr3 catalogue .Stellar effective temperature were computed with (i) the color-temperature relation of Flower(1996)[5],which only uses B-V,and (ii)Casagrande(2010)[6] calibration that incorporates [Fe/H]. Values of B-V were taken from Hipparcos[3] and its input uncertainties were prop- agated through independent Monte Carlo [4] sampling(For input uncertainties of [Fe/H] their DR3 percentiles were converted to equivalent 1 sigma error which were propagated).(Gaia Dr3[2] effective temperature and uncertainties taken as refer- ence) by examining systematic offsets,scatter,the behaviour of normalised residuals and by testing how differences depends on metalicity.We must note that B-V used here are undereddened nor systematic correction was used to rectify either of the calibrations and Gaia itself has systematic uncertainties.

How to Cite: Sayak Biswas (2025) Evaluation and Comparison of Single-Color and Multi-Parameter Photometric Temperature Calibrations with Gaia DR3 and Hipparcos Catalogue. *International Journal of Innovative Science and Research Technology*, 10(9), 2328-2336. https://doi.org/10.38124/ijisrt/25sep1452

I. INTRODUCTION

Effective temperatures are fundamental to almost all areas in astronomy and as- trophysics the accuracy of star's effective temperature directly influences a wide range of stellar parameters like determining bolometric corrections,luminosity and many more and they also set the location of a star in the Hertzsprung- Russell diagram.For large scale stellar population surveys,$T_{\rm eff}$ is most often estimated either from broad-band photometric color–temperature relations or from spectrophotometric pipelines.We explicitly note that each approach carries different systematic errors therefore its mandatory to understand their relative behaviour and uncertainty realism is very important while comparing heteroge- nous datasets. (1)Classic empirical color–temperature relations exemplified by Flower's (1996) [5]BV → $T_{\rm eff}$ calibration remain widely used because they are simple to apply and require only readily available photometry, Such single colorrelations, can introduce systematic biases because of its assumption of a standard composition of metalicity .To mitigate that,multiparameter calibration that in- cludes metalicity have been developed by Casagrande(2010) [6] (2) The Gaia Data Release 3 [2]has photometric and spectrophotometric temperature esti- mates of millions of stars and a large number of them are the same stars once studied and recorded in Hipparcos catalogue [3].And these results are used for many studies.Gaia's $T_{\rm eff}$ values are derived from BP/RP spectrophotometry and other inputs are accompanied by posterior percentiles (commonly reported as 16th and 84th percentiles giving an 68% confidence level). while Gaia pro- vides a large catalog, its temperatures are derived from models and pipelines where its very obvious to contain small systematic offsets(till 8000k Gaia's $T_{\rm eff}$ estimation contains very small systematic offsets).Thus Gaia DR3 is the most useful practical reference available for comparing two calibration method. (3)In this work we perform a systematic and quantitative comparison between two calibration methods with Gaia's Data as reference for a sample of 109 stars for $T_{\rm eff}$ less than 8000k. (a) $T_{\rm eff}$ derived with Flower (1996)[5] using un-dereddened BV color, (b) $T_{\rm eff}$ from a Casagrande-style[6] calibration that explicitly includes [Fe/H] , and (c) Gaia DR3 [2]$T_{\rm eff}$ (with 16th/84th percentiles).And it must be noted that our sample is prefiltered to the parameters where both the calibration methods are applicable for use.Important methodological choices are made here which includes propagating the input uncertainties and calibrated uncertainties using independent Monte Carlo [4] sampling and converting Gaia's percentile to 1-equivalent uncertainties where it was required for computational simplicity to do a direct comparison.A important constraint in this comparison analysis is that the B-V values are not dereddened nor systematic

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep1452

correction was used to rectify either of the calibrations(redding and extinction data were not available);note that diffrential comparisons between methods are still informative.

II. DATA AND METHODS

➢ *Catalogs and Sample Selection*

We are studying a sample of 109 stars obtained by cross-matching Hipparcos photometry with Gaia DR3 parameter tables. From Hipparcos[3] we involved the Johnson $B - V$ color and its observational uncertainty, e_{B-V} . From Gaia DR3[2] we used its available effective temperature T^{Gaia} (reported as the 16th, 50th (median), and 84th percentiles), the metallicity proxy [Fe/H]. We retained sources only if they had non-null values for $B-V$, e_{B-V}, Gaia T_{eff} percentiles, [Fe/H] and if they lie inside the validity ranges of the photometric calibrations employed (see Section 1.2). Known low-quality Gaia datas were removed using recommended quality flags (To ensure that only stars with reliable effective temperature estimates from Gaia DR3 (GSP-Phot module) are included, we applied the following quality cuts to the catalog:

- *Temperature Uncertainty:*

We needed a relative uncertainty smaller than 10%(teff gspphot error is equivalent to 1σ,which is the converted approximate symmetric uncertainty)

$$\frac{\text{teff gspphot\_error}}{\text{teff gspphot}} < 0.1.$$

- *Astrophysical Flags: Stars with no warning flags were retained,*

$$\text{flags gspphot} = \text{'0000000'}.$$

- *Astrometric Quality: Stars with poor astrometric fits were excluded by requiring*

$$\text{ruwe} < 1.4,$$

Where ruwe is the Renormalised Unit Weight Error.

- *Photometric Consistency: We applied the recommended BP/RP excess factor filter to ensure reliable color indices,*

$$1.0+0.015\,(\text{bp rp})^2 < \text{phot bp rp excess factor} < 1.3+0.06\,(\text{bp rp})^2.$$

- *Valid Temperature Range: To avoid poorly constrained solutions, only stars within*

$$3000\,K < \text{teff gspphot} < 8000\,K$$

were kept.

