



Observatório
Nacional

TESE DE DOUTORADO

ESTIMAÇÃO DE PARÂMETROS ESTELARES EM LEVANTAMENTOS
FOTOMÉTRICOS MULTIFILTROS COM APRENDIZADO DE MÁQUINA

CARLOS ANDRES GALARZA AREVALO

RIO DE JANEIRO

2022

Ministério da Ciência, Tecnologia, Inovações e Comunicações
Observatório Nacional
Programa de Pós-Graduação

Tese de Doutorado

ESTIMAÇÃO DE PARÂMETROS ESTELARES EM LEVANTAMENTOS
FOTOMÉTRICOS MULTIFILTROS COM APRENDIZADO DE MÁQUINA

por

Carlos Andres Galarza Arevalo

Tese submetida ao Corpo Docente do Programa de Pós-graduação em Astronomia do Observatório Nacional, como parte dos requisitos necessários para a obtenção do Grau de Doutor em Astronomia.

Orientador: Dra. Simone Daflon

Co-orientador: Dr. Vinicius Placco

Rio de Janeiro, RJ – Brasil

Março de 2022

G837

Galarza Arevalo, Carlos Andres

Estimação de parâmetros estelares em levantamentos
fotométricos multifiltros com aprendizado de máquina [Rio de
Janeiro] 2022.

xxii, 111 p. 29,7 cm: graf. il. tab.

Tese (doutorado) - Observatório Nacional - Rio de Janeiro,
2022.

1. palavra chave 1. 2. palavra chave 2. 3. palavra chave 3.
4. palavra chave 4. I. Observatório Nacional. II. Título.

CDU 000.000.000

“ESTIMAÇÃO DE PARÂMETROS ESTELARES EM LEVANTAMENTOS
FOTOMÉTRICOS MULTIFILTROS COM APRENDIZADO DE MÁQUINA”

CARLOS ANDRES GALARZA AREVALO

TESE SUBMETIDA AO CORPO DOCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM ASTRONOMIA DO OBSERVATÓRIO NACIONAL COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM ASTRONOMIA.

Aprovada por:

Dra. Simone Daflon – Observatório Nacional
(Orientador)

Dr. Vinicius Placco – NSF’s NOIRLab
(Co-orientador)

Dr. Alessandro Ederoclite – CEFCA

Dr. Clécio Roque De Bom – CBPF

Dr. Francisco Ferreira Souza Maia – IF/UFRJ

Dr. Claudio Bastos Pereira – Observatório Nacional

RIO DE JANEIRO, RJ – BRASIL

1 DE MARÇO DE 2022

*“Somewhere, something
incredible is waiting to be
known.” – Carl Sagan.*

Agradecimentos

Ninguém consegue nada sozinho e este caso não é a exceção. Primeiramente agradeço a minha mãe Elizabeth, a meu pai Eduardo por nunca deixar de acreditar em mim, sendo firmes quando foi necessário, mas brindando-me seu apoio inestimável tanto moral como economicamente sempre que precisei; sem esquecer obviamente da minha irmã Lina Paola que tem sido minha primeira amiga incondicional e seus conselhos. Certamente não poderia deixar de mencionar a minha orientadora Simone e o meu co-orientador Vinicius, pois sem eles o projeto não teria sido desenvolvido. Além de serem uma guia acadêmica, tornaram-se uma guia fraterna que me impulsionou a continuar nos momentos de crise e estancamento. Ao longo deste processo, também conheci pessoas muito valiosas que me brindaram com dicas importantíssimas para explorar modelos de aprendizado de máquina e sua aplicação a resolver desafios na área de astronomia e astrofísica, entre eles não posso esquecer a Nathanael Tucker quem conheci na minha participação na Open Data Science Conference e quem oficialmente me mostrou o caminho para explorar esse maravilhoso mundo e me permitiu abrir os olhos para explorar novas possibilidades. Agradeço também aos amigos que fiz desde o momento que comecei a viver na cidade maravilhosa, Carol Souza, Adriana Vale, Diego Takahashi, Nacizo Holanda me aceitaram como eu fosse um brasileiro. Esta etapa gerou uma transformação extremamente positiva na minha vida e sempre estarei eternamente agradecido pelas portas que me abriu este maravilhoso país, agradecendo especialmente também o apoio financeiro da CAPES por meio de bolsa para desenvolvimento de meu projeto de doutorado.

Enquanto a detalhes mais técnicos tais como as observações feitas com o telescópio JAST80 no Observatorio Astrofísico de Javalambre (OAJ), em Teruel, de propriedade, administrado e operado pelo Centro de Estudios de Física del Cosmos de Aragón, agradeço à Unidade de Processamento e Arquivamento de Dados OAJ (UPAD) por reduzir os dados OAJ usados neste trabalho. O financiamento para o Projeto J-PLUS foi fornecido pelos Governos da Espanha e Aragão através do Fondo de Inversiones de Teruel; o Governo de Aragão por meio dos Grupos de Pesquisa E96, E103 e *E16_17R*; o Ministério da Ciência, Inovação e Universidades de Espanha (MCIU/AEI/FEDER, UE) com bolsas PGC2018-097585-B-C21 e PGC2018-097585-B-C22; o Ministério da Economia e Competitividade da Espanha (MINECO) sob AYA2015-66211-C2-1-P, AYA2015-66211-C2-2, AYA2012-30789 e ICTS-2009-14; e financiamento FEDER europeu (FCDD10-4E-867, FCDD13-4E-2685).

As agências brasileiras FINEP, FAPESP e o Observatório Nacional do Brasil também contribuíram para este projeto.

Para finalizar não posso esquecer de mencionar a Flor Camila, quem tem sido um apoio fundamental na última etapa do doutorado, tolerando meus medos e inseguranças e brindando um apoio emocional incondicional sempre que precisei.

ESTIMAÇÃO DE PARÂMETROS ESTELARES EM LEVANTAMENTOS
FOTOMÉTRICOS MULTIFILTROS COM APRENDIZADO DE MÁQUINA

RESUMO

Os dados produzidos pelos grandes levantamentos astronômicos (fotométricos e espectroscópicos) representam um grande desafio e uma forma nova de fazer pesquisa a partir, por exemplo, de técnicas sofisticadas e precisas de aprendizado de máquina, permitindo extrair informações das variáveis observadas. Por outro lado, há uma alta demanda por telescópios e instrumentos adequados para obtenção de espectros de alta resolução, que são ideais para a análise detalhada de abundâncias químicas. Sendo assim, torna-se necessário selecionar da melhor forma possível os alvos candidatos para futuras observações espectroscópicas, baseado nos dados de grandes levantamentos. Este trabalho explora o conteúdo estelar do Javalambre Photometric Local Universe Survey (J-PLUS) Data Release 2 e mostra seu potencial para identificar estrelas de baixa metalicidade usando o *pipeline* Stellar Parameters Estimation baseado em Ensemble Methods (SPEEM). O SPEEM é uma ferramenta usada para separar fontes estelares de quasares e fornecer estimativas de parâmetros atmosféricos para estrelas, com base no exclusivo sistema fotométrico J-PLUS. A adoção de critérios de seleção adequados permite a identificação de candidatas a estrelas pobres em metais ideais para programas de acompanhamento espectroscópico. O SPEEM consiste em uma série de modelos de aprendizado de máquina que usam uma amostra de treinamento observada tanto pelo J-PLUS quanto pelo levantamento espectroscópico SEGUE. A amostra de treinamento é composta por estrelas com temperaturas efetivas entre 4.800 K e 9.000 K, valores de gravidade superficial entre 1,0 e 4,5, enquanto a metalicidade varia entre $-3,1$ e $+0,5$. A aplicação do SPEEM a uma amostra de estrelas do J-PLUS/DR2 produziu um catálogo de parâmetros atmosféricos com erros médios de $\Delta T_{\text{eff}} \sim 41$ K, $\Delta \log g \sim 0.11$ dex e $\Delta [\text{Fe}/\text{H}] \sim 0.09$ dex, quando comparados com o SEGUE. Uma subamostra de 177 estrelas potencialmente muito pobres em metais foi selecionada para acompanhamento espectroscópico. Onze estrelas dessa amostra foram observadas com o espectrógrafo ISIS no Telescópio William Herschel. A análise espectroscópica confirma que 64% das estrelas possuem $[\text{Fe}/\text{H}] < -2,5$, incluindo uma nova estrela com $[\text{Fe}/\text{H}] < -3,0$. A aplicação do SPEEM foi estendida para estrelas dos campos da Missão K2, observadas com o mesmo telescópio e câmera do J-PLUS. Neste caso, o SPEEM foi modificado para incluir dados fotométricos no infravermelho obtidos com o satélite WISE, além de resultados do satélite Gaia. A amostra de treinamento foi baseada nos levantamentos espectroscópicos GALAH e APOGEE. O SPEEM modificado foi aplicado com sucesso à amostra K2/T80, produzindo estimativa de parâmetros com

incerteza de de 92 K para T_{eff} , 0,08 para $\log g$, e 0,12 para $[\text{Fe}/\text{H}]$. O SPEEM também mostrou bons resultados na análise do aglomerado M44, permitindo obter estimativas de idade e metalicidade compatíveis com resultados da literatura, dentro das incertezas. O uso do SPEEM em combinação com o sistema de filtro J-PLUS demonstrou seu potencial na estimativa dos parâmetros atmosféricos estelares T_{eff} , $\log g$ e $[\text{Fe}/\text{H}]$.

ESTIMATION OF STELLAR PARAMETERS IN MULTIBANDS PHOTOMETRIC
SURVEYS USING MACHINE LEARNING

ABSTRACT

The data produced by large astronomical surveys (photometric and spectroscopic) represent a great challenge and a new way of doing research, based on sophisticated and precise machine learning techniques that, for example, allow new conclusions to be drawn from the observed variables. On the other hand, there is high demand for telescopes and instruments to obtain high-resolution spectra, in order to carry out detailed chemical abundance analysis. Thus, it becomes necessary to select the best potential candidates for a diversity of science cases, based on large area surveys. This work explores the stellar content of the Javalambre Photometric Local Universe Survey (J-PLUS) Data Release 2 and show its potential for identifying low-metallicity stars using the Stellar Parameters Estimation based on Ensemble Methods (SPEEM) pipeline. SPEEM is a tool tailored to separate stellar sources from quasars and to provide estimations of stellar atmospheric parameters, based on the unique J-PLUS photometric system. The adoption of adequate selection criteria allows for the identification of metal-poor star candidates that are suitable for spectroscopic follow-up investigations. SPEEM consists of a series of machine-learning models that use a training sample in common between the footprints of J-PLUS and the SEGUE spectroscopic survey. The training sample has stars with temperatures between 4.800 K and 9.000 K, surface gravities between 1.0 and 4.5, and metallicity spans from -3.1 to $+0.5$. SPEEM has been applied to a sample of stars of J-PLUS/DR2, producing a catalog of stellar parameters, with typical errors of $\Delta T_{\text{eff}} \sim 41$ K, $\Delta \log g \sim 0.11$ dex, and $\Delta [\text{Fe}/\text{H}] \sim 0.09$ dex, in comparison with SEGUE. A subsample of 177 stars has been identified as potentially very metal-poor stars, with $[\text{Fe}/\text{H}] < -2.5$, for spectroscopic follow up. Eleven stars from this subsample were observed with the ISIS spectrograph at the William Herschel Telescope. The spectroscopic analysis confirms that 64% of stars have $[\text{Fe}/\text{H}] < -2.5$, including one new star with $[\text{Fe}/\text{H}] < -3.0$. The application of the SPEEM pipeline has been extended to a sample of stars from the fields of the K2 Mission, observed in open time with the same telescope and camera used in J-PLUS. In this case, SPEEM was modified to include photometric data in the infrared obtained with the satellite WISE and results from the Gaia mission. The training sample was based on results from the Surveys GALAH and APOGEE. The successful application of the pipeline to the K2/T80 sample allows to estimate the atmospheric parameters with uncertainties of 92 K for T_{eff} , 0.08 for $\log g$, and 0.12 for $[\text{Fe}/\text{H}]$. SPEEM has shown good results from the analysis of the cluster M44, producing reliable estimates of age and

metallicity, consistent with the results from the literature. Using SPEEM in combination with the J-PLUS filter system has demonstrated its potential in estimating the stellar atmospheric parameters T_{eff} , $\log g$, and $[\text{Fe}/\text{H}]$.

Lista de Figuras

2.1	Correlação entre $\delta(U - B)$, e a velocidade W	10
2.2	Correlações entre a velocidade e , h e $\delta(U - B)$	11
2.3	Principais Passos observacionais para escolher candidatas e obter abundâncias químicas de estrelas muito pobres em metais	16
3.1	Vista aérea do Observatorio Astronômico de Javalambre (OAJ)	18
3.2	Equipamentos do OAJ	19
3.3	Sistema de filtros J-PAS	20
3.4	Sistema de filtros J-PLUS	21
3.5	Footprint do J-PLUS DR2	22
3.6	Curvas de sensibilidade dos filtros J0395 e J0430	24
3.7	Footprint do S-PLUS	25
3.8	Sequência de metalicidade de estrelas tipo F	28
3.9	Equipamentos do LAMOST	29
3.10	Distribuição de estrelas observadas no GALAH DR3	31
3.11	Exemplo de espectros observados no APOGEE	33
4.1	Estrutura do pipeline SPEEM	38
4.2	Diagrama cor-cor para os filtros WISE para separar QSOs e estrelas	40
4.3	Separação estrela-QSO baseada nas cores J-PLUS	40
4.4	Matriz de confusão da classificação QSO-estrela	41
4.5	Distribuição de T_{eff} para a amostra de treinamento J-PLUS×SEGUE	43
4.6	Distribuição de $\log g$ para a amostra de treinamento J-PLUS×SEGUE	43
4.7	Distribuição de $[\text{Fe}/\text{H}]$ para a amostra de treinamento J-PLUS×SEGUE	44
4.8	Filtros J-PLUS sobrepostos ao espectro da estrela tipo G J-PLUS 75091-15989	44
4.9	Comparação entre T_{eff} , $\log g$ e $[\text{Fe}/\text{H}]$ obtidos com SSPP e SPEEM	46
4.10	Comparação entre T_{eff} , $\log g$ e $[\text{Fe}/\text{H}]$ obtidos com SSPP e LAMOST	47
5.1	Diagrama HR da amostra de ouro usando os parâmetros estelares obtidos com SPEEM	50
5.2	Distribuição de $[\text{Fe}/\text{H}]$ a diferentes distâncias	51

5.3	Telescópio de 4.2 m William Herschel localizado nas ilhas Canárias	52
5.4	Espectros e foto-espectros dos alvos selecionados parte I	56
5.5	Espectros e foto-espectros dos alvos selecionados parte II	57
6.1	Distribuição espacial dos Campos observados na missão K2 do telescópio espacial Kepler	59
6.2	Localização dos campos F0, F4, F5 e F8 da missão K2 observados com o telescópio JAST/T80	61
6.3	Comparação entre as magnitudes medidas com os filtros estreitos obtidas para o campos K2 com o T80 e pelo levantamento J-PLUS DR2	63
6.4	Comparação entre as magnitudes medidas nos filtros g , r , i e z dos levantamentos SDSS DR12 e JPLUS DR2	64
6.5	Comparação entre os parâmetros T_{eff} , $\log g$ e $[\text{Fe}/\text{H}]$ obtidos por GALAH, LAMOST, K2-EPIC e APOGEE	66
6.6	Comparação entre os parâmetros T_{eff} , $\log g$ e $[\text{Fe}/\text{H}]$ obtidos com GALAH-APOGEE e os pipelines SPEEM	67
6.7	Distribuição do parâmetro T_{eff} para a amostra de treinamento T80×GALAH-APOGEE	68
6.8	Distribuição do parâmetro $\log g$ para a amostra de treinamento T80×GALAH-APOGEE	68
6.9	Distribuição do parâmetro $[\text{Fe}/\text{H}]$ para a amostra de treinamento T80×GALAH-APOGEE	69
6.10	Diagrama HR da amostra K2-T80 usando os parâmetros estelares obtidos com SPEEM-GALAH	70
6.11	Campo do aglomerado estelar M44	71
6.12	Análise visual de membership do aglomerado M44	72
6.13	Diagramas cor-magnitude para M44	73
6.14	Diagrama cor-magnitude das estrelas de M44 observadas com T80. Os pontos estão coloridos de acordo com $[\text{Fe}/\text{H}]$ estimada pelo pipeline SPEEM.	74
6.15	Diagramas HR para M44	75
6.16	Distribuição de metalicidades das estrelas de M44 estimadas por Gaia DR3 (painel superior) e com SPEEM modificado (painel inferior).	76
A.1	VO Services J-PLUS	100
A.2	Query Jobs VO Services J-PLUS	100
A.3	Processamento de Jobs no VO Services J-PLUS	100
A.4	J-PLUS Tap Protocol Topcat I	100
A.5	J-PLUS Tap Protocol Topcat II	101
A.6	Query Jobs com Topcat	101

B.1	Cruzamento de dados externos com Topcat	103
B.2	Cruzamento de dados externos com Topcat	104
B.3	Janela de seleção para fazer cruzamentos de dados Topcat	104

Lista de Tabelas

2.1	Nomenclatura para grupos de estrelas por $[Fe/H]$	14
3.1	Sistema de Filtros J-PLUS.	22
5.1	Dados observacionais das estrelas observadas.	53
5.2	Parâmetros estelares obtidos para as candidatas observadas	54

Sumário

Lista de Figuras	xv
Lista de Tabelas	xix
1 Introdução	1
2 Distribuição de Metalicidade no Halo Galáctico	5
2.1 Uma Breve Revisão Histórica	5
2.2 Estrutura da Via Láctea	8
2.3 Como se formou a Via Láctea?	9
2.3.1 Modelo de Eggen, Lynden-Bell e Sandage (ELS)	9
2.3.2 Modelo do Searle & Zinn	11
2.4 O que aconteceu após a formação do halo de matéria escura (DMH no inglês) e o colapso do gás?	12
2.4.1 Formação Estelar	12
2.4.2 Evolução Química de galáxias	12
2.5 Expandindo o limite de metalicidade para estrelas	13
2.6 Comentários Finais	14
3 Grandes Levantamentos Astronômicos	17
3.1 Levantamentos fotométricos multifiltros	17
3.1.1 OAJ e J-PAS	18
3.1.2 Mapeando o céu do hemisfério norte com J-PLUS	20
3.1.3 Mapeando o céu sul com S-PLUS	23
3.2 SDSS/SEGUE	26
3.3 Lamost/LEGUE	29
3.4 GALAH	30
3.5 SDSS/APOGEE	32
3.6 Comentários Finais	33
4 Metodologia: O Modelo ML SPEEM	35
4.1 Métodos de aprendizado de máquina e aplicações na astronomia	35

4.1.1	Árvores de decisão	36
4.1.2	Random Forest e XGBoost	37
4.2	Arquitetura do pipeline SPEEM	37
4.2.1	Amostras usadas no processo de treinamento	38
4.2.2	Limpeza da amostra de fontes extragalácticas e anãs brancas	39
4.3	Estimativa de parâmetros	42
4.3.1	Treinamento para regressões usando algoritmos de aprendizado de máquina	42
4.3.2	Validação da estimação de parâmetros estelares	45
4.4	Resumo e comentários finais	47
5	Busca por novas estrelas muito pobres em metais	49
5.1	Aplicação do SPEEM ao levantamento J-PLUS	49
5.2	Seleção e validação de novas candidatas	52
5.3	Resumo e comentários finais	55
6	Estendendo o uso do pipeline SPEEM	59
6.1	A Missão K2 do Telescópio Espacial Kepler	59
6.2	Dados observacionais	60
6.2.1	Calibração das magnitudes medidas com filtros de banda estreita	61
6.2.2	Magnitudes medidas com filtros de banda larga	64
6.3	Modificação e novo treinamento do SPEEM	65
6.4	Aplicação: o aglomerado M44	69
6.5	Resumo e comentários finais	72
7	Conclusões	77
	Referências Bibliográficas	79
A	Como baixar dados do J-PLUS	97
A.1	Executando a query no website J-PLUS	99
A.2	Executando a query no Topcat	99
B	Cruzamento com outras fontes	103
B.1	Carregar os dados e cruzar com bancos de dados públicos	103
C	Código Python para treinar o SPEEM	105

Capítulo 1

Introdução

A radiação eletromagnética emitida por uma fonte estelar e coletada por um telescópio permite a determinação de múltiplas grandezas físicas da fonte, como temperatura efetiva (T_{eff}), gravidade superficial ($\log g$) e metalicidade $[\text{Fe}/\text{H}]$. A análise de espectros estelares permite a determinação desses parâmetros, além da obtenção de abundâncias para muitas espécies químicas. No entanto, para estrelas que não possuem espectros disponíveis, a fotometria de banda estreita oferece uma abordagem alternativa para determinar a abundância de alguns elementos químicos, como a abundância de carbono explorada por [WHITTEN *et al.* \(2019\)](#) e [WHITTEN *et al.* \(2021\)](#).

Dentre as principais questões que o estudo da metalicidade de estrelas com diferentes massas e idades pode ajudar a responder, vale destacar as seguintes perguntas: Como se deu a formação da nossa Galáxia? Como a Galáxia evoluiu até o estado atual? A identificação de estrelas muito pobres em metais (VMP, $[\text{Fe}/\text{H}] < -2,0$) e extremamente pobres em metais (EMP, $[\text{Fe}/\text{H}] < -3,0$), de acordo com [BEERS e CHRISTLIEB \(2005\)](#), representa um passo crítico na compreensão das origens da Via Láctea. No entanto, de acordo com o banco de dados SAGA (Stellar Abundances for Galactic Archaeology Database), na sua atualização de abril de 2021, o limite inferior da função de distribuição de metalicidade do halo contém cerca de 550 EMPs, compreendendo algumas dezenas de estrelas com $[\text{Fe}/\text{H}] < -4$ confirmadas espectroscopicamente. A identificação de novos objetos desse tipo é possível por meio de diferentes abordagens baseadas em fotometria e espectroscopia. Ao contrário da espectroscopia, a fotometria oferece a vantagem de prover maiores razões sinal-ruído (S/R) para um determinado tempo de exposição e a possibilidade de coleta simultânea de dados de um grande número de fontes. Por outro lado, os dados fotométricos não são capazes de fornecer valores precisos de abundâncias químicas.

Nas últimas duas décadas, levantamentos fotométricos de banda larga de campo amplo feitos por telescópios terrestres, como o Sloan Digital Sky Survey (SDSS; [YORK *et al.*, 2000](#)) e o Panoramic Survey Telescope and Rapid Response System (Pan-STARRS; [CHAMBERS *et al.*, 2016](#)), provaram ser uma alternativa extremamente bem-sucedida

para estudar grandes amostras de objetos astronômicos.

Por outro lado, bons exemplos de levantamentos espectroscópicos empregados para procurar estrelas muito pobres em metais são: Apache Point Observatory Galactic Evolution Experiment (APOGEE; MAJEWSKI *et al.*, 2016); Sloan Extension for Galactic Understanding and Exploration (SEGUE; YANNY *et al.*, 2009); Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST; CUI *et al.*, 2012); e os futuros WHT Enhanced Area Velocity Explorer (WEAVE; DALTON *et al.*, 2012) e 4-m Multi-Object Spectroscopic Telescope (4MOST; DE JONG *et al.*, 2012).

Mais recentemente, novos projetos estão entrando em cena, como, por exemplo, o Javalambre Photometric Local Universe Survey (J-PLUS; CENARRO *et al.*, 2019) e o Javalambre Physics of the Accelerating Universe Astrophysical Survey (J-PAS; BENITEZ *et al.*, 2014), cobrindo vários milhares de graus quadrados do céu do hemisfério Norte. Os sistemas exclusivos de 12 e 59 filtros para J-PLUS e J-PAS, respectivamente, permitem estimativas precisas de parâmetros estelares, por fornecer informações sobre algumas características espectrais críticas enquadradas em alguns dos seus filtros de banda estreita. Outro projeto interessante em andamento focado no hemisfério norte é o levantamento Pristine (STARKENBURG *et al.*, 2017), que utiliza um filtro exclusivo de banda estreita centrado nas linhas Ca II H&K combinado com os filtros de banda larga do SDSS. No hemisfério sul, podemos citar o Southern Photometric Local Universe Survey (S-PLUS; MENDES DE OLIVEIRA *et al.*, 2019), com câmera e telescópio similares ao J-PLUS, e cujos dados mais recentes contam com um novo método de calibração proposto por ALMEIDA-FERNANDES *et al.* (2021). O céu do hemisfério Sul também está sendo observado sistematicamente pelo SkyMapper Southern Sky Survey (SkyMapper; KELLER *et al.*, 2007).

Os dados fotométricos e espectroscópicos se complementam, o que significa que a melhor abordagem possível para a pesquisa astronômica é combinar os dois tipos de análise de dados. Diferentes calibrações fotométricas apresentam resultados confiáveis para estimar parâmetros estelares, mas com uma faixa restrita de aplicabilidade. Por exemplo, IVEZIĆ *et al.* (2008) descobriram que a cor ($g - r$) do SDSS fornece uma estimativa precisa de T_{eff} , enquanto $[\text{Fe}/\text{H}]$ pode ser obtido através de um ajuste polinomial usando $(u - g)$ com algumas restrições em $(g - r)$, devido a uma rápida saturação da banda azul, juntamente com uma perda de precisão para estrelas com $[\text{Fe}/\text{H}] < -2,0$. Outra metodologia independente para inferir T_{eff} é o método de fluxo infravermelho (IFRM; CASAGRANDE *et al.*, 2010; CASAGRANDE, 2008; CASAGRANDE *et al.*, 2006), que pode ser usado para construir relações entre o fluxo medido no infravermelho e a fotometria BVJHK. Estas calibrações são válidas para $T_{\text{eff}} = 4000 - 8000\text{K}$, correspondendo a estrelas com tipos espectrais F, G e K, com $[\text{Fe}/\text{H}]$ variando de -5 a $+0,4$.

Em termos de observações espectroscópicas, os trabalhos de LEE *et al.* (2008a) e LEE *et al.* (2008b) apresentaram o SEGUE Stellar Parameter Pipeline (SSPP), uma

ferramenta computacional para estimar parâmetros estelares a partir de calibrações teóricas e empíricas de espectros estelares de média resolução ($R \sim 1800$) em combinação com a implementação de redes neurais e validações baseadas em bibliotecas espectrais como ELODIE (MOULTAKA *et al.*, 2004; PRUGNIEL e SOUBIRAN, 2001) e MILES (SÁNCHEZ-BLÁZQUEZ *et al.*, 2006). Espectros de alta resolução são usados para validação adicional, tal como estudado por PRIETO *et al.* (2008b). WU *et al.* (2014) e XIANG *et al.* (2015) também apresentaram códigos automáticos, ou *pipelines*, semelhantes (LASP e LSP3 respectivamente) para estimar parâmetros a partir de espectros produzidos pelo levantamento LAMOST.

Nos próximos anos, conjuntos de dados que serão produzidos por diferentes pesquisas extensas, como o Large Synoptic Survey Telescope (LSST) e os lançamentos progressivos de dados da missão Gaia (GAIA COLLABORATION *et al.*, 2016, 2018) gerarão uma quantidade descomunal de dados, de modo que serão necessárias novas estratégias eficazes para lidar com eles. Nesse sentido, algoritmos de aprendizado de máquina ou quaisquer outras ferramentas estatísticas podem ser úteis para analisar os dados e fazer previsões razoavelmente rápidas. Métodos de aprendizado de máquina, como redes neurais artificiais (RNAs), fazem parte da abordagem moderna para diversas aplicações astronômicas. Por exemplo, SINGH *et al.* (1998) apresentaram um modelo para classificar espectros estelares, enquanto, mais recentemente, WHITTEN *et al.* (2019) usaram RNAs para estimar T_{eff} e $[\text{Fe}/\text{H}]$ para dados do J-PLUS, com o objetivo de procurar estrelas de baixa metalicidade. Outros modelos baseados em outros algoritmos como *Random Forest* (doravante RF; BREIMAN, 2001) e *Extreme Gradient Boosting* (doravante XGB; CHEN e GUESTRIN, 2016; CHEN *et al.*, 2015) mostram resultados promissores de classificações morfológicas e estimativa de parâmetros físicos. MILLER *et al.* (2015) apresentaram um modelo de RF capaz de inferir T_{eff} , $\log g$ e $[\text{Fe}/\text{H}]$ baseado em cores desavermelhadas do SDSS. Mais recentemente, BAI *et al.* (2018) construíram um modelo de RF que realizou uma classificação estrela-galáxia-QSO e calculou a T_{eff} para estrelas usando dados do SDSS e LAMOST; além disso, CHAO *et al.* (2019) aplicaram um modelo baseado no algoritmo XGB que é capaz de classificar estrelas e galáxias no SDSS melhorando os resultados para fontes fracas, em particular. A busca por candidatas a VMP a partir de levantamentos fotométricos de grande campo angular (wide-angle surveys), como o J-PLUS, complementado com acompanhamentos espectroscópicos, pode se beneficiar enormemente dos métodos de aprendizado de máquina.

Este trabalho começa discutindo a importância de estudar estrelas muito pobres em metais e como elas podem ajudar a aprimorar modelos de formação e evolução da Galáxia (Capítulo 2). Outro ponto central deste trabalho é a apresentação do pipeline Stellar Parameters Estimation based on the Ensemble Methods (doravante SPEEM), que explora os benefícios dos algoritmos RF e XGB e seu potencial como uma excelente alternativa para analisar dados J-PLUS estimando três principais parâmetros físicos de fontes estelares:

T_{eff} , $\log g$ e $[\text{Fe}/\text{H}]$. O Capítulo 3 apresenta as bases de dados empregadas para produzir as amostras de treinamento e validação necessárias ao desenvolvimento do SPEEM. O Capítulo 4 cobre os detalhes sobre a arquitetura do pipeline SPEEM e o processo de limpeza centrado na remoção de contaminantes, tais como fontes extragalácticas pontuais (QSOs) e anãs brancas (WD). Posteriormente, neste mesmo capítulo se explica o processo de treinamento e validação, apontando os features (parâmetros de entrada do modelo) que foram selecionados como os indicadores mais confiáveis para estimar parâmetros estelares. O Capítulo 5 descreve como o SPEEM ajudou na seleção de estrelas muito pobres em metais adequadas para observações espectroscópicas, que resultaram em 11 candidatas a estrelas pobres em metais, com 10 estrelas recém-descobertas entre elas, juntamente com uma estrela extremamente pobre em metais. O Capítulo 6 está dedicado a apresentar a aplicação do SPEEM a estrelas tipo solar dos campos K2 do Telescópio Espacial Kepler. Por fim, o Capítulo 7 resume os resultados gerais obtidos, bem como futuras aplicações e melhorias para estender a estimativa de metalicidades para o regime $[\text{Fe}/\text{H}] < -3,0$.

Capítulo 2

Distribuição de Metalicidade no Halo Galáctico

2.1 Uma Breve Revisão Histórica

Uma boa maneira de apresentar uma breve história da formação de galáxias poderia começar com o primeiro catálogo focado na identificação de objetos diferentes de estrelas pontuais apresentado por **MESSIER (1781)**. O curioso é que, na verdade, Messier era mesmo um apaixonado “caçador de cometas”, por isso compilou um catálogo para evitar que ele e outros entusiastas observassem objetos que não fossem cometas. O catálogo original é composto por 100 objetos com uma adição posterior feita por M. Méchain¹ de quatro objetos que Messier não viu e também uma revisão moderna dando um total de 110 objetos.

Oito décadas depois, **HERSCHEL (1863)** publicou seu Catálogo Geral de Nebulosas, onde compilou 5.078 objetos juntando-se a catálogos anteriores feitos por astrônomos como William e Caroline Herschel, M. Auwers, Lacaille, e Messier (mencionado anteriormente) entre outros. Este catálogo foi organizado e complementado posteriormente por **DREYER (1888)** tomando o nome de Novo Catálogo Geral de Nebulosas e Aglomerado de Estrelas com uma extensão extra denominada Catálogos de Índice. No total, o NGC contém 7840 objetos enquanto o IC compila 5386 objetos. As notações M###, NGC#### e IC#### ainda são amplamente utilizadas na astronomia moderna.

Todos esses catálogos e compilações de observações abriram as seguintes questões: Qual era a verdadeira natureza desses objetos difusos, especialmente aqueles referidos como “nebulosas espirais”? Onde eles devem estar localizados no espaço? Para responder a essas perguntas, uma discussão entre dois dos mais conhecidos e famosos astrônomos ocorreu no ano de 1920, e foi nomeado “O Grande Debate”². De um lado, Herbert D. Curtis argumentou que nossa galáxia deveria ter um diâmetro de 30.000 anos-luz com

¹Ver <http://www.messier.seds.org/xtra/history/pm-lett.html>

²Ver https://apod.nasa.gov/debate/1920/cs_nrc.html

sua forma semelhante ou parecida com uma lente e o Sol localizado próximo ao centro. Portanto, as nebulosas espirais podem ser consideradas como universos insulares. Por outro lado, Harlow Shapley propôs que a Via Láctea deveria ter um diâmetro de cerca de 220.000 anos-luz, com o Sol localizado longe de seu centro e as nebulosas espirais pertencentes à Galáxia.

Com o mistério não resolvido da verdadeira natureza das nebulosas espirais observadas, **HUBBLE (1926)** conseguiu responder a essa pergunta. Ele conseguiu detectar variáveis Cefeidas dentro de NGC 6822, M33 e M31 estudando imagens individuais de cada uma delas. Isso lhe permitiu estimar sua distância à Terra e concluiu que algumas nebulosas (as espirais) estavam distantes demais para pertencer à Via Láctea, apoiando a ideia de “universos-ilha” que Curtis havia proposto antes. Isso pode ser considerado o início da Astronomia Extragaláctica. Ele também usou suas observações para desenvolver um sistema de classificação, conhecido agora como “diapasão” ou “sequência de Hubble”. Ele estabeleceu duas categorias: Regulares e Irregulares, e dentro dos Regulares foram identificadas duas categorias principais: Elípticas e Espirais. Hubble pensou que as galáxias evoluíram de elípticas para espirais, chamando-as de tipo inicial e tipo tardio, respectivamente, o que agora sabemos que não é verdade.

Se recuássemos uma década no tempo, notaríamos que **EINSTEIN (1916)** publicou sua teoria da Relatividade Geral que não apenas mudou nossa compreensão da força gravitacional, mas plantou as primeiras sementes para a criação de uma teoria que poderia explicar o próprio Universo como um todo. Essa teoria é conhecida como Cosmologia. No início, ele pensou que o Universo deve ser estático, então trabalhou nas soluções de suas equações de campo recentes, descobrindo surpreendentemente que as soluções sugeriam um Universo em expansão ou em contração. Para enriquecer a discussão, **FRIEDMAN** (Tradução **FRIEDMAN, 1999**) e **LEMAÎTRE (1931)** exploraram soluções estáticas e expansíveis, chegando aos mesmos resultados de forma independente. Nossa visão do Universo também foi transformada quando **HUBBLE (1929)** encontrou outro resultado notável: o Universo estava se expandindo a uma taxa constante.

ZWICKY (1935) descreveu a relação entre distância (ou *redshift*) de nebulosas extragalácticas, notando como um caso particular, que para explicar a velocidade de dispersão das galáxias observadas no aglomerado de galáxias de Coma, uma densidade média 400 vezes maior da matéria luminosa deve ser exigida. Além disso, **ZWICKY (1937)** continuou trabalhando com métodos para estimar as massas de galáxias e aglomerados de galáxias propondo que algumas galáxias massivas poderiam atuar como lentes gravitacionais produzindo múltiplas imagens de galáxias de fundo. Zwicky pode ser considerado o primeiro a introduzir o termo matéria escura para explicar o problema da estimativa das massas das galáxias. **GAMOW (1946)** discutiu sobre a origem dos elementos químicos em um cenário do Universo em expansão. Ele acreditava que a abundância dos diferentes elementos químicos era determinada pelas condições físicas de um Universo mais quente e denso,

ou seja, durante os estágios iniciais de sua expansão. No entanto, houve uma discrepância para elementos mais pesados que ele explicou argumentando que alguns processos deveriam ter ocorrido durante um intervalo de tempo limitado gerando desequilíbrio no sistema. Essa ideia evoluiu como o conceito de nucleossíntese primordial na cosmologia moderna. HOYLE (1948) zombou das ideias de George Gamow referindo-se a elas como o Hot Big Bang para argumentar a favor de um universo em estado estacionário. Além disso, Gamow, Ralph Alpher e Robert Herman percebem que deveria haver radiação térmica detectável como sinal de fundo. Tal previsão foi observada por PENZIAS e WILSON (1965) e agora é conhecida como Radiação Cósmica de Fundo em Microondas (CMB).

EGGEN *et al.* (1962) apresentaram o primeiro modelo detalhado da formação de nossa Galáxia baseado em uma análise dinâmica de uma amostra de 221 estrelas. Eles concluíram que “os objetos mais antigos foram formados em quase qualquer altura acima do plano galáctico, enquanto os mais jovens foram formados perto do plano”. Eles sugeriram que as galáxias podem ter se formado a partir do colapso de nuvens de gás nas quais elípticas e espirais diferem na rapidez com que suas populações de estrelas se formaram. Para um caso geral, suas ideias evoluíram para o modelo *Monolithic Collapse*.

PEEBLES (1965) apresentou argumentos de quais deveriam ser as condições nos estágios iniciais do Universo para favorecer o processo de formação de galáxias. Ele aceitou a ideia de que o Universo está se expandindo e evoluindo com seu campo de radiação sendo um espectro de corpo negro, e que pode ter tido um efeito considerável na evolução da matéria que acabou formando galáxias. Ele explicou como uma galáxia de “tamanho típico” se formou acumulando muitas nuvens de gás produzidas devido à instabilidade gravitacional, conforme explicado na teoria de Jeans. Este cenário hierárquico é conhecido como o modelo “bottom-up” para a formação de galáxias.

SILK (1968) sugeriu que durante o período de recombinação, no qual os elementos químicos mais leves foram formados, uma difusão de fótons deveria ter ocorrido para reduzir as anisotropias de densidade no universo primitivo, causadas por perturbações acústicas de flutuação de densidade inicial. Inicialmente, os fótons eram dispersos por prótons e elétrons que viajavam alguma distância sem serem espalhados novamente. Assim que o Universo esfriou efetivamente, a quantidade de espalhamento diminuiu e os fótons puderam viajar distâncias ainda maiores. Silk também definiu uma escala de comprimento em que esse amortecimento funciona, correspondente a 28 Mpc (Mega parsecs) com uma massa contida correspondente à massa de galáxias semelhantes à Via Láctea ($10^{12} - 10^{14} M_{\odot}$). Esses parâmetros são conhecidos como escala de Silk e massa de Silk.

Um cenário alternativo para a formação de galáxias foi proposto por ZEL'DOVICH (1970). Neste caso, as grandes estruturas (acima da escala de amortecimento) foram favorecidas ao invés de pequenas estruturas. Essas estruturas resultaram em “panquecas densas, filamentos e aglomerados compactos de matéria” (SHANDARIN e ZELDOVICH, 1989) que então se fragmentaram em nuvens menores, formando protogaláxias. Este é o

modelo “top-down” para a formação de galáxias.

GUTH e PI (1982) fizeram uma contribuição extremamente importante para nossa compreensão atual do Universo. Ele mostrou que para corrigir alguns problemas nas condições iniciais do modelo cosmológico do big-bang quente, o Universo poderia ter se expandido exponencialmente durante um breve período de tempo, resultando em um processo de super-resfriamento fundamental para corrigir o chamado problema do horizonte e o problema da planicidade.

Uma década antes (anos 1970) e 40 anos desde Zwicky, o termo matéria escura foi sugerido que poderia ser composto de objetos bariônicos luminosos mais fracos, como anãs marrons, anãs brancas e buracos negros, até partículas elementares especiais, como neutrinos massivos. Como nenhuma dessas hipóteses pôde ser confirmada, no início dos anos 1980 modelos alternativos sugeriram a existência de partículas massivas de interação fraca (MACHOs, WIMPs). Também levando em consideração o tempo necessário e qual deveria ser a verdadeira composição da matéria escura, muitos modelos de matéria escura quente (HDM), morna (WDM) ou fria (CDM) foram desenvolvidos. No início dos anos 1990, o modelo CDM já havia se estabelecido como o modelo de “melhor ajuste” para a formação de estruturas. No entanto, este modelo ainda apresentava alguns problemas em relação às estruturas de grandes escalas. Isso foi corrigido pela introdução de um modelo aberto com densidade crítica muito menor (OCDM) e a adição de uma constante cosmológica para tornar o Universo plano (Λ CDM).

