Estimation of stellar parameters using ensemble methods & ML

Andrés Galarza, Simone Daflon, Vinicius Placco, Carlos Allende Prieto

Phd candidate, Observatório Nacional, June 2020

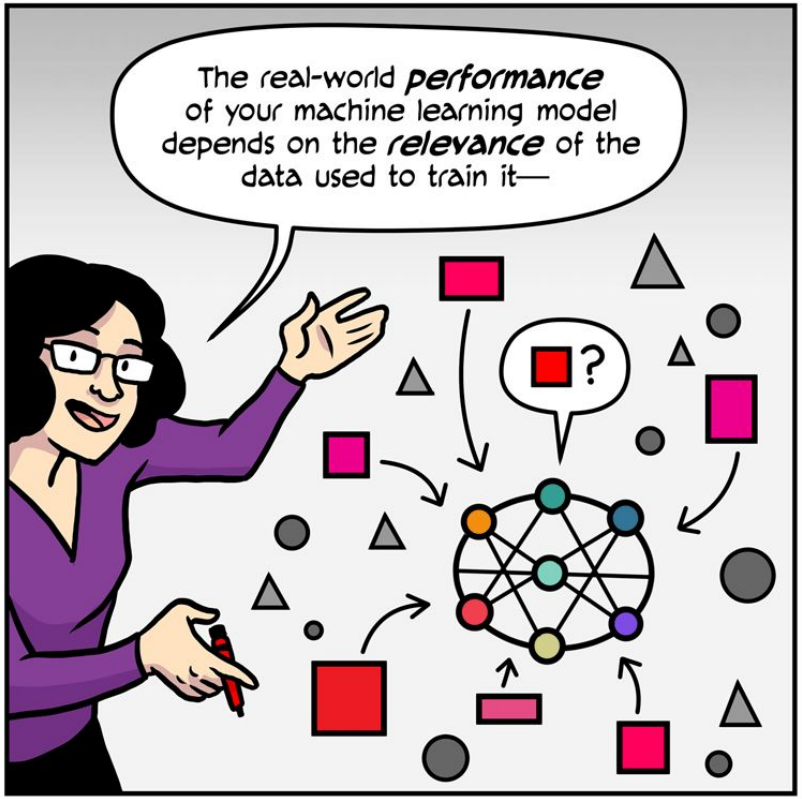
carlosgalarza@on.br

Previously...

- ★ 14th S-PLUS Meeting:
  - Searching for VMPs candidates

  https://www.youtube.com/watch?v=exhI-vxfa-I

- ★ Updated results on the Milky Way WG bi-weekly meetings.



Some weeks ago!

★ Paper in prep. using J-PLUS data.

Astronomy & Astrophysics manuscript no. output
June 2, 2021

J-PLUS: Searching for very metal-poor star candidates using the SPEEM pipeline

Carlos Andrés Galarza[1], Simone Daflon[1], Vinicius M. Placco[2], Carlos Allende-Prieto[3,4], Marcelo Borges Fernandes[1], Claudio B. Pereira[1], Haibo Yuan[5], Young Sun Lee[6], Javier Cenarro[6], David Cristóbal-Hornillos[6], Carlos Hernández-Monteagudo[6], Carlos López-Sanjuan[6], Antonio Marín-Franch[6], Mariano Moles[6], Jesús Varela[6], Héctor Vázquez Ramió[6], Jailson Alcaniz[1], Renato Dupke[1], Alessandro Ederoclite[7], Laerte Sodré Jr.[7], and Raul E. Angulo[8]

[1] Observatório Nacional - MCTI (ON), Rua Gal. José Cristino 77, São Cristóvão, 20921-400, Rio de Janeiro, Brazil
e-mail: carlosgalarza@on.br
[2] Community Science and Data Center/NSF's NOIRLab, 950 N. Cherry Ave., Tucson, AZ 85719, USA
[3] Instituto de Astrofísica de Canarias, Vía Láctea, 38205 La Laguna, Tenerife, Spain
[4] Universidad de La Laguna, Departamento de Astrofísica, 38206 La Laguna, Tenerife, Spain
[5] Department of Astronomy, Beijing Normal University, Beijing 100875, People's Republic of China
[6] Department of Astronomy and Space Science, Chungnam National University, Daejeon 34134, South Korea
[7] Centro de Estudios de Física del Cosmos de Aragón (CEFCA),Unidad Asociada al CSIC, Plaza San Juan 1, E-44001, Teruel, Spain
[8] Universidade de São Paulo, Instituto de Astronomia, Geofísica e Ciências Atmosféricas, R. do Matão 1226, 05508-090, São Paulo,Brazil
[9] Ikerbasque, Basque Foundation for Science, E-48013 Bilbao, Spain

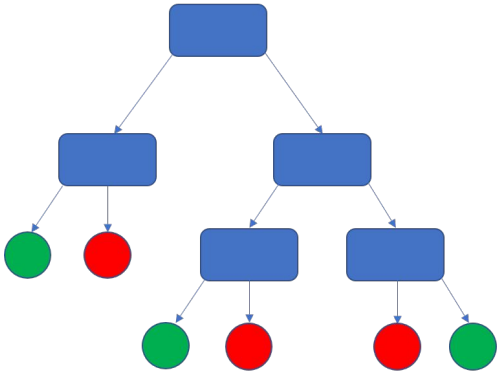
Received XXXX ?, 2021; accepted ? ?, ?

ABSTRACT

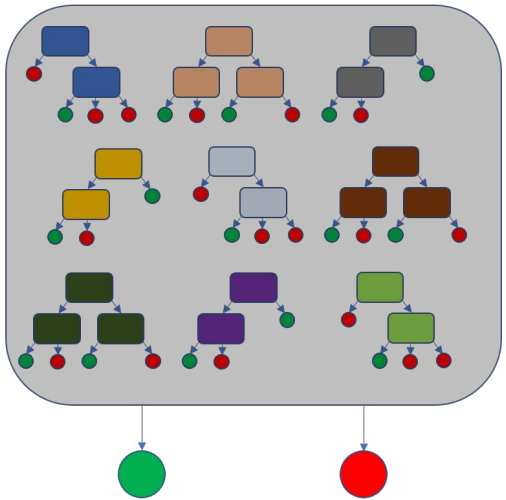
Context. We explore the stellar content of the Javalambre Photometric Local Universe Survey (J-PLUS) Data Release 2 and show its potential to identify low-metallicity stars using the SPEEM (Stellar Parameters Estimation based on Ensemble Methods) pipeline.
Aims. SPEEM is a tool that aims to provide determination of stellar atmospheric parameters ($T_{\rm eff}$, log g, and [Fe/H]) and separate stellar sources from quasars, using the unique Javalambre photometric system, which has narrow and broad band filters. The adoption of adequate selection criteria allows for the identification of metal-poor star candidates that are suitable for spectroscopic follow-up.
Methods. SPEEM consists of a series of machine learning models based on the Random Forest and Extreme Gradient Boosting algorithms, which use a training sample resulting from a cross-match between J-PLUS and the SEGUE spectroscopic survey. The parameter range of the training sample based on SEGUE covers $T_{\rm eff}$ values between 4 800 K and 9 000 K; log g between 1.0 and 4.5, and $-3.1 <$ [Fe/H]$< +0.5$. The performance of the pipeline has been tested with a sample of stars observed by the LAMOST survey within the same parameter range.
Results. The average differences between the parameters of a sample of stars observed with SEGUE and J-PLUS, which were obtained with the pipelines SSPP and SPEEM, respectively, are ∼57 K for $T_{\rm eff}$, 0.16 dex for log g and 0.16 dex for [Fe/H]. A sample of 177 stars have been identified as new candidates with [Fe/H]< -2.5 and 11 of them have been observed with the ISIS spectrograph at the William Herschel Telescope. The spectrocopic analysis confirms the low metallicity of the candidates and presents five new stars with [Fe/H]<-3.0.
Conclusions. SPEEM has shown its potential to estimate the stellar atmospheric parameters ($T_{\rm eff}$, log g, and [Fe/H]) and separate stellar sources from quasars, based on J-PLUS photometric measurements.