After applying these filters, the resulting catalog contains only stars with accu- rate and robust Gaia DR3 effective temperature measurements then we applied the parameters as specified by Torres(2010)[1].

➢ *Photometric Calibrations*

We calculated effective temperatures from photometry using two independent calibrations:

- *Flower (1996).*

We applied [5]Flower's empirical Johnson $B-V \rightarrow T_{\text{eff}}$ polynomial relations to the (un-dereddened) Hipparcos[3] $B - V$ values. Flower's calibration is single-color and does not explicitly include [Fe/H]; we therefore restrict interpretation of Flower-derived temperatures to the calibration's published color and temperature validity ranges. (Used the corrected polynomial coefficients where appropriate; see Torres 2010[7] for errata).

- *Casagrande-Style Multi-Parameter Calibration.*

We used a Casagrande(2010)[6] multi-parameter empirical relation that returns T_{eff} as a function of $B-V$, [Fe/H]. When the published polynomial form was used we applied it di- rectly; we limited results to the method's domain of validity.

All $B - V$ values in the present study are *not* dereddened nor systematic correction was used because a homogeneous extinction catalog was not available for the the sample in Gaia or Hipparcos catalog of stars we are working with. We discuss the implications of this choice in the conclusions, but it must be noted that it primarily affects absolute temperature scales while leaving many differential trends (e.g., metallicity dependence) qualitatively robust.

- *Uncertainty Symmetry Selection.*

We tested catalogued uncertainties in metallicity ([Fe/H]) for asymmetry before performing analyses that assume Gaussian symmetric errors. For each star we used the catalog-provided 16th, 50th (median) and 84th percentiles. We defined the downward and upward half-widths as

$$\sigma_- = \text{median} - p_{16}, \quad \sigma_+ = p_{84} - \text{median}.$$

Stars with $\sigma_+/\sigma_- > 1.20$ or with relative asymmetry $|\sigma_+ - \sigma_-|/((\sigma_+ + \sigma_-)/2) > 0.20$ were excluded from analyses that required symmetric errors. Uncertain- ties in ([Fe/H] are reported as median and 16/84 percentiles. Only stars with asymmetry metrics $\sigma_+/\sigma_- \leq 1.20$ (or relative asymmetry ≤ 0.20) were retained for the symmetric-error analysis.The final, cleaned working sample contains 109 (Hipparcos matched Gaia DR3)stars.

➢ *Uncertainty Propagation: Monte Carlo [4] Sampling*

To propagate input uncertainties and capture non-Gaussian, asymmetric pos- terior shapes in T_{eff} , we used

Volume 10, Issue 9, September – 2025 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep1452

the input observable distributions using a Monte Carlo sampler, executed independently for each and every star and for both the calibrations.

- *Input Distributions.*

$B - V \sim \mathrm{N}(B - V_{\mathrm{cat}},\ \sigma = e_{B-V})$.

Fe/H were sampled as Gaussians centered on the Gaia DR3 median values with standard deviation approximated by $(p_{84} - p_{16})/2$ when the 16th/84th percentiles were reported.

Most of the priors were weakly informative Gaussians (centered on catalog values) and truncation was applied only to enforce physically reasonable bounds where necessary (e.g. $T_{\mathrm{eff}} > 3000\,\mathrm{K}$).

➢ *Comparison Methodology and Statistical Measures*

We compare photometric temperature estimates to Gaia DR3[2] as our practical reference by computing, for each star and method, the posterior distribution of the difference

$$\Delta T = T_{\mathrm{eff}}^{\mathrm{method}} - T_{\mathrm{eff}}^{\mathrm{Gaia}}$$

Because both the methods and Gaia posteriors can be asymmetric, we draw independent Monte Carlo [4] samples from both posteriors and compute the distribution of ΔT sample-by-sample. From the ΔT posterior we report the median and 16th/84th percentiles.We also compute the probability $P(\Delta T > 0)$, which represents the posterior probability that the weather the method yields a hotter temperature than Gaia for a given star.

When a single normalized residual is required we compute

$$z_i = \frac{T_{\mathrm{eff},i}^{\mathrm{method}} - T_{\mathrm{eff},i}^{\mathrm{Gaia}}}{\sqrt{\sigma^2_{\mathrm{method},i} + \sigma^2_{\mathrm{Gaia},i}}}.$$

where σ is taken as $(p_{84} - p_{16})/2$ for asymmetric posteriors. We report the fraction of stars with $|z| \le 1$ and $|z| \le 2$ (expected $\approx 68\%$ and $\approx 95\%$ for correctly calibrated Gaussian errors), the reduced chi-square $\chi^2_\nu = \frac{1}{N-1}\sum_i z_i^2$, and the Pearson correlation coefficient between method and Gaia temperatures. Linear trends of ΔT vs [Fe/H] are quantified by least–squares fits; slope uncertainties are derived by bootstrapping. The final, cleaned working sample contains 109 (Hipparcos matched Gaia DR3 stars). We tested catalogued uncertainties in effective temperatures calibrated from both the method and Gaia's reported values for asymmetry before performing analyses that assume Gaussian symmetric errors. For each star we used the catalog-provided 16th, 50th (median) and 84th percentiles. We defined the downward and upward half-widths as

$$\sigma_- = \mathrm{median} - p_{16},\quad \sigma_+ = p_{84} - \mathrm{median}.$$