Além disso, desde o início dos anos 1990, com o aprimoramento da tecnologia computacional, foram desenvolvidas simulações para estudar as propriedades dos halos do CDM e a formação e evolução das galáxias dentro deles. Por tanto, duas abordagens principais foram favorecidas: modelos semi-analíticos e simulações hidrodinâmicas. A abordagem semi-analítica focou na modelagem de poços de potencial gravitacional em que as galáxias se formam e evoluem levando em conta um conhecimento prévio da estrutura dos halos do CDM (**WHITE e FRENK, 1991**) usando métodos N-Body. As simulações hidrodinâmicas, por outro lado, levam em conta processos físicos mais detalhados do gás em colapso (termodinâmica, mecânica dos fluidos, etc.)

Finalmente, uma das mudanças recentes em nossa imagem do Universo ocorreu em 1998, quando foi mostrado no estudo de **RIESS *et al.* (1998)** que a expansão do Universo está acelerando no momento atual usando a relação distância-redshift de supernovas do tipo Ia. Isso implicou uma nova interpretação da constante cosmológica Λ que Einstein havia introduzido para contrabalançar o efeito da gravidade.

2.2 Estrutura da Via Láctea

Na atualidade, a estrutura conhecida da Vía Láctea é descrita como uma galáxia espiral separada em quatro regiões diferentes, as quais indo do interior para o exterior seriam:

bojo, disco, halo estelar e halo de matéria escura. O disco, por exemplo, é conhecido por apresentar uma distribuição de estrelas ricas em metais, quando comparadas com as do halo. Esse grupo de estrelas do disco corresponde à população I. O halo galáctico contém as estrelas mais velhas e, por sua vez, as mais pobres em metais, com $[Fe/H] < -1.6$ e identificadas como estrelas de população II. Em suma, cada região possui uma distribuição característica, porém para este projeto o halo estelar será a região abordada para estudo.

Diversas observações tem mostrado que no halo Galáctico é possível encontrar objetos como estrelas sub-anãs, sendo algumas delas muito pobres em metais; gigantes dos tipos K, M, RR Lyrae, BHBs (Blue Horizontal Branch); e aglomerados globulares, formados por estrelas de baixa metalicidade. Dentre essa composição especial do halo, as estrelas muito pobres em metais são de extrema importância, não só para melhorar a nossa compreensão das primeiras estrelas formadas no Universo, senão também para aprimorar as restrições no desenvolvimento de modelos de formação e evolução da Galáxia.

2.3 Como se formou a Via Láctea?

Para estabelecer modelos sobre formação e evolução de galáxias, é preciso levar em conta algumas considerações teóricas e observacionais tais como: as perturbações do Universo nas suas primeiras fases, a física da formação estelar e o enriquecimento do meio interestelar por parte de supernovas. No caso específico da nossa Galáxia, um dos primeiros modelos desenvolvido a partir de observações foi proposto por [EGGEN *et al.*](#) (modelo ELS, 1962). Naquele modelo se concluiu que a cinemática e a metalicidade das estrelas no halo funcionam como traçadores dos processos de formação da Vía Láctea. Contudo, [SEARLE e ZINN \(1978\)](#) constataram que não existia uma correlação clara entre a metalicidade e a distância ao centro da Galáxia (R_{gc}) para objetos com $R_{gc} \geq 7kpc$, o que sugere que alguns desses objetos poderiam ter sido acretados nas primeiras fases de formação da Via Láctea. Isto ainda é uma hipótese que está sendo discutida.

2.3.1 Modelo de Eggen, Lynden-Bell e Sandage (ELS)

Este modelo foi desenvolvido através da análise das componentes de velocidade de 221 estrelas anãs, com o objetivo de calcular as excentricidades e momentos angulares das órbitas galácticas. Eles encontraram uma forte correlação entre excentricidade e excesso de ultravioleta (Figura 2.1). Assim, as estrelas com o maior excesso, ou seja, a menor abundância de metal, estão invariavelmente se movendo em órbitas altamente elípticas, enquanto as estrelas com pouco ou nenhum excesso se movem em órbitas quase circulares. Os autores também encontraram outras correlações interessantes como o excesso do ultravioleta com a componente vertical da velocidade W , e esse excesso com os momentos angulares (Figuras 2.2 - esquerda e direita).

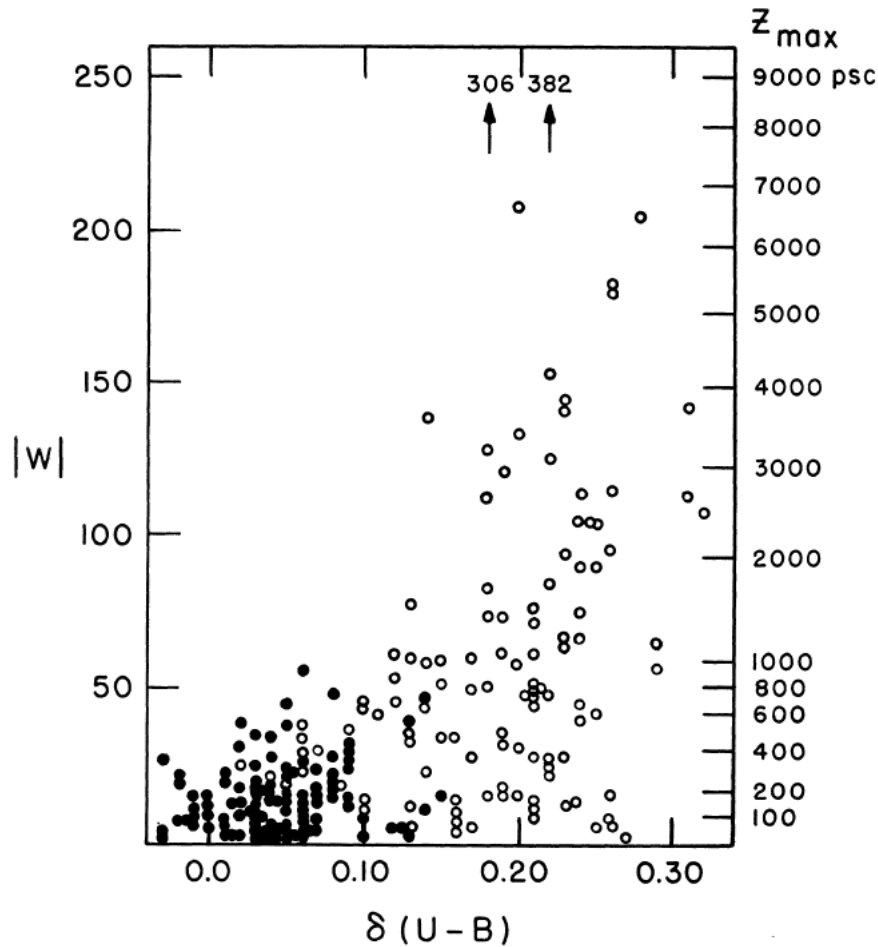


Figura 2.1: Correlação entre a velocidade W e o excesso ultravioleta para as 221 estrelas. Creditos: [EGGEN *et al.* \(1962\)](#)

O modelo ELS conclui que a cinemática e a metalicidade das estrelas no halo galáctico servem como um traçador fossilizado do processo de formação. Citando-os: “A Figura 2.2 sugere que os objetos mais antigos foram formados em quase qualquer altura acima do plano galáctico, enquanto os mais jovens foram formados muito perto do plano”. Algumas implicações que resultam do documento ELS são as seguintes:

- Protogaláxias com pouco momento angular podem colapsar em esferóides, enquanto aquelas com maior momento angular evoluiriam para sistemas de disco.
- A escala de tempo do colapso do halo galáctico é muito semelhante à escala de tempo dinâmica (relacionada ao tempo necessário para uma esfera uniforme livre de pressão colapsar em um ponto) implícita na densidade média dentro da protogaláxia.
- As maiores distâncias máximas alcançadas pelas estrelas mais pobres em metais respeito ao plano galáctico sofreram a maior influência no modelo de colapso. No caso do ELS, o rápido colapso da Galáxia está representado em menores metalicidades das estrelas do halo, devido ao pouco tempo para auto-enriquecimento.

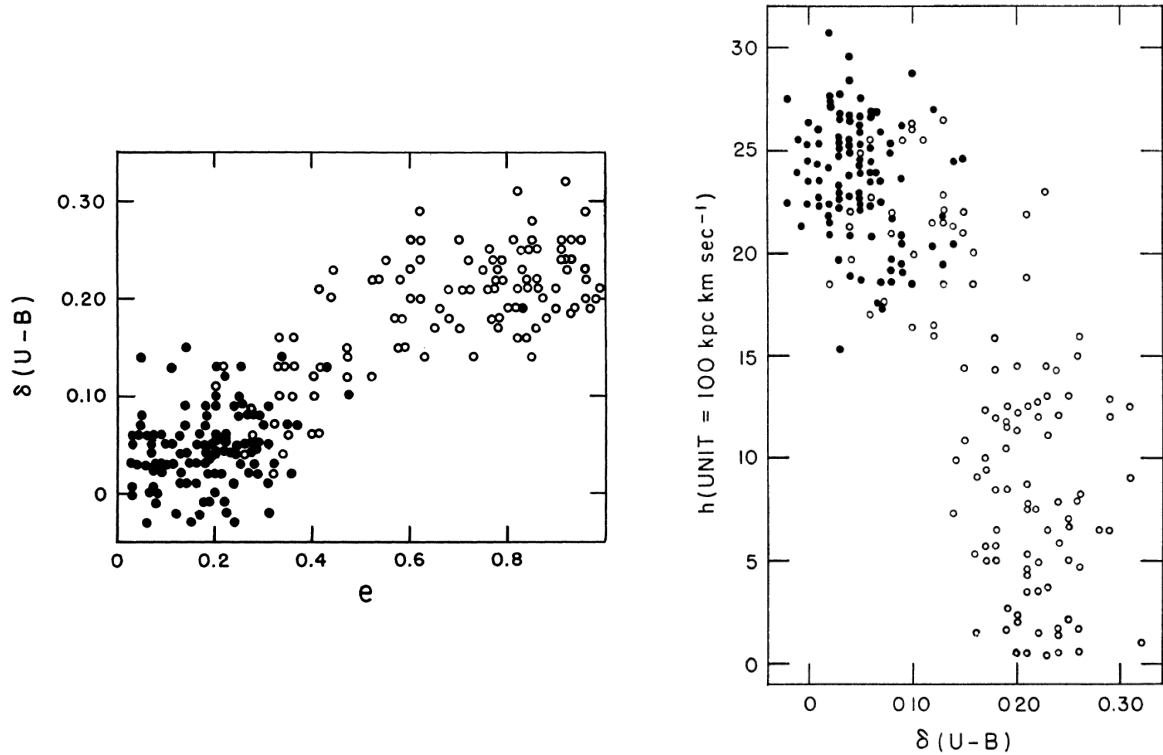


Figura 2.2: Resultados obtidos por [EGGEN *et al.* \(1962\)](#). Figuras tomadas do artigo original.

Esquerda: Correlação entre o excesso ultravioleta $\delta(U-B)$, e a excentricidade orbital e para uma amostra de 221 estrelas. *Direita:* Correlação entre o momento angular h e o excesso ultravioleta das estrelas.

2.3.2 Modelo do Searle & Zinn

Eggen, Lynden-Bell e Sandage argumentaram que as estrelas mais antigas da nossa Galáxia foram formadas em um rápido colapso de gás em direção ao plano galáctico. Em contraste, [SEARLE e ZINN \(1978\)](#) sugeriram que não há correlação entre a metalicidade e a distância ao centro galáctico (R_{gc}) para objetos além de $R_{gc} = 7 \text{ kpc}$. Eles descobriram esse resultado estimando as abundâncias de uma amostra de aglomerados globulares galácticos. Este fato implicava que um processo hierárquico deveria ser o cenário preferencial para formar o halo externo da Galáxia, através da acreção e fusão de fragmentos protogalácticos durante um longo período de tempo. Nas últimas três décadas, o consenso é um modelo composto, que consiste em um colapso dissipativo juntamente com inúmeros eventos de fusão. Ele fornece uma explicação melhor para os dados atualmente disponíveis ([MACKEY e GILMORE, 2004](#)). Os cerca de 150 aglomerados globulares que [SEARLE e ZINN \(1978\)](#) estudaram revelaram-se ferramentas importantes não só para estimar as propriedades da nossa Galáxia na época atual, mas também para analisar a sua história de formação. Em comparação com a maioria das estrelas de campo, esses objetos são geralmente mais brilhantes. Assim, podem ser observados a grandes distâncias ($\sim 100 \text{ kpc}$)

ou alternativamente, através de regiões de baixa extinção. Outro fato importante é que sua cinemática pode fornecer informações de diferentes regiões da Galáxia em diferentes posições (BELLAZZINI *et al.*, 2004; ODENKIRCHEN *et al.*, 2003).

ZINN (1993) descobriu que aglomerados globulares ricos em metais ($[Fe/H] > -0,8$) estão confinados ao bojo e disco da Galáxia, enquanto aglomerados globulares pobres em metais ($[Fe/H] < -0,8$) são geralmente localizado no halo galáctico. Uma possível explicação para isso é que as observações sugerem que a maioria dos objetos do halo antigo devem ter se formado durante um colapso dissipativo da maneira proposta por EGGEN *et al.* (1962), antes da formação do disco. Por outro lado, objetos de halo jovens e alguns dos objetos de halo velhos pobres em metal não deveriam pertencer à Galáxia em princípio. Assim, eles poderiam ter sido formados em galáxias anãs externas e acrescidos mais tarde no halo galáctico. VAN DEN BERGH (1993) chegou a uma conclusão semelhante.

2.4 O que aconteceu após a formação do halo de matéria escura (DMH no inglês) e o colapso do gás?

2.4.1 Formação Estelar

A massa produzida durante um episódio de formação estelar é distribuída de forma mais ou menos homogênea, e é descrita pela função de massa inicial (IMF; KROUPA *et al.*, 2013). Além disso, a quantidade total de massa estelar produzida em uma galáxia é distribuída principalmente em estrelas de baixa massa com as estrelas variando em uma faixa de um décimo de uma massa solar a 100 massas solares. Durante a fase de formação de estrelas, a luminosidade de uma galáxia é dominada pela taxa na qual as estrelas estão sendo produzidas. Nesse sentido, podemos inferir a quantidade de estrelas massivas observadas em um determinado momento. Além das mudanças na luminosidade, massa estelar e cor durante a fase de formação estelar, a produção de elementos mais pesados que o hélio (geralmente chamados de “metais” em astronomia) é essencial. De fato, todos os elementos encontrados na Terra foram sintetizados no interior das estrelas por fusões nucleares, e depois liberados para o meio interestelar durante explosões de supernovas entre outros processos de perda de massa, como ventos estelares, por exemplo. Consequentemente, o gás no meio interestelar mais formação de estrelas é alimentada e novas gerações de estrelas aumentam o conteúdo de metal da Galáxia (e do Universo), e assim por diante.

2.4.2 Evolução Química de galáxias

É fato que, como um dos componentes fundamentais das galáxias são as estrelas, sua evolução química é explicada pela evolução das próprias estrelas. Nesse sentido, espera-se

que a metalicidade Z aumente com a taxa de formação estelar. O livro de SCHNEIDER (2006, 2014) apresenta um modelo simples da evolução química de uma galáxia que leva a algumas conclusões interessantes, por exemplo, aquela em que parte da massa total das populações estelares recém-formadas é devolvida ao meio interestelar (ISM) por explosões de supernovas e outros processos de perda de massa. Além disso, os metais também são devolvidos ao ISM por processos de evolução estelar. Por outro lado, a energia cinética transferida para o ISM por supernovas causa um fluxo de escape, especialmente em galáxias de baixa massa, onde o gás está fracamente ligado gravitacionalmente.

No caso particular da Via Láctea, encontramos uma inconsistência conhecida como problema das anãs-G. O problema é que de acordo com o modelo simples da evolução química da Galáxia, deveriam existir mais estrelas pobres em metais dos tipos G, K do que as observadas. Possíveis soluções para o problema: acreção de gás e migração radial de estrelas. A ideia é que deve haver mais estrelas de sequência principal F e G pobres em metal que estão em discrepância com o que é observado. No entanto, há um desafio com a interpretação devido a vários fatos. Um deles, por exemplo, é que as galáxias podem agregar materiais de metalicidades diferentes. Os processos de realimentação também podem expelir gás enriquecido com metais, e as fusões podem misturar as composições dos sistemas progenitores.

2.5 Expandindo o limite de metalicidade para estrelas

Antes da década de 1980, era aceita a ideia de que o halo Galáctico não possuía um número representativo de estrelas com metalicidade $[\text{Fe}/\text{H}] < -3.0$ (que corresponde ao limite inferior de metalicidade encontrado para aglomerados globulares que habitam o halo), sendo o levantamento HK (**HK survey**) de Beers e colaboradores (BEERS *et al.*, 1985, 1992; NORRIS *et al.*, 1999) um dos únicos catálogos disponíveis que continha informação sobre VMPs (~ 1000) e EMPs (~ 100) na época. Esta ideia tem sido gradativamente abandonada com o descobrimento da sub-anã G 64-12, com metalicidade $[\text{Fe}/\text{H}] = -3.5$ reportada por CARNEY e PETERSON (1981); e a gigante CD $-38^{\circ}245$, com $[\text{Fe}/\text{H}] = -4.5$ reportada por BESSELL e NORRIS (1984). Posteriormente, as metalicidades dessas estrelas foram reportadas como $[\text{Fe}/\text{H}] = -3.2$ por RYAN *et al.* (1999); e $[\text{Fe}/\text{H}] = -3.5$ por NORRIS *et al.* (2001), respectivamente. Recentemente projetos tais como Pristine (STARKENBURG *et al.*, 2017) e SEGUE (YANNY *et al.*, 2009) vem contribuindo com o aumento desse número.

BEERS e CHRISTLIEB (2005) propuseram uma nomenclatura (ver Tabela 2.1) para estrelas pobres em metais, baseada nos limites de metalicidade. Daqui vem o termo “estrelas extremamente pobres em metais” ou EMPs (Extremely Metal-Poor Stars, em inglês), e refere-se a estrelas com $[\text{Fe}/\text{H}] < -3.0$. Na atualidade, o número de estrelas do Halo identificadas com $[\text{Fe}/\text{H}] < -3.5$ é menor que 100, embora existam vários grupos

Tabela 2.1: Nomenclatura para grupos de estrelas por $[\text{Fe}/\text{H}]$

$[\text{Fe}/\text{H}]$	Nome	Acrônimo
$> 0,5$	Super metal-rich	SMR
$\sim 0,0$	Solar	–
$< -1,0$	Metal-poor	MP
$< -2,0$	Very metal-poor	VMP
$< -3,0$	Extremely metal-poor	EMP
$< -4,0$	Ultra metal-poor	UMP
$< -5,0$	Hyper metal-poor	HMP
$< -6,0$	Mega metal-poor	MMP

dedicados à sua busca ([ABOHALIMA e FREBEL, 2018](#)).

Um dos objetivos mais interessantes na busca por estrelas pobres em metais consiste na identificação de várias estrelas que, além da baixa metalicidade, apresentam enriquecimento na razão carbono/ferro, também conhecidas como CEMPs (*carbon-enhanced metal-poor stars*).

Já que ainda não existe um acordo bem definido sobre essa nomenclatura, alguns autores adotam o limite $[C/Fe] > +1.0$, enquanto que outros preferem o valor $[C/Fe] > +0.7$ ([AOKI *et al.*, 2007](#)) para que uma estrela pobre em metais seja considerada também como CEMP. Dentro da classificação CEMP, também pode ser levada em conta a presença de elementos formados pelos processos r e s , que contribuem ao enriquecimento químico dessas estrelas.

Graças aos avanços contínuos na instrumentação óptica dos diferentes telescópios disponíveis, tem sido possível incrementar a identificação de estrelas de muito baixa metalicidade e aprimorar a análise química a partir do uso de técnicas avançadas de espectroscopia de alta resolução. Sobre esses avanços recentes vale destacar o catálogo de [SESTITO *et al.* \(2019\)](#) com 42 estrelas UMP ($[\text{Fe}/\text{H}] < -4,0$) encontradas na Galáxia; a estrela peculiar MMP ($[\text{Fe}/\text{H}] < 6,0$) encontrada por [AGUADO *et al.* \(2018\)](#) a partir de dados dos levantamentos SDSS-SEGUE (mencionado no seguinte capítulo), BOSS ([DAWSON *et al.*, 2016](#)) e LAMOST. Outros estudos como o do [PLACCO *et al.* \(2018\)](#) discutem a importância do efeito da abundância do carbono ao encontrar um incremento aproximado de 20% de estrelas tipo CEMP na faixa de $[\text{Fe}/\text{H}] < -2,0$, e 80% na faixa de $[\text{Fe}/\text{H}] < -4,0$. Esse mesmo estudo reforça a necessidade de identificar novas candidatas com magnitude $V < 12$, já que elas constituem os melhores alvos para resolver suas linhas espectrais usando espectroscopia de alta resolução em telescópios terrestres de 4 a 8 m de abertura.

2.6 Comentários Finais

Neste capítulo, se apresentou uma revisão histórica de como nossa compreensão da Galáxia e sua posição no Universo tem mudado ao longo do tempo. A questão do seu processo

de formação e evolução depende muito da correta identificação e análise química e cinemática das suas estrelas constituintes. Nesse sentido, o estudo detalhado de estrelas de baixa metalicidade no halo galáctico representa uma das linhas de pesquisa principais para estabelecer condições restritivas sobre modelos analíticos de formação e evolução da Galáxia. Desde o modelo proposto por [EGGEN *et al.* \(1962\)](#) tem aumentado o interesse e os estudos dedicados a buscar e caracterizar esse tipo de estrelas. Tal como explicado por [BEERS e CHRISTLIEB \(2005\)](#), uma metodologia muito aplicada para estudar estrelas de baixa metalicidade é apresentada na Figura 2.3, onde se apresenta uma análise preliminar da estrela HE 0107-5240 com espectroscopia de baixa resolução ($R=400$, painel superior), posteriormente focando a análise entre 3800\AA e 4500\AA para espectroscopia de média resolução ($R=2000$, painel do meio), com uma clara identificação das linhas Ca K, H δ e CH, incluindo também o espectro observado da estrela CD-38 245 que não apresenta espectro de baixa resolução. Finalmente, no painel inferior se apresentam os espectros de alta resolução no intervalo entre 3919\AA e 3940\AA das duas estrelas, onde pode-se notar claramente detalhes das linhas CH e Ca K para a estrela HE 0107-5240 e Fe I e Ca K para estrela CD-38 245, lembrando que HE 0107-5240 apresenta $[\text{Fe}/\text{H}] = -5,2$ e CD-38 245 $[\text{Fe}/\text{H}] = -4,59$ reportado no estudo de [ROEDERER *et al.* \(2014\)](#).

No seguinte capítulo se apresentam alguns dos levantamentos fotométricos e espectroscópicos que mais têm contribuído para aumentar nosso conhecimento sobre a distribuição espacial e química das estrelas da nossa Galáxia. Um dos objetivos principais deste trabalho é usar uma combinação de dados fornecidos por diferentes levantamentos como LAMOST, SEGUE e GALAH junto com J-PLUS para desenvolver catálogos de parâmetros estelares a partir da fotometria multi-banda J-PLUS, que nos permitam identificar novas candidatas potenciais VMPs ou EMPs para serem observadas com espectroscopia de média resolução. Esse é o primeiro passo para eventualmente conseguir passar da identificação fotométrica de candidatas diretamente a confirmação espectroscópica em alta resolução, sem precisar do teste intermediário de média resolução.

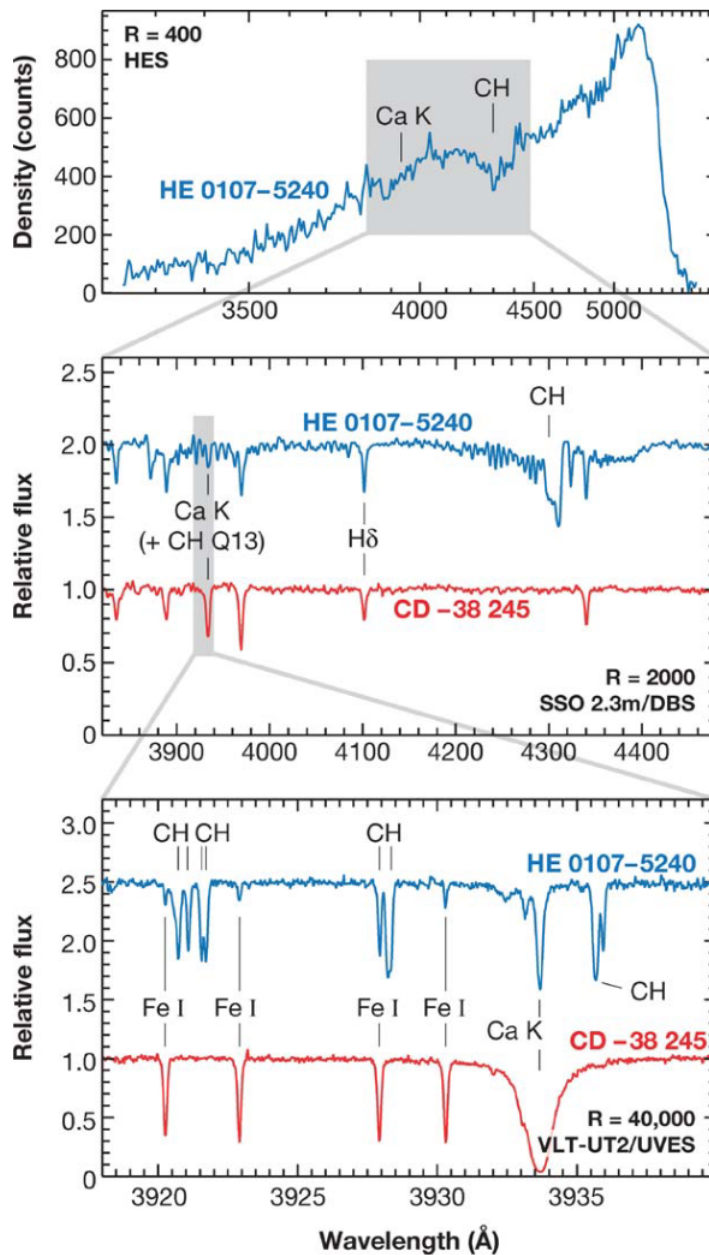


Figura 2.3: Os três principais passos observacionais para obter abundâncias elementares de estrelas pobres em metais: (a) **Wide-angle surveys** (por exemplo, levantamentos de prisma objetivo) produzem candidatas a estrelas pobres em metais; (b) seleção das candidatas por acompanhamento espectroscópico em media resolução; e (c) espectroscopia de alta resolução de candidatas confirmadas como estrelas pobres em metal. A estrela mostrada neste exemplo, HE 0107-5240, é uma das estrelas mais pobres em ferro já descobertas. As intensidades de suas linhas de absorção são comparadas com as linhas da gigante mais pobre em ferro conhecida até então, CD-38°245. Os espectros mostrados nos dois painéis inferiores foram normalizados a um continuum arbitrário. O poder de resolução, $R = \lambda/\lambda$, dos espectros é indicado ao longo do lado direito de cada painel. Espécies atômicas e moleculares proeminentes são identificadas. Créditos: **BEERS e CHRISTLIEB** (Figura 1, 2005)

Capítulo 3

Grandes Levantamentos Astronômicos

Tal como explicado por [DJORGOVSKI *et al.* \(2013\)](#), o conceito de levantamento astronômico tem mudado com o tempo. Anteriormente, estava mais associado aos chamados Atlas do Céu, mas hoje em dia refere-se a amplos catálogos de fontes e suas propriedades tais como posições, magnitudes, fluxos, morfologia, paralaxes entre outras. Além disso, o aumento considerável da quantidade de dados existentes para cada fonte observada exige um cuidado especial sobre metadados (dados que descrevem dados) bem documentados.

A pesquisa observacional moderna em astronomia e astrofísica requer um conjunto de dados em diversos comprimentos de onda, presentes em vários catálogos e bancos de dados. Analisando e combinando informações de diversos bancos de dados, é possível obter novas características de objetos previamente observados, permitindo melhorar nosso conhecimento do cosmos.

Este capítulo apresenta o levantamento fotométrico J-PLUS, cujos dados são essenciais no desenvolvimento deste projeto. Também se apresentam alguns grandes levantamentos astronômicos (fotométricos e espectroscópicos) que são úteis quando combinados com o J-PLUS para a estimação de parâmetros estelares (explicado nos Capítulos seguintes).

3.1 Levantamentos fotométricos multifiltros

Os levantamentos fotométricos multifiltros são convenientes nos casos em que a resolução espectral não é um fator limitante e a espectroscopia multiobjeto não é criticamente necessária. Algumas vantagens incluem que a seleção de objetos não é tendenciosa e o único efeito significativo é a magnitude limite.

De acordo com [BAILER-JONES \(2004\)](#), [JORDI *et al.* \(2006\)](#), e [MARTÍNEZ e MANUEL \(2006\)](#), é possível recuperar de forma confiável a distribuição espectral de energia (SED, do inglês *spectral energy distribution*) dos objetos observados a partir de um conjunto de cerca de 10-15 filtros ópticos de banda intermediária a larga. Essa proposta, em conjunto com exemplos anteriores de sucesso como COMBO-17 ([WOLF *et al.*, 2003](#)), ALHAMBRA ([MOLES *et al.*, 2008](#)), COSMOS ([ILBERT *et al.*, 2009](#)), MUSYC ([CAR-](#)

DAMONE *et al.*, 2010) e CLASH (POSTMAN *et al.*, 2012), fazem parte da motivação do desenvolvimento do projetos tais como o Javalambre Physics of the Accelerating Universe Astrophysical Survey (J-PAS) e o Javalambre-Photometric Local Universe Survey (J-PLUS).

3.1.1 O Observatório Astronômico de Javalambre e o Levantamento fotométrico J-PAS



Figura 3.1: Vista aérea do Observatorio Astronômico de Javalambre (OAJ), localizado em Teruel, Espanha, mostrando as cúpulas dos telescópios JST/T250 (acima) e JAST/T80 (abaixo).

Creditos: <https://oajweb.cefca.es/oaj/principal>

O projeto J-PAS constitui um esforço para registrar a SED de objetos astronômicos usando um sistema inédito de filtros fotométricos. O J-PAS será conduzido no Observatório Astrofísico de Javalambre em Teruel, Espanha, (OAJ, Figura 3.1, CENARRO *et al.* 2014), construído e gerenciado pelo Centro de Estudios de Física del Cosmos de Aragón (CEFCA), criado em 2008. O desenvolvimento destes projetos contam com a parceria de outras instituições científicas, dentre as quais estão o Observatório Nacional e o Instituto de Astronomia, Geofísica e Ciências Atmosféricas da Universidade de São Paulo (IAG-USP).

O levantamento J-PAS utiliza o telescópio de 2,5 m JST/T250 do OAJ (Painel esquerdo superior da Figura 3.2) acoplado à inovadora câmera JPCam, que conta com um sistema de 14 CCDs e2v de 9,2k por 9,2k pixels (Painel direito superior da Figura 3.2). O sistema de filtros do J-PAS consiste em um conjunto de 54 filtros de banda estreita com 145 Å de largura mais três filtros de banda larga, do tipo SDSS. Além disso, dois filtros de banda larga especialmente projetados na faixa de cobertura espectral óptica azul e vermelho complementam todo o sistema fotométrico (Figura 3.3).

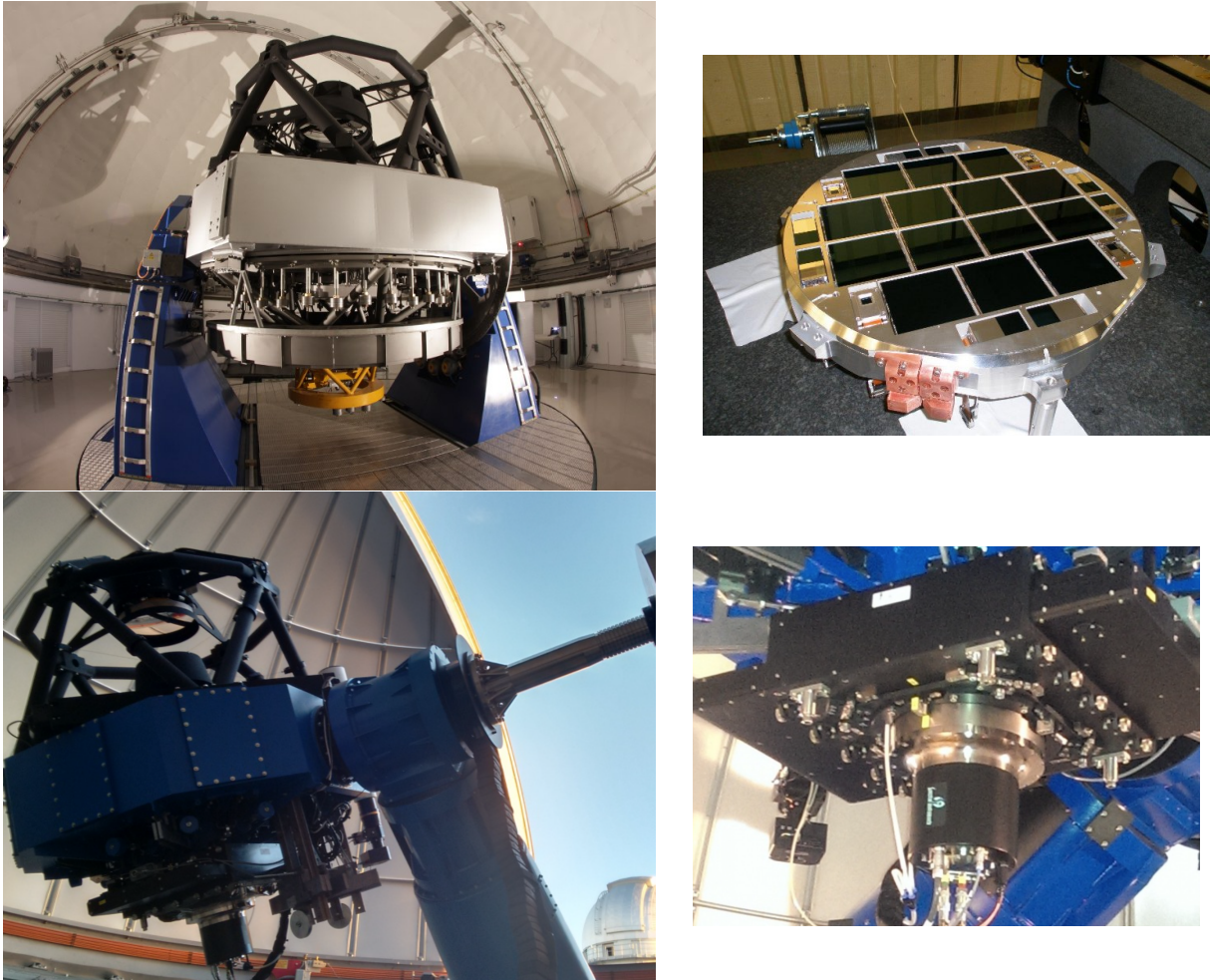


Figura 3.2: Equipamentos do OAJ: *Esquerda superior*: Telescópio JST/T250. *Direita superior*: Câmera panorâmica JPCam, de $4,7\text{deg}^2$. *Esquerda inferior*: Telescópio JAST/T80. *Direita inferior*: Câmera panorâmica T80Cam montada no foco JAST/T80 Cassegrain. Créditos: <http://www.j-pas.org/survey> e <http://www.j-plus.es/survey>

O J-PAS está projetado principalmente para observar objetos do céu profundo e atingir objetivos científicos com um viés para as áreas de cosmologia e astronomia extragaláctica, entre os quais destacam-se: redshifts fotométricos de alta precisão (photo-z), ELGs (galáxias de linha de emissão), QSOs, detecção de aglomerados e grupos de galáxias e redshifts de supernovas. A investigação do universo local inclui a população estelar do halo Galáctico, galáxias anãs e galáxias do grupo local, além de pequenos corpos do Sistema Solar.

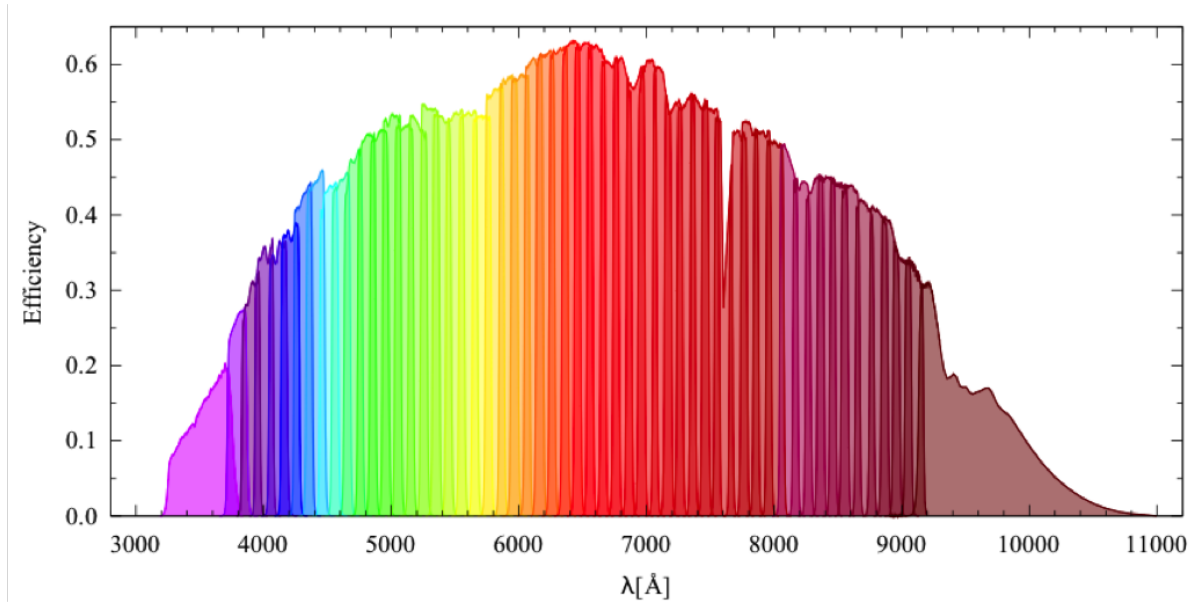


Figura 3.3: Sistema de filtros J-PAS (BONOLI *et al.*, 2021)

3.1.2 Mapeando o céu do hemisfério norte com J-PLUS

O levantamento fotométrico J-PLUS¹ foi concebido inicialmente como uma forma de calibrar as observações do J-PAS, uma vez que seu sistema fotométrico de doze filtros combina filtros de banda estreita do J-PAS, filtros do levantamento SDSS, além de filtros especialmente projetados para casos científicos específicos da astrofísica galáctica. A combinação única dos doze filtros do J-PLUS oferece grandes possibilidades na determinação de diversos parâmetros físicos tanto para galáxias como para estrelas. O J-PLUS adquiriu identidade própria, com foco principal no universo local, e suas observações podem ajudar a responder algumas importantes questões tais como, por exemplo, como é a distribuição de metalicidade do halo da Galáxia e qual é o limite de metalicidade das estrelas do campo do halo.

O J-PLUS está em andamento no OAJ, usando um telescópio de 83 cm (Javalambre Auxiliary Survey Telescope, JAST, painel inferior esquerdo da figura 3.2) e a T80Cam, uma câmera panorâmica de 9,2k × 9,2k pixels que fornece um campo de visão de 2 deg² com uma escala de pixel de 0,55 arcsec pix⁻¹ (MARÍN-FRANCH *et al.*, 2015) (painel inferior direito da figura 3.2). O sistema de filtros J-PLUS é composto por 12 bandas fotométricas, sendo 4 deles filtros de banda larga tipo SDSS (g, r, i, z), 6 de banda intermediária com uma largura entre 200-400 Å (u, J0395, J0410, J0430, J0515 e J0861), e dois de banda estreita (J0378 e J0660) como mostram a Tabela 3.1 e Figura 3.4.