Key words. very metal poor stars, machine learning

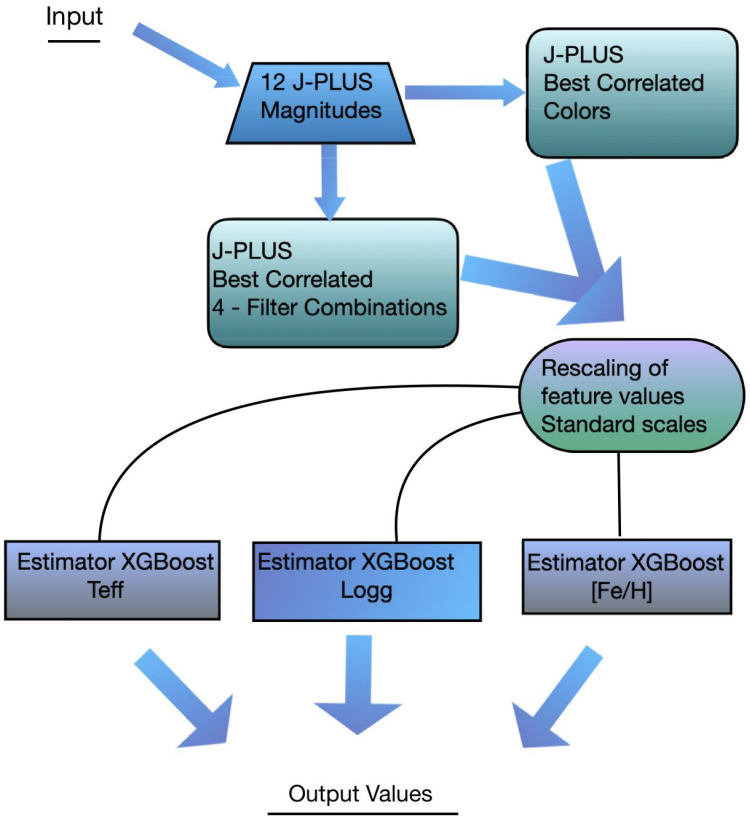
Pipeline Structure of SPEEM



Decision Tree

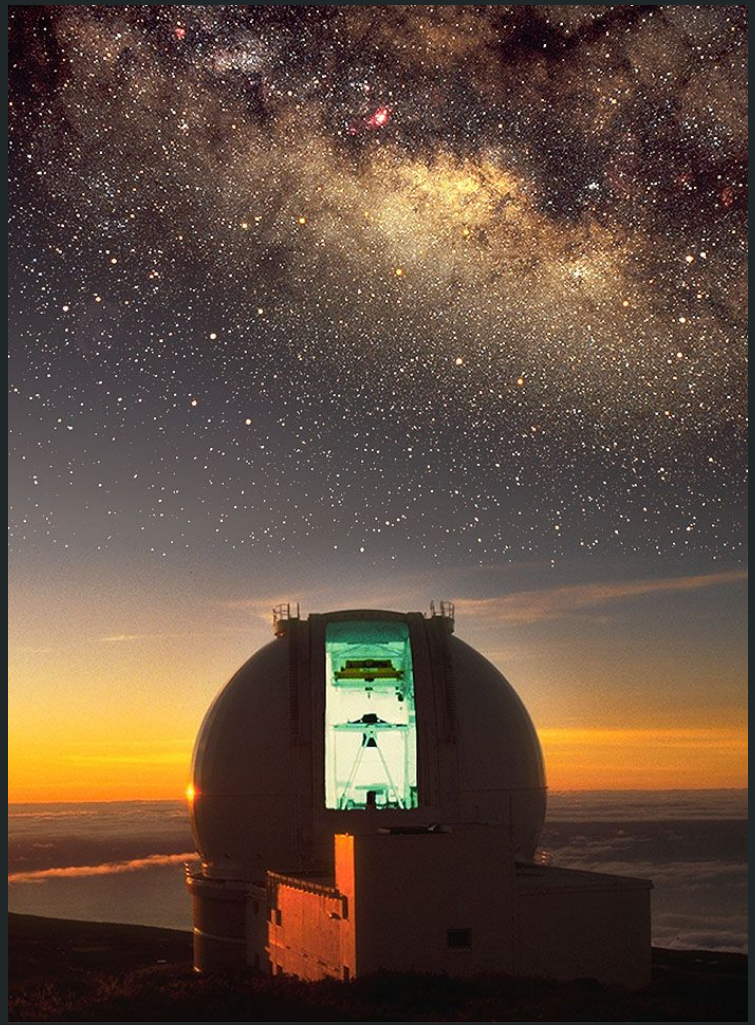


Random Forest



Spectroscopic Validation

- 11 candidates were selected with g mag [14.98, 16.75] and using J-PLUS photometry.
- Observations made by Carlos Allende with WHT and ISIS instrument.
- Standard reduction process with IRAF.



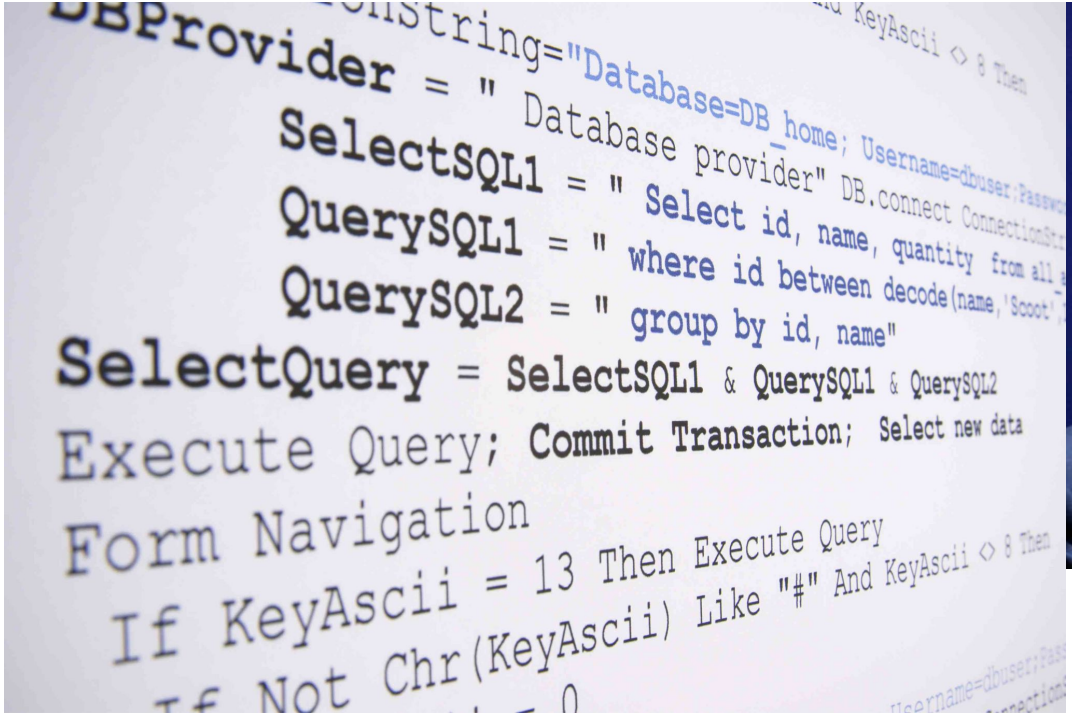
An auspicious result to search for EMPs ([Fe/H]<-3.0)