Stars with $\sigma_+/\sigma_- > 1.20$ or with relative asymmetry $|\sigma_+ - \sigma_-|/((\sigma_+ + \sigma_-)/2) > 0.20$ Uncertainties in effective temperature are reported as median and 16/84 percentiles. 70.6% of gaia's values,99.1% of Flower's calibrated values and 98.2% of Casagrande's calibrated values in the given sample has the asymmetry met- rics $\sigma_+/\sigma_- \le 1.20$ (or relative asymmetry ≤ 0.20). *Note: While the ΔT poste- rior analysis preserves asymmetric uncertainties, the normalized residual met- ric employs a symmetric approximation $\sigma = (p_{84} - p_{16})/2$ for computational convenience in aggregate statistics. This simplification allows standard metrics (fraction within —z— 1,2 and χ^2) but may slightly misrepresent significance for individual star with highly asymmetric posteriors.*

III. RESULTS AND FIGURE-BY-FIGURE ANALYSIS

Below we present all diagnostic figures produced for both photometric and spec- trophotometric calibrations and give a precise interpretation for each panel. All figures use the per-star posterior medians and 16th/84th percentiles produced by our Monte Carlo [4] independent sampling propagation. We begin with the global visual comparisons, then the residual / uncertainty diagnostics, then Monte Carlo [4] sampler diagnostics, then dependence on stellar parameters.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep1452

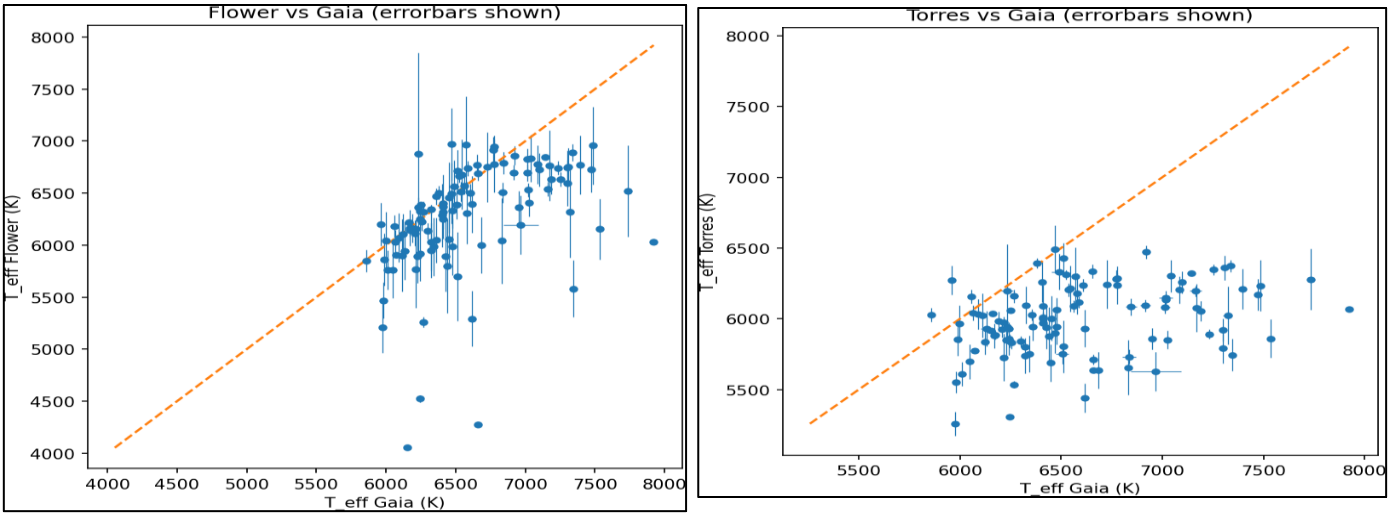

Fig 1 (a) Flower vs Gaia (Errorbars Shown). (b) Casagrande vs Gaia (Er- Rorbars Shown). The Dashed Line Shows Equality.