A estratégia observacional J-PLUS, redução de imagem e principais objetivos científicos são apresentados em CENARRO *et al.* (2019). Alguns objetivos científicos na área de astrofísica estelar e galáctica que vale a pena destacar neste trabalho, são:

- Exploração do halo galáctico e extensão das estimações de extinção interestelar (A_V)

¹www.j-plus.es

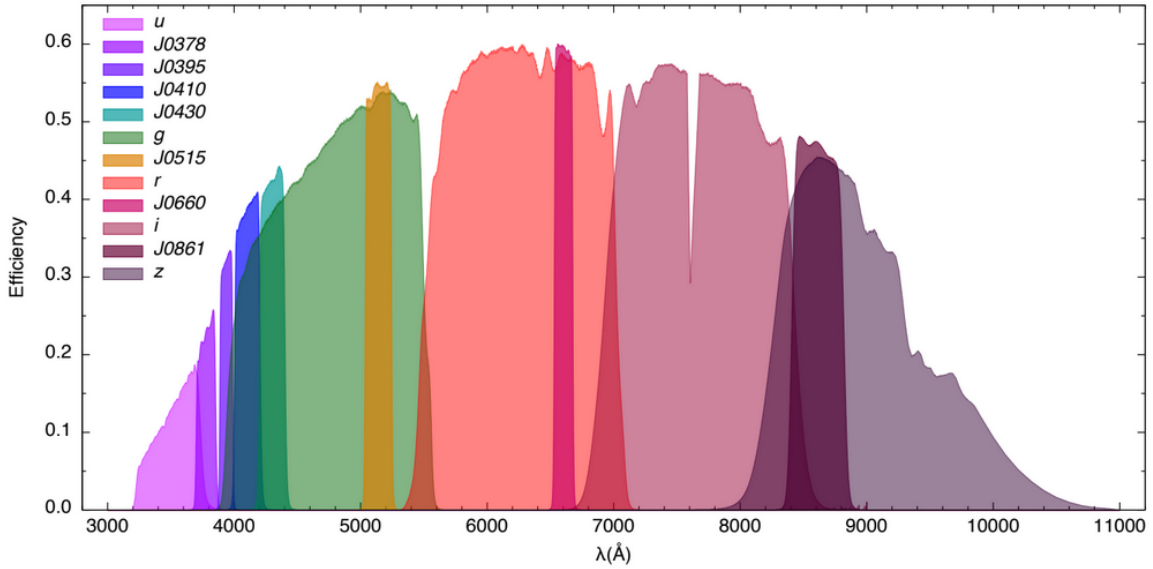


Figura 3.4: Sistema de filtros J-PLUS. [CENARRO *et al.* \(2019\)](#)

e temperatura efetiva (T_{eff}) nos limites de baixa luminosidade e metalicidade, onde o Gaia perde eficiência.

- Estimação precisa de parâmetros estelares como T_{eff} , $[\text{Fe}/\text{H}]$, $\log g$, $[\text{C}/\text{Fe}]$ e $[\text{Mg}/\text{Fe}]$ com ajuda dos filtros J0395, J0430, e J0515.
- Busca por novas estrelas UMPs ($[\text{Fe}/\text{H}] < -4.0$). Até o momento apenas foram identificadas ~ 40 estrelas nessa faixa de metalicidade ([PLACCO *et al.*, 2015](#); [SESTITO *et al.*, 2019](#)).
- Detecção e caracterização de anãs brancas (WDs).

Atualmente o J-PLUS conta com três eventos de liberação de dados (Data Releases, EDR, DR1 e DR2) disponíveis para o público geral. O próximo evento de liberação de dados (DR3) está previsto para julho de 2022. Segundo o último relatório disponível em www.j-plus.es/datareleases/data_release_dr2, já foram observados da ordem de 19.5 milhões de objetos no modo dual com magnitude $r \leq 21$. O segundo lançamento de dados do J-PLUS (DR2) compreende dados de 1088 apontamentos, cobrindo 2176 deg^2 do céu (Figura 3.5), observados e reduzidos em todas as bandas do levantamento ([VARELA e J-PLUS COLLABORATION, 2021](#)). A calibração fotométrica também tem sido aprimorada de forma progressiva em cada liberação de dados. Por exemplo, para o J-PLUS DR1 foi aplicado o método do locus estelar de anãs brancas dependente da metalicidade, enquanto que para o J-PUS DR2 foi aplicado um método alternativo baseado no locus estelar de estrelas da sequência principal levando em conta a influência de $[\text{Fe}/\text{H}]$ ([LÓPEZ-SANJUAN *et al.*, 2019a](#); [LÓPEZ-SANJUAN *et al.*, 2021](#)). As magnitudes limite (5σ , abertura de 3 segundos de arco, sistema AB) do DR2 são ~ 21 mag para todas as

bandas fotométricas. A **Point Spread Function** (PSF) e a **Full Width at Half Maximum** (FWHM) nas imagens DR2 na banda r é de 1,1 segundos de arco. A detecção das fontes foi feita na banda r usando o software **SExtractor** (BERTIN e ARNOUITS, 1996) e a medição de fluxo nas 12 bandas J-PLUS foi realizada na posição das fontes detectadas usando a abertura definida na banda r . O DR2 está disponível publicamente no site do J-PLUS².

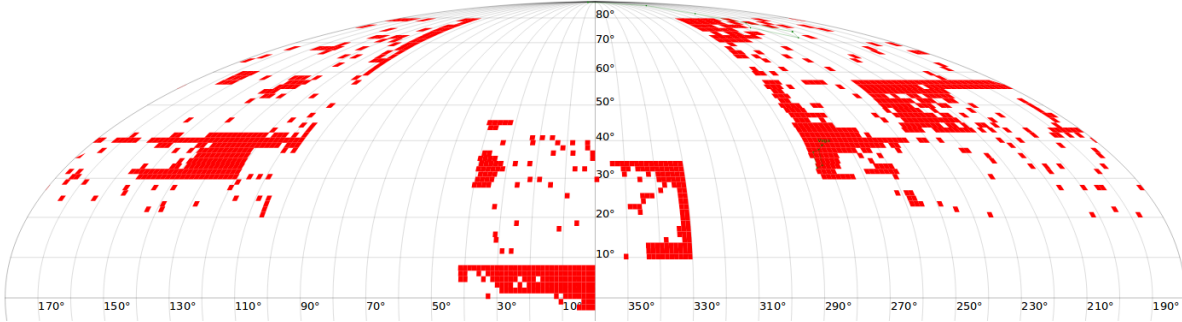


Figura 3.5: Footprint do J-PLUS DR2.

http://www.j-plus.es/datareleases/data_release_dr2

Tabela 3.1: Sistema de Filtros J-PLUS.

Filtro	Comprimento de Onda Central (Å)	FWHM (Å)	Features Principais (Comentários)
u	3 485	508	Em comum com J-PAS
J0378	3 785	168	O II, em comum com J-PAS
J0395	3 950	100	Ca II H&K
J0410	4 100	200	H δ
J0430	4 300	200	Banda G
g	4 803	1 409	
J0515	5 150	200	Tripleto Mg Ib
r	6 254	1 388	
J0660	6 600	145	H α , em comum com J-PAS
i	7 668	1 535	
J0861	8 610	400	Tripleto Ca II
z	9 114	1 409	

A fim de selecionar a amostra para nosso estudo e garantir medidas de alta qualidade em cada um dos 12 filtros, foram escolhidas as magnitudes medidas com abertura de 6 arcsec, incluindo correção de abertura, e uma configuração adequada nos parâmetros de controle de qualidade dos dados, como `MASKS_FLAGS = 0` (indicando que o alvo não está dentro de uma máscara), `FLAGS = 0` (nenhum artefato SExtractor detectado) e `NORM_WMAP_VAL > 0,8` (indicando exposição adequada). O avermelhamento interestelar foi corrigido usando o vetor de correção de extinção A_x disponível no banco de dados J-PLUS DR2. Considerando a classificação morfológica estrela/galáxia baseada na análise da fun-

ção de densidade de probabilidade (PDF) proposta por [LÓPEZ-SANJUAN *et al.* \(2019b\)](#), foram selecionados todos os alvos com probabilidade de ser uma estrela $p_{\text{star}} > 0,9$. Além disso, foram escolhidas as estrelas mais brilhantes que $g < 18$, resultando em uma amostra de 746.531 objetos (doravante denominada *amostra de ouro*).

3.1.3 Mapeando o céu sul com S-PLUS

Um projeto semelhante ao J-PLUS, denominado Southern Photometric Local Universe Survey (S-PLUS; [MENDES DE OLIVEIRA *et al.* \(2019\)](#)), está sendo realizado no Observatório Interamericano Cerro Tololo (CTIO), no Chile. Este projeto duplica a instrumentação instalada para o J-PLUS, ou seja, um sistema de 12 filtros adaptado a um telescópio de 80 cm (T80-Sul neste caso), mas concentra-se na observação do hemisfério sul.

O S-PLUS tem o potencial de complementar a espectrofotometria de baixa resolução fornecida pelo Gaia, aproveitando seu exclusivo sistema de filtro misto de banda larga, intermediária e estreita. Outro ponto importante a mencionar é que a caracterização estelar e os redshifts fotométricos obtidos de levantamentos multibanda como o S-PLUS fornecerão um recurso valioso para verificação cruzada da calibração do LSST e outros levantamentos.

Conforme referido na subsecção anterior relacionada com o J-PLUS, o S-PLUS visa também abordar diferentes interesses de pesquisa entre os quais é interessante referir os seguintes:

- Busca por correntes estelares e subestruturas ainda desconhecidas no halo galáctico a partir da identificação de estrelas azuis de ramo horizontal (BHB) e blue stragglers.
- Identificação de nebulosas planetárias, estrelas simbióticas e aglomerados globulares.
- Determinação de parâmetros estelares, tais como T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, $[\text{C}/\text{Fe}]$ e muito provável $[\alpha/\text{Fe}]$ e $[\text{Mg}/\text{Fe}]$. Nesse sentido, já existe um modelo bastante promissor capaz de estimar T_{eff} , $[\text{Fe}/\text{H}]$ e $[\text{C}/\text{Fe}]$ com boa precisão com auxílio de algoritmos de aprendizado de máquina baseados em redes neurais ([WHITTEN *et al.*, 2019](#); [WHITTEN *et al.*, 2021](#)), aproveitando especialmente a sensibilidade dos filtros J0395 e J0430 (Figura 3.6). Um modelo alternativo será objeto de discussão como parte do desenvolvimento do projeto de doutorado apresentado neste documento.
- Do S-PLUS espera-se um catálogo muito maior de estrelas muito pobres em metais com uma taxa de sucesso alta quando validadas em estudos de abundâncias de alta resolução, especialmente aquelas com $[\text{Fe}/\text{H}] < -4.0$; ou UMPs segundo a nomenclatura de [BEERS e CHRISTLIEB \(2005\)](#).
- Estudo e caracterização de fontes variáveis.

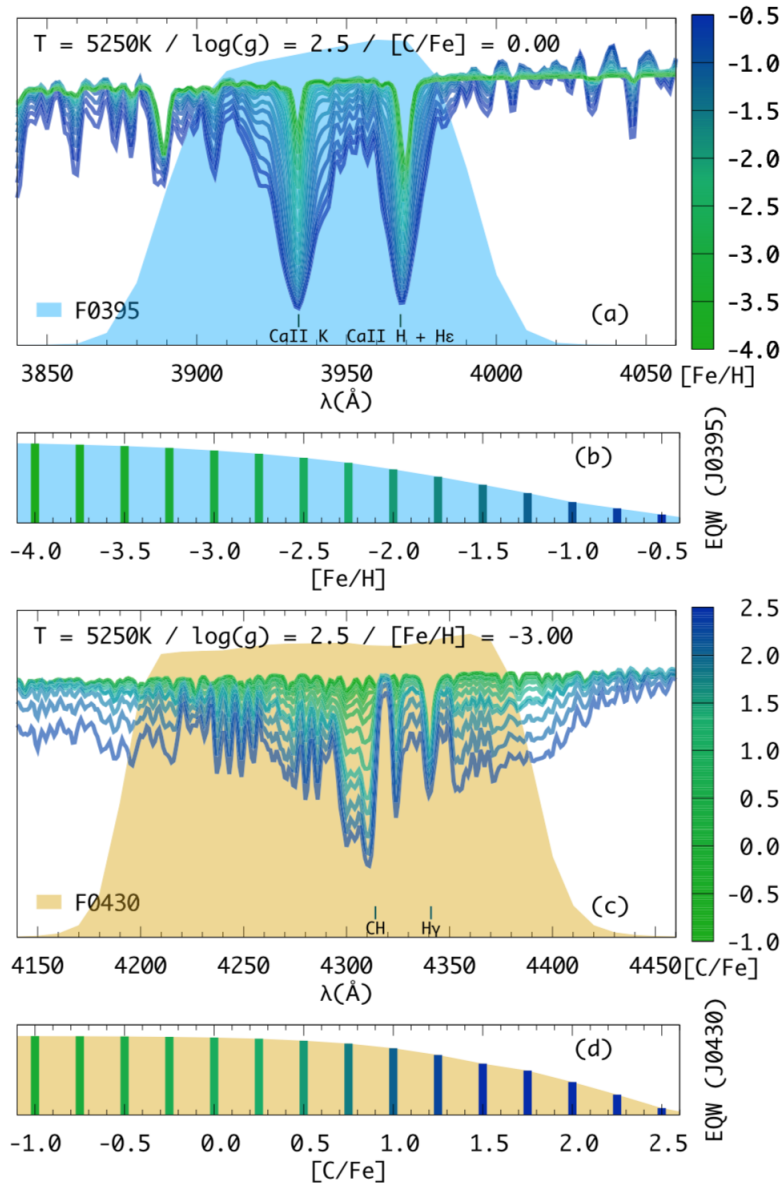


Figura 3.6: a) Curva de sensibilidade do filtro J0395, comparada com espectros sintéticos de diferentes metalicidades. (b) Comportamento do fluxo integrado na área J0395 para os espectros sintéticos mostrados em (a). (c) Curva de sensibilidade do filtro J0430, comparada com espectros sintéticos de diferentes abundâncias de carbono. (d) Comportamento do fluxo integrado na área J0430 para os espectros sintéticos mostrados em (c).

Creditos: [MENDES DE OLIVEIRA *et al.* \(2019\)](#) Fig. 10

Um dos grandes desafios referente à estimação de parâmetros e seleção de estrelas candidatas a muito baixa metalicidade tem a ver com o aumento da taxa de sucesso para encontrar estrelas com $[\text{Fe}/\text{H}] \leq -4,0$. Tradicionalmente, a identificação da maioria desse tipo raro de estrelas tem seguido a metodologia proposta por [BEERS e CHRISTLIEB \(2005\)](#) e referenciada na seção 2.6, que parte de uma pre-seleção de candidatas a partir de observações com espectrógrafos de baixa resolução conforme explicado nos trabalhos de [PLACCO *et al.* \(2013\)](#); [PLACCO *et al.* \(2016\)](#). Neste sentido, projetos tais como J-PLUS,

S-PLUS e J-PAS pretendem substituir a pre-seleção com espectroscopia de baixa resolução por informação da distribuição de energia espectral SED com sistemas multifiltro (banda estreita + banda larga). Outros levantamentos fotométricos, por exemplo, SkyMapper (KELLER *et al.*, 2007; WOLF *et al.*, 2018) e Pristine (STARKENBURG *et al.*, 2017) também aparecem como alternativas viáveis a espectrógrafos de baixa resolução, contando com limites de saturação levemente maiores, porém com um sistema de filtros de banda estreita menor que J-PLUS ou S-PLUS. Os primeiros resultados que exploram a taxa de sucesso que o S-PLUS indicam que o levantamento é capaz de identificar novas candidatas e estimar suas metalicidades e abundâncias de carbono, como apresentado em WHITTEN *et al.* (2021) e brevemente discutido na Figura 3.6 (Figura 10 do artigo de MENDES DE OLIVEIRA *et al.* (2019)) que mostra o efeito das linhas espectrais Ca II H+K, CH e H_γ nas curvas de sensibilidade dos filtros J0395 e J0430 respectivamente (painéis (a)-(d)). Nota-se que esses filtros são capazes de capturar mudanças até $\sim -3,0$ para $[\text{Fe}/\text{H}]$ e $\sim +0,5$ para $[\text{C}/\text{Fe}]$, sugerindo que os sistemas multifiltros de 12 bandas J-PLUS e S-PLUS são muito mais eficientes para a identificação de esse tipo de estrelas.

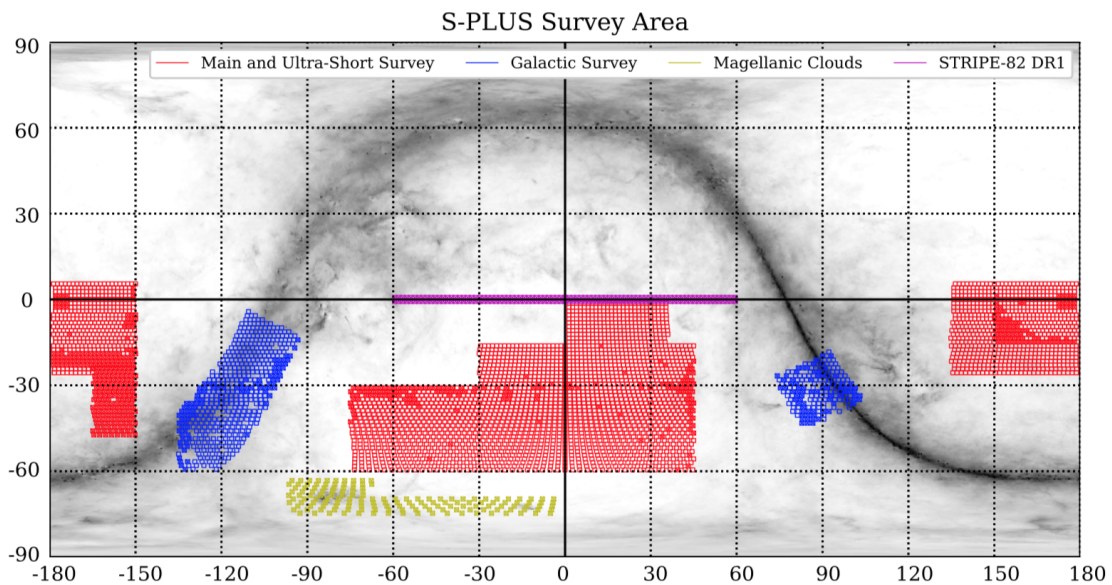


Figura 3.7: Footprint de três dos cinco sub-levantamentos S-PLUS, sobreposto ao mapa de extinção de SCHLEGEL *et al.* (1998) em projeção cartesiana. Os quadrados vermelhos mostram os levantamentos principal e ultracurto, que compartilham a mesma área. Os quadrados azuis mostram os campos galácticos. Os quadrados amarelos destacam a área das Nuvens de Magalhães, que estão incluídas no Levantamento Principal. As áreas preenchidas já foram observadas no momento da redação do artigo S-PLUS, em março de 2019. Magenta é a área do Stripe 82 contida no DR1 que faz parte da Pesquisa Principal, mas destacando-a com uma cor diferente para maior clareza.

Creditos: MENDES DE OLIVEIRA *et al.* (2019) Fig. 8

Para atingir os diferentes objetivos, as observações (footprint apresentado na figura 3.7) foram classificadas e organizadas em 5 sub-levantamentos diferentes, sendo eles:

1. O levantamento principal (Main Survey; MS - sigla em inglês). O MS cobrirá uma área de $\sim 8000 \text{ deg}^2$ com uma única observação de época de cada campo, por filtro, sob condições fotométricas e *seeing* de 0,8 a 2,0 segundos de arco. No momento da redação desta tese, existem três versões de dados disponíveis em <https://splus.cloud/documentation/datareleases>.
2. O levantamento de muita curta exposição (Ultra Short Survey; USS). Compartilha o mesmo footprint que o MS, mas com tempos de exposição que são 1/12 dos valores, o que significa que o limite de saturação é mais brilhante em todos os 12 filtros.
3. O levantamento de campos variáveis (Variable Fields Survey; VFS). O VFS realizará observações repetidas com uma cadência definida pela frequência de noites não fotométricas, abrangendo vários campos já observados pelo MS. Pelo menos 30 por cento do tempo total da pesquisa será dedicado ao VFS. Seus dados serão inspecionados em busca de novos asteróides e outros objetos em movimento.
4. O levantamento galáctico (Galactic Survey; GS). O GC cobrirá uma área de cerca de 1420 deg^2 no plano da Via Láctea em todos os 12 filtros, incluindo regiões do Bojo ($-10^\circ < l < 10^\circ$ e $-15^\circ < b < +5^\circ$, para um total de $\sim 400 \text{ deg}^2$) e o Disco ($220^\circ < l < 278^\circ$ e $-15^\circ < b < +5^\circ$, para um total de $\sim 1020 \text{ deg}^2$)
5. O levantamento do campo Marble (Marble Field Survey; MFS). O MFS é composto por um conjunto de campos específicos que serão revisitados com a maior frequência possível sob noites escuras ou cinzentas e condições fotométricas, quando a visão é muito ruim para observações do MS, ou seja, maior que 2,0 segundos de arco.

Nesse contexto, dado que as candidatas do MS serão mais fracas que $r = 12 \text{ mag}$, devido aos efeitos de saturação, o S-PLUS USS está sendo desenvolvido para encontrar candidatas a estrelas brilhantes de baixa metalicidade adequadas para acompanhamento espectroscópico de alta resolução e estudos no ultravioleta próximo usando o Telescópio Espacial Hubble. Estudos de acompanhamento já foram feitos para um número limitado de estrelas brilhantes de baixa metalicidade (PLACCO *et al.*, 2014, 2015). De importância central, o S-PLUS USS fornecerá alvos para estudos espectroscópicos de alta resolução subsequentes necessários para separar as subclasses UMP e CEMP.

3.2 Sloan Extension for Galactic Understanding and Exploration - SEGUE

O levantamento SEGUE (YANNY *et al.*, 2009, Sloan Extension for Galactic Understanding and Exploration) faz parte da segunda fase do levantamento Sloan (SDSS-II) e consiste em um conjunto de 250.000 espectros de estrelas de média resolução observadas

em 3.500 graus^2 do céu do hemisfério norte, excluindo regiões em baixa latitude galáctica ($|l| < 30^\circ$). A cobertura de comprimento de onda está entre 3900\AA e 9000\AA em $R \sim 1800$, com a maioria das estrelas observadas apresentando tipos espectrais de A a M. Em resumo, SEGUE tomou espectros de estrelas na faixa de magnitude $14 < g < 20,3$, uma vez que para $g \sim 14$, os espectrógrafos SDSS saturam em uma exposição de 300 s. Espectros de objetos mais fracos que $r = 18,5$ podem ser obtidos com relação sinal-ruído $(S/N) > 30$, suficiente para velocidade radial com incerteza de 4 km s^{-1} e medições de metalicidade $[\text{Fe}/\text{H}]$ com precisão de 0,2 dex para uma ampla variedade de tipos espectrais (A–K).

Para a seleção de alvos foi usada a cor fotométrica $u - g$ de estrelas tipo F, G, e K com a condição adicional de $(g - r)_0 < 0,8$ lembrando a ideia de que as estrelas mais azuis tendem a ter menor metalicidade (LENZ *et al.*, 1998). Além disso, outro item a ser considerado é que enquanto as estrelas anãs K e M de baixa metalicidade vivem por muito mais tempo na sequência principal do que as estrelas F e G, a magnitude limite da amostra dessas estrelas é dominada por estrelas de disco que estão próximas do sol. Como as estrelas F mais evoluídas são significativamente mais brilhantes (elas podem ser 1 – 2 mag mais brilhantes do que a luminosidade da sequência principal de idade zero) e, portanto, podem ser vistas a distâncias maiores, muitas das candidatas de mais baixas metalicidade $[\text{Fe}/\text{H}] < -3,0$ identificadas até o momento foram encontrados com cores de estrelas F evoluídas. Uma discussão mais aprofundada sobre algumas características de estrelas F no ponto de desligamento da sequência principal localizadas no halo é feita por PRIETO *et al.* (2008b) e CAROLLO *et al.* (2007).

O software que automatiza a redução e análise chama-se SSPP (SEGUE Stellar Parameter Pipeline), e estima os parâmetros estelares e as abundâncias de elementos selecionados dos espectros SEGUE. Ele emprega uma combinação de várias técnicas, como ajuste espectral (ALLENDE PRIETO *et al.*, 2006a) e minimização de χ^2 dentro de grades de modelos atmosféricos, análise da linha K do Ca II, funções de autocorrelação (BEERS *et al.*, 1999), calibrações de para o tripleto do Ca II (CENARRO *et al.*, 2001a,b), implementação de redes neurais artificiais treinadas com espectros observados e sintéticos (FIORENTIN *et al.*, 2007) e previsões semiempíricas usando a cor $g - r$ (válido apenas para T_{eff}).

Como parte do plano de observação SEGUE observou 30.998 candidatos a estrelas com baixa metalicidade. Cerca de 12% (4600) apresentam $[\text{Fe}/\text{H}] < -2,0$. Cerca de 0,1% (32) possuem $[\text{Fe}/\text{H}] < -3,0$. As candidatas de menor metalicidade precisarão ser acompanhadas em telescópios maiores em uma resolução mais alta. A Figura 3.8 mostra um conjunto de espectros de estrelas com temperaturas efetivas semelhantes e metalicidades variando de $[\text{Fe}/\text{H}] < -3,0$ a super-solar. O SEGUE produziu vários resultados importantes, incluindo a identificação de estrelas em $[\text{Fe}/\text{H}] < -3,0$ (AOKI *et al.*, 2012; PLACCO *et al.*, 2015), o estudo da estrutura do disco espesso Galáxia e halo (DE JONG *et al.*,

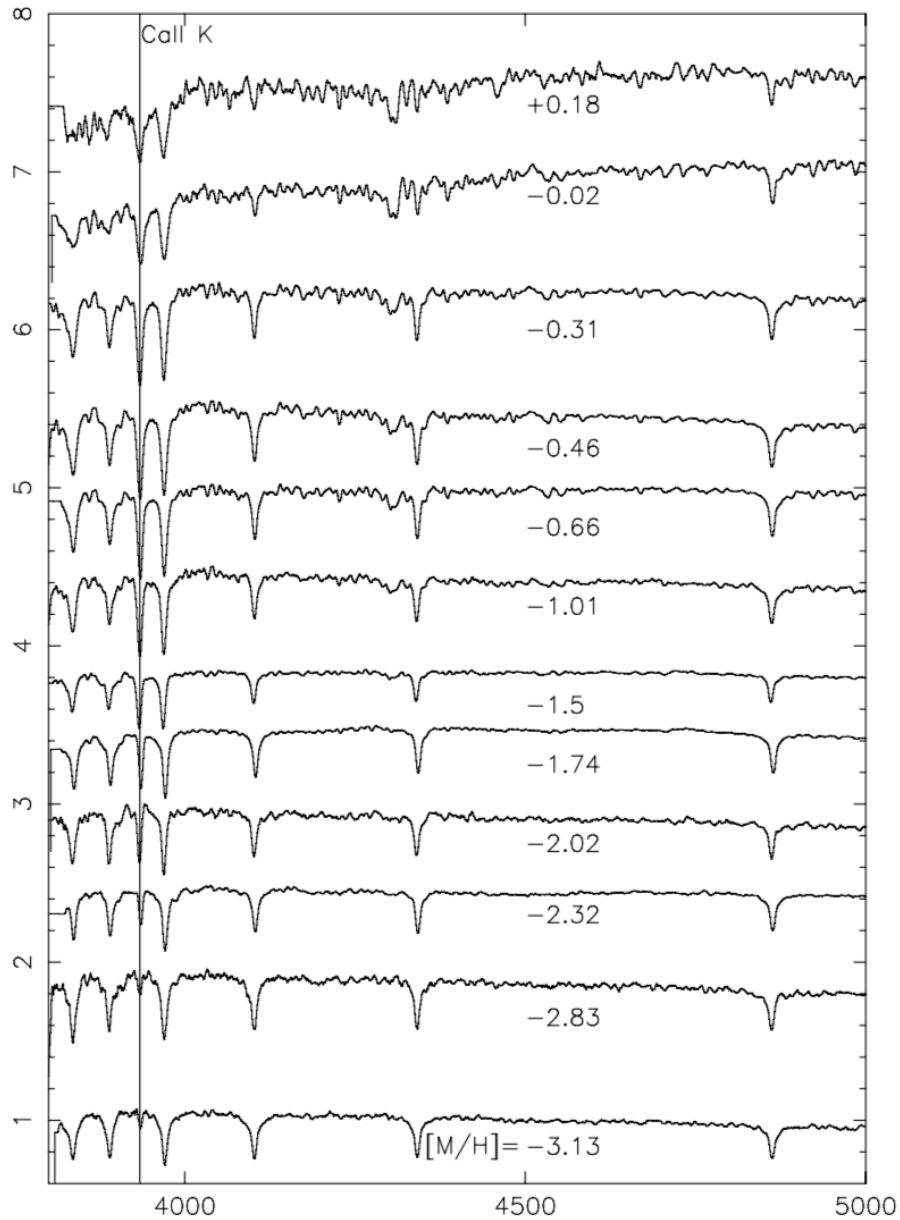


Figura 3.8: Sequência de metalicidade de estrelas tipo F: um conjunto de espectros de estrelas F, selecionadas para mostrar a faixa de metalicidades. Todas as 13 estrelas têm temperaturas efetivas semelhantes, perto de 6500 K, mas a intensidade da linha K do Ca II em $\lambda 3933$ indica metalicidades variando de menos de 0,001 até 1,5 vezes o valor solar. Credits: [YANNY *et al.* \(2009\)](#) Fig. 7

2010; [KIM *et al.*, 2019, 2021](#); [LEE *et al.*, 2017, 2019](#)), a análise da razão $[\alpha/\text{Fe}]$ de anãs G do disco galáctico ([LEE *et al.*, 2011](#)), e o estudo da razão carbono-ferro de estrelas da Via Láctea ([LEE *et al.*, 2013](#)).

3.3 LAMOST Experiment for Galactic Understanding and Exploration - LEGUE

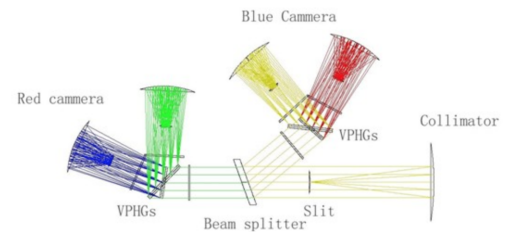


Figura 3.9: Equipamentos do LAMOST: *Esquerda*: Telescópio Schmidt localizado na Estação Xinglong do Observatório Astronômico Nacional, China. *Direita*: Sistema de 16 espectrógrafos de baixa resolução do levantamento LAMOST. Cada um dos 16 espectrógrafos pode acomodar até 250 fibras e podem obter espectros de 4.000 objetos celestes simultaneamente. A faixa de comprimento de onda de cada espectrógrafo é de 370nm a 900nm, com resolução espectral entre 1000 e 5000 dependendo da configuração das grades e câmeras utilizadas.

Créditos: <http://www.lamost.org/public/instrument?locale=en>.

LEGUE (DENG *et al.*, 2012) é um levantamento espectroscópico da estrutura galáctica dentro do projeto LAMOST (CUI *et al.*, 2012; ZHAO *et al.*, 2012). O espectrógrafo multi-objetos usado no projeto permite obter espectros com $R \sim 1.800$ para 4.000 alvos simultaneamente. Em última análise, LEGUE fornecerá uma amostra de 5 milhões de estrelas com cobertura limitada do céu. Um dos principais objetivos científicos da LEGUE é a busca de EMPs.

A partir dos dados para um grande número de estrelas do levantamento espectroscópico conduzido pelo telescópio de 4 metros de abertura LAMOST, é possível obter parâmetros relacionados com a cinemática estelar e a função de distribuição de metalicidade (MDF) na Galáxia. Tais resultados podem fornecer importantes restrições aos modelos atuais da estrutura galáctica, história de formação, evolução cinemática e dinâmica e evolução química na Via Láctea.

O LAMOST Stellar Parameter Pipeline (LASP; WU *et al.*, 2014) foi aplicado para estimar parâmetros estelares de dados LAMOST, usando o Software de Análise Espectroscópica da Université de Lyon (ULYSS; KOLEVA *et al.*, 2009; WU *et al.*, 2011) em conjunto com a biblioteca ELODIE. Este código foi testado e aplicado com sucesso para estimar velocidades radiais e parâmetros estelares para os dados do LAMOST. Os catálogos de parâmetros estão disponíveis com os lançamentos de dados oficiais (LUO *et al.*,

2015, Data releases). O pipeline de parâmetros estelares LAMOST na Universidade de Pequim LSP3 (XIANG *et al.*, 2015) é um pipeline alternativo útil para calcular parâmetros estelares, que também são testados com dados do LAMOST. Ele usa a biblioteca MILES além de parâmetros médios que melhor correspondem aos modelos de espectro e valores gerados pela minimização χ^2 . O LAMOST é um projeto altamente influenciado pelo sucesso obtido pelo SDSS e oferece alternativas para complementar estimativas de T_{eff} , $[\text{Fe}/\text{H}]$, $\log g$ e $[\alpha/\text{Fe}]$ além de procurar fluxos cinemáticos em velocidade radial (RV) e posição que são contíguos no céu e conectar fragmentos de estruturas que de outra forma não seria possível associar uns aos outros.

O LAMOST atualmente está na oitava liberação de dados DR8, acessível somente para membros da colaboração, enquanto o DR7 ² está disponível para a comunidade astronômica em geral. Alguns dados significativos a respeito do DR7 são:

- Catálogo disponível de 10.640.255 de espectros de baixa resolução, sendo 9.881.260 pertencentes a estrelas (dos quais 8.922.696 possuem $S/N > 10$ nas bandas g ou i). Neste catálogo é possível encontrar estimativas de T_{eff} , $[\text{Fe}/\text{H}]$, $\log g$ obtidas pelo pipeline LASP.
- Um total de 6.159.427 espectros observados de estrelas com tipos AFGK e 711.936 a estrelas tipo M. Adicionalmente existe um catálogo de 601,186 estrelas tipo A mais quentes.
- Catálogo adicional de 3.854.244 espectros de média resolução, sendo 2.260.387 espectros com $S/N > 10$. Neste catálogo é possível encontrar estimativas de T_{eff} , $[\text{Fe}/\text{H}]$, $\log g$ e abundâncias de $[\text{C}/\text{Fe}]$, $[\text{N}/\text{Fe}]$, $[\text{O}/\text{Fe}]$, $[\text{Mg}/\text{Fe}]$, $[\text{Al}/\text{Fe}]$, $[\text{Si}/\text{Fe}]$, $[\text{S}/\text{Fe}]$, $[\text{Ca}/\text{Fe}]$, $[\text{Ti}/\text{Fe}]$, $[\text{Cr}/\text{Fe}]$, $[\text{Ni}/\text{Fe}]$, $[\text{Cu}/\text{Fe}]$ a partir da aplicação de um modelo de redes neurais convolucionais (CNN no inglês) explicado no website oficial³.
- Do total de espectros observados, 198.393 correspondem a galáxias e 66.406 correspondem a QSOs.

3.4 GALactic Archeology with Hermes - GALAH

GALAH é um levantamento espectroscópico conduzido pelo telescópio anglo-australiano (AAT em inglês) de 3.9m junto com espectrógrafo HERMES, capaz de atingir uma resolução $R \sim 28000$ e uma observação simultânea de 392 estrelas. Da mesma forma que em outros projetos, GALAH visa esclarecer o cenário de formação da nossa Galáxia, contribuindo para o aumento de resultados de composição química e dinâmica de estrelas, consideradas um dos seus constituintes principais. Vale destacar que, devido à dificuldade

²<http://dr7.lamost.org/v1.3/catalogue>

³<http://dr7.lamost.org/v2.0/doc/mr-data-production-description>

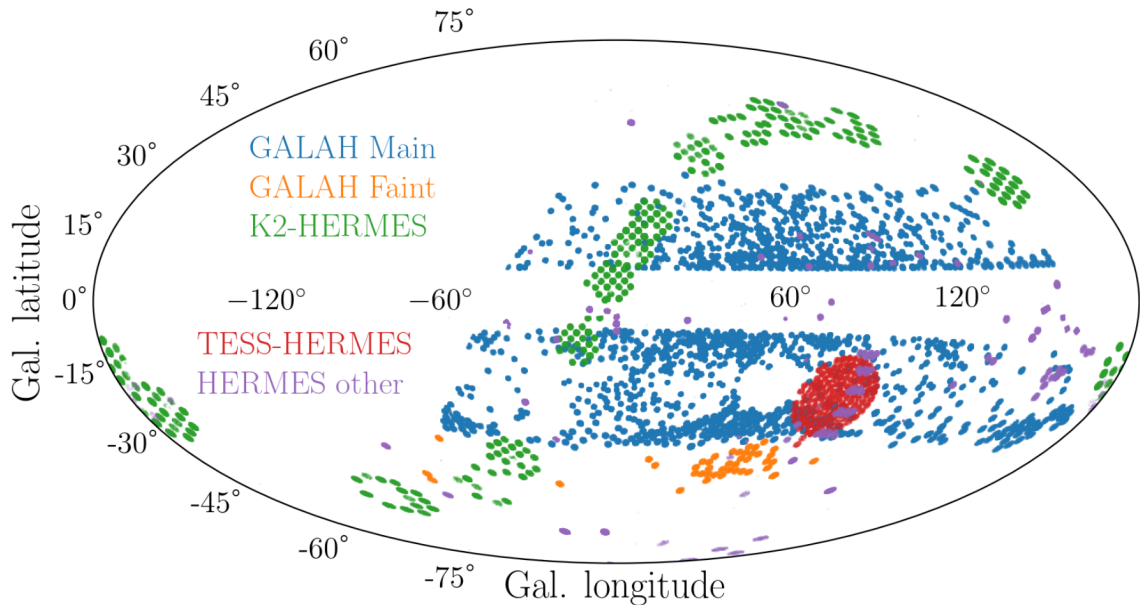


Figura 3.10: Visão geral da distribuição de estrelas observadas em coordenadas galácticas com o centro da Galáxia na origem. São mostrados os alvos principais (azul) e fracos (laranja) do GALAH, que evitam o plano galáctico. Os alvos do acompanhamento K2-HERMES (verde) se enquadram nas campanhas K2 ao longo da eclíptica e mostram o padrão característico do telescópio Kepler. As observações do TESS-HERMES (vermelho) estão focadas na Zona de Visualização Contínua Sul do TESS. Outros alvos HERMES (roxo) estão distribuídos pelo céu e foram observados durante vários programas independentes. Credits: [BUDER *et al.* \(2021\)](#) Fig. 1

de estimar com precisão a distância das estrelas, até poucas décadas atrás só era possível fazer estudos detalhados da vizinhança solar. Porém, graças ao aprimoramento de técnicas de observação multi-objeto e à missão Gaia ([GAIA COLLABORATION *et al.*, 2016, 2018](#)), atualmente existem ferramentas adicionais para o estudo da formação e evolução da Via Láctea.

No momento da escrita dessa tese, o GALAH encontra-se sua terceira liberação de dados (DR3), cuja distribuição espacial das observações pode ser apreciada na Figura 3.10. Para a geração do DR3, se utilizou informação combinada de fotometria como as magnitudes absolutas do levantamentos Two Micron All Sky Survey (2MASS⁴), no infravermelho, e Gaia, no óptico, junto a movimentos próprios aportados pelo levantamento UCAC4⁵ ([ZACHARIAS *et al.*, 2013](#)) para quebrar degenerescências e vieses que ocorrem quando apenas certas linhas de absorção são consideradas para a estimativa de parâmetros estelares. Quanto à precisão atingida na estimativa desses parâmetros, vale mencionar que a temperatura efetiva T_{eff} é estimada exclusivamente pelo melhor ajuste do espectro sintético sobre o modelo observado sem considerar calibrações fotométricas. Por outro lado, $\log g$ é estimada por uma relação semi-empírica com a Massa M , T_{eff} , e a lumino-

⁴<https://irsa.ipac.caltech.edu/Missions/2mass.html>

⁵<https://irsa.ipac.caltech.edu/data/UCAC4/ucac4.html>

sidade bolométrica (ver equação 1, [BUDER *et al.* \(2021\)](#)) e mostra um bom acordo com os valores reportados pelo **Gaia FGK benchmark stars (GBS)** e algumas delas também com valores estimados por asterosismologia. Finalmente para $[\text{Fe}/\text{H}]$, a estimação é feita também com ajuste de espectros sintéticos sobre espectros observados usando grades MARCS.

Os dados produzidos pelo GALAH DR3 tem beneficiado a compreensão dos exoplanetas através de suas estrelas hospedeiras graças à disponibilidade de parâmetros estelares aprimorados tal como apresentado nos estudos de [CLARK *et al.* \(2021\)](#), desenvolvimento de simulações mais realistas sobre formação de planetas como as desenvolvidas por [BITSCH e BATTISTINI \(2020\)](#) e a possibilidade de explorar estrelas hospedeiras de exoplanetas em um contexto quimio-cinâmico ou -dinâmico tal como discutido por [CARRILLO *et al.* \(2020\)](#).