Table 2: VMPs candidates: Stellar parameters obtained from SPEEM and from the spectroscopic analysis

J-PLUS ID	gSDSS	$T_{\rm eff}$ (K) SPEEM	[Fe/H] SPEEM	$T_{\rm eff}$(K) n-SSPP	log g n-SSPP	[Fe/H] n-SSPP	
72863-2745	16.780	5 033	−3.04	4 803	0.39	−3.13	✓
66430-46822	16.698	5 198	−2.69	5 016	1.07	−2.64	
72875-16592	16.788	5 219	−2.57	4 857	1.09	−3.18	✓
73136-5464	16.916	5 419	−2.52	5 498	2.15	−2.80	
73259-21448	16.514	5 217	−2.63	5 186	2.66	−3.09	✓
73039-16081	16.541	5 077	−2.72	5 194	2.44	−2.23	
75091-15989	14.955	5 326	−2.58	5 153	1.68	−3.25	✓
71582-10685	14.962	5 166	−2.70	5 315	3.01	−2.63	
66416-5807	16.709	5 204	−2.67	5 113	0.47	−2.94	
66723-1757	15.740	5 118	−2.95	4 805	1.10	−3.10	✓

Galarza et al. (in prep.)

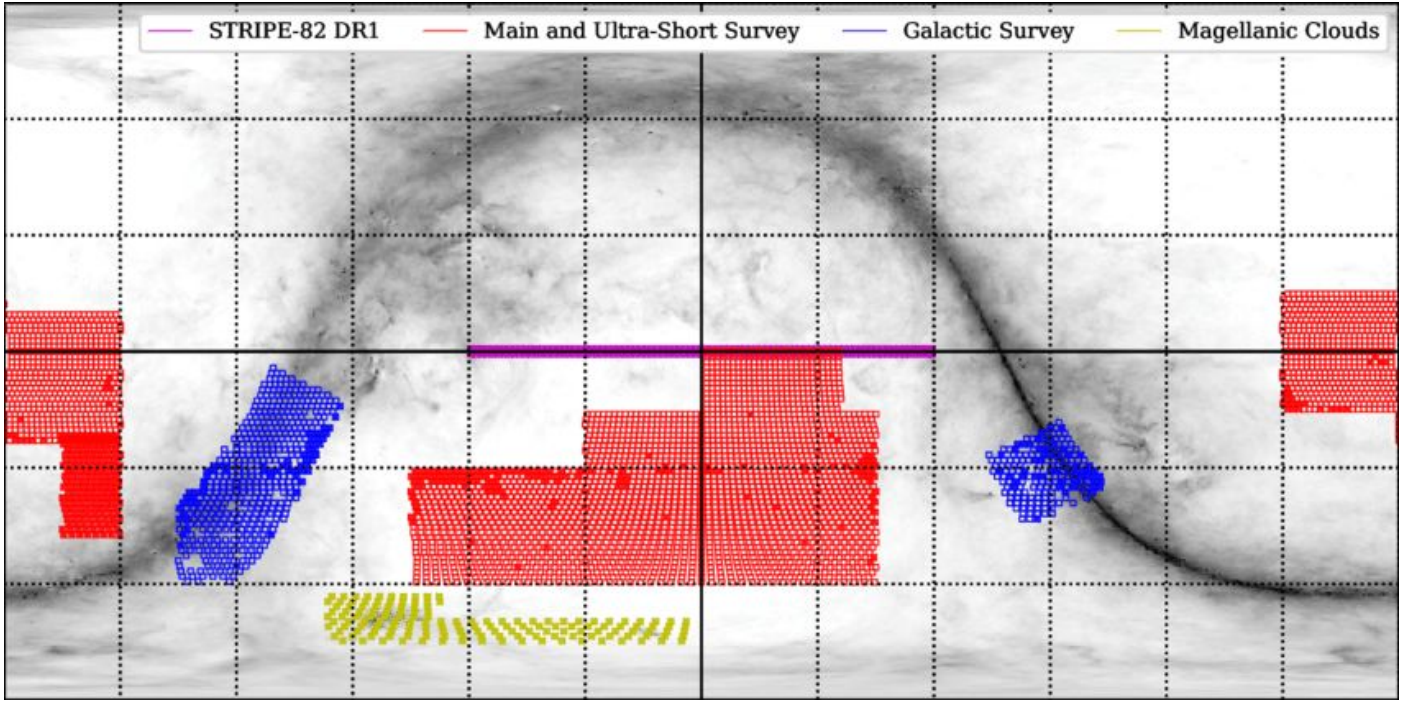
Downloading S-PLUS DR2 Data



Gustavo Schwarz talk and Brainstorming session B on Day 1.