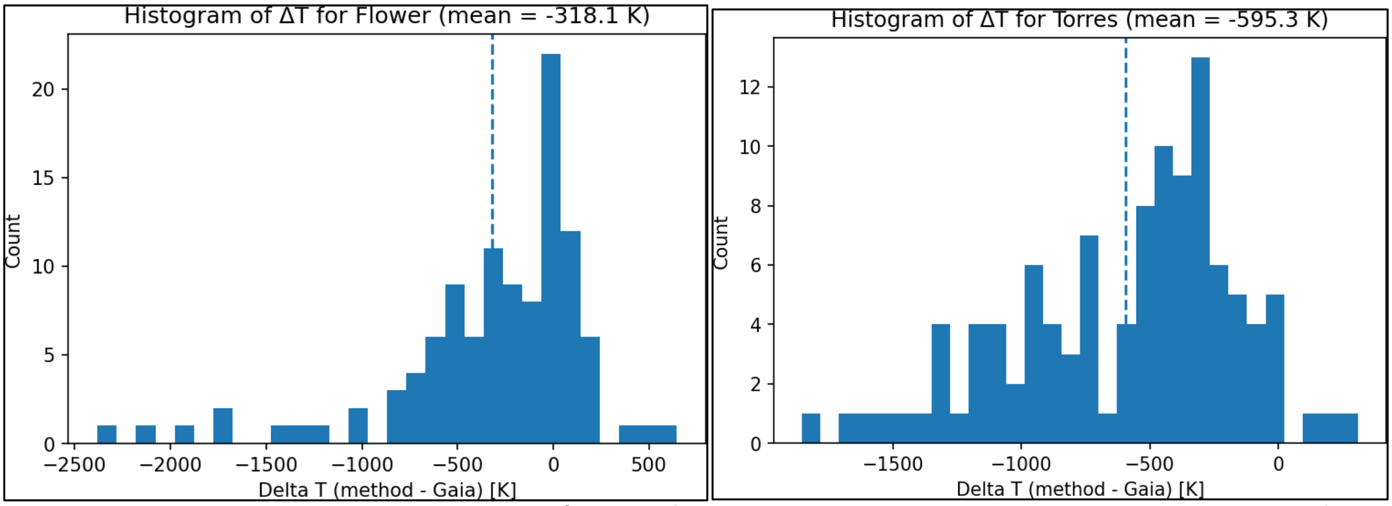

Fig 2 (a) Flower: Histogram of $\Delta T = T^{\text{Flower}} - T^{\text{Gaia}}$. (b) Casagrande: Histogram of $\Delta T = T^{\text{Torres}} - T^{\text{Gaia}}$.

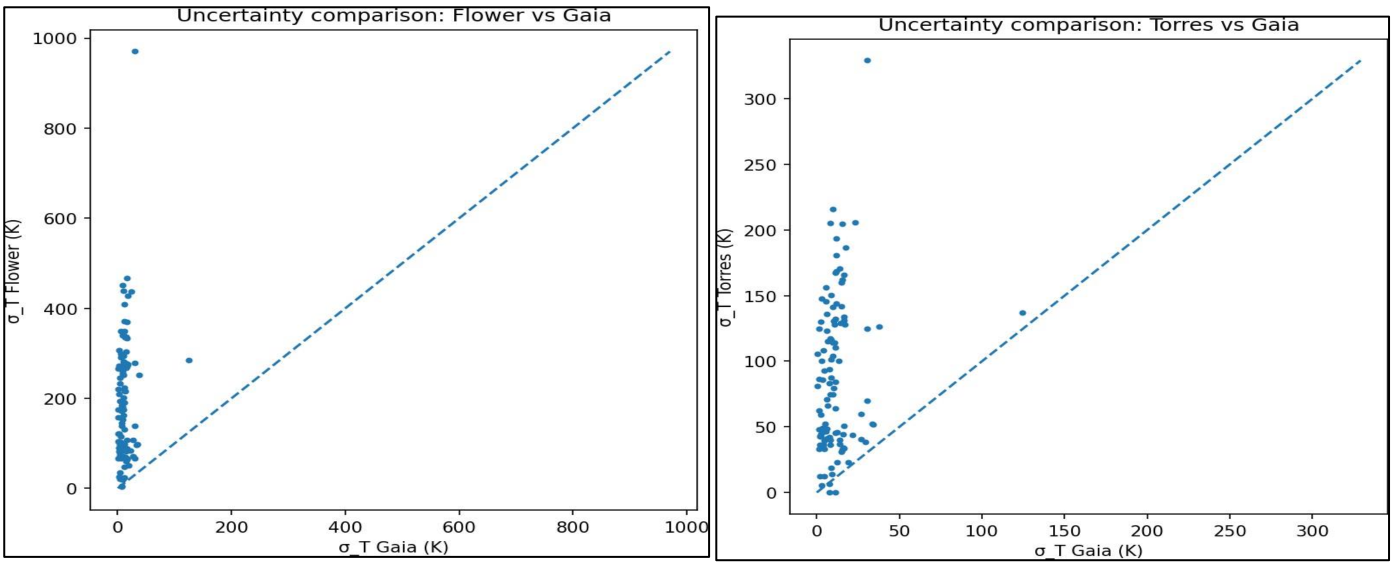

Fig 3 (a) Flower: Reported σ_{method} vs σ_{Gaia}. (b) Casagrande: Reported σ_{method} vs σ_{Gaia}. Points Above the 1:1 Line Indicate Method Uncertainties Larger than Gaia's.

Volume 10, Issue 9, September – 2025 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep1452

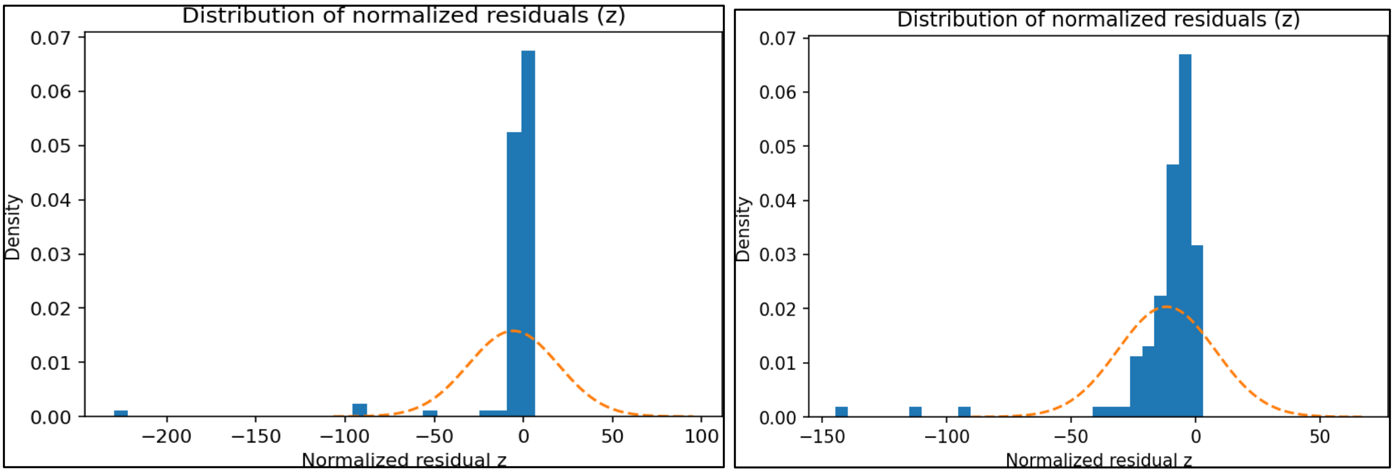

Fig 4 (a) Flower: normalized-residual histogram $z = (T_{\text{method}} - T_{\text{Gaia}})/\sqrt{\sigma_{\text{method}}^2 + \sigma_{\text{Gaia}}^2}$. (b) Casagrande: same. For well-calibrated Gaussian errors, ≈ 68% of values fall within $|z| \leq 1$.