3.5 The Apache Point Observatory Galactic Evolution Experiment - APOGEE

De uma forma similar a SEGUE, o APOGEE ([MAJEWSKI *et al.*, 2016](#); [MAJEWSKI *et al.*, 2009](#); [PRIETO *et al.*, 2008a](#)) é um outro projeto dentro do grande projeto Sloan que faz parte da fase III (SDSS-III). Neste caso, o catálogo consiste de 657.000 espectros de alta resolução ($R \sim 22,500$) e com razão sinal-ruído típica $S/N \geq 100$. As observações concentram-se no infravermelho e os espectros são registrados em três CCDs, nos intervalos espectrais $1.647 - 1.696 \mu\text{m}$, $1.585 - 1.644 \mu\text{m}$, $1.514 - 1.581 \mu\text{m}$, como mostra a figura 3.11. As observações foram feitas com o Sloan Foundation Telescope at the Apache Point Observatory, nos Estados Unidos. Este experimento teve uma continuação dentro da fase IV do SDSS (APOGEE-2), que passou a contar também com observações do hemisfério Sul a partir do du Pont Telescope, Las Campanas Observatory, no Chile.

Atualmente, o APOGEE encontra-se na liberação de dados numero 17. No caso deste levantamento, a estimação de parâmetros estelares é feita com o pipeline ASCAP ([GARCÍA PÉREZ *et al.*, 2016](#); [HOLTZMAN *et al.*, 2018](#); [JÖNSSON *et al.*, 2020](#)) que faz um ajuste geral do espectro observado a partir de uma biblioteca de espectros sintéticos calculados com o código FERRE ([ALLENDE PRIETO *et al.*, 2006b](#)), mais um segundo ajuste em algumas linhas específicas para obter as abundâncias químicas de um conjunto de 18 elementos (C, N, O, Na, Mg, Al, Si, S, K, Ca, Ti, V, Cr, Mn, Fe, Co, Ni, e Ce), dependendo da qualidade do espectro observado. Embora os dados produzidos pelo APOGEE possam

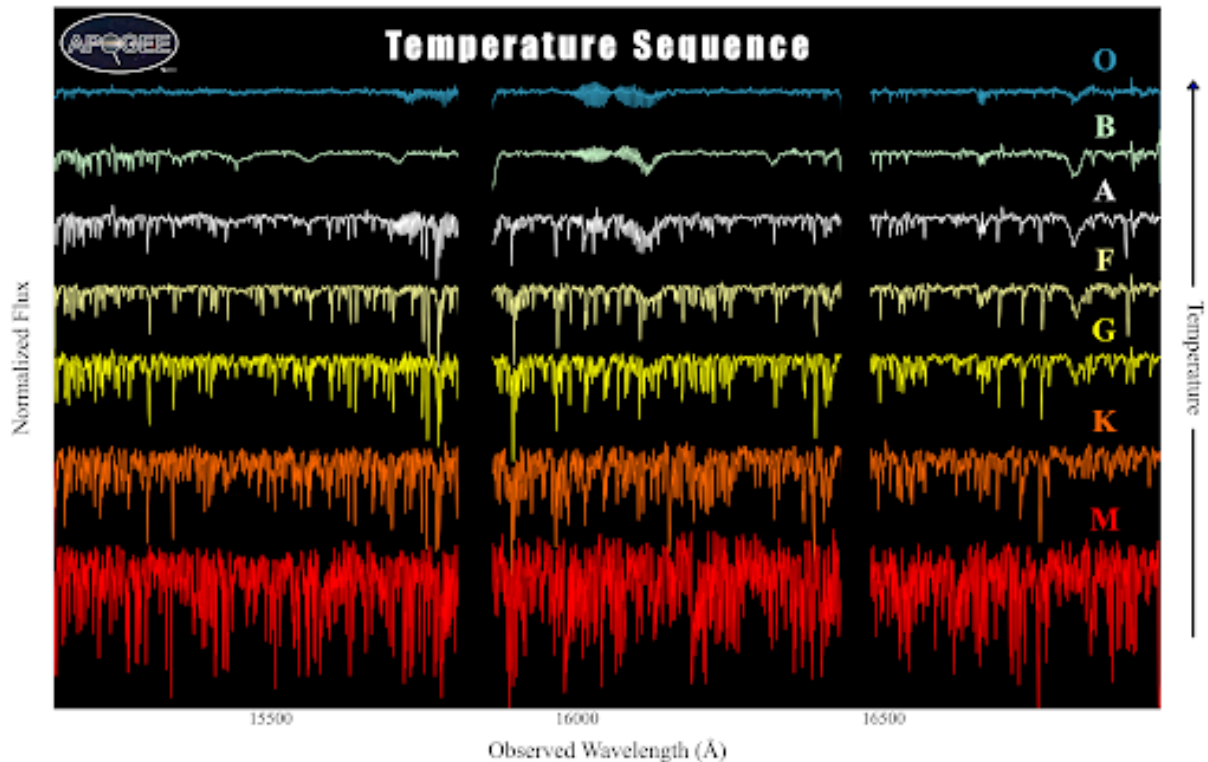


Figura 3.11: Exemplo de alguns espectros observados pelo APOGEE para diferentes tipos espectrais em três bandas diferentes. Créditos: https://www.sdss3.org/images/apogee/temp_sequence.png

3.6 Comentários Finais

Neste capítulo, se apresentaram alguns dos levantamentos fotométricos e espectroscópicos mais conhecidos e com dados disponíveis para serem usados em diferentes pesquisas observacionais. Todas as subseções apresentadas neste capítulo resumem alguns dos detalhes técnicos e objetivos científicos centrais, todos devidamente referenciados com discussões mais aprofundadas nos artigos fonte.

Um dos objetivos principais é mostrar e aproveitar ao máximo as vantagens do levantamento fotométrico multifiltro J-PLUS, especialmente com alguns filtros únicos entre os quais destacam-se J0378, J0395, J0515 e J0861 como fortes indicadores de $[\text{Fe}/\text{H}]$ e $\log g$ junto às cores $g - i$ e $g - r$, bem conhecidas como bons indicadores de T_{eff} desde o levantamento fotométrico SDSS. A ideia de trazer alguns desses pontos nesse capítulo e fazer um contexto curto porém adequado de como a combinação de diferentes fontes de dados permite criar modelos de estimação de parâmetros estelares basados em algoritmos modernos de aprendizado de máquina, tal como será discutido no capítulo seguinte.

Capítulo 4

Metodologia: O Modelo ML SPEEM

Nesta seção se apresenta uma discussão sobre alguns algoritmos de aprendizado de máquina, tais como **decision trees (árvores de decisão)**, **random forest** e **extreme gradient boost - XGB**, mostrando como eles foram aproveitados para desenvolver o pipeline SPEEM, que permite estimar parâmetros estelares usando dados do levantamento fotométrico J-PLUS.

Em particular, SPEEM é uma ferramenta que consiste num conjunto de modelos de aprendizado de máquina desenvolvidos a partir de uma abordagem de treinamento supervisionado, começando com uma subamostra de objetos com parâmetros previamente conhecidos, como classificação estrela/QSO, tipo espectral (A, F, G, K...), temperatura efetiva T_{eff} , gravidade superficial $\log g$ e metalicidade $[\text{Fe}/\text{H}]$. O processo de treinamento explora as relações estatísticas entre os parâmetros de entrada e as grandezas medidas, como magnitudes e índices de cor com as variáveis que se pretendem estimar, ou seja, os parâmetros físicos estelares. Uma discussão estendida de técnicas de aprendizagem supervisionada por máquina está disponível em [KOTSIANTIS *et al.* \(2007\)](#). Em particular, essas relações nos permitem prever os parâmetros estelares das estrelas da amostra de ouro.

4.1 Métodos de aprendizado de máquina e aplicações na astronomia

Os métodos de aprendizado de máquina têm como objetivo ajudar a processar um grande volume de dados de uma forma eficiente, permitindo a busca por padrões ou relações matemáticas entre um conjunto de variáveis de entrada e outro de saída.

Diversos projetos observacionais na astronomia, sejam levantamentos fotométricos ou espectroscópicos, constituem uma enorme fonte de dados que precisam ser analisados. Geralmente, o primeiro passo nessa análise consiste em resolver um problema de classificação

(como, por exemplo, separando estrelas de galáxias ou quasares a partir da informação de suas magnitudes fotométricas) ou regressão (caso da estimação de parâmetros físicos, tais como temperatura efetiva T_{eff} , metalicidade $[\text{Fe}/\text{H}]$ e gravidade superficial $\log g$). Justamente esses problemas são aqueles que os algoritmos de aprendizado de máquina pretendem resolver e resulta daí que a relação entre essa área das ciências da computação e a astronomia seja muito estreita.

Já que o objetivo deste trabalho não é fazer uma apresentação extensa de todos os algoritmos disponíveis, mas apenas manter o foco naqueles que foram usados no desenvolvimento desta pesquisa, vamos iniciar com uma introdução curta às árvores de decisão.

4.1.1 Árvores de decisão

As árvores de decisão são parte das ferramentas mais úteis e intuitivas usadas para implantar modelos de aprendizado de máquina para tarefas de classificação e regressão, geralmente chamadas de árvores de classificação e regressão (CARTs; BREIMAN *et al.*, 1984), que são fáceis de treinar e interpretar. No entanto, essas ferramentas tendem a sobreajustar em alguns casos, por exemplo, na aplicação a conjuntos de dados com classes não balanceadas, produzindo estimativas menos precisas (STROBL *et al.*, 2009).

Uma estrutura CART é hierárquica (nós de cima para baixo) e composta por um subconjunto da amostra de treinamento derivada recursivamente. Esse processo de divisão continua até que o subconjunto em um nó não seja mais estatisticamente significativo quando comparado à relação entre a entrada e as variáveis de destino. Para construir uma árvore de classificação, é necessário definir um parâmetro de ganho de informação a ser otimizado pelo algoritmo para obter estimativas precisas.

$$G(X_n) = \sum_k p_{nk}(1 - p_{nk}), \quad (1)$$

$$H(X_n) = \sum_k p_{nk} \log(p_{nk}), \quad (2)$$

Este parâmetro geralmente vem da impureza de Gini ou coeficientes de entropia de Shannon, expressos pelas Equações (1) e (2) respectivamente, onde X_n são os dados de treinamento no nó n e p_{nk} é a proporção dos valores observados da classe k no nó n . Para fins de regressão, o critério a ser minimizado geralmente é o erro quadrático médio (MSE), expresso por

$$\text{MSE} = \frac{1}{n} \sum_{i=0}^{n-1} \Delta^2, \quad (3)$$

Onde $\Delta = X_i^{\text{Target}} - X_i^{\text{Model}}$ é a diferença entre o valor de referência X_i^{Target} e o valor estimado pelo aprendizado de máquina X_i^{Model} , e n é o tamanho da amostra.

4.1.2 Random Forest e XGBoost

O algoritmo RF (random forest ou floresta aleatória) representa uma evolução das árvores de decisão, uma vez que combina um número arbitrário de árvores, cada uma delas alimentada com uma subamostra aleatória dos dados de entrada e equipada com subconjuntos aleatórios de recursos para evitar seleção enviesada da subamostra de controle e evitar o sobre-ajuste das estimações, oferecendo estimativas mais robustas. Por outro lado, o Extreme Gradient Boost (XGB) pode superar os modelos de RF, permitindo a possibilidade de penalizar as árvores no sentido de aprender os dados mais difíceis, usando uma combinação de computação paralela e sequencial (ou seja, com as últimas árvores ou floresta aprendendo com os erros cometidos por execuções anteriores).

Uma preocupação crítica no uso de árvores de decisão está na distribuição de parâmetros da amostra estudada. Por exemplo, a distribuição de metalicidade obtida no levantamento Pristine tem um pico em $[\text{Fe}/\text{H}] = -1.6$ para o halo Galáctico interno (YOUAKIM *et al.*, 2020) e o número de estrelas diminui para valores mais baixos de $[\text{Fe}/\text{H}]$ com uma inclinação de $\Delta(\log N)/\Delta[\text{Fe}/\text{H}] = 1.0 \pm 0.1$. Como estamos principalmente interessados em estrelas com $[\text{Fe}/\text{H}] < -2.2$, que são mais escassas do que a maior parte da distribuição, os conjuntos de dados de treinamento para diferentes intervalos de metalicidade são desequilibrados. A adoção de árvores de decisão em tais conjuntos de dados desequilibrados pode ser uma desvantagem, portanto é provável que a RF seja uma alternativa razoável. Apesar das RFs terem um bom desempenho na estimativa dos parâmetros, decidimos aplicar o algoritmo XGB ao processo de regressão dos parâmetros fotométricos. Isso resultou em uma melhoria significativa no tratamento de conjuntos de dados desequilibrados, como é o caso da distribuição de metalicidade do halo galáctico. As seções a seguir discutem a precisão e a exatidão obtidas pelo SPEEM.

4.2 Arquitetura do pipeline SPEEM

A arquitetura SPEEM, conforme mostrado na Figura 4.1, consiste em três partes: a seleção de recursos, redimensionamento dos recursos (normalização) e o regressor de múltiplas saídas, que fornece a saída final. Todos os algoritmos usados no desenvolvimento do SPEEM estão disponíveis no pacote `Scikit-learn` para a linguagem de programação Python (PEDREGOSA *et al.*, 2011).

A primeira parte do pipeline, a seleção de recursos, é ativada pela aplicação do módulo "união de recursos", que cria um conjunto personalizado de recursos composto de magnitudes fotométricas, cores e uma combinação de cores. Esses recursos são usados pelo pipeline como entrada para fazer a classificação do tipo espectral e a estimativa dos parâmetros estelares. Em seguida, a função "escalador padrão" ou **standard scaler**, transforma a distribuição dos parâmetros em distribuições normais centradas em torno

de 0, com um desvio padrão de 1, para evitar quaisquer vieses. O modelo considera a contribuição de cada variável sem pressupostos estatísticos prévios devido à diferença nos valores da escala. Finalmente, todos os recursos escalados são passados pela terceira parte do pipeline, o regressor de múltiplas saídas, que consiste em três regressores RF paralelos, cada um deles treinado para calcular um dos parâmetros estelares considerados.

Desta forma, SPEEM requer apenas uma execução para receber as entradas (as 12 magnitudes J-PLUS) e retornar T_{eff} , $\log g$ e $[\text{Fe}/\text{H}]$ como saídas.

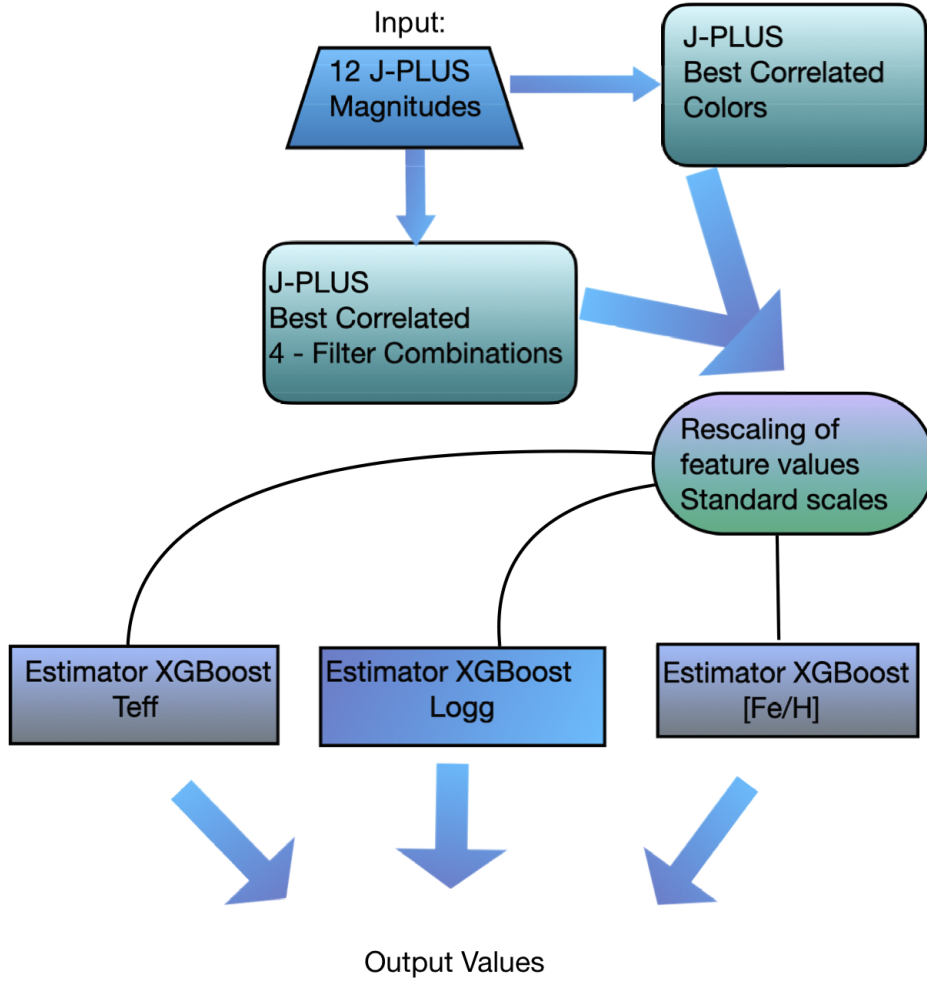


Figura 4.1: Estrutura do pipeline SPEEM composta de três partes: seleção das variáveis de entrada, normalização e aplicação do regressor multi-saída.

4.2.1 Amostras usadas no processo de treinamento

Para conseguir treinar o pipeline SPEEM é preciso uma amostra de treinamento bem definida. Neste sentido, a amostra de ouro J-PLUS foi cruzada com os levantamentos SEGUE e LAMOST DR5 para recuperar informações sobre parâmetros físicos e classificação morfológica para os alvos em comum. Esse processo produziu os seguintes conjuntos de dados:

- J-PLUS×SEGUE, com 6 794 alvos em comum, contém informações sobre parâmetros estelares (T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$) estimados pelo pipeline SSPP.
- J-PLUS×LAMOST, com 99 184 alvos com $S/N > 10$. A análise apresentada neste artigo está ancorada em estimativas de parâmetros atmosféricos obtidas com o pipeline n-SSPP (BEERS *et al.*, 2014; BEERS *et al.*, 2017), que é uma modificação do pipeline original SSPP do SEGUE, que faz uma estimativa de parâmetros usando uma grade de espectros com minimização de χ^2 para escolher o melhor ajuste ao espectro observado, fornecendo T_{eff} , $\log g$ e $[\text{Fe}/\text{H}]$.

4.2.2 Limpeza da amostra de fontes extragalácticas e anãs brancas

A estimativa de parâmetros estelares requer uma amostra de treinamento livre de contaminantes, como quasares (QSO) ou anãs brancas (WD). Esses objetos podem aumentar a taxa de candidatos falso-positivos quando se trata de pesquisar novos candidatos a VMP. Esses contaminantes não resolvidos podem imitar a estimativa fotométrica de estrelas VMP, conforme mostrado em trabalhos anteriores usando dados do SEGUE (DAWSON *et al.*, 2012) ou do Baryon Oscillation Spectroscopic Survey (BOSS; BOLTON *et al.*, 2012).

Para este propósito, construímos um classificador de dados usando como parâmetro objetivo o flag SPECTYPE_CLASS fornecido pelo pipeline SSPP do SEGUE para estimar o tipo espectral do objeto observado. Usamos o conjunto de dados J-PLUS × SEGUE como amostra de treinamento para fazer uma classificação de tipo espectral preliminar da amostra de ouro, permitindo-nos identificar e remover objetos tipo QSO ou WD que podem se assemelhar a candidatos VMP no espaço de cores J-PLUS.

A separação entre QSO e estrelas foi realizada a partir de um modelo de aprendizado de máquina treinado com as features fotométricas de maior importância relativa, selecionados por RF em um processo de duas etapas a partir de uma lista de 12 magnitudes J-PLUS e 66 cores J-PLUS.

Embora as magnitudes WISE (Wide-field Infrared Survey Explorer) estejam no regime infravermelho, enquanto as magnitudes J-PLUS correspondam ao óptico, a separação QSO-estrela de nossa amostra foi inspirada no diagrama $W1 - W2 \times W2 - W3$, apresentado por WRIGHT *et al.* (2010) e SCARINGI *et al.* (2013), como uma ferramenta para separar QSOs (com $W1 - W2 > 0,5$) e estrelas (com $W1 - W2 < 0,5$). A figura 4.2 apresenta o diagrama cor-cor para as estrelas da amostra J-PLUS×SEGUE que possuem dados obtidos com os filtros WISE. Os QSO estão representados por círculos rosas, localizados na região em que $W1 - W2 > 0,5$, enquanto as estrelas, representadas por diferentes cores, de acordo com o tipo espectral, apresentam $W1 - W2 < 0,5$.

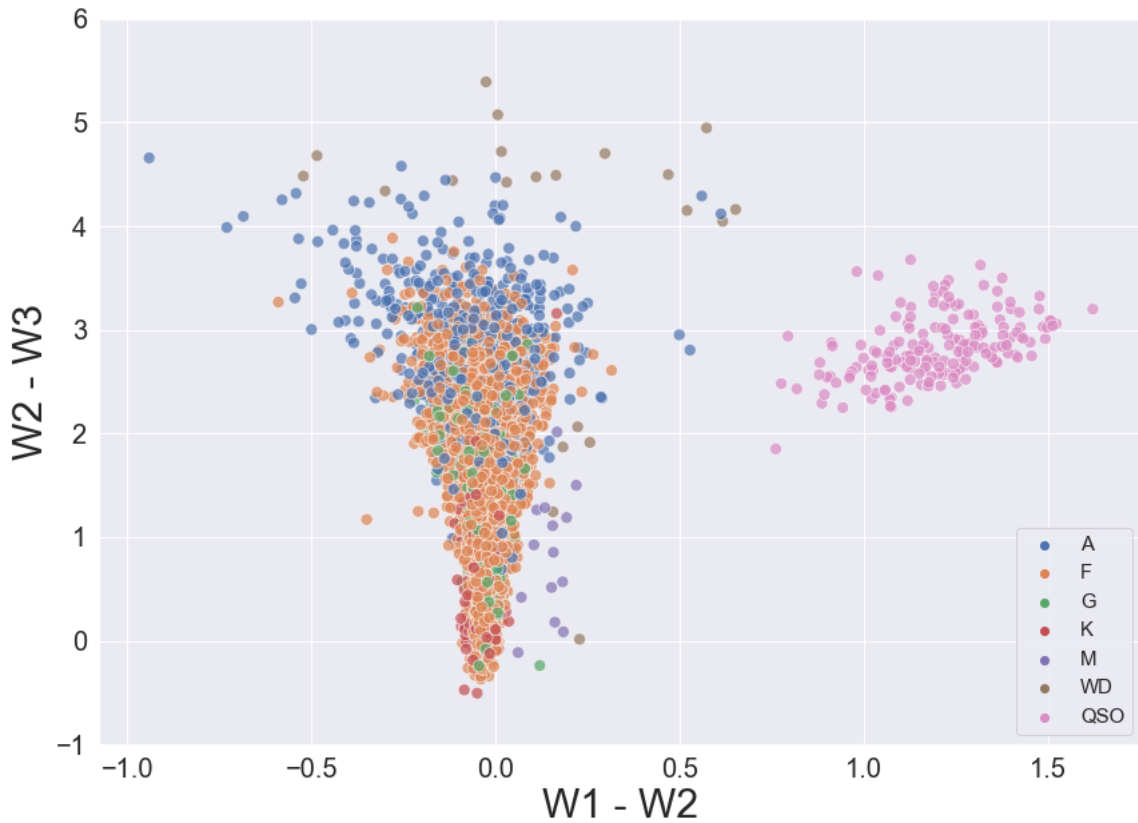


Figura 4.2: Diagrama cor-cor para os filtros WISE para separar QSOs (círculos rosas, com $W1 - W2 > 0,5$) e estrelas (com $W1 - W2 < 0,5$).

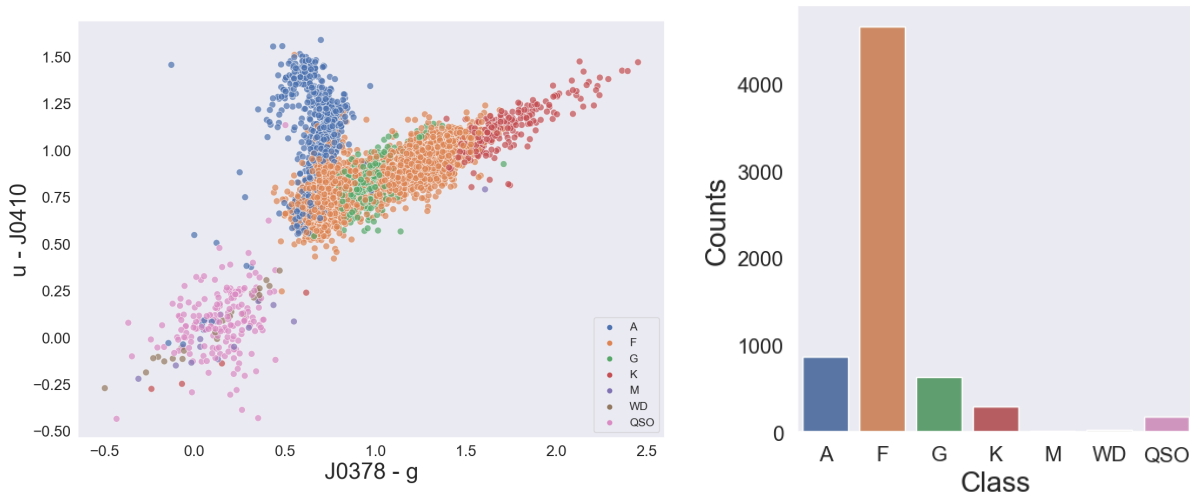


Figura 4.3: Classificação com SPEEM. *Esquerda*: separação estrela-QSO baseada nas cores J-PLUS $u - J0410$ e $J0378 - g$ para uma amostra de 6 794 estrelas. Uma linha em $u - J0410 = 0,5$ pode ser usada para separar os QSOs (círculos rosas) e as estrelas (círculos coloridos de acordo com o tipo espectral). *Direita*: Distribuição das classes (tipos espectrais estelares e QSO) designadas por SEGUE para a amostra estudada.

No sistema fotométrico J-PLUS, construímos um diagrama baseado nas duas cores essenciais e não correlacionadas, $(u - J0410) \times (J0378 - g)$, da lista de importância relativa das features usadas no algoritmo RF para separar QSOs de estrelas. Um exemplo

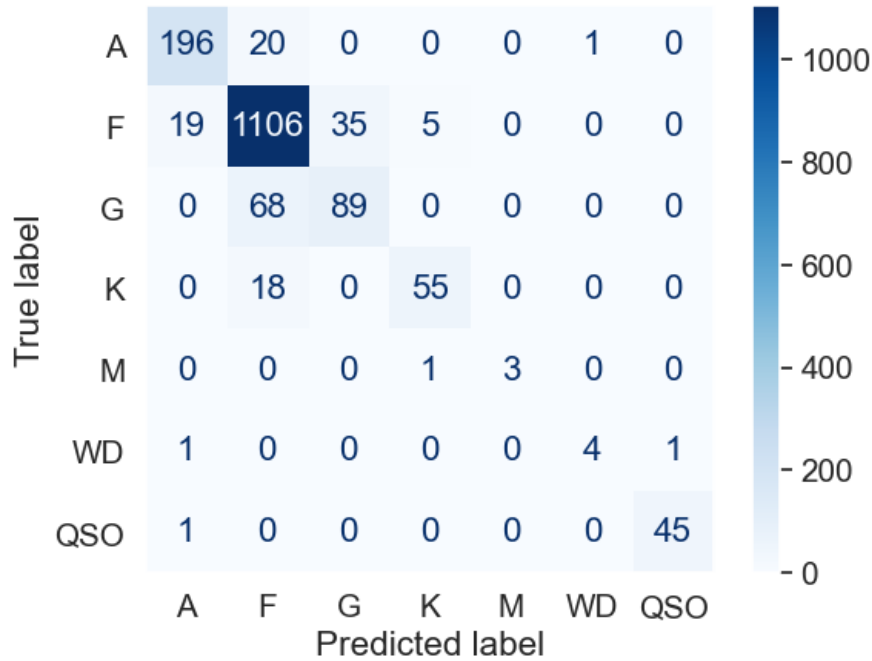


Figura 4.4: Matriz de confusão da classificação aplicada a 1668 objetos da amostra J-PLUS×SEGUE que não foram usados no processo de treinamento. A barra azul indica o número de objetos em cada classe.

desta separação estrela-QSO aplicada à amostra J-PLUS × SEGUE é mostrado no painel esquerdo da Figura 4.3, sugerindo que há um limite em $(u - J0410) = 0.5$ como uma restrição preliminar para separar QSOs (representados como círculos rosas) de estrelas, e círculos coloridos de acordo com o tipo espectral atribuído pelo SEGUE. O painel direito da Figura 4.3 mostra a distribuição das classes usadas nos dados de treinamento ingeridos no modelo.

Em seguida, a amostra J-PLUS × SEGUE foi dividida aleatoriamente em duas subamostras com uma proporção de 0,75/0,25, com um total de 5004 objetos para treinamento e 1668 objetos para testar as previsões. A Figura 4.4 apresenta a matriz de confusão do pipeline SPEEM para o processo de classificação, codificada por cores pelo número de objetos em cada classe. A precisão do classificador, ou a taxa geral de classificação correta, é de 0,91. Conforme mostrado na matriz de confusão apresentada na Figura 4.4, a taxa de identificação de QSOs e estrelas M foi de 98% e 75%, respectivamente. As estrelas do tipo espectral A, F e K (196 de 217, 1107 de 1165 e 55 de 73, respectivamente) apresentaram uma taxa de recuperação satisfatória. Por outro lado, a métrica mais baixa correspondente às estrelas G (88 de 157 estrelas classificadas corretamente) é provavelmente devido à classificação incorreta de estrelas do tipo F e K. Finalmente, a taxa de identificação de WDs foi em torno de 67%.

Nesta fase da nossa análise, estamos interessados em excluir os possíveis QSOs e WDs de nossa amostra, resultando em uma amostra quase pura, ideal para treinar SPEEM para estimar os parâmetros estelares.

4.3 Estimativa de parâmetros

4.3.1 Treinamento para regressões usando algoritmos de aprendizado de máquina

A estimativa de T_{eff} , $\log g$ e $[\text{Fe}/\text{H}]$ foi baseada em uma amostra de treinamento com 4299 estrelas com tipos espectrais F e G que foram selecionados aleatoriamente de J-PLUS \times SEGUE (após a remoção de QSOs, WDs e espectros de baixa razão sinal-ruído) logo de aplicar o classificador construído, conforme descrito na Seção 4.2.2. A amostra de treinamento apresenta uma distribuição de temperaturas efetivas entre 4800 K e 9000 K, gravidades superficiais entre 1.0 e 4.5 e metalicidades entre -3.0 e $+0.5$, estimadas pelo SEGUE, conforme mostrado nos painéis esquerdos das Figuras 4.5 - 4.7. A distribuição de T_{eff} (Fig. 4.5) é consistente com as temperaturas correspondentes aos tipos espectrais selecionados F e G, com possível contaminação de algumas estrelas A classificadas incorretamente, o que estende a distribuição de T_{eff} até 9000 K. A Figura 4.6 indica que as estrelas da sequência principal dominam a amostra de treinamento, mas também contém algumas estrelas evoluídas. A distribuição de metalicidade na Figura 4.7 sugere que a amostra de treinamento contém estrelas do disco espesso, correspondendo ao pico em $[\text{Fe}/\text{H}] \sim -0.6$, bem como estrelas do halo interno, correspondendo a $[\text{Fe}/\text{H}] \sim -1.6$ (CAROLLO *et al.*, 2010).

Neste trabalho, estamos principalmente interessados em derivar os parâmetros atmosféricos e em buscar novos candidatos para estrelas VMP, portanto a estimativa fotométrica de $[\text{Fe}/\text{H}]$ é crucial. No entanto, a determinação da metalicidade estelar com base em medições fotométricas é especialmente desafiadora na resolução dos dados J-PLUS: para uma dada temperatura, o contraste entre os perfis de absorção das linhas metálicas e o contínuo diminui significativamente em valores de metalicidade mais baixos. Portanto, é crucial definir uma amostra de treinamento cobrindo uma ampla faixa de valores de $[\text{Fe}/\text{H}]$. Por esta razão, o conjunto de dados J-PLUS \times SEGUE, que contém estrelas de $[\text{Fe}/\text{H}]$ entre -3.17 e 0.5 em uma distribuição bimodal, como mostrado na Figura 4.7, é mais adequado para treinar o modelo.

A análise de correlação estatística do conjunto completo de cores e todas as combinações possíveis de filtros J-PLUS com os parâmetros estelares T_{eff} , $\log g$ e $[\text{Fe}/\text{H}]$ nos permite selecionar o melhor conjunto de recursos (ou seja, aquelas features com o maior coeficiente de correlação de Pearson) para alimentar RF e obter as previsões mais precisas. O RF fornece a importância relativa atribuída aos parâmetros de entrada para cada grandeza estimada (como mostrado nos painéis direitos das Figuras 4.5 a 4.7) como um parâmetro para testar o desempenho do modelo e precisão das estimativas. A importância relativa dos indicadores corresponde à média de cada árvore de decisão dentro do modelo treinado (HASTIE *et al.*, 2009). Por exemplo, no caso de T_{eff} , as características

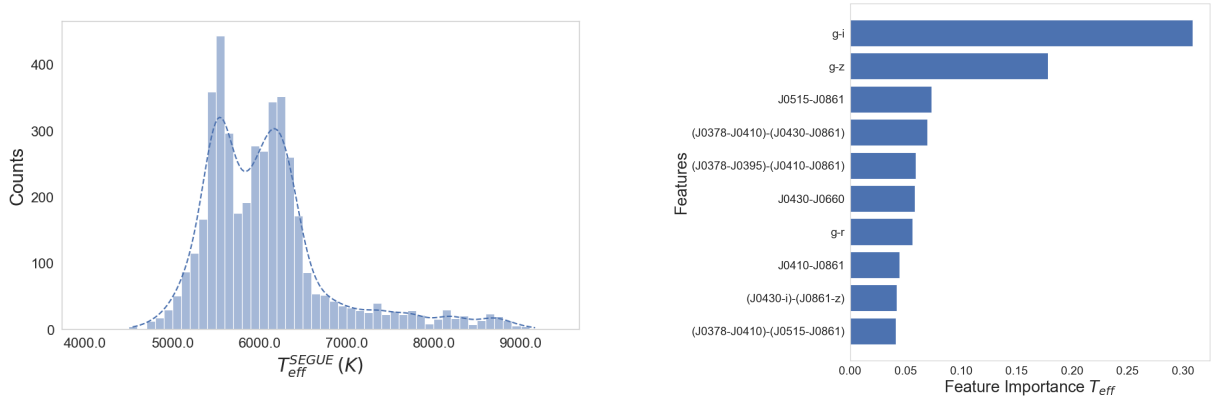


Figura 4.5: *Esquerda:* Distribuição do parâmetro T_{eff} para a amostra de treinamento J-PLUS \times SEGUE composta por 4 299 estrelas. *Direita:* Importância relativa das características espectrais para a estimação da T_{eff} . As características mais importantes como indicadores de T_{eff} são as cores $g-i$, $g-z$, e a cor J0515–J0861 baseada nos filtros de banda estreita que contém os tripletos de Mg I e de Ca II.

essenciais incluem as cores $g-i$ e $g-z$ e a cor J0515 – J0861, com base nos filtros de banda estreita centrados no tripleto Mg I e no tripleto Ca II, respectivamente. Para $\log g$, o melhor indicador é a cor (J0378 – J0410) – (J0430 – J0861), que contém os perfis de H δ , banda G e o tripleto Ca II. Finalmente, para $[\text{Fe}/\text{H}]$, o índice (J0395 – J0430) – (J0660 – z), baseado em filtros contendo as características espectrais das linhas Ca II H&K, da banda G e de H α , destaca-se como bom indicador de metalicidade.

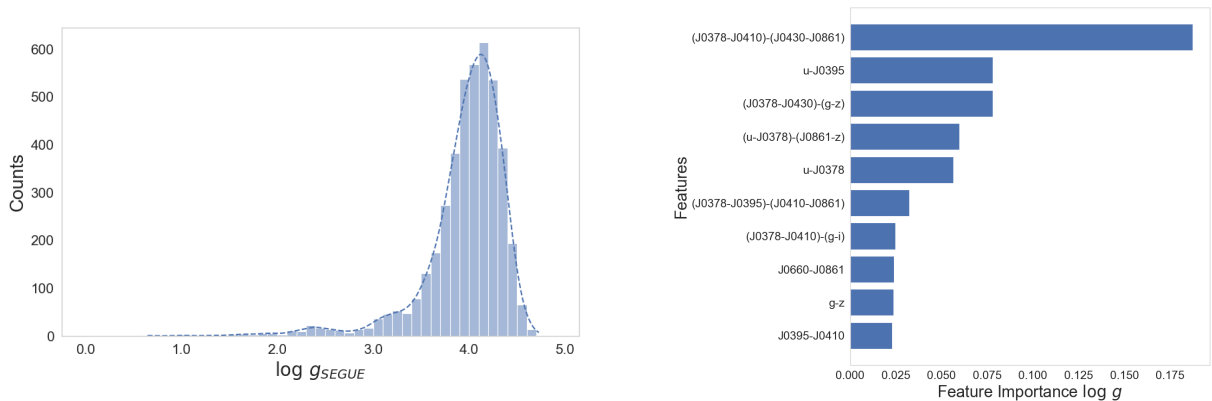


Figura 4.6: *Esquerda:* Distribuição do parâmetro $\log g$ para a amostra de treinamento J-PLUS \times SEGUE composta por 4 299 estrelas. *Direita:* Importância relativa das características espectrais para a estimação da $\log g$. A característica mais importante como indicador de $\log g$ é a pseudo-cor (J0378–J0410)–(J0430–J0861)

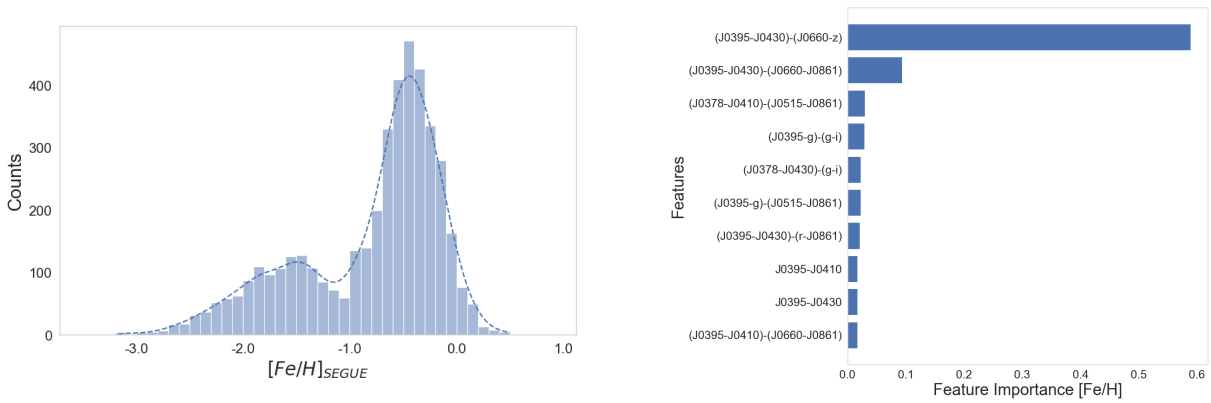


Figura 4.7: *Esquerda:* Distribuição do parâmetro $[Fe/H]$ para a amostra de treinamento J-PLUS \times SEGUE composta por 4299 estrelas. *Direita:* Importância relativa das características espectrais para a estimação da $[Fe/H]$. A característica mais importante como indicador de $[Fe/H]$ é a pseudo-cor $(J0395-J0430)-(J0660-z)$ baseada no filtros de banda estreita que contém os perfis de Ca II H & K, G-band, and $H\alpha$ e o filtro de banda larga z.