Special thanks to Felipe Almeida and Pierre Augusto

Downloading S-PLUS DR2 Data



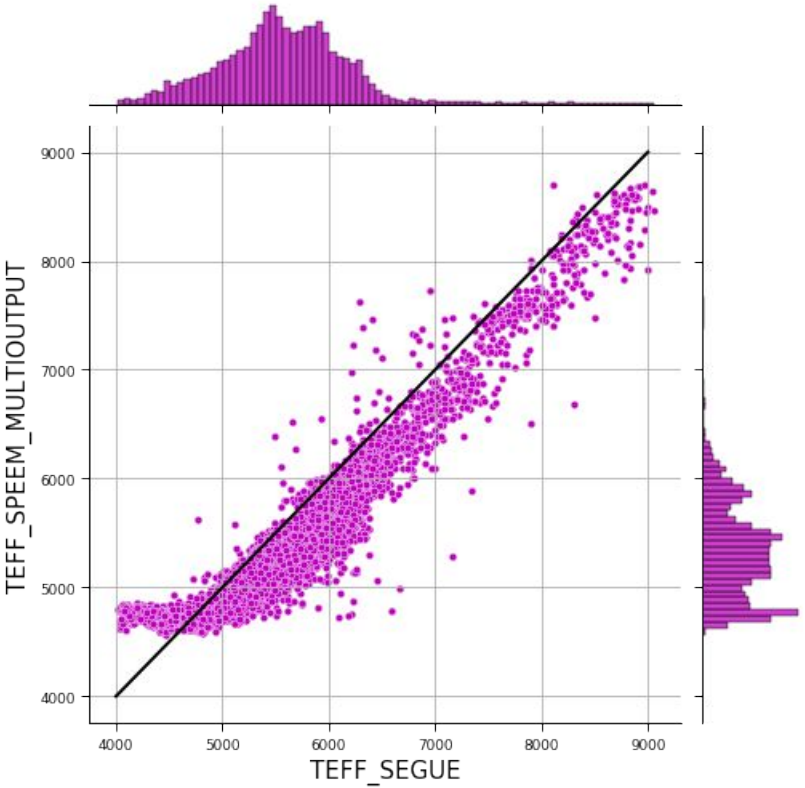
1,190,553 stars in total

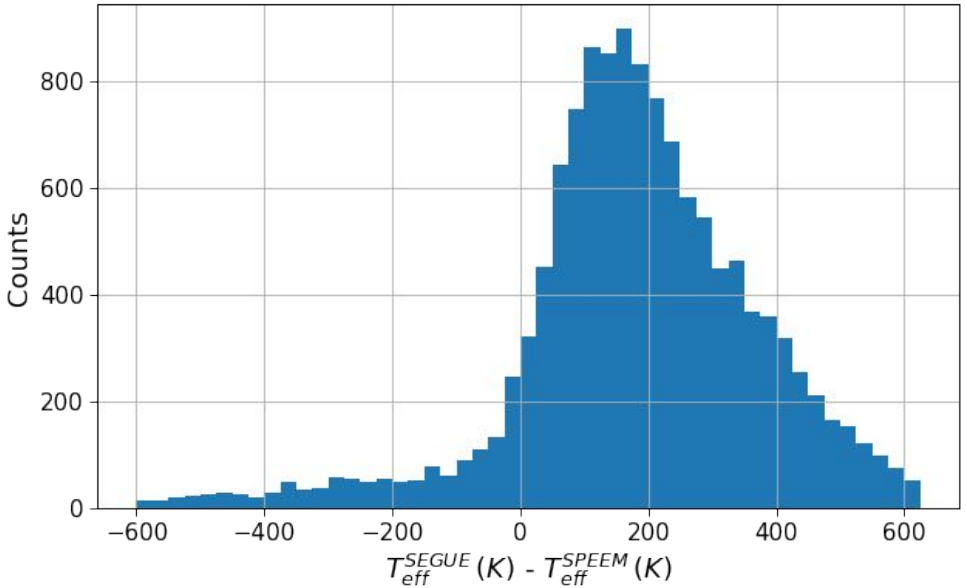
224485 stars with mags < 22 and g < 18 PSTotal Photometry

13050 stars in common with SEGUE and good stellar parameters. No -9999 values.

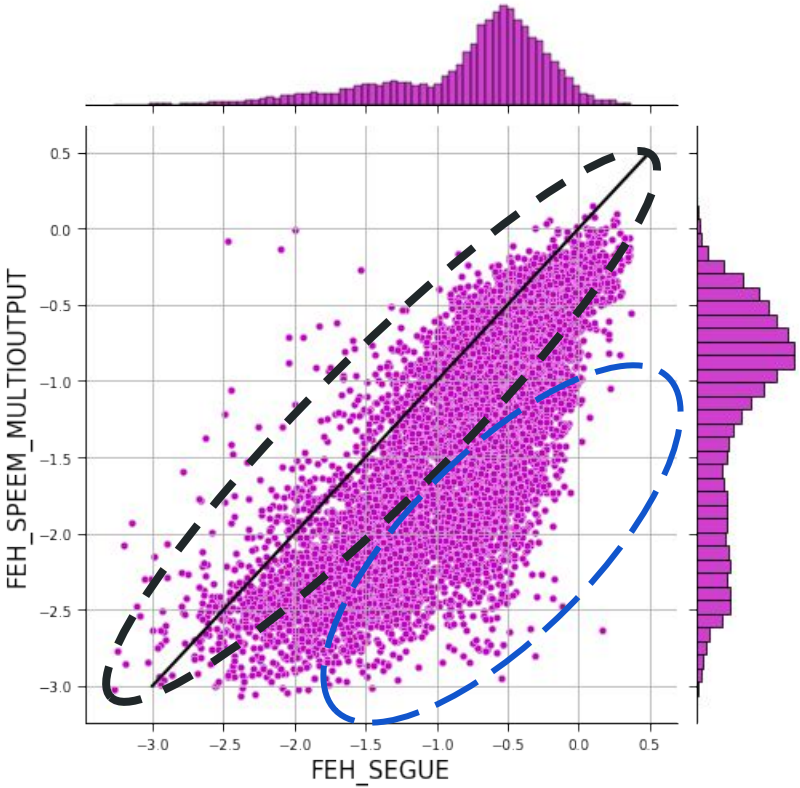
Special thanks to Felipe Almeida and Pierre Augusto

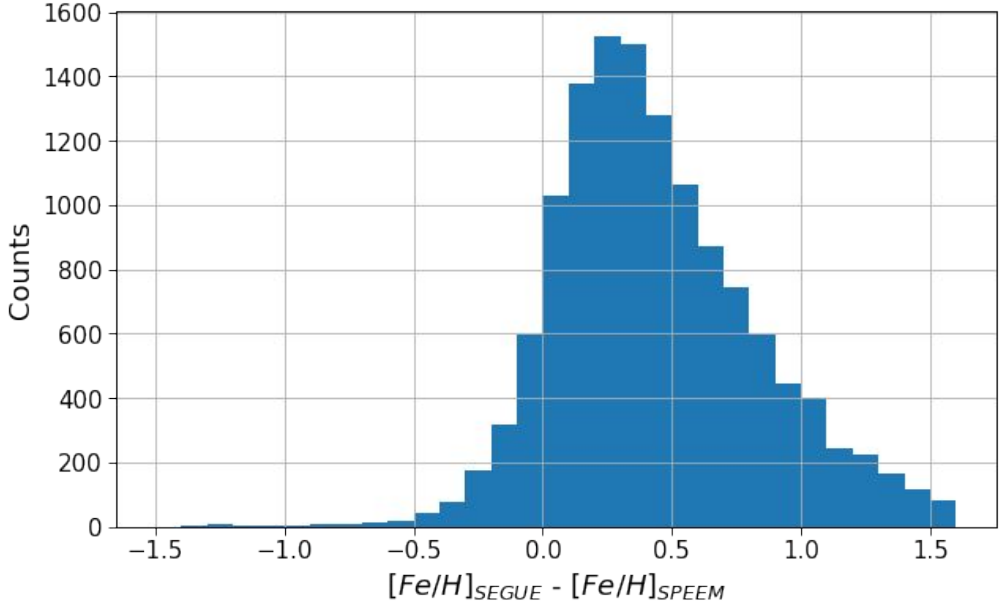
First Attempt - Applying SPEEM on S-PLUS Data