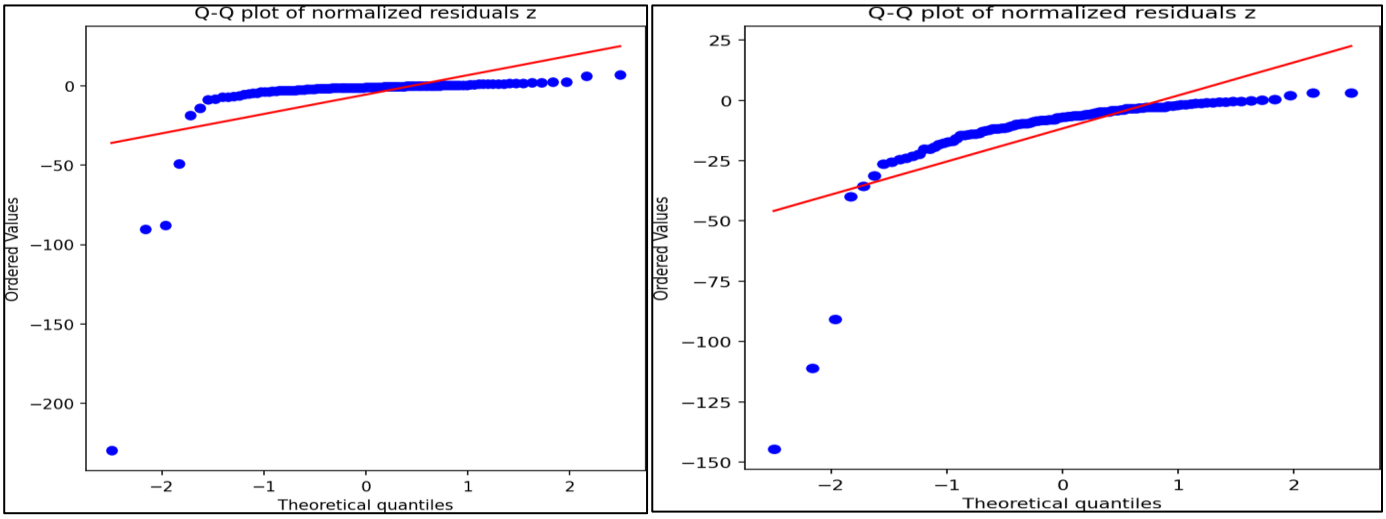

Fig 5 (a) Flower: Q–Q Plot of Normalized Residuals vs the Standard Normal. (b) Casagrande: Q–Q Plot. Deviations from the Diagonal Indicate Non-Normality (e.g., Skew or Heavy Tails).

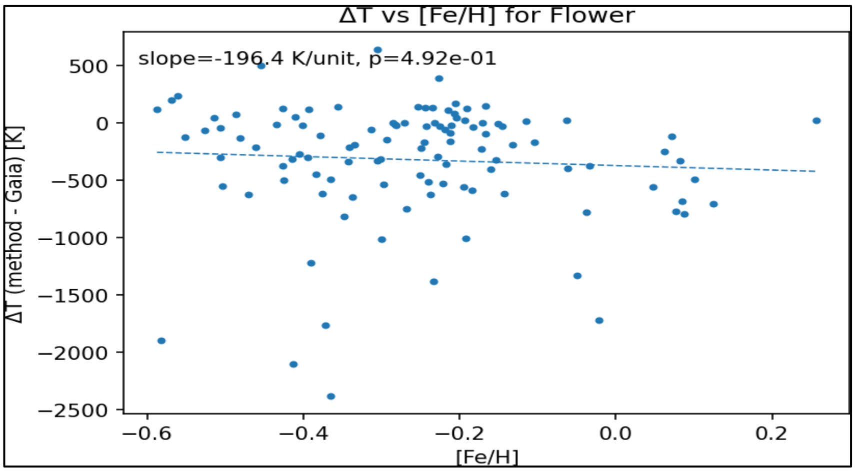

Fig 6 (a) Flower: ΔT vs [Fe/H].Linear Fits (Dashed) and Slope Value is Reported in the Figure Panels.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep1452