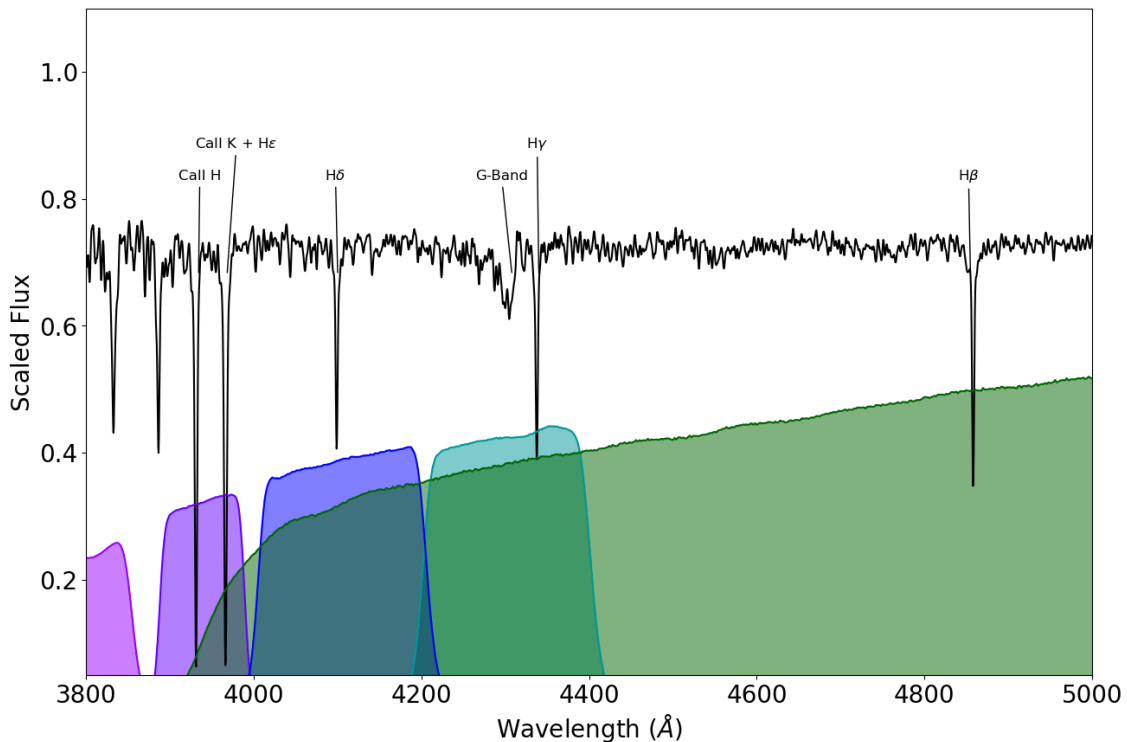


Figura 4.8: Alguns filtros J-PLUS sobrepostos ao espectro da estrela tipo G J-PLUS 75091-15989, com $[Fe/H] = -3.09$. Os filtros J0395 e J0410 contém os perfis de Ca II H & K e a linha de hidrogênio $H\delta$, respectivamente. As linhas $H\gamma$ e a banda G são visíveis no filtro J0430.

A Figura 4.8 mostra os filtros J-PLUS J0395, J0410, J0430 e g sobrepostos ao espectro observado de uma estrela G (no capítulo 5 se abordaram mais detalhes sobre as observações e redução dos dados espectroscópicos das candidatas selecionadas), com a identificação de algumas características espectrais essenciais como as linhas Ca II H&K, as linhas Balmer $H\beta$, $H\gamma$ e $H\delta$, além da banda G. Essas características espectrais, no entanto, podem ser sensíveis a mais de um parâmetro atmosférico simultaneamente, introduzindo uma degenerescência na determinação dos parâmetros. Por exemplo, a Figura 4 de WHITTEN *et al.* (2019) exibe a sensibilidade das linhas Ca II H&K com T_{eff} e $[\text{Fe}/\text{H}]$. Por outro lado, as linhas H são principalmente sensíveis à temperatura efetiva para estrelas com tipos espectrais posteriores a F. Assim, a análise baseada em uma combinação de filtros contendo características espectrais com diferentes sensibilidades aos parâmetros da atmosfera pode resolver esta degenerescência.

4.3.2 Validação da estimacão de parâmetros estelares

Depois que os diferentes modelos SPEEM concluíram o processo de aprendizagem, testamos o pipeline em uma subamostra de 1668 estrelas (correspondendo a 25% da amostra J-PLUS \times SEGUE não usada no processo de treinamento) para comparar os valores estimados por SEGUE (espectroscopia de resolução média) com os previstos por SPEEM como mostrado na Figura 4.9. A incerteza das previsões corresponde ao desvio absoluto mediano (MAD) e ao desvio absoluto normalizado (NMAD).

Os valores T_{eff} estimados pelo SPEEM estão em boa concordância com os estimados pelo pipeline SSPP do SEGUE (painel superior esquerdo da Figura 4.9), obtendo um erro médio de 41 ± 61 K. A maioria das estrelas na amostra de teste tem $T_{\text{eff}} < 7000$ K, uma vez que é altamente dominada por estrelas do tipo F (Seção 4.2.2, painel direito da Figura 4.3). O histograma no painel inferior esquerdo da Figura 4.9) representa a distribuição das diferenças entre o T_{eff} obtido com SSPP e SPEEM, e não mostra nenhuma diferença sistemática entre as duas escalas de temperatura.

O painel do meio da Figura 4.9 mostra a correlação entre os valores SSPP e as previsões SPEEM para $\log g$ com $\text{MAD} = 0.11 \pm 0.17$ dex. Os valores $\log g$ estimados por SPEEM concordam com aqueles derivados por SSPP dentro de 0.05 dex para estrelas da sequência principal. Por outro lado, a diferença média é 0,26 dex para estrelas mais evoluídas, com $\log g \leq 3.5$. Até mesmo o uso de cores diferentes como indicadores de gravidade superficial (Figura 4.6) não fornece uma precisão geral alta o suficiente para fazer estimativas mais precisas de $\log g$. O estudo da determinação fotométrica da gravidade superficial ancorada na asteroseismologia é discutida no capítulo 6 desta Tese.

Finalmente, o painel superior direito da Figura 4.9 apresenta uma correlação razoavelmente boa entre os valores de $[\text{Fe}/\text{H}]$ derivados por SSPP e SPEEM para $[\text{Fe}/\text{H}] > -1.0$. No entanto, há uma dispersão mais significativa no regime de menor metalicidade. Os

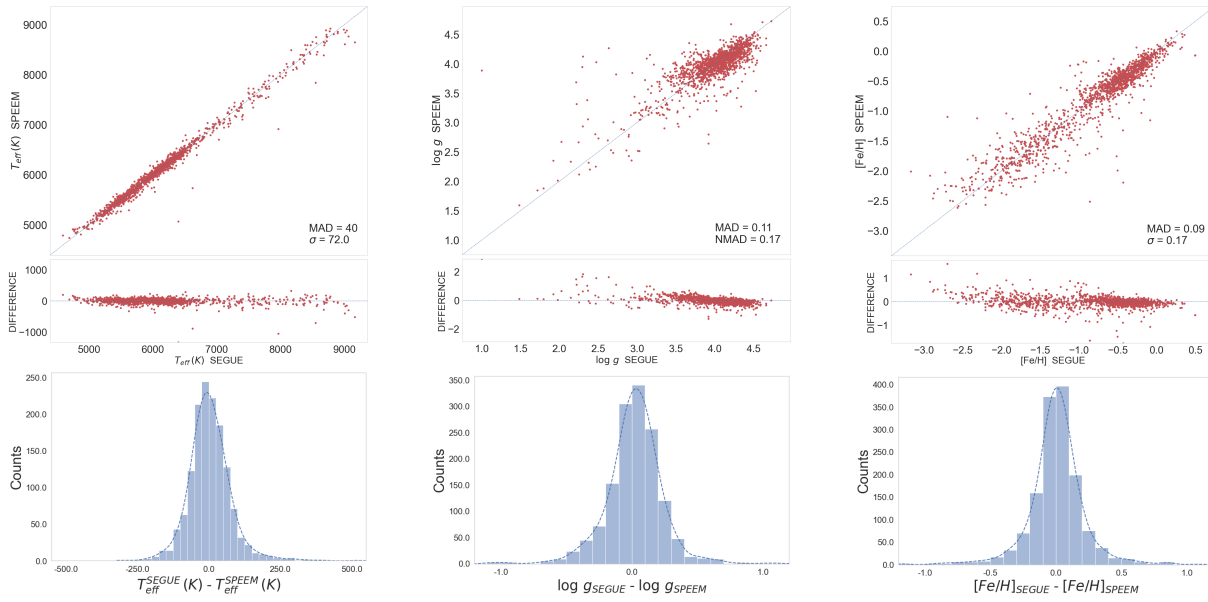


Figura 4.9: Comparação entre parâmetros T_{eff} (painel esquerdo superior), $\log g$ (painel do meio superior) e $[\text{Fe}/\text{H}]$ (painel direito superior) obtidos com os pipelines SSPP e SPEEM para 25% da amostra J-PLUS \times SEGUE. Os parâmetros MAD e NMAD são apresentados em cada painel. Os painéis inferiores mostram as distribuições das diferenças entre os respectivos parâmetros obtidos com SSPP e SPEEM. As linhas tracejadas dos painéis superiores marcam a guia de comparação para estabelecer quando o parâmetro estimado coincide com o parâmetro determinado pelo SEGUE. Nas distribuições das diferenças dos painéis inferiores as linhas tracejadas correspondem a ajustes gaussianos.

valores de MAD e NMAD são 0.09 ± 0.14 dex, e as diferenças mostram uma ligeira inclinação para $[\text{Fe}/\text{H}] < -2.0$, sugerindo que SPEEM pode superestimar o valor $[\text{Fe}/\text{H}]$ na faixa inferior da distribuição de metalicidade. Essa tendência provavelmente se deve à distribuição da metalicidade e à falta de dados necessários para um processo de aprendizado adequado nesta faixa de metalicidade.

Um teste adicional da capacidade de predição do SPEEM foi realizado com a amostra J-PLUS \times LAMOST (livre de QSOs, WDs e espectros com baixo S/N) descrita na seção 4.2.1. Os painéis superiores da Figura 4.10 mostram as comparações entre os parâmetros estelares estimados pelo SPEEM e aqueles obtidos aplicando o pipeline n-SSPP aos espectros LAMOST. A dispersão típica entre os valores de T_{eff} dos dois conjuntos de dados é inferior a 150 K. A dispersão é maior do que em comparação com SEGUE mostrado na Figura 4.9, embora o intervalo de T_{eff} se estende para temperaturas mais baixas, chegando a 4 500 K. A estimativa da gravidade superficial mostra uma maior dispersão com uma diferença média de $0,22 \pm 0,33$ dex enquanto a comparação para metalicidade tem $\text{MAD} = 0,09 \pm 0,13$. As distribuições do diferenças entre os parâmetros estimados pelo SPEEM e pelo n-SSPP para a amostra são mostradas nas respectivas painéis inferiores. A diferença $[\text{Fe}/\text{H}]_{\text{SPEEM}} - [\text{Fe}/\text{H}]_{\text{LAMOST}}$ mostra uma tendência com $T_{\text{eff,SPEEM}} - T_{\text{eff,LAMOST}}$ de forma que uma mudança de 100 K em T_{eff} produz uma variação de $\sim 0,2$ dex em $[\text{Fe}/\text{H}]$. Como

resultado, essa nova amostra de treinamento reforça o desequilíbrio na distribuição de metalicidade e introduz uma incerteza mais considerável na estimativa de baixas metalicidades. Por esse motivo, decidimos manter o J-PLUS×SEGUE como amostra de treinamento, pois em comparação J-PLUS×SEGUE com um maior número de estrelas com $[\text{Fe}/\text{H}] > -2,5$ na distribuição de metalicidade.

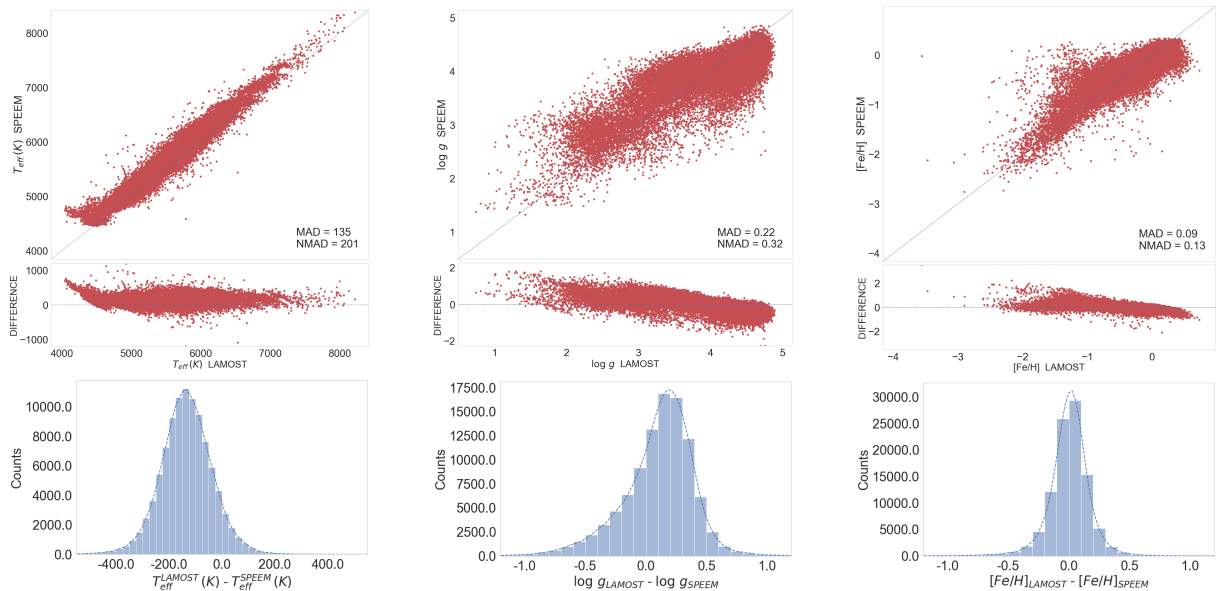


Figura 4.10: Comparação entre parâmetros T_{eff} (painel esquerdo superior), $\log g$ (painel do meio superior) e $[\text{Fe}/\text{H}]$ (painel direito superior) obtidos com os pipelines n-SSPP e SPEEM para 25% da amostra J-PLUS×LAMOST. Os parâmetros MAD e NMAD são apresentados em cada painel. Os painéis inferiores mostram as distribuições das diferenças entre os respectivos parâmetros obtidos com SSPP e SPEEM. As linhas tracejadas dos painéis superiores marcam a guia de comparação para estabelecer quando o parâmetro estimado coincide com o parâmetro determinado pelo SEGUE. Nas distribuições das diferenças dos painéis inferiores as linhas tracejadas correspondem a ajustes gaussianos.

4.4 Resumo e comentários finais

Neste capítulo se explica a necessidade de aproveitar algumas ferramentas computacionais tais como alguns algoritmos de aprendizado de máquina (RF e XGB) para conseguir desenvolver modelos estatísticos baseados em dados cruzados entre J-PLUS e outros levantamentos espectroscópicos (SEGUE e LAMOST) para conseguir obter estimativas muito mais precisas quando comparadas com calibrações fotométricas clássicas.

Os diferentes testes apresentados ao longo deste capítulo sugerem que o SPEEM é capaz de estimar T_{eff} , $\log g$ e $[\text{Fe}/\text{H}]$ com erros médios de 40 K, 0,11 e 0,09 respectivamente quando comparados com SEGUE (ver Figura 4.9). No seguinte capítulo se discute a aplicação do SPEEM à amostra de ouro inteira para gerar um catálogo estelar de parâmetros físicos (VAC) a fim de selecionar potenciais candidatas a VMPs e EMPs.

Capítulo 5

Busca por novas estrelas muito pobres em metais

Este capítulo apresenta as vantagens de buscar estrelas de baixa metalicidade a partir do uso de modelos de aprendizado de máquina para gerar catálogos de valores precisos e usá-los como fonte para caracterizar a distribuição de estrelas da amostra de interesse. Desta forma, os pedidos de tempo para obter espectros de média e alta resolução podem ser beneficiados e de fato ter maior chance de sucesso, não só na questão de validar os parâmetros obtidos a partir da fotometria de banda estreita, como também no estudo de abundâncias de outras espécies químicas, como o carbono. Para o caso específico deste trabalho, a validação independente do SPEEM consistiu em selecionar candidatas adequadas a serem observadas com espectroscopia de média resolução e comparar os valores dos parâmetros obtidos com as duas técnicas: aprendizado de máquina (ML no inglês) e espectroscopia. Neste caso, a validação espectroscópica das candidatas resultou na identificação de nove VMPs e uma EMP que ainda não possuíam determinações publicadas na literatura.

5.1 Aplicação do SPEEM ao levantamento J-PLUS

Depois de concluir o processo de validação descrito na seção 4.3.2, o SPEEM foi aplicado a toda a *amostra de ouro* (746 531 estrelas) para obter T_{eff} , $\log g$ e $[\text{Fe}/\text{H}]$. A Figura 5.1 mostra o diagrama de Hertzsprung-Russell para a *amostra de ouro*, com os parâmetros estelares T_{eff} e $\log g$ obtidos com o código SPEEM. Para facilitar a visualização dos dados, os pontos foram agrupados em várias sub-regiões hexagonais (hexbin plot) codificadas por cores de acordo com o número de estrelas no espaço de parâmetros T_{eff} e $\log g$, tal como indicado no eixo auxiliar $\log_{10}(N)$, com N o número de estrelas expresso em notação científica. O diagrama obtido neste caso sugere que a amostra de ouro está dominada por as estrelas anãs da sequência principal ($3,5 < \log g \leq 4,5$) e com tipos espectrais F e G

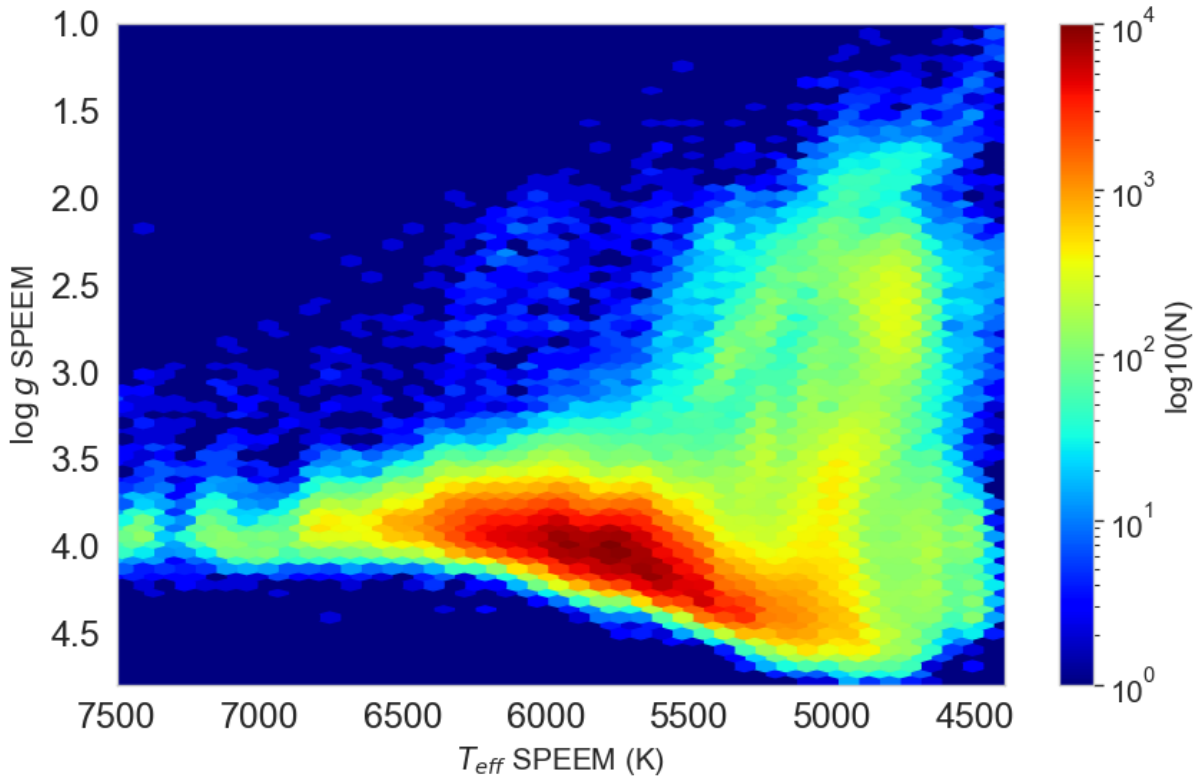


Figura 5.1: Diagrama HR da amostra de ouro usando os parâmetros estelares obtidos com SPEEM. A cor dos pontos está codificada de acordo com o número de estrelas, indicando que a amostra é composta principalmente por estrelas anãs da sequência principal com tipos espectrais F e G.

dados também pelo SPEEM ($5000 < T_{\text{eff}} \leq 6500$) (região vermelha da figura 5.1). Vale destacar que os valores estimados pelo SPEEM para a amostra de ouro correspondem com o espaço de parâmetros da amostra de treinamento, em outras palavras, o SPEEM, como qualquer outro modelo de aprendizado de máquina, faz um ótimo trabalho interpolando valores dentro de um certo intervalo de valores no conjunto de parâmetros, mas não consegue fazer boas extrapolações (ver Capítulo 4).

A Figura 5.2 mostra a distribuição de metalicidade obtida com o SPEEM para a amostra de ouro. As contagens no eixo y se encontram em escala exponencial para facilitar a visualização. O histograma do painel superior esquerdo apresenta uma bimodalidade com regiões de alta concentração de estrelas que podem ser relacionados aos dois picos das populações estelares do disco espesso em $[\text{Fe}/\text{H}] \sim -0,4$ e o halo em $[\text{Fe}/\text{H}] \sim -1,6$. Nota-se também uma forte queda no número de estrelas com $[\text{Fe}/\text{H}] < -2,5$, fato já estudado na literatura (LI *et al.*, 2010; SCHÖRCK *et al.*, 2009, entre outros), e que evidencia um grande desafio na estimação de $[\text{Fe}/\text{H}]$ a partir de cálcio presentes no filtro J0395 e em cores associadas como por exemplo $(J0395 - J0430) - (J0660 - z)$. No regime de altas metalicidades, o corte abrupto em $[\text{Fe}/\text{H}] \sim +0,4$ deve-se ao efeito de escala das contagens combinado ao fato do SPEEM não ter sido otimizado para estimar parâmetros de estrelas ricas em metais relacionado com o processo de ajuste de hiperparâmetros aplicado

nos algoritmos XGBoost implementados no SPEEM. A distribuição de metalicidade do halo obtida para a amostra de J-PLUS×SEGUE a mesma tendência amplamente discutida na literatura como o pico do halo interno encontrado por [CAROLLO *et al.* \(2007\)](#) a partir de um estudo de $\sim 20\,000$ estrelas associadas a diferentes componentes no halo galáctico. Graças aos recentes dados produzidos por Gaia DR2 e DR3 junto com J-PLUS é possível aprimorar o estudo da distribuição de metalicidade do halo galáctico usando $[\text{Fe}/\text{H}]$ de estrelas estimada por SPEEM junto com informação das distâncias estimadas por [BAILER-JONES *et al.* \(2021\)](#), especialmente a componente Z que nos permite determinar a posição da estrela respeito ao plano galáctico. As diferentes sub-figuras da Figura 5.2 mostram claramente um deslocamento do pico de $[\text{Fe}/\text{H}]$ na direção das metalicidades mais baixas à medida que aumenta a distância ao plano da Galáxia.

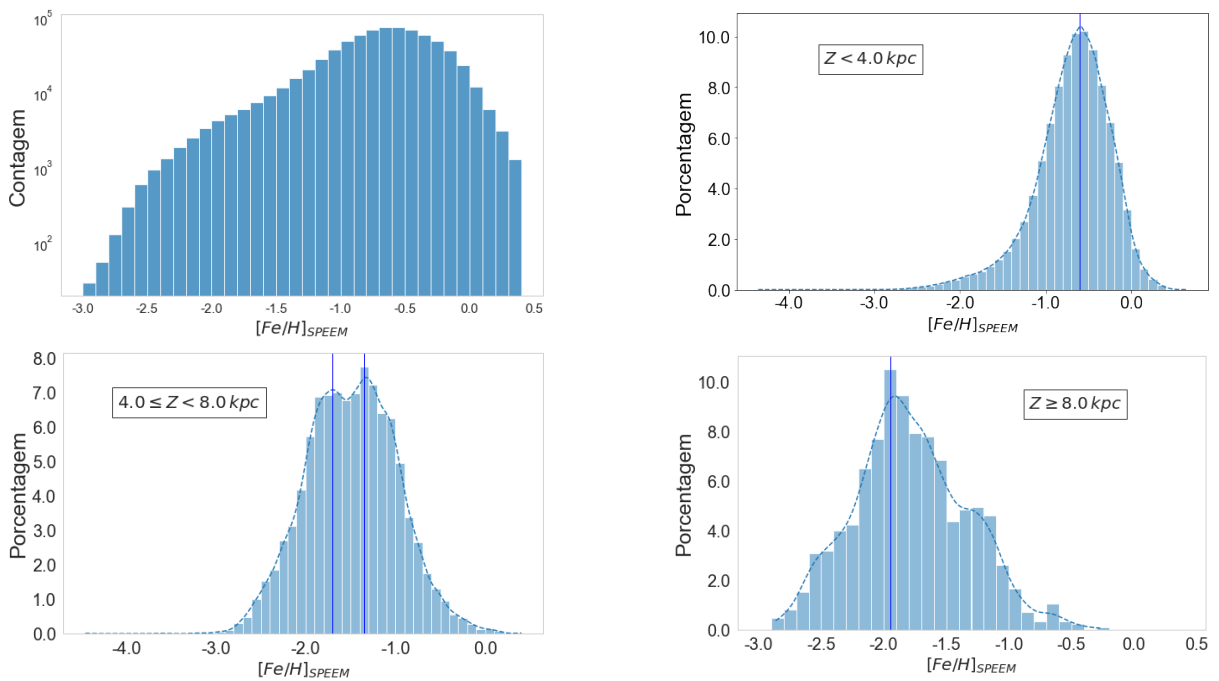


Figura 5.2: Distribuição de $[\text{Fe}/\text{H}]$ a diferentes distâncias. *Esquerda superior*: Distribuição de $[\text{Fe}/\text{H}]$ obtida com SPEEM para a amostra de ouro. A bimodalidade da distribuição esta relacionada aos dois picos (reportados na literatura, [CAROLLO *et al.* \(2010\)](#)) das populações estelares do disco espesso e do halo interno, sendo $[\text{Fe}/\text{H}] \sim -0,4$, e $[\text{Fe}/\text{H}] \sim -1,6$ respectivamente. *Direita superior*: Distribuição de $[\text{Fe}/\text{H}]$ para estrelas até 4 kpc do plano da Galáxia. *Esquerda inferior*: Distribuição de $[\text{Fe}/\text{H}]$ para estrelas da região central do halo interno da Galáxia. *Direita inferior*: Distribuição de $[\text{Fe}/\text{H}]$ para estrelas de regiões afastadas do halo interno da Galáxia. As linhas tracejadas correspondem a um ajuste de densidade de kernel da distribuição usando a função `displot()` da livraria Seaborn do Python.

5.2 Seleção e validação de novas candidatas

Para fazer a busca de novas candidatas a estrelas VMP/EMP, foi utilizada uma subamostra da amostra de ouro de 575 600 estrelas que não tinham sido identificadas previamente por outros levantamentos como SEGUE, APOGEE ou LAMOST. A fim de definir uma lista para o acompanhamento espectroscópico, foram levados em conta os seguintes critérios de seleção: $T_{\text{eff}} < 5500$ K (que corresponde a estrelas de tipos espectrais G, K e M) e $[\text{Fe}/\text{H}] < -2,5$ estimadas pelo SPEEM com propósito de estudar a capacidade do modelo para encontrar novas candidatas nesse range de baixa metalicidade. A magnitude limite foi restringida ainda mais, para $g < 17$, a fim de garantir observações com telescópios de 4 m de abertura com tempos de exposições de até 30 minutos.



Figura 5.3: *A Esquerda:* Telescópio de 4.2 m William Herschel localizado nas ilhas Canárias, cujo tempo de observação está dedicado em um 70% ao levantamento WEAVE. *A Direita:* Montagem do espectrógrafo de braço duplo ISIS (Intermediate-dispersion Spectrograph Imaging System).

Créditos: <https://www.ing.iac.es/astronomy/telescopes/wht/>.

A aplicação desses critérios de seleção gerou uma lista de 177 candidatas a estrelas de muito baixa metalicidade. Desta lista, foram selecionadas 11 estrelas com coordenadas ótimas para observações com o Telescópio William Herschel (WHT, no inglês) de 4,2 m no Observatório Roque de Los Muchachos, La Palma (Figura 5.3). As observações com o WHT utilizaram o espectrógrafo ISIS com uma fenda de 1" que fornece espectros (no braço azul) na faixa de 3900–5100Å e resolução nominal $\lambda/\Delta\lambda \sim 2000$ em 4000Å. Esse tempo de observação foi possível graças à colaboração com Dr. Carlos Allende-Prieto, do IAA.

Na Tabela 5.1 são listados os dados observacionais para as candidatas observadas com o WHT: coordenadas equatoriais, datas de observação, tempos de exposição, magnitudes na banda g , e razão sinal-ruído S/R medida em $\sim 4500\text{Å}$. A redução de dados seguiu o procedimento padrão de subtração de *bias*, *flat-field*, extração dos espectros unidimensi-

Tabela 5.1: Dados observacionais das estrelas observadas: coordenadas, data de observação, magnitude g , tempos de exposição e razão S/N dos espectros obtidos

J-PLUS ID	RA (J2000)	DEC (J2000)	Date	g	t_{exp} (s)	S/N
66416-5807	18:27:22.020	+41:03:51.51	2019-09-28	16,36	1800	40
66430-46822	18:37:21.190	+42:04:11.34	2019-09-28	16,47	1800	55
66723-1757	15:58:17.750	+42:22:39.39	2019-09-28	15,66	1200	55
71582-10685	00:13:06.280	+02:30:46.71	2019-09-29	14,85	900	70
72863-2745	01:34:37.850	+06:54:11.02	2019-09-29	16,63	2200	40
72875-16592	01:35:24.540	+07:40:09.70	2019-09-28	16,56	1800	70
73039-16081	22:11:14.900	+10:34:53.30	2019-09-28	16,28	1800	60
73136-5464	22:54:43.580	+09:50:23.51	2019-09-29	16,64	1800	45
73259-21448	22:48:39.880	+12:11:39.52	2019-09-29	16,23	1800	60
75091-15989	02:35:22.300	+43:27:51.80	2019-09-28	14,66	600	90
75293-17425	14:57:17.890	+54:12:10.54	2019-09-28	16,49	1200	40

onais e calibração em comprimento de onda usando o IRAF¹. Note-se que as observações foram realizadas durante as noites do 28 e 29 de setembro de 2019, antes que o conjunto de dados J-PLUS DR2 fosse liberado oficialmente. Isso foi possível uma vez que as primeiras versões do SPEEM foram testadas no J-PLUS DR1 e a versão final foi atualizada e testada novamente com o J-PLUS DR2, recuperando a mesmas candidatas escolhidas previamente. Esses resultados foram publicados em [GALARZA, CARLOS ANDRÉS *et al.* \(2022\)](#).

Os parâmetros atmosféricos estelares foram determinados usando o pipeline n-SSPP ([BEERS *et al.*, 2014](#); [BEERS *et al.*, 2017](#)), que corresponde a uma adaptação do pipeline original SSPP para analisar espectros não observados pelo levantamento SEGUE. A aplicação do n-SSPP à lista de candidatas forneceu os valores de T_{eff} , $\log g$ e $[\text{Fe}/\text{H}]$ listados na Tabela 5.2. Também listamos as estimativas obtidas com o SPEEM para esses parâmetros, além de temperaturas efetivas obtidas do catálogo Gaia DR2. A estimativa dos parâmetros estelares pelo n-SSPP é baseada no espectro observado de cada estrela, além de alguns dados de entrada adicionais tais como dados fotométricos, índices de linha e correspondência com um banco de dados de espectros sintéticos. Mais detalhes sobre os procedimentos podem ser encontrados em [LEE *et al.* \(2008a\)](#); [LEE *et al.* \(2008b\)](#).

Todas as candidatas selecionadas foram também observadas pelo satélite Gaia e têm estimativas de T_{eff} em Gaia DR2 conforme listado na Tabela 5.2. As diferentes estimativas T_{eff} de Gaia, SPEEM e n-SSPP cobrem um intervalo entre 4,800 K e 5 600 K e são todas consistentes dentro de 2σ . As diferenças médias em relação aos resultados de Gaia são $T_{\text{eff,nSSPP}} - T_{\text{eff,Gaia}} = 213 \pm 14$ K e $T_{\text{eff,SPEEM}} - T_{\text{eff,Gaia}} = 165 \pm 13$ K. As diferenças entre as temperaturas efetivas estimadas com SPEEM e aquelas derivadas da análise

¹IRAF é um pacote computacional clássico para redução de dados astronômicos antigamente distribuído pelo Observatório Nacional de Astronomia Óptica (NOAO, em inglês) e atualmente mantido pela comunidade (<https://iraf-community.github.io/>), operado pela Associação de Universidades para Pesquisa em Astronomia, Inc., sob acordo de cooperação com a National Science Foundation.

Tabela 5.2: Parâmetros estelares obtidos para as candidatas observadas

J-PLUS ID	T_{eff} (K) Gaia	T_{eff} (K) SPEEM	$\log g$	[Fe/H]	T_{eff} (K)	$\log g$ n-SSPP	[Fe/H]
66416-5807	4994	5176	1,94	-2,57	5257±55	2,82±0,55	-2,54±0,12
66430-46822	5038	5268	2,18	-2,76	5376±87	3,45±0,46	-2,82±0,13
66723-1757	5058	5112	2,37	-2,92	5032±69	2,59±0,21	-2,57±0,14
71582-10685	5105	5207	2,52	-2,62	5231±50	1,72±0,10	-2,34±0,10
72863-2745	4936	4921	1,47	-2,83	4989±44	1,12±0,21	-2,78±0,18
72875-16592	4860	5201	2,62	-2,49	4973±103	2,29±0,55	-2,84±0,15
73039-16081	4911	5092	1,81	-2,67	5156±48	2,43±0,45	-2,08±0,09
73136-5464	5125	5384	2,71	-2,64	5595±50	2,80±0,23	-2,30±0,14
73259-21448	4968	5076	2,37	-2,69	5295±114	3,60±0,52	-2,71±0,24
75091-15989	5058	5319	2,63	-2,67	5289±77	2,87±0,28	-3,09±0,11
75293-17425	5012	5123	1,97	-2,87	5221±35	1,19±0,10	-2,46±0,16

espectroscópica variam entre -211 K e $+228$ K, com uma diferença média de 49 ± 128 K.

Os painéis da esquerda das Figuras 5.4 e 5.5 mostram os espectros observados obtidos para as candidatas. Os valores de T_{eff} e [Fe/H] obtidos com o n-SSPP são indicados em cada painel. Os painéis direitos das Figuras 5.4 e 5.5 mostram os foto-espectros J-PLUS para as estrelas da amostra. As estimativas de T_{eff} e [Fe/H] obtidas com SPEEM são indicadas em cada painel. As distribuições de energia espectral (SED) das estrelas selecionadas são consistentes com a SED de uma estrela do tipo G típica. Os pontos são codificados por cores de acordo com os filtros J-PLUS, como em [CENARRO *et al.* \(2019\)](#). Os quadrados representam os filtros de banda larga e os círculos representam os filtros de banda estreita. O espectro de média resolução obtido por SEGUE está disponível para apenas uma estrela de nossa amostra, J-PLUS ID 66723-1757, representada pela linha cinza no painel inferior direito da Figura 5.5.

Validando novamente nossa busca de candidatas com a base de dados em comum observadas com SEGUE, resultou que uma das candidatas observadas com ISIS/WHT já havia sido analisada por SEGUE, fornecendo uma atualização com relação aos valores $T_{\text{eff,SEGUE}}=5115$ K e $[\text{Fe}/\text{H}]_{\text{SEGUE}} = -2,96$ para a estrela J-PLUS ID 66723-1757. Os parâmetros estimados para esta estrela com SPEEM são $T_{\text{eff,SPEEM}}=5112$ K e $[\text{Fe}/\text{H}]_{\text{SPEEM}} = -2,92$, enquanto a análise espectroscópica indica valores ligeiramente mais baixos ($T_{\text{eff,nSSPP}}=5032$ K e $[\text{Fe}/\text{H}]_{\text{nSSPP}} = -2,57$), e a temperatura efetiva listada no catálogo Gaia DR2 é 5058 K.

As diferenças em metalicidade derivadas de SPEEM e n-SSPP são, em média, $+0,11 \pm 0,33$, e a maioria das diferenças está dentro de $\pm 0,4$ dex. Todas as onze candidatas selecionadas apresentam $[\text{Fe}/\text{H}] < -2,0$ e sete estrelas têm metalicidade espectroscópica $[\text{Fe}/\text{H}] < -2,5$, representando uma taxa de sucesso de $64_{-29}^{+21}\%$ ². Este resultado dá suporte

²Intervalo de confiança de Wilson ([WILSON, 1927](#))

ao aplicativo SPEEM como ferramenta para obtenção de estimativas fotométricas dos parâmetros estelares fundamentais.

Novas calibrações fotométricas como as exploradas pelo levantamento Pristine (STAR-KENBURG *et al.*, 2017, e referências associadas) junto com técnicas sofisticadas de aprendizado de máquina podem melhorar significativamente a acurácia das estimações de parâmetros estelares, especialmente no caso de $[\text{Fe}/\text{H}]$ e $\log g$ além de permitir explorar estimações de abundâncias outras espécies químicas como $[\text{C}/\text{Fe}]$, $[\alpha/\text{Fe}]$ e $[\text{Mg}/\text{Fe}]$. De fato, ao momento da escrita deste trabalho já existem alguns avanços ao respeito como os trabalhos desenvolvidos por GALARZA, CARLOS ANDRÉS *et al.* (2022); WHITTEN *et al.* (2019); WHITTEN *et al.* (2021) e YANG, LIN *et al.* (2022); porém, ainda não é claro qual seria o limite inferior extremo ainda confiável para estimar tais abundâncias químicas a partir de fotometria de banda estreita.

5.3 Resumo e comentários finais

Neste capítulo foi demonstrada a eficiência do SPEEM, aplicando o modelo à amostra de ouro do J-PLUS para estimar parâmetros físicos e selecionar novas candidatas a estrelas VMP não reportadas antes na literatura. A distribuição de metalicidade estimada pelo SPEEM atinge um limite mínimo de $[\text{Fe}/\text{H}] > -3,0$ (ver estimativa da estrela 66723-1757 da tabela 5.2), sendo que a forte queda para $[\text{Fe}/\text{H}] < -2,5$ coincide com o resultado reportado na literatura. De igual forma, pode-se notar um aumento da proporção de estrelas de muito baixa metalicidade em relação com modelos de formação de galáxias.

O cenário ideal seria poder ter uma validação espectroscópica de um grande número de estrelas candidatas, porém a baixa disponibilidade de tempo dos grandes telescópios nos permitiu testar apenas uma lista reduzida de 11 candidatas, uma das quais já tinha sido observada pelo SEGUE. No entanto, esse teste preliminar foi um sucesso para mostrar que de fato o SPEEM pode ser usado para caracterizar amostras de estrelas que tenham sido observadas por J-PLUS DR2 e os próximos lançamentos de dados. Todas as estrelas analisadas foram confirmadas como VMP, sendo uma de elas uma EMP não reportada anteriormente na literatura.

No seguinte capítulo se exploram outros usos potenciais do SPEEM, relacionando dados obtidos do sistema de filtros T80 para observar alguns campos da missão K2 do telescópio Kepler com o objetivo de aprimorar a estimativa de $\log g$ aproveitando novos cruzamentos com bases de dados espectroscópicos em alta resolução, como por exemplo, dados dos levantamentos GALAH e APOGEE.