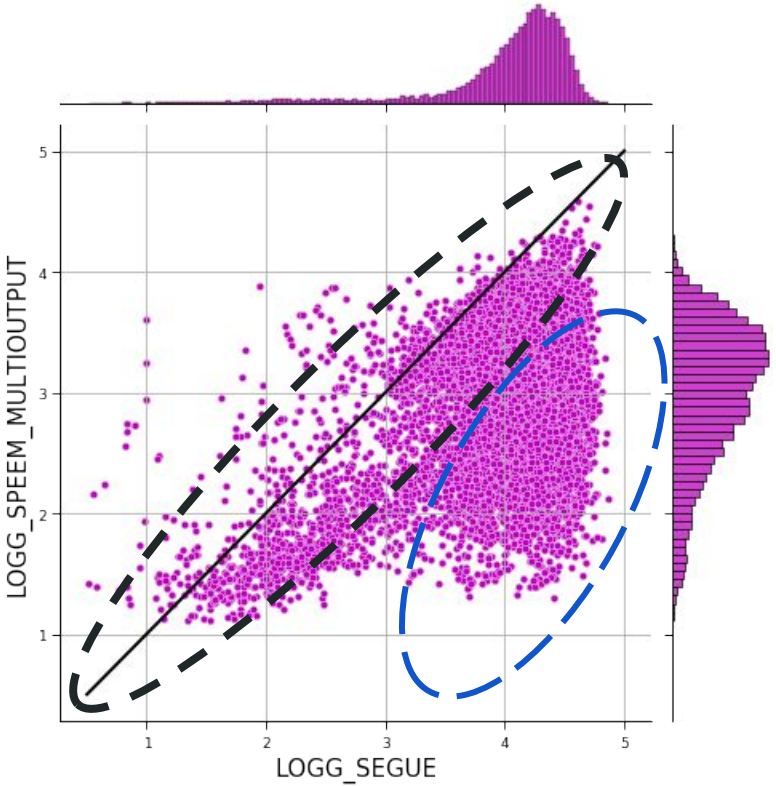


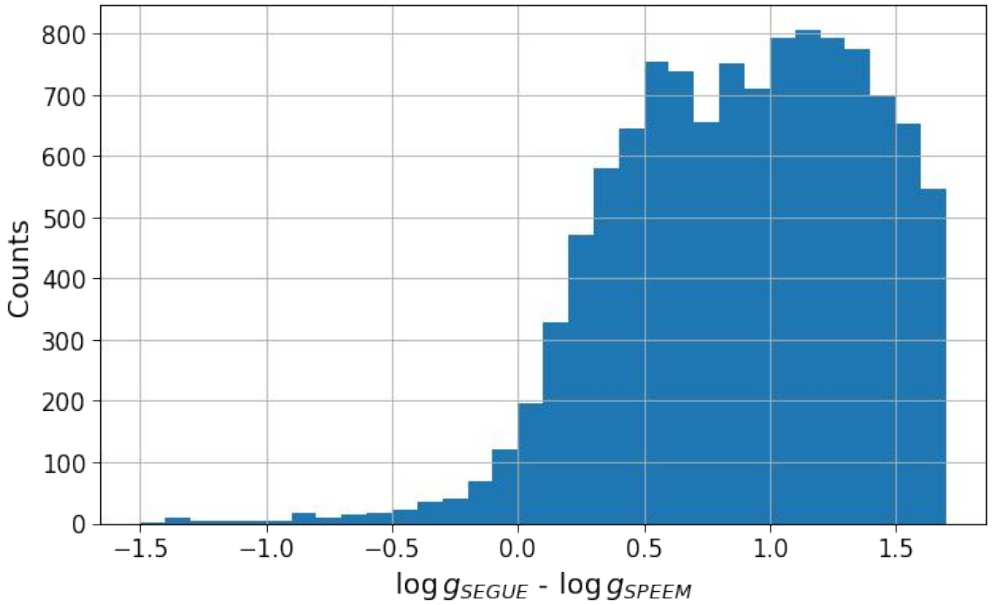
First Attempt - Applying SPEEM on S-PLUS Data



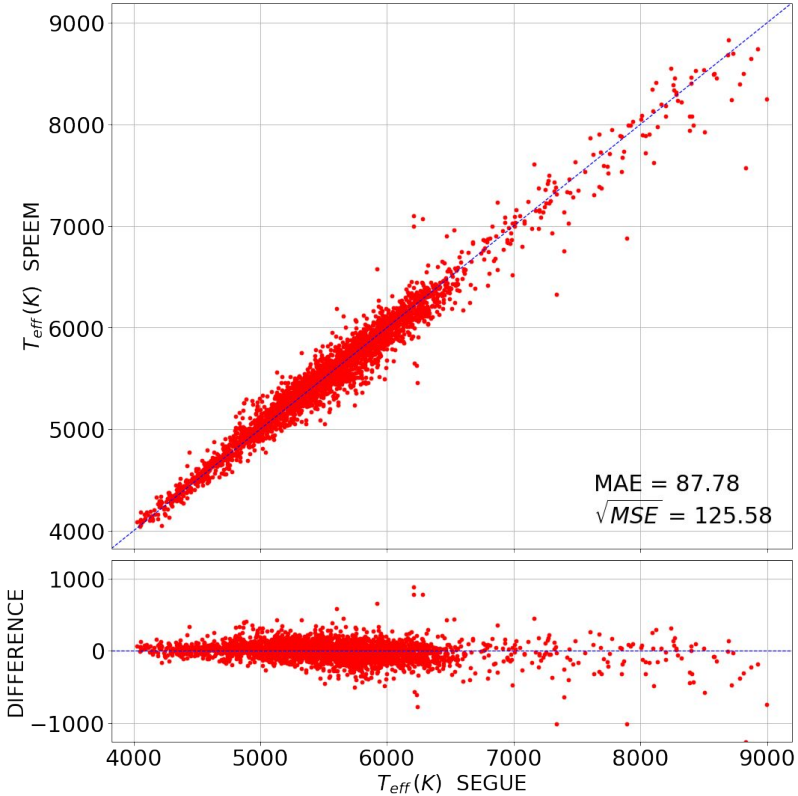


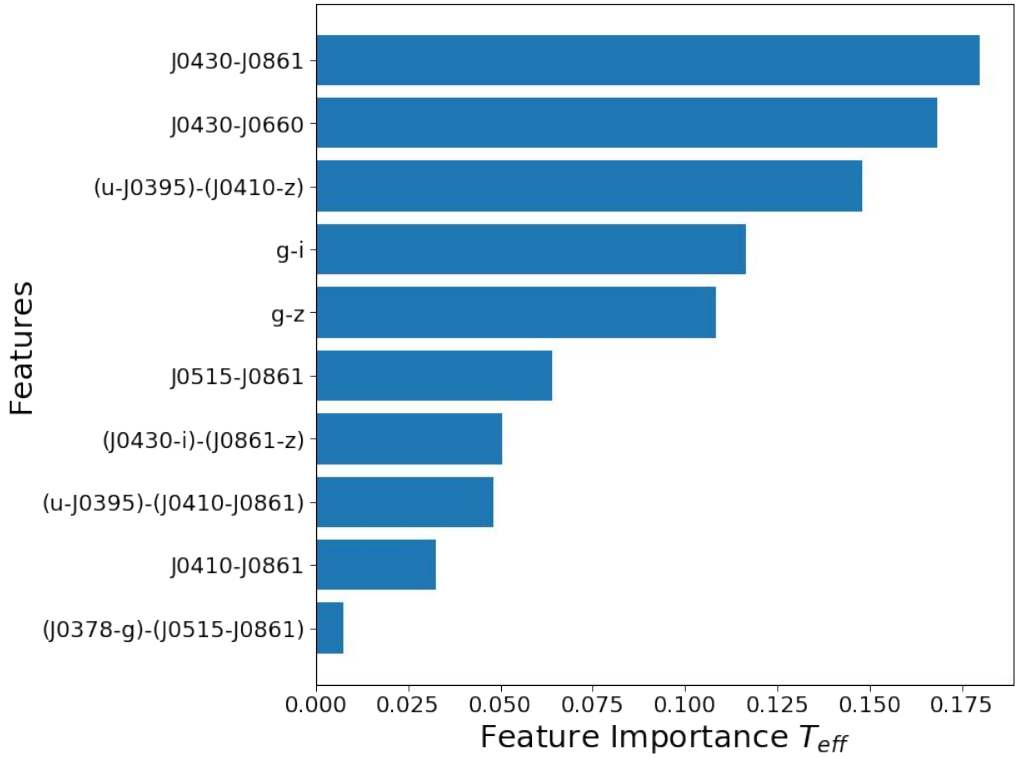
First Attempt - Applying SPEEM on S-PLUS Data