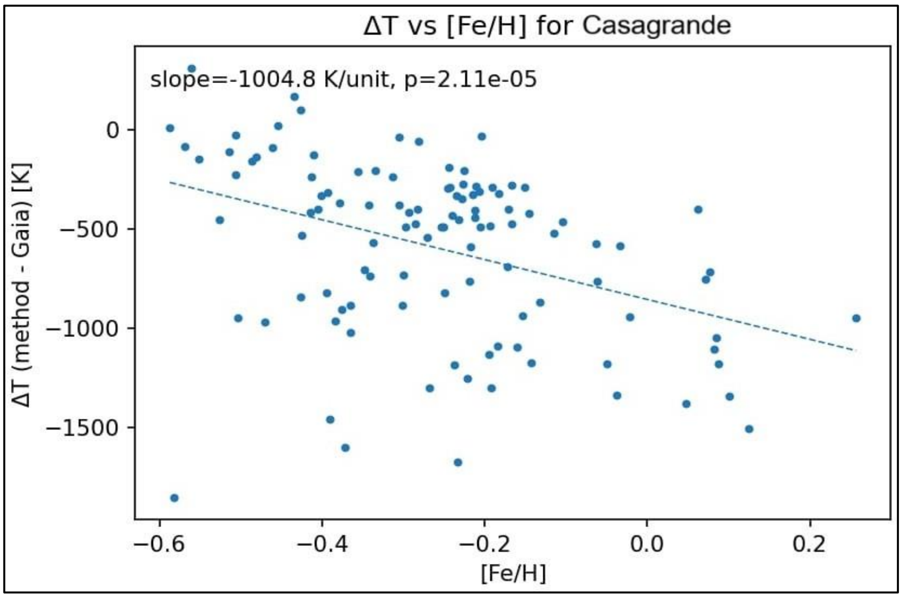

Fig 7 (a) Casagrande: ΔT vs [Fe/H]. Note the Strong Parameter Dependence Visible in the Panel.

IV. SUMMARY STATISTICS AND COMPARISON WITH DI- AGNOSTIC FIGURES

Table 1 Sample Summary (Median and 16th–84th Range Or Min–Max).

Quantity	Median	Range (16th–84th or min–max)
N (stars)	109	–
$T_{\rm eff}^{\rm Gaia}$ (K)	6479	6170–7131
$B-V$ (mag)	0.487	0.395–0.598 Fe/H¿Fe/H
(dex)	-0.25	-0.43– -0.12
$\log g$ (dex)	3.77	3.77–4.11
median e_{B-V} (mag)	0.0370	0.0150–0.0720

Table 2 Ensemble Summary Statistics Computed for the Cleaned Sample (N = 109). Values are Calculated from Per-Star Posterior Medians and Propagated Uncertainties.

Statistic	Flower	Casagrande	Units / remark
N	109	109	number of stars
mean(ΔT)	−318.1231	−595.3207	K (method − Gaia)
stderr	49.0770	42.2161	K (standard error of mean)
std	512.3792	440.7494	K (sample standard deviation)
RMS	510.0234	438.7229	K
Pearson r (method vs Gaia)	0.4522	0.3182	correlation
Pearson p	7.9221×10^{-7}	7.4610×10^{-4}	two-sided
frac($\|z\| \leq 1$)	0.3577	0.0733	fraction within 1 combined σ
frac($\|z\| \leq 2$)	0.6422	0.1284	fraction within 2 combined σ
χ^2 KS stat	669.6116 0.3749	521.8590 0.8227	reduced chi-square (very large) KS statistic vs normal
KS p	2.8540×10^{-14}	1.0755×10^{-81}	effectively zero
t (mean=0)	−6.4821	−14.1017	one-sample t-statistic
t-p	2.7804×10^{-9}	3.0654×10^{-26}	highly significant

The table above reproduces our computed ensemble diagnostics exactly. In the text below we compare these numbers to the diagnostic figures we produced and give precise interpretations.

V. HIGH-LEVEL INTERPRETATION

The ensemble diagnostics show three robust facts: 1. Systematic cool bias for both photometric methods relative to Gaia: Flower's method[5] has mean $\Delta T \approx -318.1$ K and Casagrande's method[6]has mean $\Delta T \approx -595.3$ K. Both one-sample t-tests reject the null hypothesis of zero mean (Flower $p \sim 2.8\times10^{-9}$; Casagrande $p \sim 3.1 \times 10^{-26}$). 2. Large object-to-object scatter: sample stan- dard deviations and RMS are very large (std $\approx$ 440–510 K). This implies that for individual stars photometric estimates often differ from Gaia by several hundred K (plots: both scatter panels(Figure:1) and T histograms(Figure:2)). 3. Uncertainty realism fails: very low fractions of normalized residuals within 1σ (Flower $\approx$ 36%, Casagrande $\approx$ 7%) and very large reduced χ^2_ν plus KS tests with $p \ll 10^{-10}$ mean the combined quoted uncertainties (method + Gaia) do not explain the observed residual distribution—the residuals are non- Gaussian with heavy negative tails (confirmed by the Q-Q plots(Figure:4)and T histograms(Figure:2)).

VI. DETAILED FIGURE-BY-FIGURE ANALYSIS

➢ *(Figure:1)Scatter: Method vs Gaia*

What the figures show: • Both panels show a positive correlation with Gaia (Pearson $r_{F\ lower} = 0.452$, $r_{Casagrande} = 0.318$), so the methods partially track Gaia trends.

In both panels the cloud of points is displaced below the 1:1 line (method¡ Gaia), consistent with the negative mean offsets above; Torres is displaced farther.

Individual stars often lie hundreds of K off the 1:1 line even when errorbars are small: this visually demonstrates that systematic offsets are not explained by single-star quoted uncertainties alone.

Overview of Results (scatter): "Both Flower and Casagrande correlate with Gaia (Pearson $r \approx 0.45$ and 0.32, respectively), but the point clouds are offset below the 1:1 line, corresponding to mean offsets $\Delta T_{F\ lower} = -318.1$ K and $\Delta T_{Casagrande} = -595.3$ K. The scatter (RMS ~440–510 K) indicates large object-to-object discrepancies beyond quoted per-star uncertainties."

➢ *T Histograms(Figure:2)*

What the figures show: • Flower: histogram peaked near −300 K with a long negative tail; distribution skewed left.