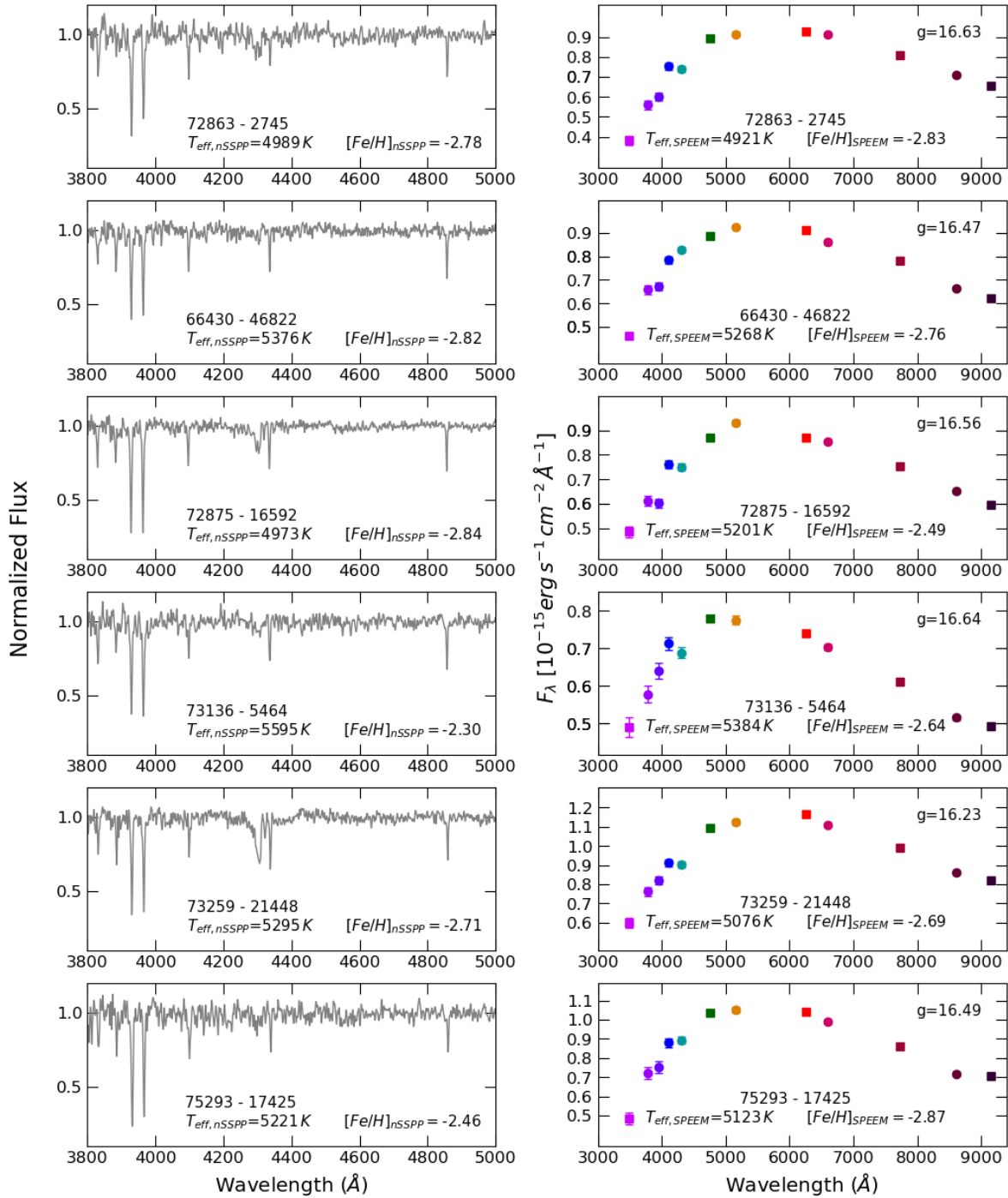


Figura 5.4: Espectros e foto-espectros dos alvos. *Painéis esquerdos*: espectros WHT/ISIS das candidatas identificadas na Tabela 5.2. As linhas H & K do Ca II e as linhas de hidrogênio $H\beta$, $H\gamma$ e $H\delta$, bem como a banda G, são claramente visíveis. *Painéis direitos*: foto-espectros J-PLUS das candidatas. Os pontos são codificados por cores de acordo com os filtros J-PLUS; os quadrados representam os filtros de banda larga e os círculos representam os filtros de banda estreita.

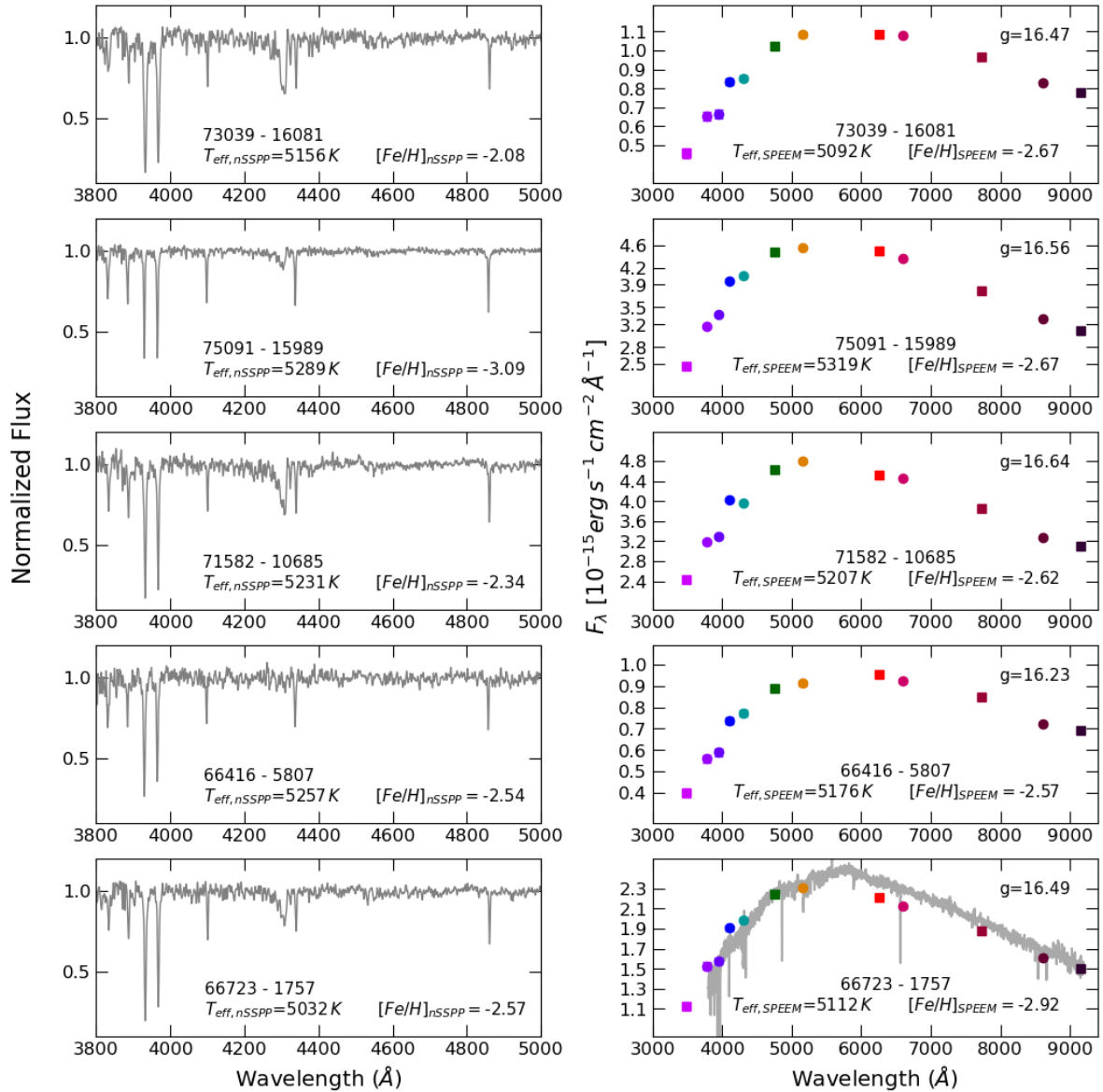


Figura 5.5: Espectros e foto-espectros dos alvos. *Painéis esquerdos*: espectros WHT/ISIS de candidatas identificadas na Tabela 5.2. As linhas H & K do Ca II e as linhas de hidrogênio H β , H γ e H δ , bem como a banda G, são claramente vistas. *Painéis direitos*: foto-espectros J-PLUS das candidatas. Os pontos são codificados por cores de acordo com os filtros J-PLUS; os quadrados representam os filtros de banda larga e os círculos representam os filtros de banda estreita. A linha cinza representa o espectro SEGUE da estrela 66723-1757 superposta ao seu foto-espectro.

Capítulo 6

Estendendo o uso do pipeline SPEEM

Neste capítulo, apresentamos uma aplicação do pipeline SPEEM a dados fotométricos obtidos com o telescópio JAST/T80 e o sistema de filtros J-PLUS para uma amostra de estrelas tipo solar localizadas em campos observados pela missão K2 do Telescópio Espacial Kepler, da NASA.

6.1 A Missão K2 do Telescópio Espacial Kepler

O Telescópio Espacial Kepler, lançado pela NASA em 2009, tinha como objetivo principal observar estrelas do tipo Solar em busca de exoplanetas, além de obter informações de asterosismologia para um grande número de estrelas gigantes. Kepler foi equipado com um fotômetro com o objetivo de obter curvas de luz precisas o suficiente para permitir a detecção de trânsitos. A estratégia observacional da missão consistiu em apontar continuamente para a região do céu de Cygnus-Lyra a fim de obter curvas de luz super precisas. Em 2013, devido a problemas na estabilidade do satélite, o projeto foi redesenhado e uma nova missão, chamada K2, observou uma sequência de 20 campos ao longo da eclíptica, entre 2014 e 2018, como mostra a Figura 6.1.

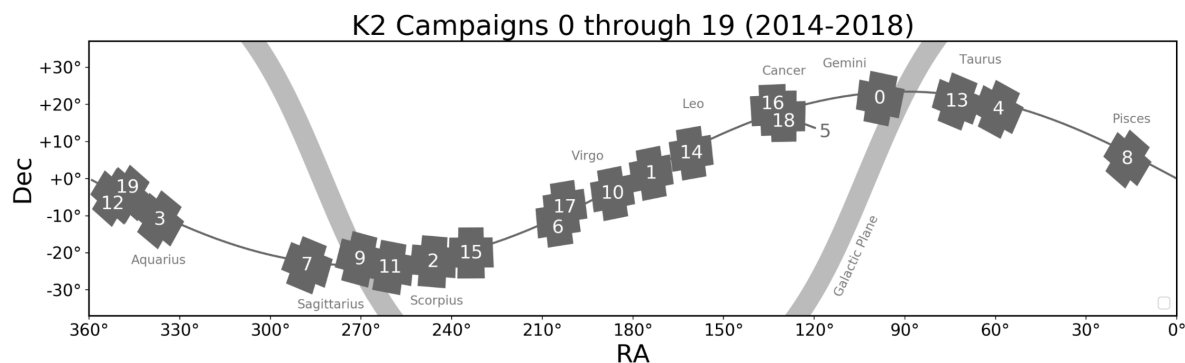


Figura 6.1: Distribuição espacial dos Campos observados na missão K2 do telescópio espacial Kepler. Crédito: <https://archive.stsci.edu/missions-and-data/k2/campaign-fields>

A missão K2 usa o método de trânsito para a detecção de exoplanetas. Para a maioria dos alvos do K2, apenas a fotometria de banda larga de levantamentos de todo o céu como 2MASS ou o APASS (AAVSO Photometric All-Sky Survey; HENDEN e MUNARI, 2014) está disponível. A pesquisa de exoplanetas necessita de parâmetros atmosféricos precisos para as estrelas hospedeiras e a fotometria de banda larga é limitada para esse objetivo. Além disso, medidas precisas dos raios dependem sensivelmente da metalicidade estelar para anãs frias (MANN *et al.*, 2015). A caracterização estelar via asterosismologia do K2 só é possível para subamostras de anãs para as quais a cadência curta está disponível, portanto, as observações com JAST/T80 podem desempenhar um papel fundamental na caracterização das estrelas anãs do tipo solar que abrigam exoplanetas.

A combinação entre os dados de asterosismologia obtidos com a missão K2 e parâmetros atmosféricos obtidos através da espectrofotometria (T_{eff} , $\log g$ e $[\text{Fe}/\text{H}]$) pode fornecer estimativas confiáveis de massa, raio e idade (CHAPLIN *et al.*, 2011; MIGLIO *et al.*, 2013; STELLO *et al.*, 2009). Sendo assim, obtivemos dados de espectrofotometria usando JAST/T80 com o sistema de filtros do J-PLUS para todos os alvos K2 com ascensão reta adequada e magnitudes na faixa $9 < V < 15$ mag durante o semestre 2016B com o objetivo de caracterizar a amostra. Combinando os dados obtidos com o telescópio JAST/T80 e a astrometria fornecida pelo satélite Gaia é possível determinar parâmetros estelares precisos para uma grande amostra de estrelas tipo solar.

6.2 Dados observacionais

Os dados deste projeto foram obtidos com o JAST/T80 para uma amostra de aproximadamente 28.000 estrelas localizadas em 4 campos observados pela missão K2: F0, F4, F5 e F8, mostrados na Figura 6.2). As observações foram realizadas em 18 apontamentos distintos, usando os oito filtros disponíveis de banda estreita e intermediária do sistema J-PLUS montados na T80Cam: u , J0378, J0395, J0410, J0430, J0515, J0660 e J0861. As estrelas dos campos K2 observadas com o JAST/T80 compõem a amostra K2/T80, que constitui a base da pesquisa apresentada neste capítulo. Os campos observados têm as seguintes características:

1. Os campos F4 e F8 também foram extensivamente mapeados pelos levantamentos GALAH e APOGEE e os dados espectroscópicos obtidos podem ser usados para calibrar e testar os filtros estreitos do J-PLUS e sua importância na determinação de parâmetros estelares fundamentais. Além disso, a asterosismologia pode fornecer valores precisos de gravidades superficiais, geralmente conhecidas dentro de 0,01 dex. Isso pode ser muito útil para testar gravidades superficiais derivadas dos filtros J-PLUS.
2. O campo F0 é muito próximo ao plano galáctico, portanto, espera-se que o averme-

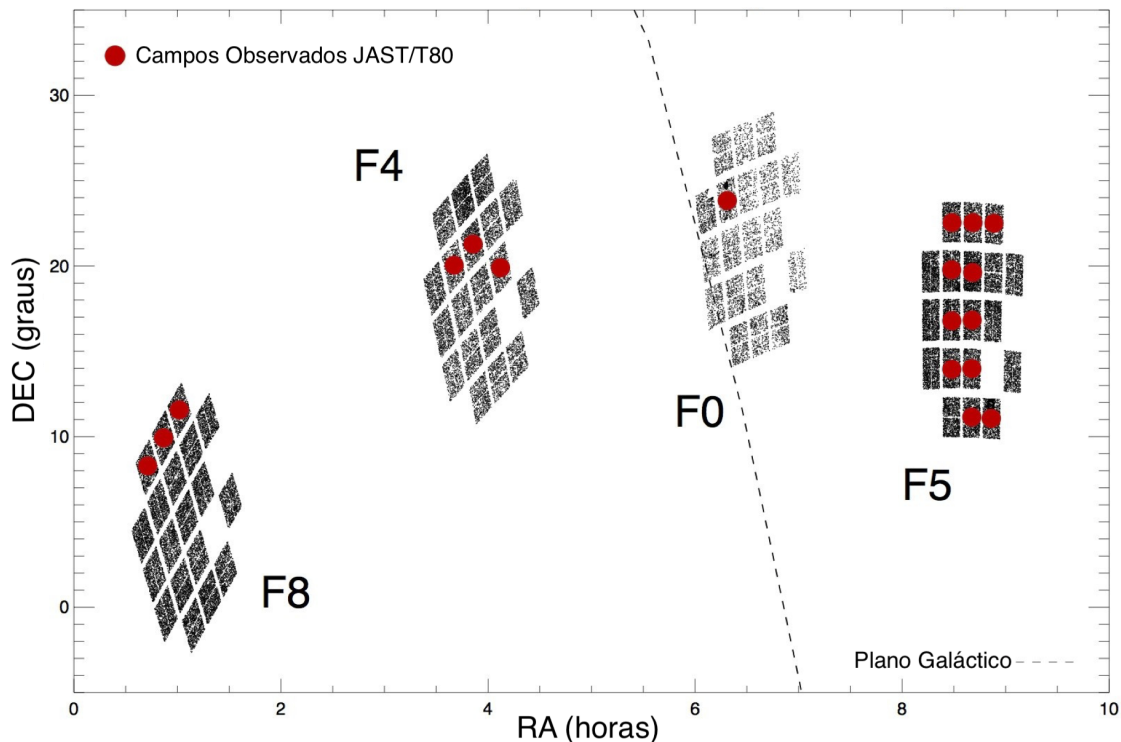


Figura 6.2: Localização dos campos F0, F4, F5 e F8 da missão K2 observados com o telescópio JAST/T80. As regiões observadas estão marcadas em vermelho.

lhamento seja significativo. O aglomerado aberto NGC 2158, presente neste campo, pode ser usado não só para garantir determinações de metalicidade, mas também para testar o efeito do avermelhamento nos parâmetros estelares em regiões fortemente afetadas pela extinção interestelar (CASAGRANDE *et al.*, 2014).

3. O campo F5 contém um aglomerado jovem (M 44) e outro com idade intermediária (M 67) que podem ser muito úteis para calibrações de parâmetros estelares. Os membros desses aglomerados também desempenharão um papel fundamental na determinação das idades das estrelas usando informações sísmicas.

6.2.1 Calibração das magnitudes medidas com filtros de banda estreita

Os dados da amostra K2/T80 obtidos em 2016 foram recalibrados aplicando o método de regressão de cor estelar (SCR; HUANG e YUAN, 2022; YUAN *et al.*, 2015), baseado em dados espectroscópicos do LAMOST DR7 (LUO *et al.*, 2015) e os dados fotométricos corrigidos do Gaia EDR3 (YANG, LIN *et al.*, 2022) (Haibo Yuan, comunicação privada). Este método foi usado para calibrar dados de outros levantamentos fotométricos, incluindo o J-PLUS DR1/DR2 (Yuan *et al.* em preparação), Gaia DR2 (NIU *et al.*, 2021a), Gaia EDR3 (NIU *et al.*, 2021b; YANG *et al.*, 2021), Stripe 82 (HUANG e YUAN, 2022), SMSS

DR2 (HUANG *et al.*, 2021) e PS1 DR1 (XIAO e YUAN, 2022). Uma abordagem muito semelhante à usada neste trabalho foi adotada para os dados J-PLUS e PS1.

Os painéis da Figura 6.3 mostram uma comparação entre as magnitudes nos filtros estreitos obtidas pelo levantamento J-PLUS DR2, no eixo x, e as medidas obtidas com os filtros estreitos do sistema fotométrico da T80cam para a amostra K2/T80, no eixo y, para 4266 estrelas em comum entre as duas amostras. Os pontos azuis representam os dados originais, reduzidos e calibrados segundo o método adotado em 2016. As linhas pontilhadas representam as retas x:y. É evidente que alguns filtros apresentam diferenças sistemáticas que podem chegar a quase 1 mag, indicando, portanto, que é imprescindível recalibrar os dados do pedido de tempo feito em 2016 para que sejam comparáveis ao método adotado no J-PLUS DR2.

Existem 10.960 estrelas comuns entre a amostra K2/T80 e os catálogos LAMOST DR7 e Gaia EDR3. Apenas estrelas que atendem aos seguintes critérios foram usadas na calibração de magnitudes fotométricas dos filtros da T80Cam:

1. $\log g > 3,8$, para excluir estrelas evoluídas;
2. $5000 < T_{\text{eff}}(\text{K}) < 6500$;
3. $[\text{Fe}/\text{H}] > -1$;
4. $E(G_{BP} - G_{RP}) < 1,3 + 0,6 \times (G_{BP} - G_{RP})^2$;
5. $\text{err}(u\text{JAVA}/\text{J0378}/\text{J0395}) < 0,03$, $\text{err}(\text{J0410}/\text{J0430}/\text{J0515}) < 0,02$, e $\text{err}(\text{J0660}/\text{J0861}) < 0,01$.

Para estrelas de calibração de um determinado filtro (por exemplo, uJAVA) de um determinado campo, temos a seguinte equação,

$$G_{BP} - u\text{JAVA} = (G_{BP} - u\text{JAVA})_0 + R \times E(G_{BP} - G_{RP})$$

onde G_{BP} são as magnitudes corrigidas de Gaia EDR3; $(G_{BP} - u\text{JAVA})_0$ são as cores intrínsecas que podem ser estimadas usando seus valores T_{eff} e $[\text{Fe}/\text{H}]$, através de uma relação obtida ao calibrar os dados do J-PLUS DR2; $E(G_{BP} - G_{RP})$ são valores de avermelhamento que foram estimados usando a técnica de par de estrelas (YUAN *et al.*, 2013); R é o coeficiente de avermelhamento em relação a $E(G_{BP} - G_{RP})$, que também é conhecido ao calibrar os dados J-PLUS DR2. Portanto, as magnitudes $u\text{JAVA}$ previstas podem ser obtidas. Então, ajustando $u\text{JAVA} - u\text{JAVA}_{\text{inst}}$ como uma função das posições (X,Y) no CCD para as estrelas de calibração, e usando um polinômio bidimensional, obtemos os valores do ponto zero em todo o plano focal do CCD, com erros de *flat-field* levados em conta. A precisão interna da calibração é de cerca de 5 mmag em filtros uJAVA/J0378/J0395 e cerca de 2 mmag nos outros filtros. É importante notar que as

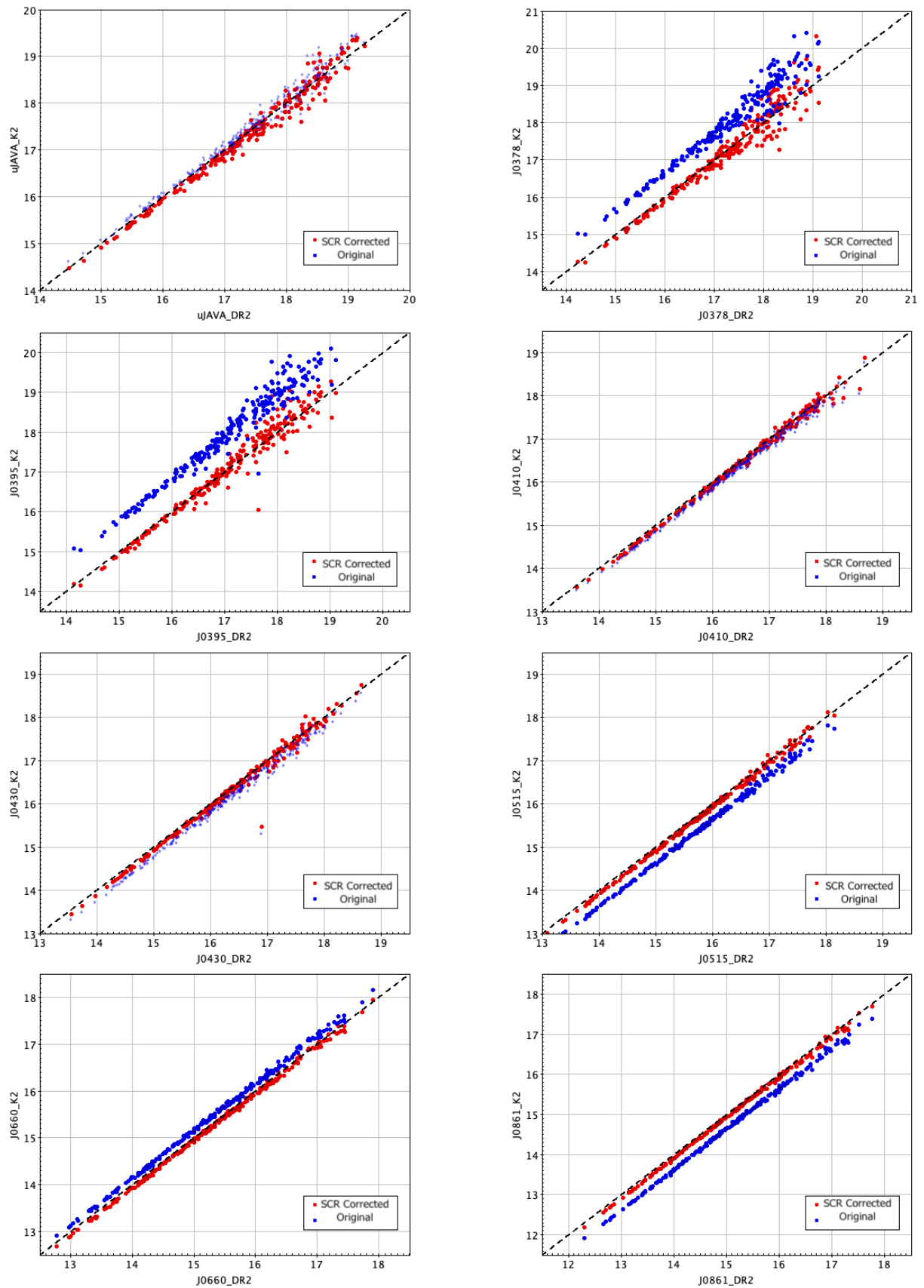


Figura 6.3: Comparação entre as magnitudes medidas com os filtros estreitos obtidas para o campos K2 com o T80 (eixo y) e pelo levantamento J-PLUS DR2 (eixo x). Os pontos azuis representam os dados originais dos campos K2, reduzidos e calibrados segundo o método adotado em 2016. Os pontos vermelhos representam os dados recalibrados. As linhas pontilhadas representam as retas $x:y$. As diferenças sistemáticas observadas nos dados originais foram removidas após a recalibração dos dados usando RSC.

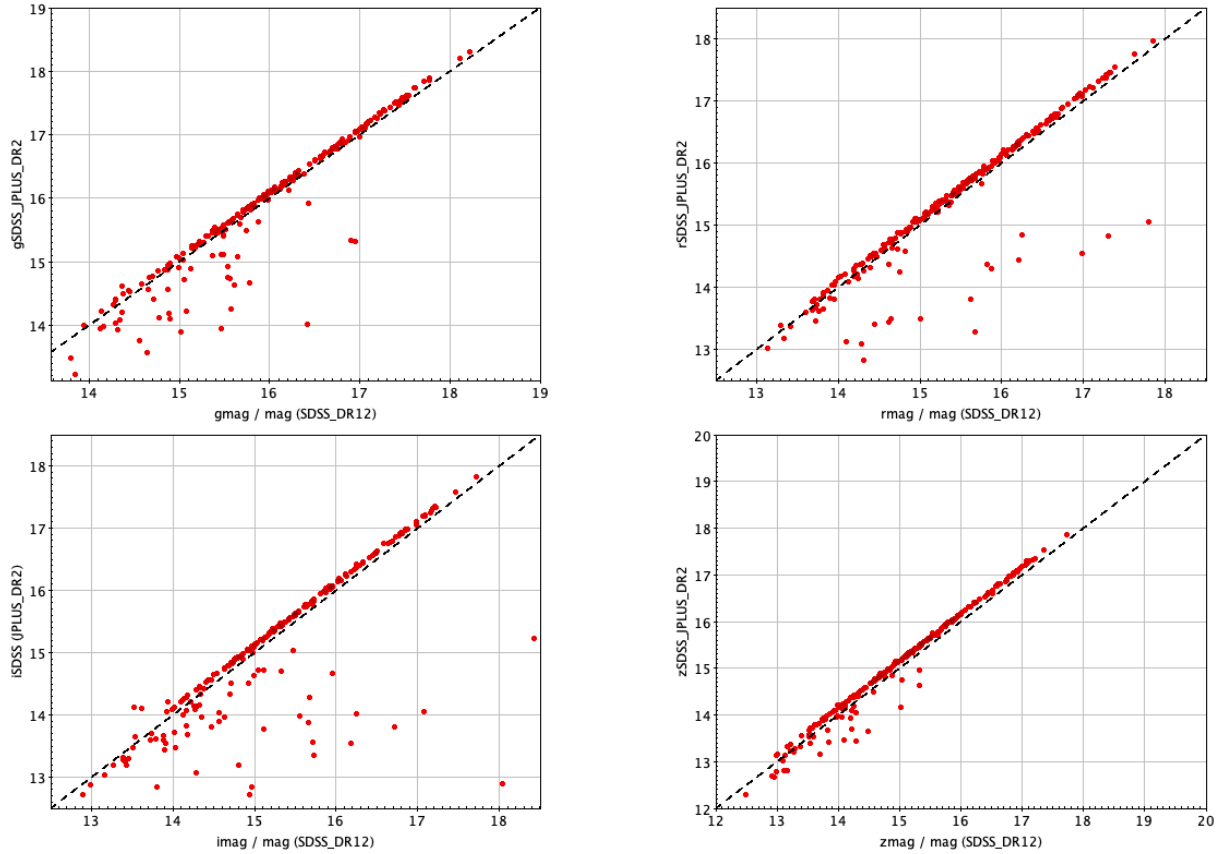


Figura 6.4: Comparação entre as magnitudes medidas nos filtros g , r , i e z dos levantamentos SDSS DR12 e JPLUS DR2, para uma subamostra de 222 estrelas em comum entre as amostra K2/T80 e JPLUS DR2.

magnitudes Gaia G_{BP} são usadas para calibrar todos os filtros, exceto J0660 e J0861, para os quais são usadas magnitudes G_{RP} .

Os dados recalibrados são representados como pontos vermelhos na Figura 6.3, onde pode-se observar que as diferenças sistemáticas foram corrigidas, e as medidas obtidas em 2016, após a correção, são consistentes com as medidas obtidas no levantamento J-PLUS, para as estrelas em comum nas duas amostras.

6.2.2 Magnitudes medidas com filtros de banda larga

As observações da amostra K2/T80 não incluíram os filtros de banda larga da T80Cam, g , r , i , e z , que são comuns aos sistemas fotométricos do SDSS e J-PLUS. Todas as estrelas da amostra K2/T80 foram observadas também com o levantamento SDSS, portanto seria possível combinar os dados de banda estreita obtidos com os dados de banda larga do SDSS DR12 (ALAM *et al.*, 2015) a fim de completar o sistema fotométrico do J-PLUS. Verificamos a consistência das magnitudes dos filtros de banda larga nos levantamentos SDSS DR12 e J-PLUS DR2 através da comparação entre as magnitudes medidas nos 4 filtros para uma subamostra de 222 estrelas em comum entre a amostra K2/T80 e o

catálogo J-PLUS DR2, como mostram os painéis da Figura 6.4.

6.3 Modificação e novo treinamento do SPEEM

A necessidade de treinar o SPEEM com um novo conjunto de dados surge logo após fazer um primeiro teste aplicando o modelo original desenvolvido e otimizado para detectar estrelas muito pobres em metais tal como explicado no Capítulo 4. Um teste preliminar mostrou que o modelo original não consegue fazer boas estimativas de $[\text{Fe}/\text{H}]$ e $\log g$ para a amostra de estrelas K2/T80. Portanto, a busca por novas amostras de treinamento foi inevitável, levando em conta que no pedido de tempo original não foram utilizados os filtros g , r , i e z . A Figura 6.5 mostra as correlações existentes entre os parâmetros estelares T_{eff} , $\log g$ e $[\text{Fe}/\text{H}]$ determinados pelos levantamentos GALAH (nos eixos y) e LAMOST, K2-EPIC (catálogo de entrada do K2, que não faz uma estimativa própria, mas trata-se de um compilado de valores iniciais) e APOGEE. A Figura 6.5 permite concluir que os parâmetros determinados pelo GALAH apresentam maior correlação com os parâmetros obtidos pelo APOGEE.

Este novo treinamento difere do anterior apresentado no Capítulo 4 pela amostra de treinamento usada e pelo conjunto de magnitudes usados como indicadores. A amostra de treinamento agora consiste em 4266 estrelas em comum entre as amostras K2/T80 e campos em comum dos dados APOGEE e GALAH. A estimação dos parâmetros foi baseada nas magnitudes das bandas estreitas e intermediárias do T80, J , H , K , $W1$, $W2$, $W3$ do ALLWISE, G_{BP} , G e G_{RP} do Gaia EDR3 e distâncias estimadas por [BAILER-JONES *et al.* \(2021\)](#). Adicionalmente foram considerados parâmetros estelares estimados por dados espectroscópicos de alta resolução dos levantamentos GALAH DR3 e APOGEE DR17, que abrangem as regiões do óptico e infravermelho, respectivamente, do espectro electromagnético.

É importante deixar claro que a arquitetura do pipeline SPEEM tal como apresentada na Figura 4.1 foi mantida, mudando somente os dados de entrada para incluir os novos parâmetros mencionados no parágrafo anterior e o fato de que a ordem de predição dos parâmetros a estimar torna-se importante para este novo modelo. Neste caso específico, o modelo primeiro estima $\log g$, fixando esse valor para estimar T_{eff} e logo fixando os dois parâmetros anteriores para finalmente estimar $[\text{Fe}/\text{H}]$. Das 4266 estrelas, 3199 foram usadas para o ajuste, deixando as 1067 restantes como amostra de teste para calcular as métricas estatísticas e estimar a acurácia e precisão para esta nova versão do SPEEM. Da Figura 6.6 é possível notar uma melhora significativa da acurácia dos valores estimados, especialmente no caso de $\log g$ quando comparados com as Figuras 4.9 e 4.10.

Vale destacar que a variável de entrada que mais contribui para a estimação precisa da T_{eff} é a pseudo-cor (uJAVA–J0378)–(J0395–J0660), como mostra o painel esquerdo da Figura 6.7, enquanto o painel direito mostra a distribuição de T_{eff} obtida para a amostra

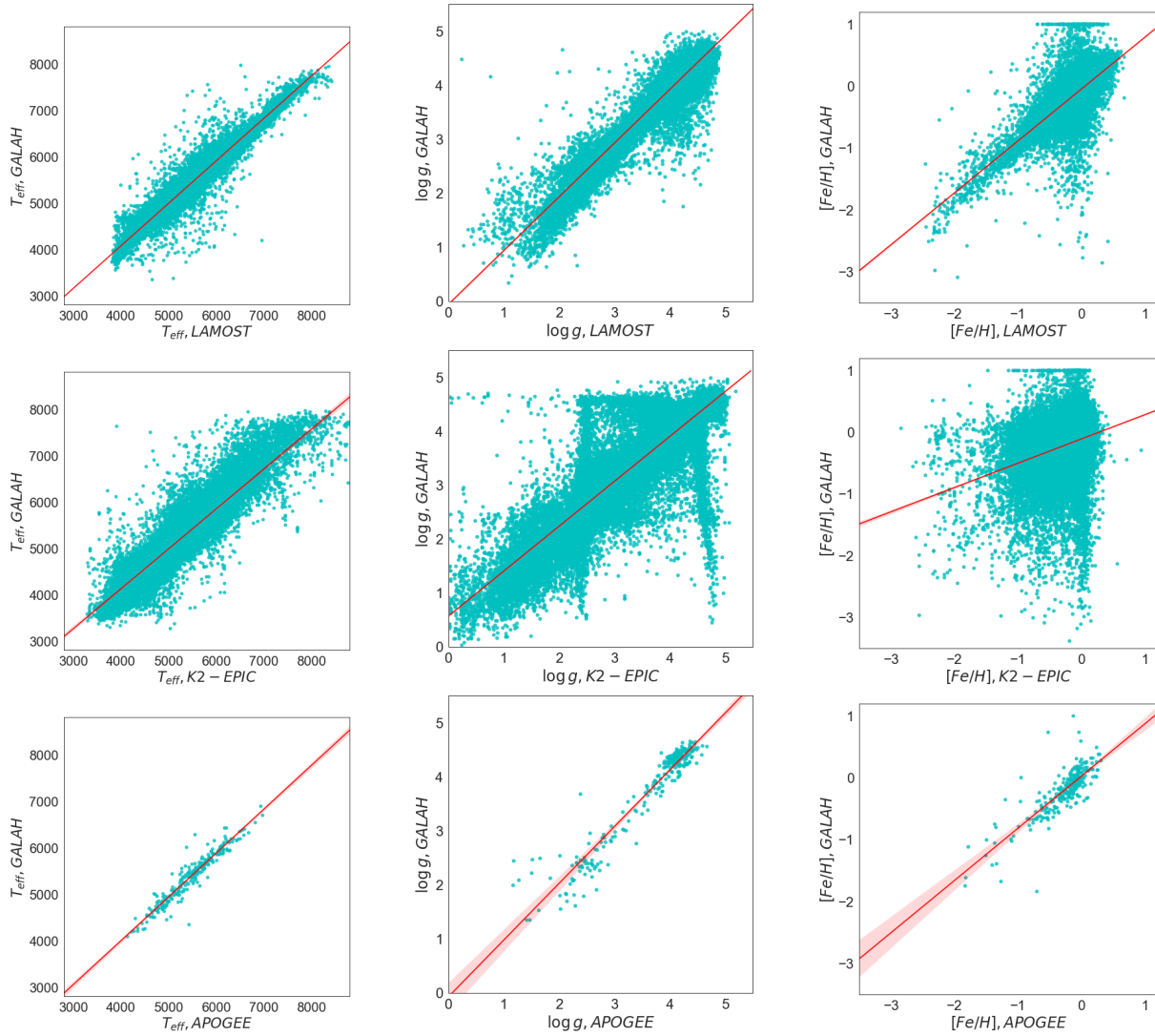


Figura 6.5: Comparação entre os parâmetros T_{eff} (*painéis da coluna esquerda*), $\log g$ (*painéis da coluna central*) e $[\text{Fe}/\text{H}]$ (*painéis da coluna direita*) obtidos por GALAH e os valores listados pelo levantamento LAMOST (*painéis superiores*), catálogo K2-EPIC (*painéis do meio*) e pelo APOGEE (*painéis inferiores*). As linhas vermelhas correspondem a um ajuste de regressão lineal entre as variáveis do eixo X e o eixo Y para estimar o grau de correlação entre os diferentes parâmetros.

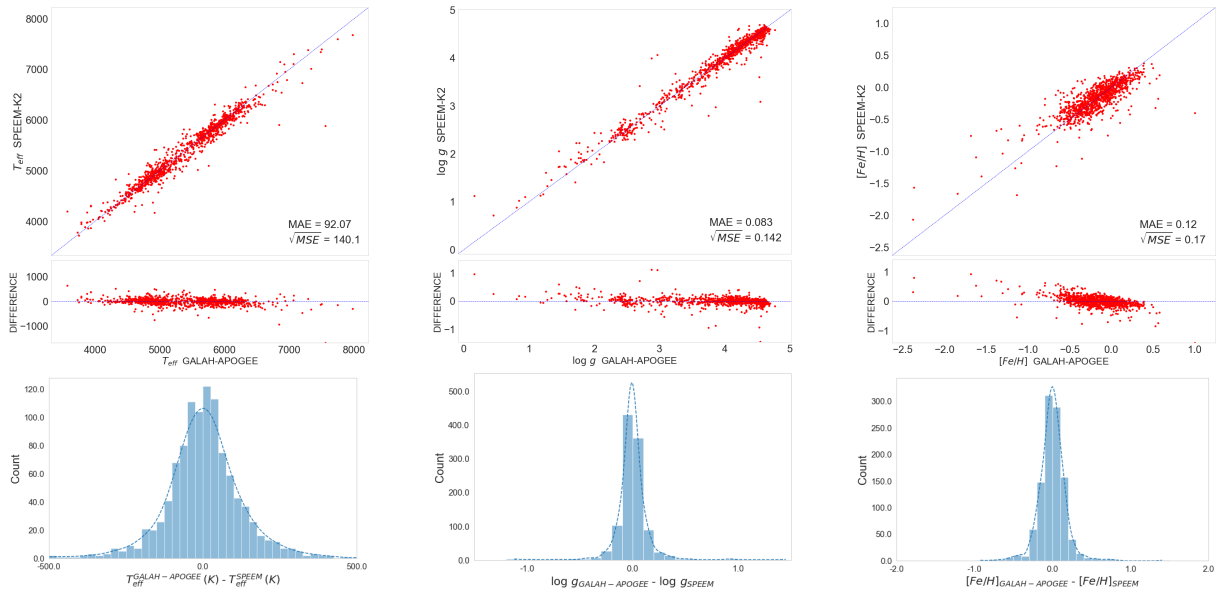


Figura 6.6: Comparação entre os parâmetros T_{eff} (*painel superior esquerdo*), $\log g$ (*painel central superior*) and $[\text{Fe}/\text{H}]$ (*painel superior esquerdo*) obtidos com GALAH-APOGEE e os pipelines SPEEM para 0,25% da amostra T80×GALAH×APOGEE atuando como amostra de teste. O desvio absoluto mediano (MAD) e o desvio absoluto mediano normalizado correspondente (NMAD) são mostrados em cada painel. Os painéis inferiores mostram as distribuições das diferenças entre os respectivos parâmetros obtidos com GALAH-APOGEE e SPEEM. As linhas tracejadas dos painéis superiores marcam a guia de comparação para estabelecer quando o parâmetro estimado coincide com o parâmetro determinado pelo SEGUE. Nas distribuições das diferenças dos painéis inferiores as linhas tracejadas correspondem a ajustes gaussianos.