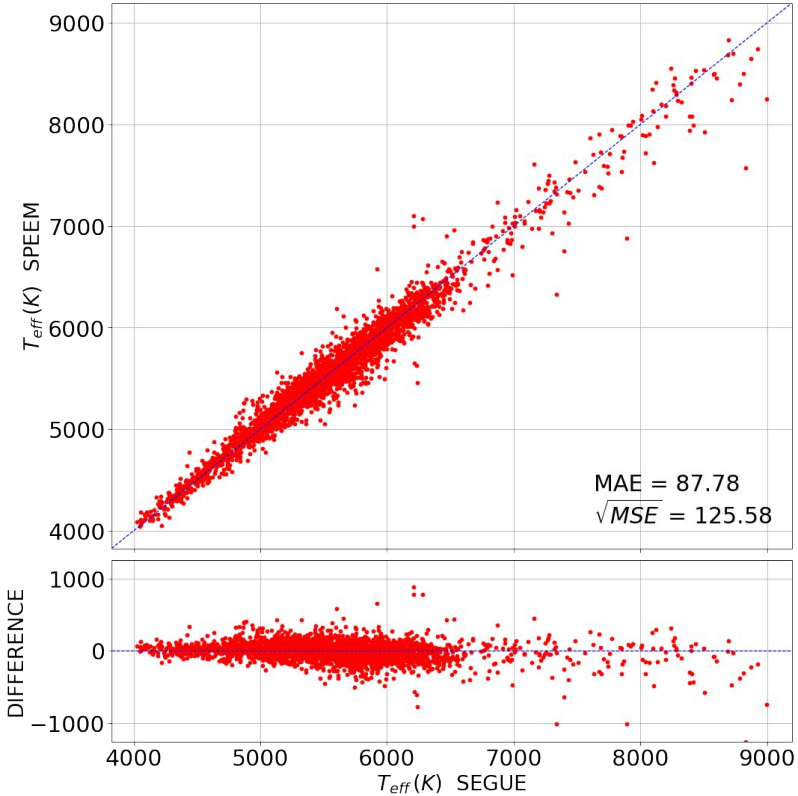


Retraining SPEEM using S-PLUS Data



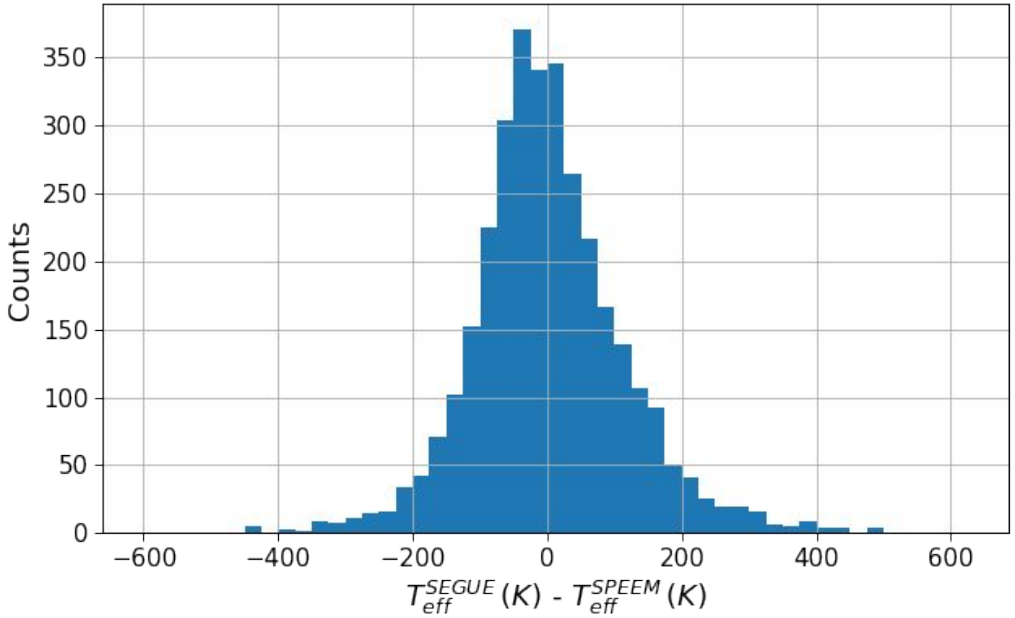


Retraining SPEEM using S-PLUS Data

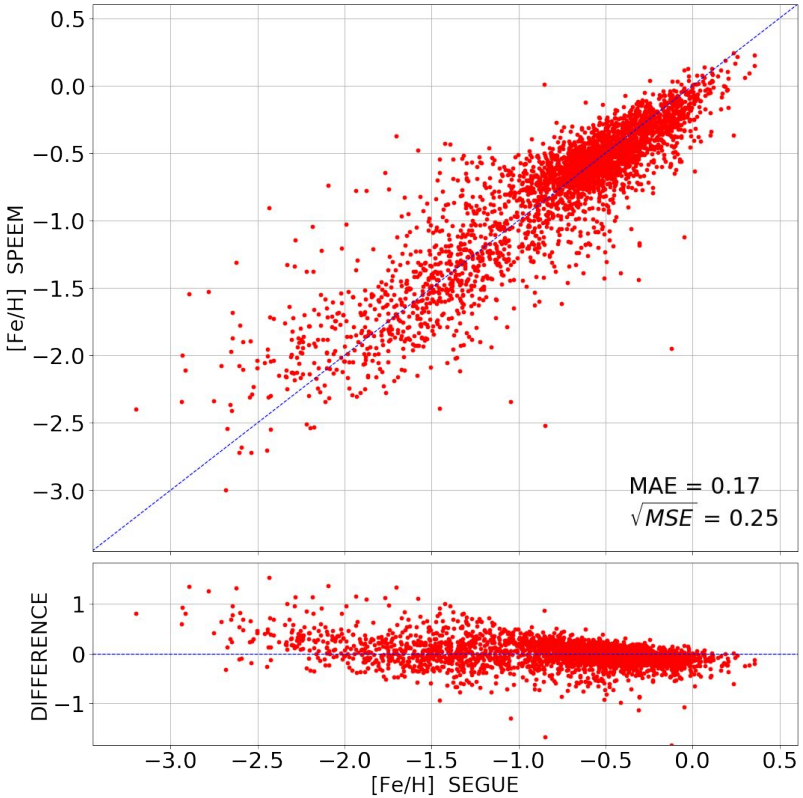


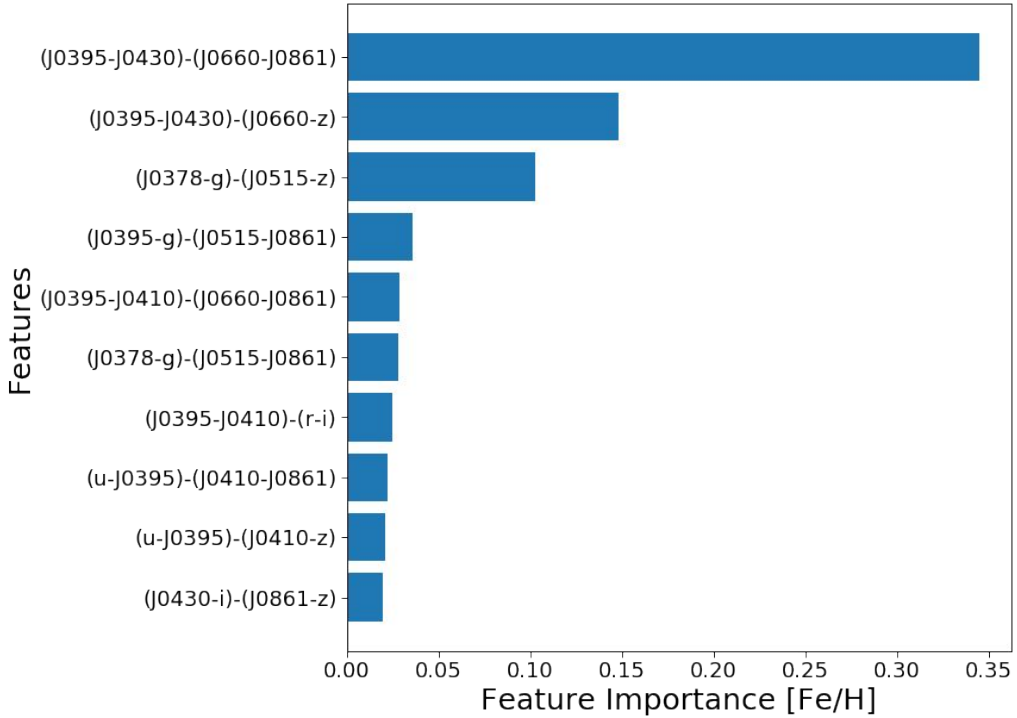
Training Sample = 9787 stars

Test Sample = 3263 stars

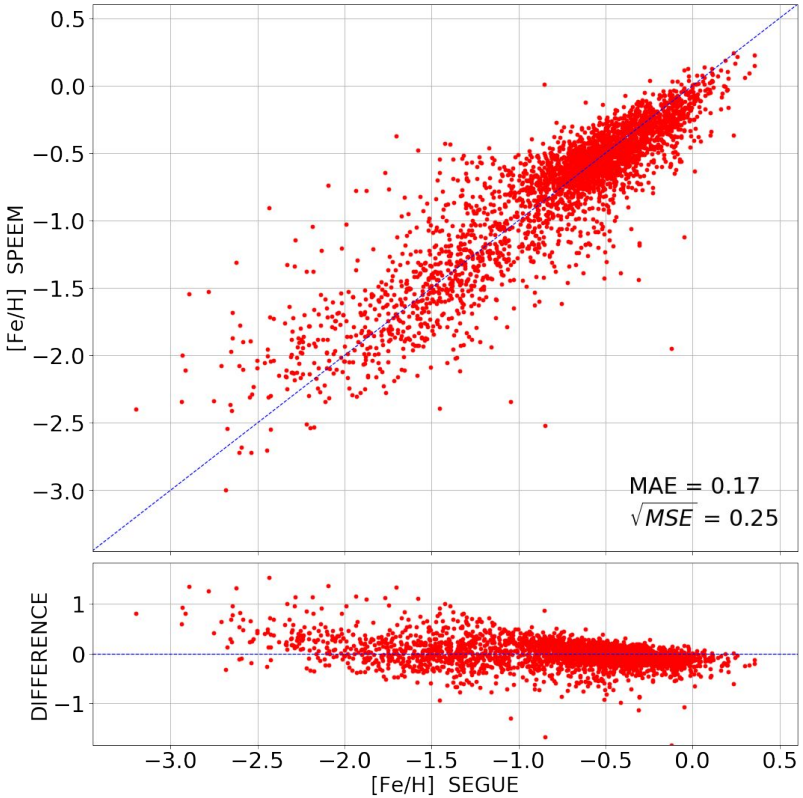


Retraining SPEEM using S-PLUS Data



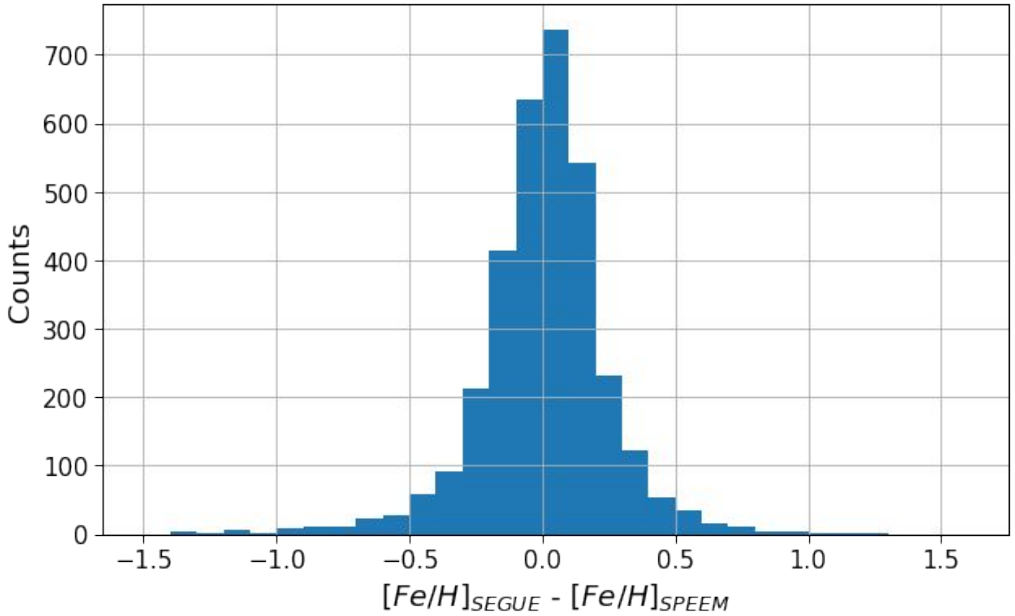