Casagrande: histogram peaked near −600 K with a stronger left skew and heavier tail—many stars with $\Delta T < -500$ K.

Overview of Results (histograms): "Flower's residuals centered near −318.1 K with substantial left-skew and a heavy tail, while Casagrande residuals are centered near −595.3 K and show an even more pronounced negative skew. These histograms corroborate the ensemble means and emphasize the asym- metric, non-Gaussian nature of the residuals."

➢ *Uncertainty Comparison(Figure:3)*

What the figures show: • For most stars σ_{method} is comparable to Gaia's σ, but both methods show a small population with very large σ_{method} (hundreds to thousands of K).

The tail of large method uncertainties corresponds to stars with poorly constrained posteriors—these will appear as vertical errorbars much larger than typical Gaia errors.

Overview of Results (uncertainty comparison): "Comparison of per-star 1σ credible intervals: while the bulk of stars have comparable photometric and Gaia uncertainties, a non-negligible tail of stars exhibits very large photometric uncertainties (hundreds to > 1000 K). These objects should be handled sepa- rately."

➢ *Normalized Residuals(Figure:4) and Q-Q(Figure:5)*

What the figures show: • Histograms of z are not centered at 0 nor shaped like a standard normal. Flower: only $\approx$ 35.8% of z satisfy $|z| \leq 1$. Casagrande: only $\approx$ 7.3%.

Q–Q plots show severe deviations in the left tail (many extreme negative residuals). KS statistics are large and p-values $\ll 10^{-10}$ (see Table 2), confirming non-normality.

Overview of Results (z diagnostics): "Normalized-residual diagnostics show that the combined uncertainties do not describe the residuals: only ~ 36% (Flower) and ~ 7% (Casagrande) of stars lie within $|z| \leq 1$ (expected ~ 68%). Q-Q plots and KS tests (Table 2) demonstrate heavy non-Gaussian tails dominated by large negative deviations, implying underestimated uncertainties and/or systematic offsets."

➢ *Trends vs [Fe/H]*

What the figures show and how to relate to the table: • Flower(Figure:6): visu- ally and statistically ΔT shows little or no dependence on [Fe/H](no significant slope; earlier diagnostics reported slope −196.4 K/dex with $p \approx 0.49$ for [Fe/H]). This is consistent with Flower being a single-color calibration whose main problem here is an overall zero-point offset and scatter rather than a composition trend.

Casagrande(Figure:7): strong, statistically significant trends with both [Fe/H] were observed in the trend panels you supplied (reported slopes on the order of −1004.8 K/dex for [Fe/H]). These trends explain why Casagrande has such a poor fraction within 1σ—its systematic errors vary across parameter space.

Overview of Results (parameter trends): "Parameter-space trends (Fig. W) reveal no significant [Fe/H] dependence for Flower, whereas Casagrande exhibits strong and significant metallicity dependencies (estimated slopes ∼ −1000 K/dex for [Fe/H]). This multi-parameter bias renders Torres unreliable across the full sample without recalibration or domain restriction."

VII. SYNTHESIS, INTERPRETATION

Synthesis: • Flower: smaller systematic bias (mean ≈ −318.1 K) and no strong dependence on [Fe/H] in your sample—thus relatively more reliable as a simple photometric estimator here. Nevertheless, Flower shows large scatter (std ∼ 512 K) and underestimates the ensemble scatter compared to observed residuals.

Casagrande: larger systematic bias (mean ≈ −595.3 K) and strong parameter- dependent systematics (metallicity). Casagrande is therefore not trustworthy across the sample without re-fitting or restricting to the calibration domain.

In both cases reddening (un-dereddened $B - V$), calibration-domain mis- match, and input parameter systematics are the most plausible causes of the observed offsets and non-Gaussian residuals.

VIII. CONCLUSIONS

We have compared effective temperatures derived from two photometric calibra- tions (Flower 1996; a Casagrande's-style multi-parameter calibration) with Gaia DR3 effective temperatures for a cleaned sample of 109 Hipparcos–Gaia cross- matched stars. Our main quantitative conclusions are: • Systematic offsets. The Flower method shows a mean offset $\langle \Delta T \rangle = T_{F\,lower} - T_{Gaia} = -318.1$ K (stderr = 49.1 K), i.e., Flower underestimates Gaia temperatures on average by ∼ 320K. The Casagrande's method shows a larger mean offset of $\langle \Delta T \rangle = -595.3$ K (stderr = 42.2 K), i.e., it underestimates Gaia temperatures by ∼ 590 K on average.

Scatter / precision. Both methods exhibit large scatter: $\sigma_{\Delta T} \approx 512$ K (Flower) and ≈ 441 K (Casagrande), and similar RMS (∼ 440–510 K). Thus
single-star photometric predictions from these calibrations can deviate by several hundred K from Gaia values.

Correlation with Gaia. Flower has a stronger Pearson correlation with Gaia ($r \approx 0.452$, $p \ll 10^{-5}$) than Casagrande($r \approx 0.318$), indicating Flower tracks Gaia temperatures somewhat better in a rank sense.

Uncertainty calibration (normalized residuals). The fraction of stars with $|z| \le 1$ (where $z = \Delta T / \sqrt{\sigma^2_{method} + \sigma^2_{Gaia}}$) is 0.358 for Flower and only ∼ 0.073 for Casagrande—both well below the expected ∼ 0.68 for correctly-calibrated Gaussian 1σ errors. The reduced chi square values are extremely large (Flower $\chi^2 \approx 670$, Casagrande $\chi^2 \approx 522$), indicating that the reported uncertainties are underestimates for this sample, or that significant systematic errors exist.