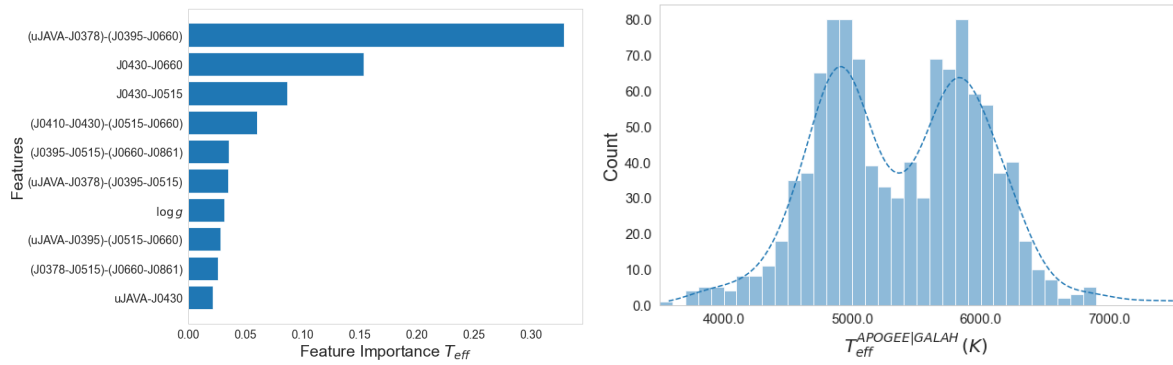


Figura 6.7: *Esquerda:* Distribuição do parâmetro T_{eff} para a amostra de treinamento $T80 \times \text{GALAH-APOGEE}$ composta por 3 199 estrelas. *Direita:* Importância relativa das características espectrais para a estimação da T_{eff} . As características mais importantes como indicadores de T_{eff} são a pseudo-cor (uJAVA-J0378)-(J0395-J0660) e as cores J0430-J0660 e J0430-J0515. As linhas tracejadas correspondem a um ajuste de densidade de kernel da distribuição usando a função `displot()` da livreria Seaborn do Python.

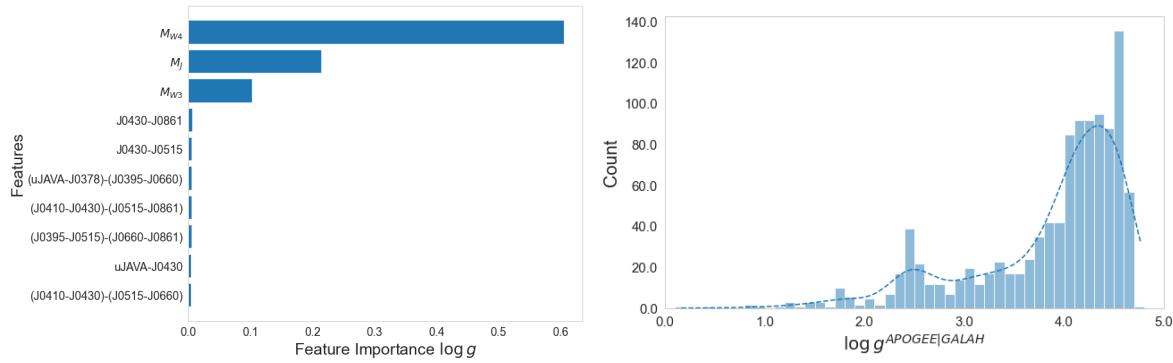


Figura 6.8: *Esquerda:* Distribuição do parâmetro $\log g$ para a amostra de treinamento $T80 \times \text{GALAH-APOGEE}$ composta por 3 199 estrelas. *Direita:* Importância relativa das características espectrais para a estimação da $\log g$. As características mais importantes como indicadores de $\log g$ são as magnitudes absolutas M_{W4} , M_J e M_{W3} . As linhas tracejadas correspondem a um ajuste de densidade de kernel da distribuição usando a função `displot()` da livreria Seaborn do Python.

analisada, com dois picos correspondentes a temperaturas típicas de estrelas de tipos espectrais F e G. Para $\log g$ os 3 melhores indicadores são magnitudes no infravermelho M_{W4} , J e M_{W3} , como mostra o painel esquerdo da Figura 6.8. A distribuição de $\log g$ é apresentada no painel direito da Figura 6.8 e mostra que a amostra é fortemente dominada por estrelas da sequência principal. Por fim, o painel esquerdo da Figura 6.9 mostra que as características mais importantes para a determinação da metalicidade são as pseudo-cores (J0378–J0430)–(J0660–J0861) e (uJAVA–J0430)–(J0660–J0861), além do valor de $\log g$ estimado no passo anterior do SPEEM. A distribuição de metalicidade obtida para a amostra é apresentada no painel direito da Figura 6.9 e mostra que a amostra é dominada por estrelas com metalicidade no intervalo $[-0.5, +0.5]$.

Após validar as previsões do novo modelo SPEEM, o seguinte passo natural consiste

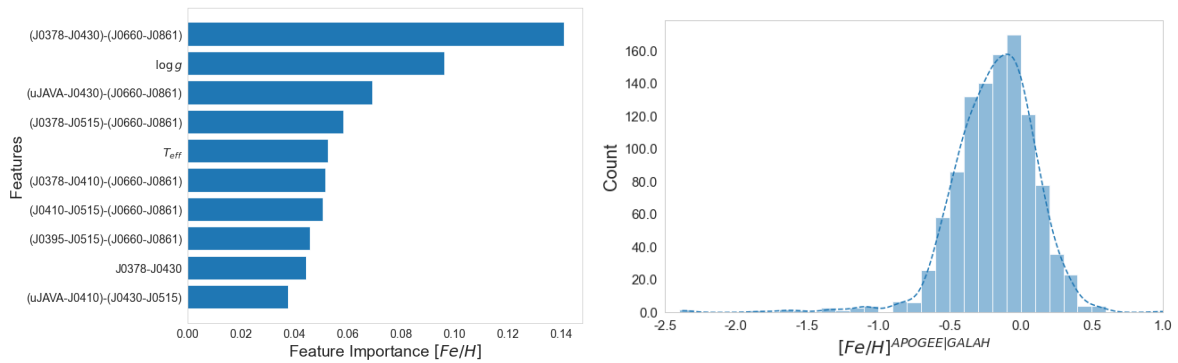


Figura 6.9: *Esquerda:* Distribuição do parâmetro $[Fe/H]$ para a amostra de treinamento T80×GALAH-APOGEE composta por 3 199 estrelas. *Direita:* Importância relativa das características espectrais para a estimação da $[Fe/H]$. As características mais importantes como indicadores de $[Fe/H]$ são (J0378-J0430)-(J0660-J0861), o valor de $\log g$ estimado previamente pelo SPEEM e (uJAVA-J0430)-(J0660-J0861). As linhas tracejadas correspondem a um ajuste de densidade de kernel da distribuição usando a função `displot()` da livreria Seaborn do Python.

em aplicar o modelo para a amostra completa do K2/T80 de 28.490 estrelas. O diagrama HR (Figura 6.10) obtido para a amostra K2/T80 indica que a maior parte das estrelas encontram-se na sequência principal, mas também mostra o ramo das gigantes e o *red clump* bem definidos. Isso indica uma boa caracterização da amostra para gerar um catálogo estelar que possa ser aproveitado, por exemplo, para o estudo de estrelas com exoplanetas.

6.4 Aplicação: o aglomerado M44

O aglomerado M44, (Figura 6.11) também conhecido como NGC 2632, Cr 189, Beehive ou Praesepe, é um aglomerado aberto muito populoso, que contém cerca de 685 estrelas segundo CANTAT-GAUDIN *et al.* (2020). M44 está localizado na constelação de Câncer, sendo um dos aglomerados mais próximos do sistema solar, a uma distância estimada de 187 pc (GAO, 2019) e $\log(\text{idade}) \sim 8,863^1$ (GOSSAGE *et al.*, 2018).

O catálogo obtido a partir da aplicação do novo pipeline SPEEM permite recuperar parâmetros de membros do aglomerado aberto M44 observado pelo JAST/T80. A identificação efetiva das estrelas membros do aglomerado foi feita usando os valores de movimento próprio e paralaxe do Gaia DR2, como apresentado Figura 6.12 pelos pontos azuis. Uma confirmação extra pode ser obtida com o cruzamento da tabela de membros obtida no estudo de CANTAT-GAUDIN *et al.* (2020) a partir de dados do Gaia. O diagrama cor×magnitude (CMD, do inglês) para as estrelas membros de M44 baseado nos dados do Gaia é apresentado no painel superior da Figura 6.13), enquanto o painel inferior mostra o CMD obtido no sistema de filtros da T80Cam, usando a cor J0378–J0861

¹https://webda.physics.muni.cz/cgi-bin/ocl_page.cgi?cluster=M44

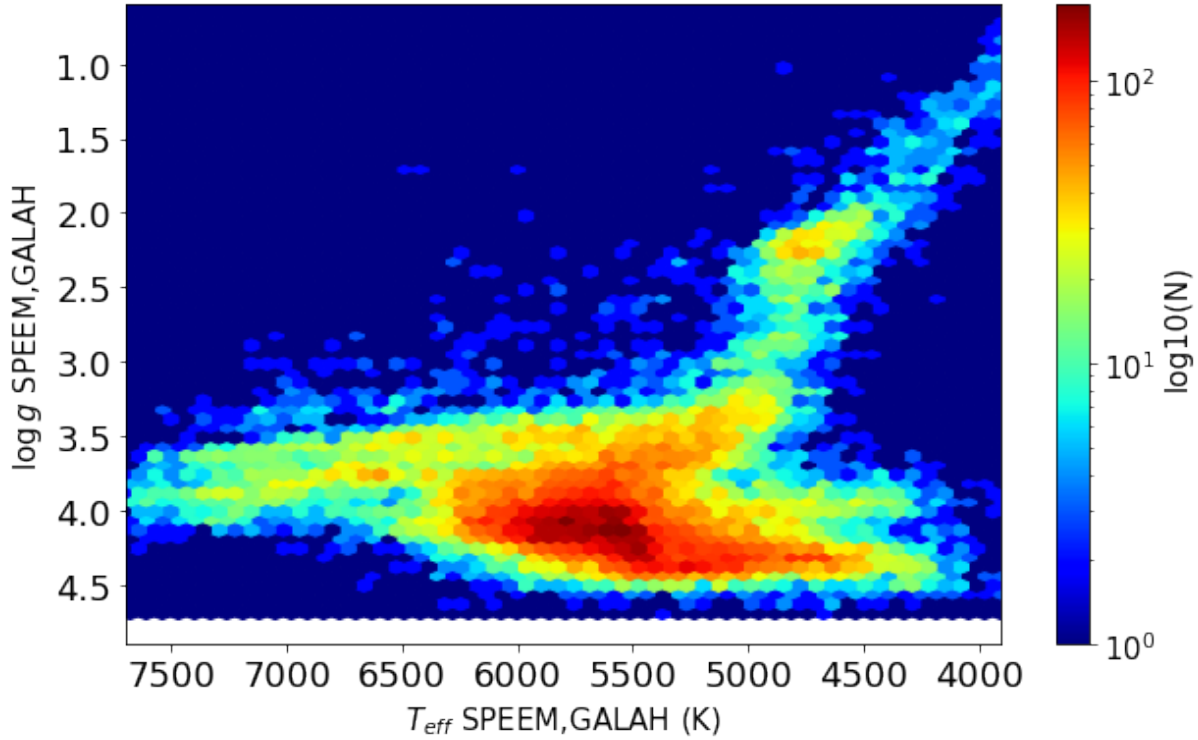


Figura 6.10: Diagrama HR da amostra K2-T80 usando os parâmetros estelares obtidos com SPEEM-GALAH. A cor dos pontos está codificada de acordo com o número de estrelas, indicando que a amostra é composta principalmente por estrelas anãs da sequência principal com tipos espectrais F e G.

e a magnitude na banda J0378, conforme sugerido por [BONATTO *et al.* \(2019\)](#) e as isocronas do [BRESSAN *et al.* \(2012\)](#). Note-se que apenas 128 estrelas pertencentes a M44 do total de 685 propostas por [CANTAT-GAUDIN *et al.* \(2020\)](#) foram observadas com T80. Adicionalmente a Figura 6.14 mostra um diagrama cor-magnitude com filtros T80 e incluindo o efeito de $[\text{Fe}/\text{H}]$ com uma concentração de estrelas de baixa metalicidade no canto superior esquerdo do diagrama, que coincide com uma região de estrelas mais quentes e evoluídas.

Com a recente liberação do catálogo de parâmetros estelares estimados a partir dos espectros de muito baixa resolução do Gaia ([FOUESNEAU *et al.*, 2022](#)), foi possível fazer uma comparação entre os parâmetros atmosféricos obtidos pelo Gaia e pelo SPEEM. Os painéis da Figura 6.15 mostram os diagramas HR para o aglomerado M44 obtidos com os dados de Gaia DR3 (painel superior) e estimados pelo SPEEM (painel inferior). Os pontos estão coloridos de acordo com a escala de metalicidade na lateral dos painéis. Os dois diagramas apresentam um bom acordo, com um desvio sistemático de $\sim 0,2$ dex para $\log g$. Vale mencionar que existe um pequeno conjunto de estrelas que apresentam uma subestimação em T_{eff} quando comparadas com os valores estimados pelo Gaia para $T_{\text{eff}} \geq 7000\text{K}$. Este conjunto de estrelas coincide com parte das estrelas da amostra que apresentam metalicidades menores que $[\text{Fe}/\text{H}] \sim -0.5$. Este efeito é provavelmente decorrente da degenerência entre os parâmetros T_{eff} e $[\text{Fe}/\text{H}]$.



Figura 6.11: Campo do aglomerado estelar M44, também chamado Praesepe, que contém cerca de 700 estrelas.

Fonte: <https://www.jpl.nasa.gov>

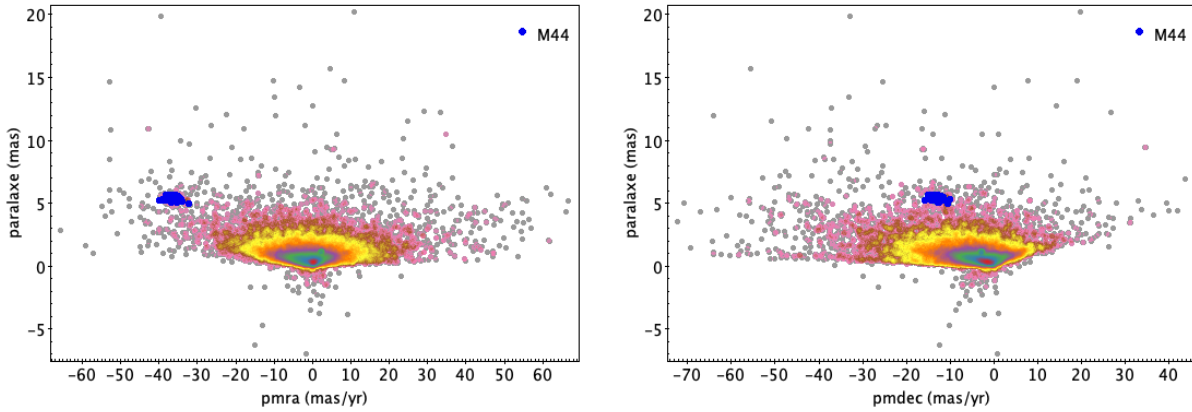


Figura 6.12: Análise visual de **membership** do M44 usando informação de movimentos próprios e paralaxes estimadas por Gaia DR2. As estrelas do aglomerado M44 estão representadas pelos pontos azul escuro.

O painel superior da Figura 6.16 apresenta a distribuição de $[\text{Fe}/\text{H}]$ obtida para o aglomerado M44, usando os valores reportados por Gaia DR3, indicando um valor médio e mediana de $-0,024$ e $-0,026$ respectivamente. A distribuição de $[\text{Fe}/\text{H}]$ das estrelas de M44 estimada pelo pipeline SPEEM é apresentada no painel inferior da Figura 6.16 com um pico ao redor de $0,05$. As duas distribuições mostram um conjunto de estrelas com metalicidade $< -0,5$, que são as estrelas mais quentes da amostra. O valor médio de $[\text{Fe}/\text{H}]$ para a amostra é de $-0,1$ enquanto a mediana (mais robusta a valores atípicos) é de $0,02$ com um valor de NMAD (desvio estatístico robusto a valores atípicos) de $0,15$. Este resultado é consistente, dentro das incertezas, com os valores de metalicidade de M44 reportados na literatura de $[\text{Fe}/\text{H}] = 0,16$ (NETOPIL *et al.*, 2016) e $[\text{Fe}/\text{H}] = 0,20$ (HEITER *et al.*, 2014).

6.5 Resumo e comentários finais

Neste capítulo foi discutido a modificação do SPEEM e sua aplicação a uma amostra de estrelas dos campos F0, F4, F5 e F8 da missão K2, observadas com o telescópio T80 e a câmera T80cam. A análise preliminar indicou um problema na calibração fotométrica dos filtros T80, motivo pelo qual foi preciso fazer uma recalibração dos dados usando o método SCR (YUAN *et al.*, 2015). Feita a recalibração, as magnitudes corrigidas foram usadas no pipeline SPEEM modificado, considerando um nova amostra de treinamento baseada nos levantamentos GALAH e APOGEE, e adicionando novos features tais como magnitudes absolutas do catálogo AllWISE mantendo a mesma estrutura apresentada na Figura 4.1.

Os resultados obtidos da aplicação do modelo SPEEM modificado indicam uma melhora significativa na estimação de parâmetros tipo solar, especialmente o parâmetro $\log g$ quando comparados com a versão anterior. Adicionalmente, como caso de validação, o

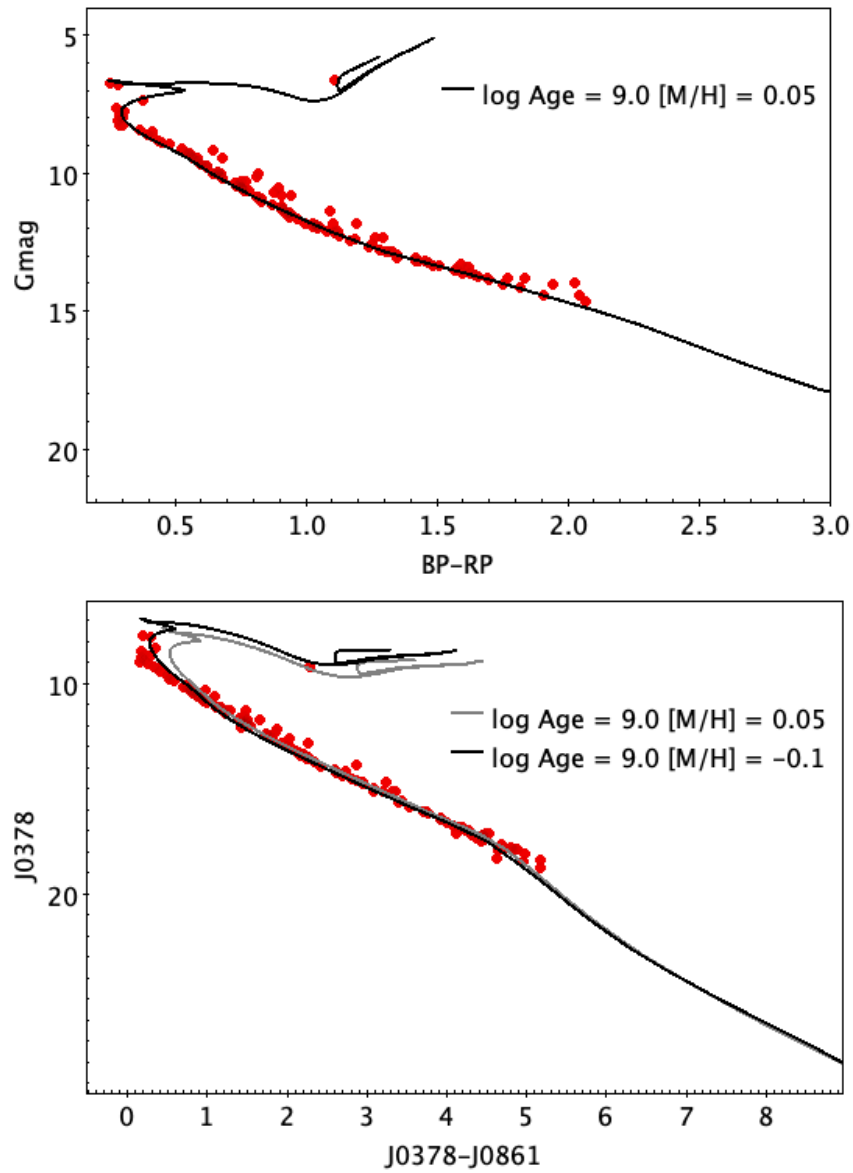


Figura 6.13: *Painel superior*: Diagrama cor-magnitude para M44 com dados do Gaia DR2. *Painel inferior*: Diagrama cor-magnitude para M44 com dados obtidos com o T80. As isócronas de ajuste correspondem aos cálculos de [BRESSAN et al. \(2012\)](#).

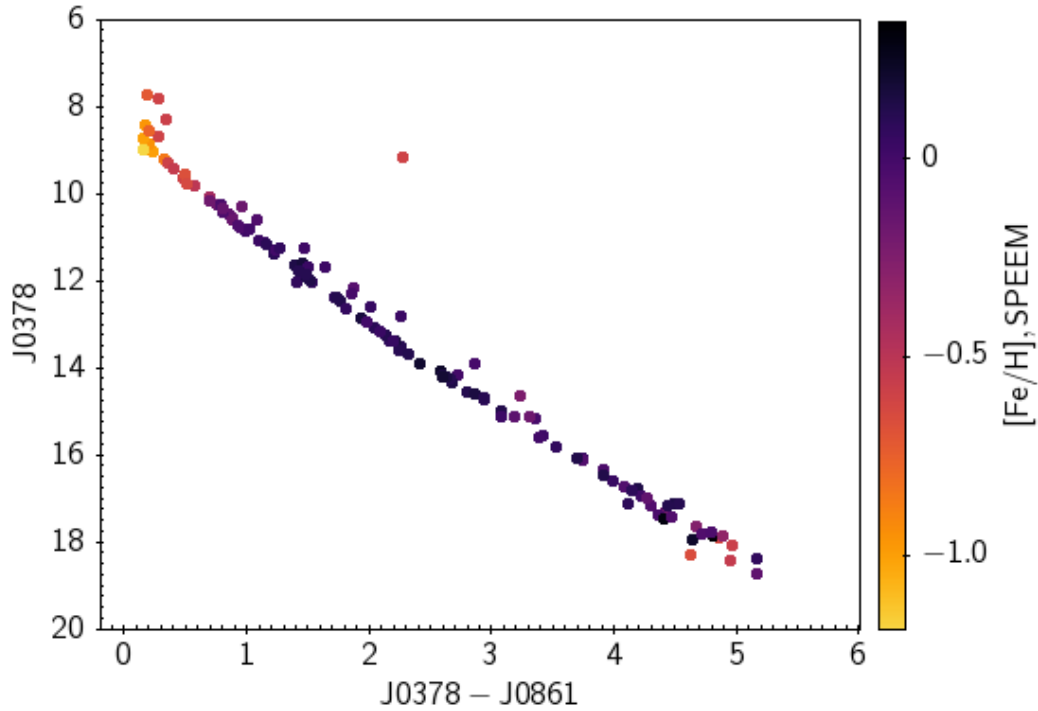


Figura 6.14: Diagrama cor-magnitude das estrelas de M44 observadas com T80. Os pontos estão coloridos de acordo com $[\text{Fe}/\text{H}]$ estimada pelo pipeline SPEEM.

SPEEM modificado foi usado para estimar os parâmetros das estrelas pertencentes ao aglomerado aberto M44, levando em conta que para identificar a lista de membros foi usado o estudo do [CANTAT-GAUDIN *et al.* \(2020\)](#). A idade estimada a partir de um ajuste de isócronas se aproxima muito ao valor reportado pela literatura enquanto a distribuição de $[\text{Fe}/\text{H}]$ também se assemelha bastante à distribuição obtida do recente liberação do catálogo de parâmetros estelares estimados pelo espectroscopia de baixa resolução observada pelo Gaia.

Futuras liberações de dados dos levantamentos J-PLUS, S-PLUS e J-PAS, integrados com os dados do Gaia, permitirão aprimorar os modelos desenvolvidos neste trabalho com a possibilidade inclusive de estimar de forma precisa abundâncias como $[\alpha/\text{Fe}]$ e $[\text{Mg}/\text{Fe}]$.

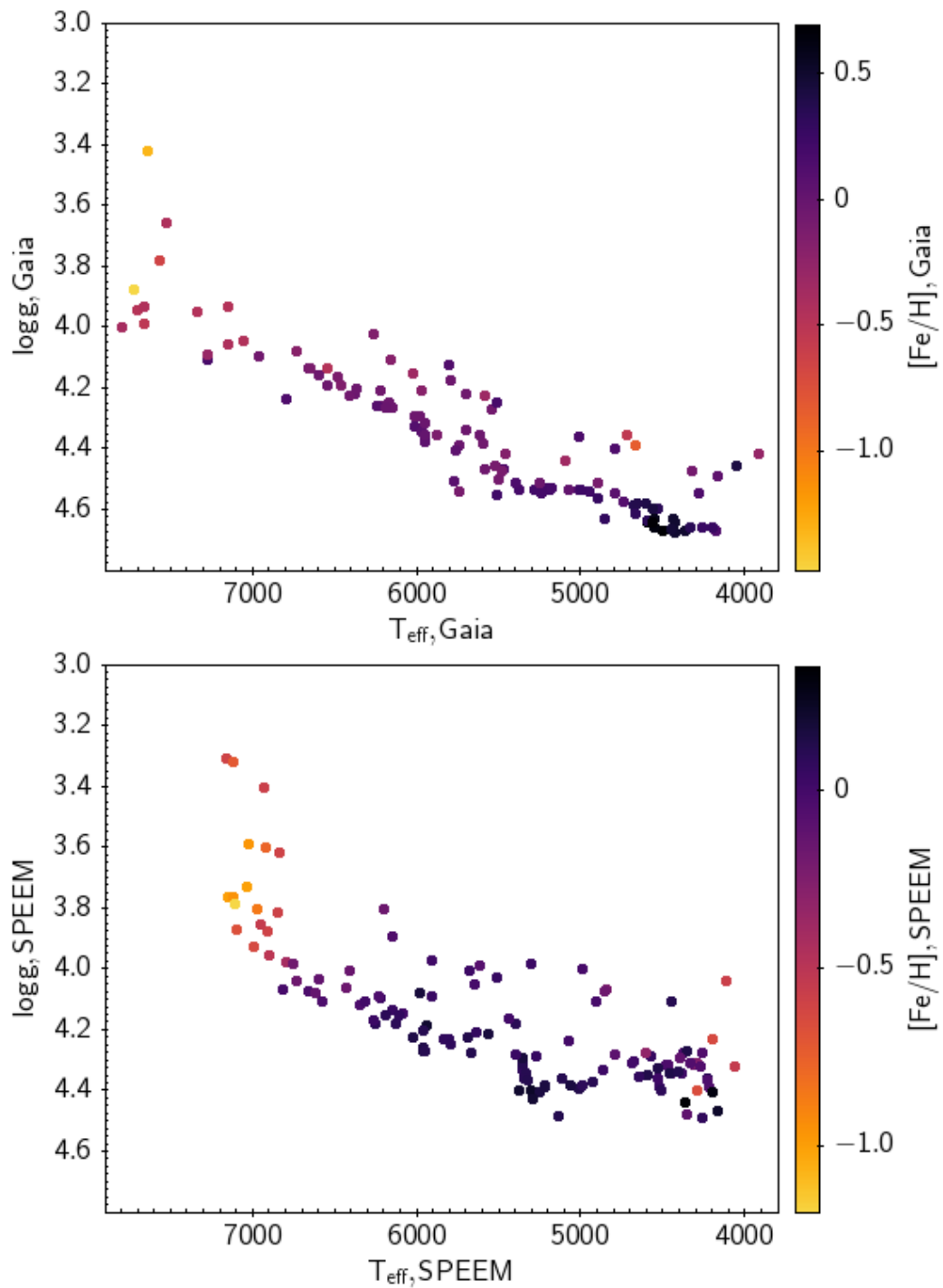


Figura 6.15: Diagramas HR das estrelas de M44 observadas com T80 com os parâmetros estelares de Gaia DR3 (*Painel superior*) e os parâmetros estimados pelo pipeline SPEEM modificado (*Painel inferior*).

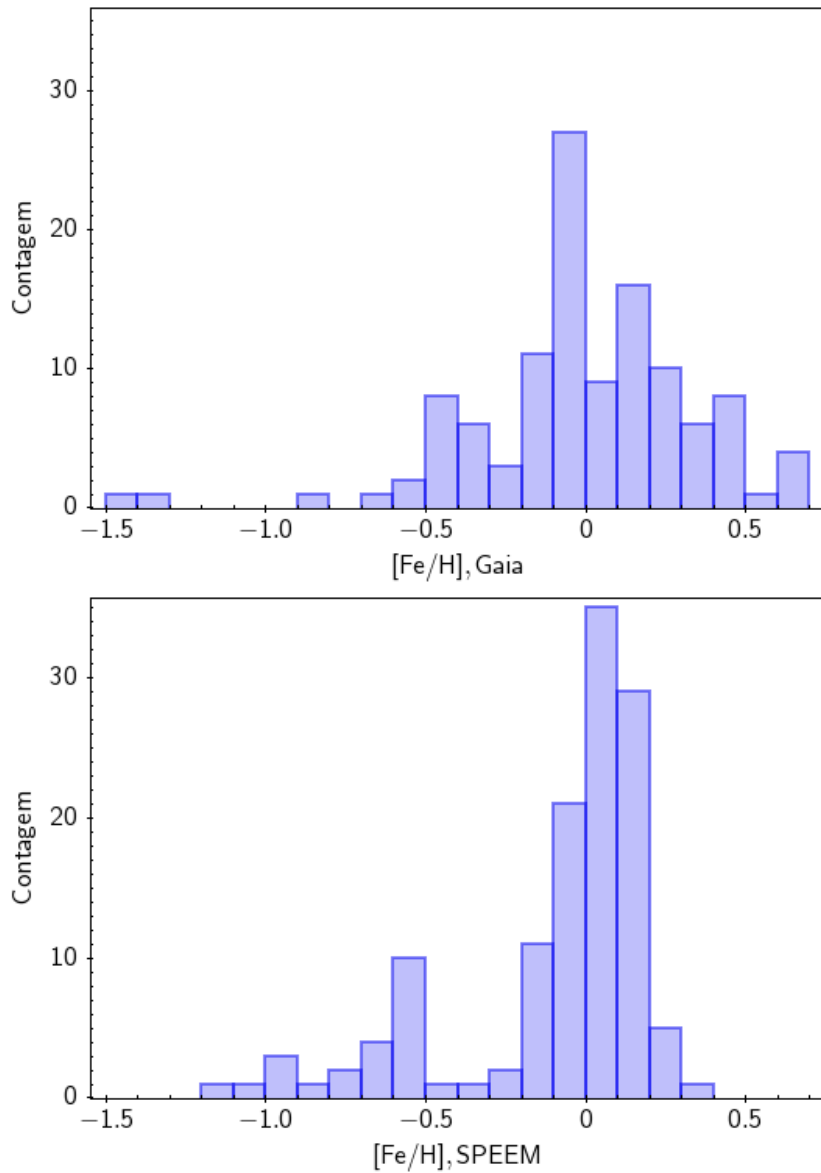


Figura 6.16: Distribuição de metalicidades das estrelas de M44 estimadas por Gaia DR3 (painel superior) e com SPEEM modificado (painel inferior).

Capítulo 7

Conclusões

Neste projeto se apresentou o pipeline SPEEM como uma ferramenta útil para estabelecer a classificação $\text{star} \times \text{QSO}$ e obter os três parâmetros atmosféricos estelares fundamentais (T_{eff} , $\log g$ e $[\text{Fe}/\text{H}]$), com base em medidas fotométricas no sistema J-PLUS. O SPEEM emprega algoritmos de aprendizado de máquina RF e XGB, treinados com amostras selecionadas de dados cruzados com o levantamento espectroscópico SEGUE. O pipeline pode recuperar os parâmetros estimados pelo SEGUE dentro de um desvio de 41 ± 61 K para T_{eff} ; $0,11 \pm 0,17$ para $\log g$ e $0,09 \pm 0,14$ para $[\text{Fe}/\text{H}]$.

O teste de desempenho do pipeline SPEEM inclui uma amostra de estrelas em comum com o LAMOST. Embora a amostra J-PLUS \times LAMOST seja maior que J-PLUS \times SEGUE, a distribuição da metalicidade da amostra anterior é mais desequilibrada no sentido de conter poucas estrelas com $[\text{Fe}/\text{H}] \leq -2,5$ e uma altíssima concentração de estrelas com $[\text{Fe}/\text{H}] \geq -0,8$, o que pode atrapalhar um bom ajuste de parâmetros para esse tipo escasso de estrelas. Como resultado, existe uma tendência na diferença de metalicidade em função da diferença de temperatura, representando uma variação de ~ 0.2 dex em $[\text{Fe}/\text{H}]$ para ~ 100 K em T_{eff} , embora o valor de $[\text{Fe}/\text{H}]$ obtido não apresente diferenças significativas em relação às estimativas do LAMOST.

A “amostra de ouro” analisada nesta Tese compreende uma subamostra do J-PLUS DR2, selecionada de acordo com critérios baseados na qualidade das medidas fotométricas e classificação estrela/galáxia a partir da análise da PDF. A aplicação do SPEEM à “amostra de ouro” produziu parâmetros estelares para 746.531 estrelas. Uma lista de 177 estrelas com $[\text{Fe}/\text{H}] < -2,5$, $T_{\text{eff}} < 5500$ K, e magnitude $g < 17,0$ foi selecionada como candidatas potenciais para estrelas muito pobres em metais que são brilhantes o suficiente para um follow-up espectroscópico.

Onze candidatas foram observadas com o WHT+ISIS e os espectros resultantes foram analisados usando o n-SSPP, produzindo valores espectroscópicos de T_{eff} , $\log g$ e $[\text{Fe}/\text{H}]$. A comparação entre os parâmetros estimados pelo SPEEM e aqueles derivados espectroscopicamente mostra que todas as estrelas estudadas são estrelas muito pobres em metal com $[\text{Fe}/\text{H}] < -2,0$, incluindo sete estrelas com $[\text{Fe}/\text{H}] < -2,5$ e uma nova extremamente

pobre em metais, com $[\text{Fe}/\text{H}] = -3.09$.

Do estudo das candidatas a VMP, o SPEEM apresentou uma taxa de sucesso de $64^{+21}_{-29}\%$ para validar a busca de estrelas com $[\text{Fe}/\text{H}] < -2.5$. Para comparação, o acompanhamento espectroscópico realizado pela colaboração Pristine tem uma taxa de sucesso de 70% para $[\text{Fe}/\text{H}] < -2.5$ (YOUAKIM *et al.*, 2017) em uma amostra de 149 estrelas. Além disso, o Best & Brightest Survey (SCHLAUFMAN e CASEY, 2014) encontrou $\sim 32\%$ estrelas com $-3,0 < [\text{Fe}/\text{H}] < -2,0$ em seu estudo de acompanhamento espectroscópico. Amostras espectroscópicas futuras de estrelas de baixa metalicidade no regime $[\text{Fe}/\text{H}] < -3.0$ serão usadas para estender as capacidades do SPEEM em aplicação para futuros lançamentos de dados do levantamento J-PLUS.

O pipeline SPEEM foi aplicado também a uma amostra de estrelas tipo solar observadas com o telescópio JAST/T80, usando 8 filtros de banda estreita do sistema fotométrico J-PLUS. Essas estrelas estão distribuídas nos campos F0, F4, F5 e F8 da missão K2 do Telescópio Espacial Kepler. Os resultados obtidos para os parâmetros estelares, no entanto, mostraram um grande dispersão, provavelmente introduzida por dois fatores: (i) falta das medidas dos filtros g , r , i , e z ; e (ii) a amostra de treinamento original é enviesada para estrelas de baixa metalicidade.

Sendo assim, o SPEEM foi modificado para se basear em um conjunto de filtros que inclui os filtros T80, e os filtros dos levantamentos Gaia e ALLWISE, além de adotar uma nova amostra de treinamento baseada em resultados do GALAH e APOGEE. Aproveitando informação de paralaxes e distâncias estimadas pelo Gaia para calcular magnitudes absolutas dos filtros, pode-se concluir uma melhora significativa na capacidade de predição do modelo, especialmente quanto à determinação de $\log g$.

Tanto os resultados obtidos nos Capítulos 5 e 6 mostram o potencial da aplicação de modelos de aprendizado de máquina para caracterização e estimação de parâmetros estelares de uma forma rápida e precisa em grandes volumes de dados produzidos por grandes levantamentos fotométricos. Especialmente, as estimações feitas pelo SPEEM sobre os dados J-PLUS não só beneficiam a busca de novas candidatas a estrelas de baixa metalicidade mas podem contribuir de forma positiva sobre outro tipo de pesquisas que requeram de um bom conhecimento da distribuição dos parâmetros físicos de estrelas FGK no halo galáctico. As análises apresentadas nesta Tese demonstram a versatilidade do SPEEM para determinação de parâmetros atmosféricos baseada em fotometria multifiltros.

Referências Bibliográficas

- ABOHALIMA, A., FREBEL, A., 2018, “JINAbase—A Database for Chemical Abundances of Metal-poor Stars”, *Astrophys. J. Suppl.*, 238(2):36. doi: 10.3847/1538-4365/aadfe9.
- AGUADO, D. S., PRIETO, C. A., HERNÁNDEZ, J. I. G., et al., 2018, “J0023+ 0307: A Mega Metal-poor Dwarf Star from SDSS/BOSS”, *The Astrophysical Journal Letters*, v. 854, n. 2, pp. L34.
- ALAM, S., ALBARETI, F. D., PRIETO, C. A., et al., 2015, “The eleventh and twelfth data releases of the Sloan Digital Sky Survey: final data from SDSS-III”, *The Astrophysical Journal Supplement Series*, v. 219, n. 1, pp. 12.
- ALLENDE PRIETO, C., BEERS, T. C., WILHELM, R., et al., 2006a, “A Spectroscopic Study of the Ancient Milky Way: F- and G-Type Stars in the Third Data Release of the Sloan Digital Sky Survey”, *The Astrophysical Journal*, v. 636, n. 2 (jan), pp. 804–820. doi: 10.1086/498131. Disponível em: <<https://doi.org/10.1086/498131>>.
- ALLENDE PRIETO, C., BEERS, T. C., WILHELM, R., et al., 2006b, “A Spectroscopic Study of the Ancient Milky Way: F- and G-Type Stars in the Third Data Release of the Sloan Digital Sky Survey”, *Astrophys. J.*, v. 636, n. 2 (jan.), pp. 804–820. doi: 10.1086/498131.
- ALMEIDA-FERNANDES, F., SAMPEDRO, L., HERPICH, F. R., et al., 2021, “Data Release 2 of S-PLUS: accurate template-fitting based photometry covering ~1000 square degrees in 12 optical filters”, *arXiv e-prints*, art. arXiv:2104.00020.
- AOKI, W., BEERS, T. C., CHRISTLIEB, N., et al., 2007, “Carbon-enhanced metal-poor stars. I. Chemical compositions of 26 stars”, *The Astrophysical Journal*, v. 655, n. 1, pp. 492.
- AOKI, W., BEERS, T. C., LEE, Y. S., et al., 2012, “High-resolution spectroscopy of extremely metal-poor stars from SDSS/SEGUE. I. Atmospheric parameters and chemical compositions”, *The Astronomical Journal*, v. 145, n. 1, pp. 13.