Retraining SPEEM using S-PLUS Data

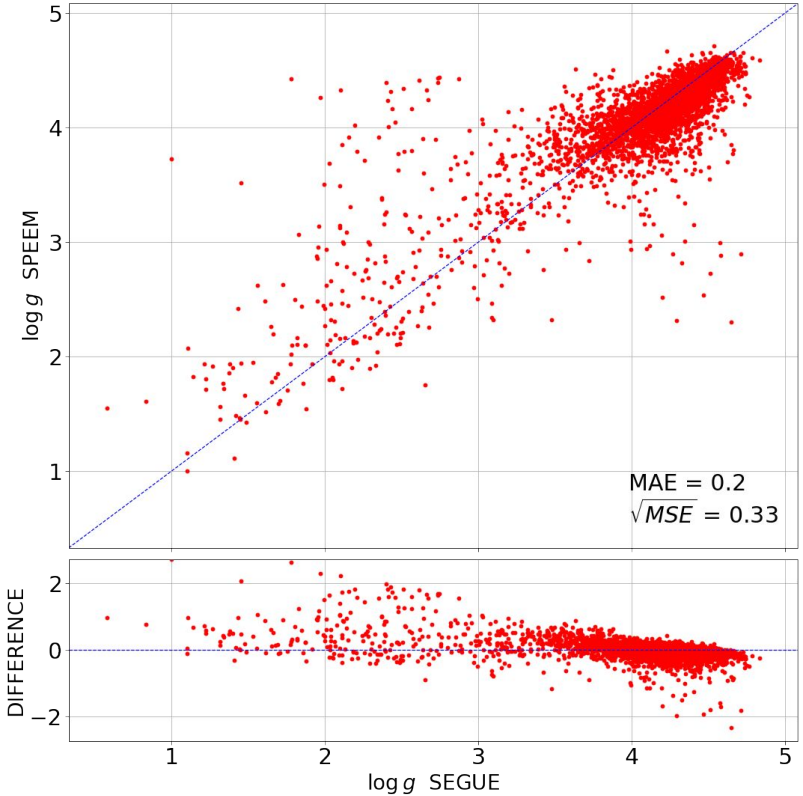


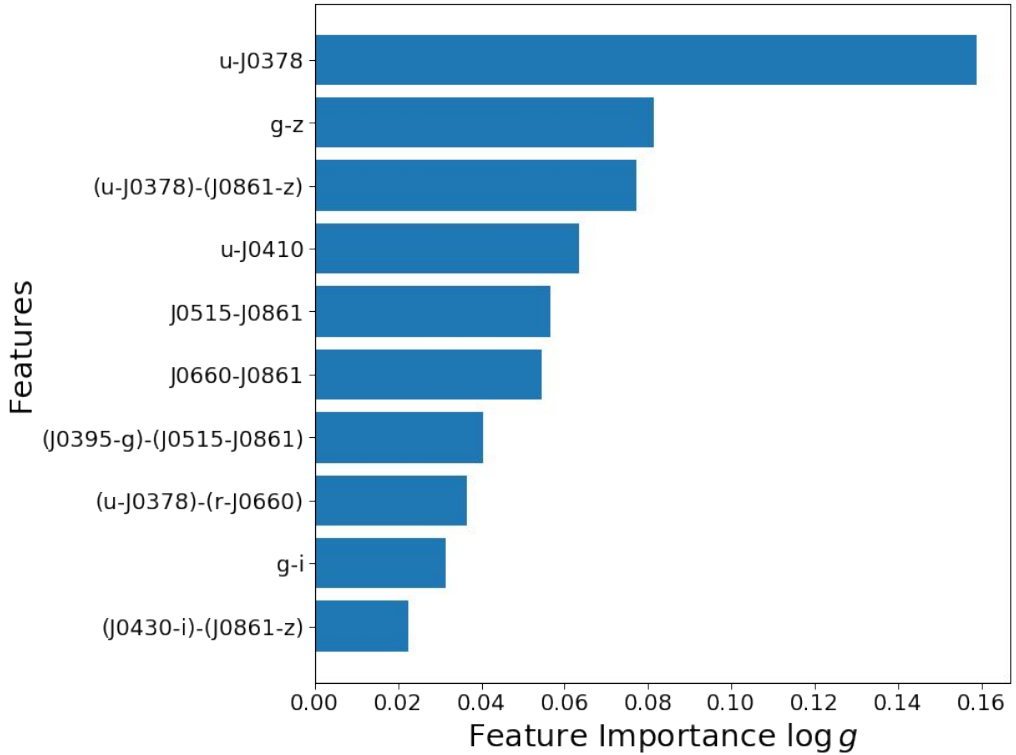
Training Sample = 9787 stars

Test Sample = 3263 stars

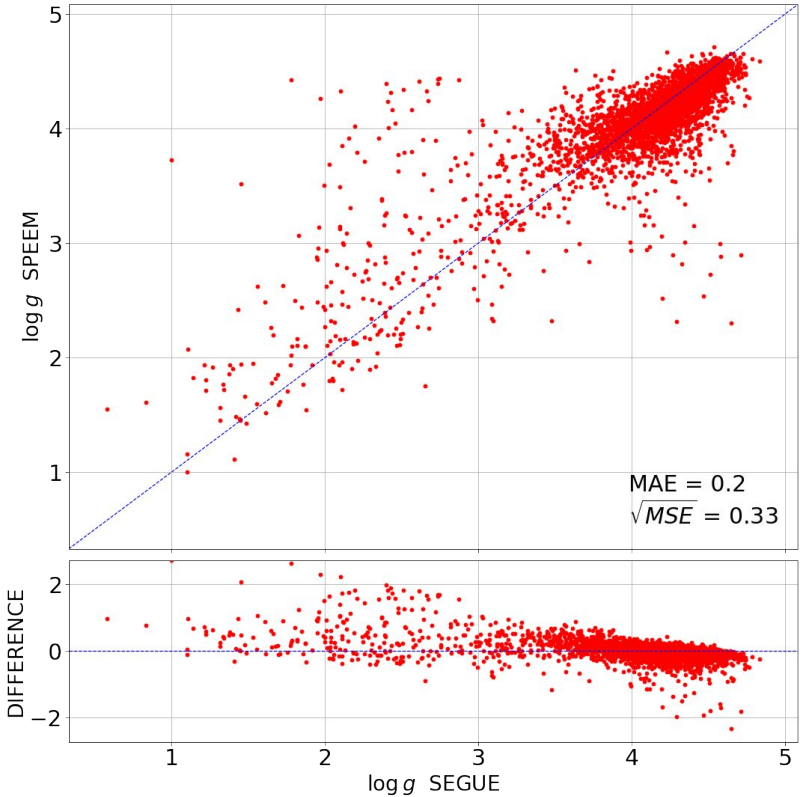


Retraining SPEEM using S-PLUS Data



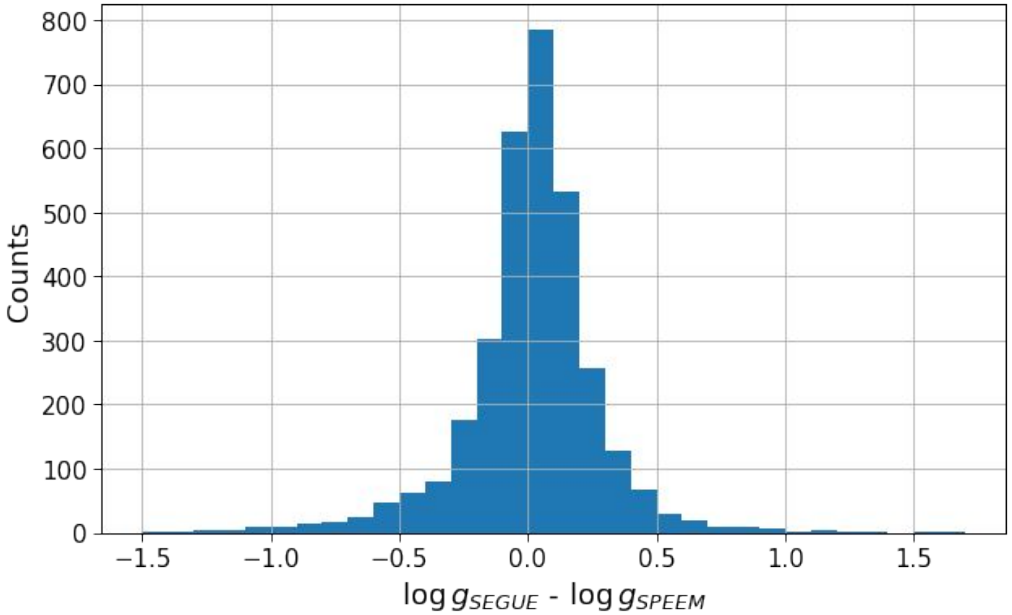


Retraining SPEEM using S-PLUS Data



Training Sample = 9787 stars

Test Sample = 3263 stars



Conclusions

- Differences in photometry quality may severely impact the machine learning estimations.
- SPEEM allows an overall characterization of stars in the range of (4000K - 9000K) for Teff, (-3.5 - 0.5) dex for [Fe/H] and (1.0 - 5.0) dex for Logg.
- Not always vast amounts of data are required to train accurate machine learning models.

Thank you!

Any question, comment or suggestion will be appreciated!

carlosgalarza@on.br

Stay tune for Timothy Beers next talk!