Distributional tests and significance. KS tests strongly reject normality for residuals (Flower: $D = 0.375$, $p \ll 10^{-10}$; Casagrande : $D = 0.823$, $p \approx 0$). One-sample t-tests show the mean offsets are highly significant (Flower $t = -6.48$, $p \ll 10^{-5}$; Casagrande $t = -14.10$, $p \ll 10^{-20}$).

Dependence on stellar parameters. From the regression and plotted diag- nostics (see Figure:6, Figure:7): the Casagrande's-calibration shows pronounced, highly significant trends of ΔT with both [Fe/H], whereas the Flower calibra- tion shows much weaker or no statistically significant trends with [Fe/H] in this sample.

➢ *The Zero Point Calibration.*

The zero-point calibration of the Gaia GSP-Phot effective temperatures ($T_{\rm eff,GSP}$) was performed using a sample of 109 stars with $(B-V)$ colors. A second-degree polynomial was fit to the relationship between $T_{\rm eff,GSP}$ and $(B-V)$, from which a synthetic $(B-V)_{\rm pred}$ color was calculated. After a σ-clipping process that refined the sample to 100 stars, the median residual between the predicted and observed color was found to be: $\Delta(B - V)_{\rm zp} = (BV_{\rm pred} - BV_{\rm obs})_{\rm median} = +0.0151(71) mag$.

This color zero-point offset was converted to an effective temperature offset using the sensitivity of the Flower [5] $T_{\rm eff}$ scale to changes in $(B - V)$ color, calculated to be $dT_{\rm eff}/d(B-V) = -4139.4 K\, mag^{-1}$ at the median color of $(B-V)_{\rm med} = 0.4870 mag$. The resulting systematic correction to the Gaia GSP-Phot temperature scale is: $\Delta T_{\rm eff,zp} = -62.6(29.2) K$. However, we have used the raw data; neither dereddening nor systematic correction was applied to rectify either of the calibrations.

OVERALL ASSESSMENT

The comparative analysis reveals that both the Flower and Casagrande photo- metric temperature scales exhibit significant limitations when applied to indi- vidual stars in our Hipparcos-matched Gaia DR3 sample.

The combination of three key findings—(1) statistically significant system- atic biases of −318 K and −595 K for Flower and Casagrande respectively, (2) substantial object-to-object scatter exceeding 440 K, and (3) severely underesti- mated uncertainties with only 7–36% of residuals within 1σ versus the expected 68%—indicates that neither method provides reliable temperature estimates for individual stars when compared against Gaia DR3 as a reference.

The photometric scales demonstrate systematic offsets that are astrophysi- cally significant (several hundred Kelvin) and statistically robust ($p \ll 0.001$). More critically, the large scatter and non-Gaussian residual distributions suggest that these methods cannot precisely constrain temperatures on a star-by-star basis. The extreme reduced chi-square values ($\chi^2 \approx$ 670 and 522) and KS test rejections of normality further confirm that the reported uncertainties are substantial underestimates for this sample.

Notably, the Casagrande calibration shows pronounced dependencies on metallicity , while the Flower calibration exhibits weaker trends, suggesting differential systematic effects between the methods. The additional finding of a +0.0151 mag color zero-point in Gaia GSP-Phot temperatures, corresponding to a −62.6 K offset, provides context for these comparisons but does not alter the fundamental limitations identified.

While these photometric relations may retain value for population-level stud- ies where systematic biases can be calibrated and random errors average out, they are not recommended for applications requiring accurate temperatures for individual stars. The results emphasize the superior precision of Gaia DR3's spectrophotometric temperatures, which demonstrate better internal consis- tency and uncertainty calibration for the stars in our sample.

ACKNOWLEDGEMENTS

The author acknowledge the use of ChatGPT (OpenAI) and DeepSeek for as- sistance with language polishing and copy-editing only. All scientific content, data, analyses, interpretations, and conclusions are the sole responsibility of the authors; no AI tool has been credited as an author.

REFERENCES

[1]. G. Torres, J. Andersen, A. Gim´enez, "Accurate masses and radii of normal stars: modern results," *Astronomy and Astrophysics Review*, vol. 18 (2010), pp. 67–126.

[2]. Gaia Collaboration, A. Vallenari, A. G. A. Brown, et al., "Gaia Data Re- lease 3: Summary of the content and survey properties," *Astronomy & Astrophysics*, vol. 674 (2023), A1.

[3]. M. A. C. Perryman, L. Lindegren, J. Kovalevsky, et al., "The Hipparcos Catalogue," *Astronomy & Astrophysics*, vol. 323 (1997), L49–L52.

[4]. W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1 (1970), pp. 97–109.

[5]. P. J. Flower, "Transformations from theoretical Hertzsprung–Russell dia- grams to color–magnitude diagrams: effective temperatures, B–V colors, and bolometric corrections," *Astrophysical Journal*, vol. 469, pp. 355–365, 1996. doi:10.1086/177785

[6]. L. Casagrande, I. Ram´ırez, J. Mel´endez, M. Bessell, and M. Asplund, "An absolutely calibrated *T eff* scale from the infrared flux method. Dwarfs and subgiants," *Astronomy & Astrophysics*, vol. 512, p. A54, 2010. doi:10.1051/0004-6361/200913204

[7]. G. Torres, "On the use of empirical bolometric corrections for stars," *The Astronomical Journal*, vol. 140, pp. 1158–1162, 2010. doi:10.1088/0004-6256/140/5/1158