- BAI, Y., LIU, J., WANG, S., et al., 2018, “Machine Learning Applied to Star–Galaxy–QSO Classification and Stellar Effective Temperature Regression”, *The Astronomical Journal*, v. 157, n. 1, pp. 9.
- BAILER-JONES, C. A. L., 2004, “Evolutionary design of photometric systems and its application to Gaia”, *Astron. Astrophys.*, v. 419 (maio), pp. 385–403. doi: 10.1051/0004-6361:20035779.
- BAILER-JONES, C. A. L., RYBIZKI, J., FOUESNEAU, M., et al., 2021, “VizieR Online Data Catalog: Distances to 1.47 billion stars in Gaia EDR3 (Bailer-Jones+, 2021)”, *VizieR Online Data Catalog*, art. I/352.
- BEERS, T. C., CHRISTLIEB, N., 2005, “The Discovery and Analysis of Very Metal-Poor Stars in the Galaxy”, *Ann. Rev. Astron. Astrophys.*, v. 43, n. 1 (set.), pp. 531–580. doi: 10.1146/annurev.astro.42.053102.134057.
- BEERS, T. C., PRESTON, G. W., SHECTMAN, S. A., 1985, “A search for stars of very low metal abundance. I”, *The Astronomical Journal*, v. 90, pp. 2089–2102.
- BEERS, T. C., PRESTON, G. W., SHECTMAN, S. A., 1992, “A search for stars of very low metal abundance. II”, *The Astronomical Journal*, v. 103, pp. 1987–2034.
- BEERS, T. C., ROSSI, S., NORRIS, J. E., et al., 1999, “Estimation of stellar metal abundance. II. A recalibration of the Ca II K technique, and the autocorrelation function method”, *The Astronomical Journal*, v. 117, n. 2, pp. 981.
- BEERS, T. C., NORRIS, J. E., PLACCO, V. M., et al., 2014, “POPULATION STUDIES. XIII. A NEW ANALYSIS OF THE BIDELMAN-MACCONNELL “WEAK-METAL” STARS—CONFIRMATION OF METAL-POOR STARS IN THE THICK DISK OF THE GALAXY”, *The Astrophysical Journal*, v. 794, n. 1 (sep), pp. 58. doi: 10.1088/0004-637x/794/1/58. Disponível em: <<https://doi.org/10.1088/0004-637x/794/1/58>>.
- BEERS, T. C., PLACCO, V. M., CAROLLO, D., et al., 2017, “Bright Metal-Poor Stars from the Hamburg/ESO Survey. II. A Chemodynamical Analysis”, *Astrophys. J.*, 835(1):81. doi: 10.3847/1538-4357/835/1/81.
- BELLAZZINI, M., IBATA, R., MONACO, L., et al., 2004, “Detection of the Canis Major galaxy at $(l; b) = (244^\circ -8^\circ)$ and in the background of Galactic open clusters”, *Mon. Not. Roy. Astron. Soc.*, v. 354, n. 4 (nov.), pp. 1263–1278. doi: 10.1111/j.1365-2966.2004.08283.x.
- BENITEZ, N., DUPKE, R., MOLES, M., et al., 2014, “J-PAS: the Javalambre-physics of the accelerated universe astrophysical survey”, *arXiv preprint arXiv:1403.5237*.

- BERTIN, E., ARNOUITS, S., 1996, “SExtractor: Software for source extraction.” *Astron. Astrophys. Suppl.*, v. 117 (jun.), pp. 393–404. doi: 10.1051/aas:1996164.
- BESSELL, M. S., NORRIS, J., 1984, “The ultra-metal-deficient (population III) red giant CD -38 245 ?” *Astrophys. J.*, v. 285 (out.), pp. 622–636. doi: 10.1086/162539.
- BITSCH, B., BATTISTINI, C., 2020, “Influence of sub-and super-solar metallicities on the composition of solid planetary building blocks”, *Astronomy & Astrophysics*, v. 633, pp. A10.
- BOLTON, A. S., SCHLEGEL, D. J., AUBOURG, É., et al., 2012, “Spectral classification and redshift measurement for the SDSS-III baryon oscillation spectroscopic survey”, *The Astronomical Journal*, v. 144, n. 5, pp. 144.
- BONATTO, C., CHIES-SANTOS, A. L., COELHO, P. R., et al., 2019, “J-PLUS: A wide-field multi-band study of the M 15 globular cluster-Evidence of multiple stellar populations in the RGB”, *Astronomy & Astrophysics*, v. 622, pp. A179.
- BONOLI, S., MARÍN-FRANCH, A., VARELA, J., et al., 2021, “The miniJPAS survey: A preview of the Universe in 56 colors”, *Astronomy & Astrophysics*, v. 653, pp. A31.
- BREIMAN, L., 2001, “Random Forests”, *Machine Learning*, v. 45, n. 1 (Oct), pp. 5–32. ISSN: 1573-0565. doi: 10.1023/A:1010933404324. Disponível em: <<https://doi.org/10.1023/A:1010933404324>>.
- BREIMAN, L., FRIEDMAN, J., STONE, C. J., et al., 1984, *Classification and regression trees*. CRC press.
- BRESSAN, A., MARIGO, P., GIRARDI, L., et al., 2012, “PARSEC: stellar tracks and isochrones with the PAdova and TRieste Stellar Evolution Code”, *Mon. Not. Roy. Astron. Soc.*, v. 427, n. 1 (nov.), pp. 127–145. doi: 10.1111/j.1365-2966.2012.21948.x.
- BUDER, S., SHARMA, S., KOS, J., et al., 2021, “The GALAH+ survey: Third data release”, *Mon. Not. Roy. Astron. Soc.*, v. 506, n. 1 (set.), pp. 150–201. doi: 10.1093/mnras/stab1242.
- CANTAT-GAUDIN, T., ANDERS, F., CASTRO-GINARD, A., et al., 2020, “VizieR Online Data Catalog: Portrait Galactic disc (Cantat-Gaudin+, 2020)”, *VizieR Online Data Catalog*, pp. J–A+.

- CARDAMONE, C. N., VAN DOKKUM, P. G., URRY, C. M., et al., 2010, “THE MULTI-WAVELENGTH SURVEY BY YALE–CHILE (MUSYC): DEEP MEDIUM-BAND OPTICAL IMAGING AND HIGH-QUALITY 32-BAND PHOTOMETRIC REDSHIFTS IN THE ECDF-S”, *The Astrophysical Journal Supplement Series*, v. 189, n. 2 (jul), pp. 270–285. doi: 10.1088/0067-0049/189/2/270. Disponível em: <<https://doi.org/10.1088/0067-0049/189/2/270>>.
- CARNEY, B. W., PETERSON, R. C., 1981, “Abundance analyses of subdwarfs of the remote halo.” *Astrophys. J.*, v. 245 (abr.), pp. 238–246. doi: 10.1086/158804.
- CAROLLO, D., BEERS, T. C., LEE, Y. S., et al., 2007, “Two stellar components in the halo of the Milky Way”, *Nature*, v. 450, n. 7172 (dez.), pp. 1020–1025. doi: 10.1038/nature06460.
- CAROLLO, D., BEERS, T. C., CHIBA, M., et al., 2010, “Structure and Kinematics of the Stellar Halos and Thick Disks of the Milky Way Based on Calibration Stars from Sloan Digital Sky Survey DR7”, *Astrophys. J.*, v. 712, n. 1 (mar.), pp. 692–727. doi: 10.1088/0004-637X/712/1/692.
- CARRILLO, A., HAWKINS, K., BOWLER, B. P., et al., 2020, “Know thy star, know thy planet: chemo-kinematically characterizing TESS targets”, *Monthly Notices of the Royal Astronomical Society*, v. 491, n. 3, pp. 4365–4381.
- CASAGRANDE, L., RAMÍREZ, I., MELENDEZ, J., et al., 2010, “An absolutely calibrated Teff scale from the infrared flux method-Dwarfs and subgiants”, *Astronomy & Astrophysics*, v. 512, pp. A54.
- CASAGRANDE, L., SILVA AGUIRRE, V., STELLO, D., et al., 2014, “Strömgren Survey for Asteroseismology and Galactic Archaeology: Let the SAGA Begin”, *Astrophys. J.*, 787(2):110. doi: 10.1088/0004-637X/787/2/110.
- CASAGRANDE, L., 2008, “Infrared flux method and colour calibrations”, *Physica Scripta*, v. 2008, n. T133, pp. 014020.
- CASAGRANDE, L., PORTINARI, L., FLYNN, C., 2006, “Accurate fundamental parameters for lower main-sequence stars”, *Mon. Not. Roy. Astron. Soc.*, v. 373, n. 1 (nov.), pp. 13–44. doi: 10.1111/j.1365-2966.2006.10999.x.
- CENARRO, A. J., MOLES, M., MARÍN-FRANCH, A., et al., 2014, “The Observatorio Astrofísico de Javalambre: current status, developments, operations, and strategies”. In: *Observatory Operations: Strategies, Processes, and Systems V*, v. 9149, Proc. SPIE, p. 91491I, ago. doi: 10.1117/12.2055455.

- CENARRO, A. J., MOLES, M., CRISTÓBAL-HORNILLOS, D., et al., 2019, “J-PLUS: The Javalambre Photometric Local Universe Survey”, *A&A*, v. 622, pp. A176. doi: 10.1051/0004-6361/201833036. Disponível em: <<https://doi.org/10.1051/0004-6361/201833036>>.
- CENARRO, A., CARDIEL, N., GORGAS, J., et al., 2001a, “Empirical calibration of the near-infrared Ca II triplet—I. The stellar library and index definition”, *Monthly Notices of the Royal Astronomical Society*, v. 326, n. 3, pp. 959–980.
- CENARRO, A., GORGAS, J., CARDIEL, N., et al., 2001b, “Empirical calibration of the near-infrared Ca ii triplet—II. The stellar atmospheric parameters”, *Monthly Notices of the Royal Astronomical Society*, v. 326, n. 3, pp. 981–994.
- CHAMBERS, K. C., MAGNIER, E. A., METCALFE, N., et al., 2016, “The Pan-STARRS1 Surveys”, *arXiv e-prints*, art. arXiv:1612.05560.
- CHAO, L., WEN-HUI, Z., JI-MING, L., 2019, “Study of star/Galaxy classification based on the XGBoost algorithm”, *Chinese Astronomy and Astrophysics*, v. 43, n. 4, pp. 539–548.
- CHAPLIN, W. J., KJELDSEN, H., CHRISTENSEN-DALSGAARD, J., et al., 2011, “Ensemble Asteroseismology of Solar-Type Stars with the NASA Kepler Mission”, *Science*, v. 332, n. 6026 (abr.), pp. 213. doi: 10.1126/science.1201827.
- CHEN, T., GUESTRIN, C., 2016, “Xgboost: A scalable tree boosting system”. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794.
- CHEN, T., HE, T., BENESTY, M., et al., 2015, “Xgboost: extreme gradient boosting”, *R package version 0.4-2*, v. 1, n. 4.
- CLARK, J. T., CLERTÉ, M., HINKEL, N. R., et al., 2021, “The GALAH survey: using galactic archaeology to refine our knowledge of TESS target stars”, *Monthly Notices of the Royal Astronomical Society*, v. 504, n. 4, pp. 4968–4989.
- CUI, X.-Q., ZHAO, Y.-H., CHU, Y.-Q., et al., 2012, “The large sky area multi-object fiber spectroscopic telescope (LAMOST)”, *Research in Astronomy and Astrophysics*, v. 12, n. 9, pp. 1197.
- DALTON, G., TRAGER, S. C., ABRAMS, D. C., et al., 2012, “WEAVE: the next generation wide-field spectroscopy facility for the William Herschel Telescope”. In: McLean, I. S., Ramsay, S. K., Takami, H. (Eds.), *Ground-based and Airborne Instrumentation for Astronomy IV*, v. 8446, *Society of Photo-Optical*

- Instrumentation Engineers (SPIE) Conference Series*, p. 84460P, set. doi: 10.1117/12.925950.
- DAWSON, K. S., SCHLEGEL, D. J., AHN, C. P., et al., 2012, “The Baryon oscillation spectroscopic survey of SDSS-III”, *The Astronomical Journal*, v. 145, n. 1, pp. 10.
- DAWSON, K. S., KNEIB, J.-P., PERCIVAL, W. J., et al., 2016, “The SDSS-IV extended Baryon Oscillation Spectroscopic Survey: overview and early data”, *The Astronomical Journal*, v. 151, n. 2, pp. 44.
- DE JONG, J. T. A., YANNY, B., RIX, H.-W., et al., 2010, “Mapping the Stellar Structure of the Milky Way Thick Disk and Halo Using SEGUE Photometry”, *Astrophys. J.*, v. 714, n. 1 (maio), pp. 663–674. doi: 10.1088/0004-637X/714/1/663.
- DE JONG, R. S., BELLIDO-TIRADO, O., CHIAPPINI, C., et al., 2012, “4MOST: 4-metre multi-object spectroscopic telescope”. In: *Ground-based and Airborne Instrumentation for Astronomy IV*, v. 8446, p. 84460T. International Society for Optics and Photonics.
- DENG, L.-C., NEWBERG, H. J., LIU, C., et al., 2012, “LAMOST Experiment for Galactic Understanding and Exploration (LEGUE)—The survey’s science plan”, *Research in Astronomy and Astrophysics*, v. 12, n. 7, pp. 735.
- DJORGOVSKI, S. G., MAHABAL, A., DRAKE, A., et al., 2013, “Sky surveys”, *Planets, Stars and Stellar Systems*, v. 2, pp. 223–281.
- DREYER, J. L. E., 1888, “A New General Catalogue of Nebulæ and Clusters of Stars, being the Catalogue of the late Sir John F. W. Herschel, Bart, revised, corrected, and enlarged”, *Mem. Roy. Astron. Soc.*, v. 49 (jan.), pp. 1.
- EGGEN, O. J., LYNDEN-BELL, D., SANDAGE, A. R., 1962, “Evidence from the motions of old stars that the Galaxy collapsed.” *Astrophys. J.*, v. 136 (nov.), pp. 748. doi: 10.1086/147433.
- EINSTEIN, A., 1916, “Die Grundlage der allgemeinen Relativitätstheorie”, *Annalen der Physik*, v. 354, n. 7, pp. 769–822. doi: <https://doi.org/10.1002/andp.19163540702>. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19163540702>>.
- FIORENTIN, P. R., BAILER-JONES, C., LEE, Y. S., et al., 2007, “Estimation of stellar atmospheric parameters from SDSS/SEGUE spectra”, *Astronomy & Astrophysics*, v. 467, n. 3, pp. 1373–1387.

- FOUESNEAU, M., FRÉMAT, Y., ANDRAE, R., et al., 2022. “Gaia Data Release 3: Apsis II – Stellar Parameters”. Disponível em: <<https://arxiv.org/abs/2206.05992>>.
- FRIEDMAN, A., “Über die Krümmung des Raumes”, *Zeitschrift für Physik*, v. 10, pp. 377–386.
- FRIEDMAN, A., 1999, “On the Curvature of Space”, *General Relativity and Gravitation*, v. 31, pp. 1991–2000.
- GAIA COLLABORATION, PRUSTI, T., DE BRUIJNE, J. H. J., et al., 2016, “The Gaia mission”, *A&A*, v. 595, pp. A1. doi: 10.1051/0004-6361/201629272. Disponível em: <<https://doi.org/10.1051/0004-6361/201629272>>.
- GAIA COLLABORATION, BROWN, A. G. A., VALLENARI, A., et al., 2018, “Gaia Data Release 2. Summary of the contents and survey properties”, *Astron. Astrophys.*, 616:A1. doi: 10.1051/0004-6361/201833051.
- GALARZA, CARLOS ANDRÉS, DAFLON, SIMONE, PLACCO, VINICIUS M., et al., 2022, “J-PLUS: Searching for very metal-poor star candidates using the SPEEM pipeline”, *A&A*, v. 657, pp. A35. doi: 10.1051/0004-6361/202141717. Disponível em: <<https://doi.org/10.1051/0004-6361/202141717>>.
- GAMOW, G., 1946, “Expanding Universe and the Origin of Elements”, *Physical Review*, v. 70, n. 7-8 (out.), pp. 572–573. doi: 10.1103/PhysRev.70.572.2.
- GAO, X.-H., 2019, “Membership and fundamental parameters of the Praesepe cluster based on Gaia-DR2”, *Mon. Not. Roy. Astron. Soc.*, v. 486, n. 4 (jul.), pp. 5405–5413. doi: 10.1093/mnras/stz1213.
- GARCÍA PÉREZ, A. E., ALLENDE PRIETO, C., HOLTZMAN, J. A., et al., 2016, “ASPCAP: The APOGEE Stellar Parameter and Chemical Abundances Pipeline”, *Astron. J.*, 151(6):144. doi: 10.3847/0004-6256/151/6/144.
- GOSSAGE, S., CONROY, C., DOTTER, A., et al., 2018, “Age Determinations of the Hyades, Praesepe, and Pleiades via MESA Models with Rotation”, *The Astrophysical Journal*, v. 863, n. 1 (aug), pp. 67. doi: 10.3847/1538-4357/aad0a0. Disponível em: <<https://doi.org/10.3847/1538-4357/aad0a0>>.
- GUTH, A. H., PI, S. Y., 1982, “Fluctuations in the New Inflationary Universe”, *Phys. Rev. Lett.*, v. 49, n. 15 (out.), pp. 1110–1113. doi: 10.1103/PhysRevLett.49.1110.

- HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J., 2009, *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.
- HEITER, U., SOUBIRAN, C., NETOPIL, M., et al., 2014, “On the metallicity of open clusters. II. Spectroscopy”, *Astron. Astrophys.*, 561:A93. doi: 10.1051/0004-6361/201322559.
- HENDEN, A., MUNARI, U., 2014, “The APASS all-sky, multi-epoch BVgri photometric survey”, *Contributions of the Astronomical Observatory Skalnaté Pleso*, v. 43, n. 3 (mar.), pp. 518–522.
- HERSCHEL, J. F. W., 1863, “A General Catalogue of Nebulae and Clusters of Stars for the Year 1860.0, with Precessions for 1880.0. [Abstract]”, *Proceedings of the Royal Society of London*, v. 13, pp. 1–3. ISSN: 03701662. Disponível em: <<http://www.jstor.org/stable/111986>>.
- HOLTZMAN, J. A., HASSELQUIST, S., SHETRONE, M., et al., 2018, “APOGEE Data Releases 13 and 14: Data and Analysis”, *Astron. J.*, 156(3):125. doi: 10.3847/1538-3881/aad4f9.
- HOYLE, F., 1948, “A New Model for the Expanding Universe”, *Mon. Not. Roy. Astron. Soc.*, v. 108 (jan.), pp. 372. doi: 10.1093/mnras/108.5.372.
- HUANG, B., YUAN, H., 2022, “Photometric Recalibration of the SDSS Stripe 82 to a Few Millimagnitude Precision with the Stellar Color Regression Method and Gaia EDR3”, *Astrophys. J. Suppl.*, 259(1):26. doi: 10.3847/1538-4365/ac470d.
- HUANG, Y., YUAN, H., LI, C., et al., 2021, “Milky Way Tomography with the SkyMapper Southern Survey. II. Photometric Recalibration of SMSS DR2”, *Astrophys. J.*, 907(2):68. doi: 10.3847/1538-4357/abca37.
- HUBBLE, E. P., 1926, “Extragalactic nebulae.” *Astrophys. J.*, v. 64 (dez.), pp. 321–369. doi: 10.1086/143018.
- HUBBLE, E., 1929, “A Relation between Distance and Radial Velocity among Extragalactic Nebulae”, *Proceedings of the National Academy of Science*, v. 15, n. 3 (mar.), pp. 168–173. doi: 10.1073/pnas.15.3.168.
- ILBERT, O., CAPAK, P., SALVATO, M., et al., 2009, “Cosmos Photometric Redshifts with 30-Bands for 2-deg²”, *Astrophys. J.*, v. 690, n. 2 (jan.), pp. 1236–1249. doi: 10.1088/0004-637X/690/2/1236.

- IVEZIĆ, Ž., SESAR, B., JURIC, M., et al., 2008, “The Milky Way Tomography with SDSS. II. Stellar Metallicity”, *Astrophys. J.*, v. 684, n. 1 (set.), pp. 287–325. doi: 10.1086/589678.
- JÖNSSON, H., HOLTZMAN, J. A., ALLENDE PRIETO, C., et al., 2020, “APOGEE Data and Spectral Analysis from SDSS Data Release 16: Seven Years of Observations Including First Results from APOGEE-South”, *Astron. J.*, 160(3):120. doi: 10.3847/1538-3881/aba592.
- JORDI, C., HØG, E., BROWN, A. G. A., et al., 2006, “The design and performance of the Gaia photometric system”, *Mon. Not. Roy. Astron. Soc.*, v. 367, n. 1 (mar.), pp. 290–314. doi: 10.1111/j.1365-2966.2005.09944.x.
- KELLER, S. C., SCHMIDT, B. P., BESSELL, M. S., et al., 2007, “The SkyMapper Telescope and The Southern Sky Survey”, *Publ. Astron. Soc. Australia*, v. 24, n. 1 (maio), pp. 1–12. doi: 10.1071/AS07001.
- KIM, Y. K., LEE, Y. S., BEERS, T. C., 2019, “Dependence of Galactic Halo Kinematics on the Adopted Galactic Potential”, *The Astrophysical Journal*, v. 882, n. 2, pp. 176.
- KIM, Y. K., LEE, Y. S., BEERS, T. C., et al., 2021, “Evidence for Multiple Accretion Events in the Gaia-Sausage/Enceladus Structures”, *The Astrophysical Journal Letters*, v. 911, n. 2, pp. L21.
- KOLEVA, M., PRUGNIEL, P., BOUCHARD, A., et al., 2009, “ULySS: a full spectrum fitting package”, *Astronomy & Astrophysics*, v. 501, n. 3, pp. 1269–1279.
- KOTSIANTIS, S. B., ZAHARAKIS, I., PINTELAS, P., 2007, “Supervised machine learning: A review of classification techniques”, *Emerging artificial intelligence applications in computer engineering*, v. 160, n. 1, pp. 3–24.
- KROUPA, P., WEIDNER, C., PFLAMM-ALTENBURG, J., et al., 2013, “The Stellar and Sub-Stellar Initial Mass Function of Simple and Composite Populations”. In: Oswalt, T. D., Gilmore, G. (Eds.), *Planets, Stars and Stellar Systems. Volume 5: Galactic Structure and Stellar Populations*, v. 5, p. 115. doi: 10.1007/978-94-007-5612-0_4.
- LEE, Y. S., BEERS, T. C., SIVARANI, T., et al., 2008a, “The SEGUE stellar parameter pipeline. I. Description and comparison of individual methods”, *The Astronomical Journal*, v. 136, n. 5, pp. 2022.

- LEE, Y. S., BEERS, T. C., SIVARANI, T., et al., 2008b, “The SEGUE Stellar Parameter Pipeline. II. Validation with Galactic Globular and Open Clusters”, *Astron. J.*, v. 136, n. 5 (nov.), pp. 2050–2069. doi: 10.1088/0004-6256/136/5/2050.
- LEE, Y. S., BEERS, T. C., AN, D., et al., 2011, “Formation and Evolution of the Disk System of the Milky Way: $[\alpha/\text{Fe}]$ Ratios and Kinematics of the SEGUE G-dwarf Sample”, *Astrophys. J.*, 738(2):187. doi: 10.1088/0004-637X/738/2/187.
- LEE, Y. S., BEERS, T. C., MASSERON, T., et al., 2013, “Carbon-enhanced metal-poor stars in SDSS/SEGUE. I. Carbon abundance estimation and frequency of CEMP stars”, *The Astronomical Journal*, v. 146, n. 5, pp. 132.
- LEE, Y. S., BEERS, T. C., KIM, Y. K., et al., 2017, “Chemical Cartography. I. A Carbonicity Map of the Galactic Halo”, *The Astrophysical Journal*, v. 836, n. 1, pp. 91.
- LEE, Y. S., BEERS, T. C., KIM, Y. K., 2019, “Chemical Cartography. II. The Assembly History of the Galactic Stellar Halo Traced by Carbon-enhanced Metal-poor Stars”, *The Astrophysical Journal*, v. 885, n. 2, pp. 102.
- LEMAÎTRE, G., 1931, “Expansion of the universe, A homogeneous universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae”, *Mon. Not. Roy. Astron. Soc.*, v. 91 (mar.), pp. 483–490. doi: 10.1093/mnras/91.5.483.
- LENZ, D. D., NEWBERG, J., ROSNER, R., et al., 1998, “Photometric Separation of Stellar Properties Using SDSS Filters”, *Astrophys. J. Suppl.*, v. 119, n. 2 (dez.), pp. 121–140. doi: 10.1086/313155.
- LI, H. N., CHRISTLIEB, N., SCHÖRCK, T., et al., 2010, “The stellar content of the Hamburg/ESO survey. VI. Metallicity distribution of main-sequence turn-off stars in the Galactic halo”, *Astron. Astrophys.*, 521:A10. doi: 10.1051/0004-6361/201014797.
- LÓPEZ-SANJUAN, C., VARELA, J., CRISTÓBAL-HORNILLOS, D., et al., 2019a, “J-PLUS: photometric calibration of large-area multi-filter surveys with stellar and white dwarf loci”, *A&A*, v. 631, pp. A119. doi: 10.1051/0004-6361/201936405. Disponível em: <<https://doi.org/10.1051/0004-6361/201936405>>.
- LÓPEZ-SANJUAN, C., VÁZQUEZ RAMIÓ, H., VARELA, J., et al., 2019b, “J-PLUS: Morphological star/galaxy classification by PDF analysis”, *A&A*, v. 622, pp. A177. doi: 10.1051/0004-6361/201732480. Disponível em: <<https://doi.org/10.1051/0004-6361/201732480>>.

- LÓPEZ-SANJUAN, C., YUAN, H., VÁZQUEZ RAMIÓ, H., et al., 2021, “J-PLUS: Systematic impact of metallicity on photometric calibration with the stellar locus”, *arXiv e-prints*, art. arXiv:2101.12407.
- LUO, A. L., ZHAO, Y.-H., ZHAO, G., et al., 2015, “The first data release (DR1) of the LAMOST regular survey”, *Research in Astronomy and Astrophysics*, 15(8): 1095. doi: 10.1088/1674-4527/15/8/002.
- LUO, A.-L., ZHAO, Y.-H., ZHAO, G., et al., 2015, “The first data release (DR1) of the LAMOST regular survey”, *Research in Astronomy and Astrophysics*, v. 15, n. 8, pp. 1095.
- MACKEY, A. D., GILMORE, G. F., 2004, “Comparing the properties of local globular cluster systems: implications for the formation of the Galactic halo”, *Mon. Not. Roy. Astron. Soc.*, v. 355, n. 2 (dez.), pp. 504–534. doi: 10.1111/j.1365-2966.2004.08343.x.
- MAJEWSKI, S. R., APOGEE TEAM, APOGEE-2 TEAM, 2016, “The Apache Point Observatory Galactic Evolution Experiment (APOGEE) and its successor, APOGEE-2”, *Astronomische Nachrichten*, v. 337, n. 8-9 (set.), pp. 863. doi: 10.1002/asna.201612387.
- MAJEWSKI, S. R., WILSON, J. C., HEARTY, F., et al., 2009, “The Apache point observatory galactic evolution experiment (APOGEE) in sloan digital sky survey III (SDSS-III)”, *Proceedings of the International Astronomical Union*, v. 5, n. S265, pp. 480–481.
- MANN, A. W., FEIDEN, G. A., GAIDOS, E., et al., 2015, “How to Constrain Your M Dwarf: Measuring Effective Temperature, Bolometric Luminosity, Mass, and Radius”, *Astrophys. J.*, 804(1):64. doi: 10.1088/0004-637X/804/1/64.
- MARÍN-FRANCH, A., TAYLOR, K., CENARRO, J., et al., 2015, “T80Cam: a wide field camera for the J-PLUS survey”. In: *IAU General Assembly*, v. 29, p. 2257381, Aug.
- MARTÍNEZ, C., MANUEL., J., 2006, “Diseño y caracterización del sistema fotométrico de la misión GAIA de la Agencia Espacial Europea”. .
- MENDES DE OLIVEIRA, C., RIBEIRO, T., SCHOENELL, W., et al., 2019, “The Southern Photometric Local Universe Survey (S-PLUS): improved SEDs, morphologies, and redshifts with 12 optical filters”, *Mon. Not. Roy. Astron. Soc.*, v. 489, n. 1 (out.), pp. 241–267. doi: 10.1093/mnras/stz1985.

- MESSIER, C., 1781. “Catalogue des Nébuleuses et des Amas d’Étoiles (Catalog of Nebulae and Star Clusters)”. *Connaissance des Temps ou des Mouvements Célestes*, for 1784, p. 227-267, jan.
- MIGLIO, A., CHIAPPINI, C., MOREL, T., et al., 2013, “Galactic archaeology: mapping and dating stellar populations with asteroseismology of red-giant stars”, *Mon. Not. Roy. Astron. Soc.*, v. 429, n. 1 (fev.), pp. 423–428. doi: 10.1093/mnras/sts345.
- MILLER, A. A., BLOOM, J. S., RICHARDS, J. W., et al., 2015, “A MACHINE-LEARNING METHOD TO INFER FUNDAMENTAL STELLAR PARAMETERS FROM PHOTOMETRIC LIGHT CURVES”, *The Astrophysical Journal*, v. 798, n. 2 (jan), pp. 122. doi: 10.1088/0004-637x/798/2/122. Disponível em: <<https://doi.org/10.1088/0004-637x/798/2/122>>.
- MOLES, M., BENÍTEZ, N., AGUERRI, J. A. L., et al., 2008, “THE ALHAMBRA SURVEY: A LARGE AREA MULTIMEDIUM-BAND OPTICAL AND NEAR-INFRARED PHOTOMETRIC SURVEY”, *The Astronomical Journal*, v. 136, n. 3 (aug), pp. 1325–1339. doi: 10.1088/0004-6256/136/3/1325. Disponível em: <<https://doi.org/10.1088/0004-6256/136/3/1325>>.
- MOULTAKA, J., ILOVAISKY, S., PRUGNIEL, P., et al., 2004, “The ELODIE archive”, *Publications of the Astronomical Society of the Pacific*, v. 116, n. 821, pp. 693.
- NETOPIL, M., PAUNZEN, E., HEITER, U., et al., 2016, “On the metallicity of open clusters. III. Homogenised sample”, *Astron. Astrophys.*, 585:A150. doi: 10.1051/0004-6361/201526370.
- NIU, Z., YUAN, H., LIU, J., 2021a, “Correction to the Photometric Colors of the Gaia Data Release 2 with the Stellar Color Regression Method”, *Astrophys. J.*, 909 (1):48. doi: 10.3847/1538-4357/abdbac.
- NIU, Z., YUAN, H., LIU, J., 2021b, “Correction to the Photometric Colors of Gaia Early Data Release 3”, *Astrophys. J. Lett.*, 908(1):L14. doi: 10.3847/2041-8213/abe1c2.
- NORRIS, J. E., RYAN, S. G., BEERS, T. C., 1999, “A Search for Stars of Very Low Metal Abundance. III. UBV Photometry of Metal-weak Candidates”, *Astrophys. J. Suppl.*, v. 123, n. 2 (ago.), pp. 639–686. doi: 10.1086/313248.
- NORRIS, J. E., RYAN, S. G., BEERS, T. C., 2001, “Extremely Metal-Poor Stars. VIII. High-Resolution, High Signal-to-Noise Ratio Analysis of Five Stars with $[\text{Fe}/\text{H}] < -3.5$ ”, *Astrophys. J.*, v. 561, n. 2 (nov.), pp. 1034–1059. doi: 10.1086/323429.

- ODENKIRCHEN, M., GREBEL, E. K., DEHNEN, W., et al., 2003, “The Extended Tails of Palomar 5: A 10° Arc of Globular Cluster Tidal Debris”, *Astron. J.*, v. 126, n. 5 (nov.), pp. 2385–2407. doi: 10.1086/378601.
- PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., et al., 2011, “Scikit-learn: Machine Learning in Python”, *Journal of Machine Learning Research*, v. 12, pp. 2825–2830.
- PEEBLES, P. J. E., 1965, “The Black-Body Radiation Content of the Universe and the Formation of Galaxies.” *Astrophys. J.*, v. 142 (nov.), pp. 1317. doi: 10.1086/148417.
- PENZIAS, A. A., WILSON, R. W., 1965, “A Measurement of Excess Antenna Temperature at 4080 Mc/s.” *Astrophys. J.*, v. 142 (jul.), pp. 419–421. doi: 10.1086/148307.
- PLACCO, V. M., FREBEL, A., BEERS, T. C., et al., 2013, “METAL-POOR STARS OBSERVED WITH THE MAGELLAN TELESCOPE. I. CONSTRAINTS ON PROGENITOR MASS AND METALLICITY OF AGB STARS UNDERGOING C/N -PROCESS NUCLEOSYNTHESIS”, *The Astrophysical Journal*, v. 770, n. 2 (may), pp. 104. doi: 10.1088/0004-637x/770/2/104. Disponível em: <<https://doi.org/10.1088/0004-637x/770/2/104>>.
- PLACCO, V. M., BEERS, T. C., ROEDERER, I. U., et al., 2014, “HUBBLE SPACE TELESCOPE NEAR-ULTRAVIOLET SPECTROSCOPY OF THE BRIGHT CEMP-NO STAR BD44°493”, *The Astrophysical Journal*, v. 790, n. 1 (jun), pp. 34. doi: 10.1088/0004-637x/790/1/34. Disponível em: <<https://doi.org/10.1088/0004-637x/790/1/34>>.
- PLACCO, V. M., BEERS, T. C., IVANS, I. I., et al., 2015, “HUBBLE SPACE TELESCOPE NEAR-ULTRAVIOLET SPECTROSCOPY OF BRIGHT CEMP-sSTARS”, *The Astrophysical Journal*, v. 812, n. 2 (oct), pp. 109. doi: 10.1088/0004-637x/812/2/109. Disponível em: <<https://doi.org/10.1088/0004-637x/812/2/109>>.
- PLACCO, V. M., FREBEL, A., LEE, Y. S., et al., 2015, “Metal-poor Stars Observed with the Magellan Telescope. III. New Extremely and Ultra Metal-poor Stars from SDSS/SEGUE and Insights on the Formation of Ultra Metal-poor Stars”, *Astrophys. J.*, 809(2):136. doi: 10.1088/0004-637X/809/2/136.
- PLACCO, V. M., FREBEL, A., BEERS, T. C., et al., 2016, “Observational Constraints on First-Star Nucleosynthesis. II. Spectroscopy of an Ultra metal-poor CEMP-no Star”, *Astrophys. J.*, 833(1):21. doi: 10.3847/0004-637X/833/1/21.

- PLACCO, V. M., BEERS, T. C., SANTUCCI, R. M., et al., 2018, “Spectroscopic Validation of Low-metallicity Stars from RAVE”, *The Astronomical Journal*, v. 155, n. 6, pp. 256.
- POSTMAN, M., COE, D., BENÍTEZ, N., et al., 2012, “The Cluster Lensing and Supernova Survey with Hubble: An Overview”, *Astrophys. J. Suppl.*, 199(2):25. doi: 10.1088/0067-0049/199/2/25.
- PRIETO, C. A., MAJEWSKI, S., SCHIAVON, R., et al., 2008a, “APOGEE: the Apache point observatory galactic evolution experiment”, *Astronomische Nachrichten: Astronomical Notes*, v. 329, n. 9-10, pp. 1018–1021.
- PRIETO, C. A., SIVARANI, T., BEERS, T. C., et al., 2008b, “The SEGUE stellar parameter pipeline. III. Comparison with high-resolution spectroscopy of SDS-S/SEGUE field stars”, *The Astronomical Journal*, v. 136, n. 5, pp. 2070.
- PRUGNIEL, P., SOUBIRAN, C., 2001, “A database of high and medium-resolution stellar spectra”, *Astronomy & Astrophysics*, v. 369, n. 3, pp. 1048–1057.
- RIESS, A. G., FILIPPENKO, A. V., CHALLIS, P., et al., 1998, “Observational evidence from supernovae for an accelerating universe and a cosmological constant”, *The Astronomical Journal*, v. 116, n. 3, pp. 1009.
- ROEDERER, I. U., PRESTON, G. W., THOMPSON, I. B., et al., 2014, “A Search for Stars of Very Low Metal Abundance. VI. Detailed Abundances of 313 Metal-poor Stars”, *Astron. J.*, 147(6):136. doi: 10.1088/0004-6256/147/6/136.
- RYAN, S. G., NORRIS, J. E., BEERS, T. C., 1999, “The Spite Lithium Plateau: Ultrathin but Postprimordial”, *Astrophys. J.*, v. 523, n. 2 (out.), pp. 654–677. doi: 10.1086/307769.
- SCARINGI, S., GROOT, P., VERBEEK, K., et al., 2013, “Spectroscopic identifications of blue-H α -excess sources in the Kepler field of view”, *Monthly Notices of the Royal Astronomical Society*, v. 428, n. 3, pp. 2207–2215.
- SCHLAUFMAN, K. C., CASEY, A. R., 2014, “The Best and Brightest Metal-poor Stars”, *Astrophys. J.*, 797(1):13. doi: 10.1088/0004-637X/797/1/13.
- SCHLEGEL, D. J., FINKBEINER, D. P., DAVIS, M., 1998, “Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds”, *The Astrophysical Journal*, v. 500, n. 2 (jun), pp. 525–553. doi: 10.1086/305772. Disponível em: <<https://doi.org/10.1086/305772>>.

- SCHNEIDER, P., 2006, *Extragalactic Astronomy and Cosmology: An Introduction*. Springer. ISBN: 9783540331742. Disponível em: <<https://books.google.com.co/books?id=uP1Hz-6sHaMC>>.
- SCHNEIDER, P., 2014, *Extragalactic Astronomy and Cosmology: An Introduction*. Springer Berlin Heidelberg. ISBN: 9783642540837. Disponível em: <<https://books.google.com.co/books?id=kcrHBAAQBAJ>>.
- SCHÖRCK, T., CHRISTLIEB, N., COHEN, J. G., et al., 2009, “The stellar content of the Hamburg/ESO survey. V. The metallicity distribution function of the Galactic halo”, *Astron. Astrophys.*, v. 507, n. 2 (nov.), pp. 817–832. doi: 10.1051/0004-6361/200810925.
- SEARLE, L., ZINN, R., 1978, “Composition of halo clusters and the formation of the galactic halo.” *Astrophys. J.*, v. 225 (out.), pp. 357–379. doi: 10.1086/156499.
- SESTITO, F., LONGEARD, N., MARTIN, N. F., et al., 2019, “Tracing the formation of the Milky Way through ultra metal-poor stars”, *Mon. Not. Roy. Astron. Soc.*, v. 484, n. 2 (abr.), pp. 2166–2180. doi: 10.1093/mnras/stz043.
- SHANDARIN, S. F., ZELDOVICH, Y. B., 1989, “The large-scale structure of the universe: Turbulence, intermittency, structures in a self-gravitating medium”, *Reviews of Modern Physics*, v. 61, n. 2, pp. 185.
- SILK, J., 1968, “Cosmic Black-Body Radiation and Galaxy Formation”, *Astrophys. J.*, v. 151 (fev.), pp. 459. doi: 10.1086/149449.
- SINGH, H. P., GULATI, R. K., GUPTA, R., 1998, “Stellar spectral classification using principal component analysis and artificial neural networks”, *Monthly Notices of the Royal Astronomical Society*, v. 295, n. 2, pp. 312–318.
- STARKENBURG, E., MARTIN, N., YOUAKIM, K., et al., 2017, “The Pristine survey - I. Mining the Galaxy for the most metal-poor stars”, *Mon. Not. Roy. Astron. Soc.*, v. 471, n. 3 (nov.), pp. 2587–2604. doi: 10.1093/mnras/stx1068.
- STELLO, D., CHAPLIN, W. J., BRUNTT, H., et al., 2009, “Radius Determination of Solar-type Stars Using Asteroseismology: What to Expect from the Kepler Mission”, *Astrophys. J.*, v. 700, n. 2 (ago.), pp. 1589–1602. doi: 10.1088/0004-637X/700/2/1589.
- STROBL, C., MALLEY, J., TUTZ, G., 2009, “An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests.” *Psychological methods*, v. 14, n. 4, pp. 323.

- SÁNCHEZ-BLÁZQUEZ, P., PELETIER, R. F., JIMÉNEZ-VICENTE, J., et al., 2006, “Medium-resolution Isaac Newton Telescope library of empirical spectra”, *Monthly Notices of the Royal Astronomical Society*, v. 371, n. 2 (08), pp. 703–718. ISSN: 0035-8711. doi: 10.1111/j.1365-2966.2006.10699.x. Disponível em: <<https://doi.org/10.1111/j.1365-2966.2006.10699.x>>.
- VAN DEN BERGH, S., 1993, “The Proto-Galaxy, Globular Clusters, and Quasars”, *Astrophys. J.*, v. 411 (jul.), pp. 178. doi: 10.1086/172817.
- VARELA, J., J-PLUS COLLABORATION, 2021, “J-PLUS: second data release”, *in preparation*.
- WHITE, S. D. M., FRENK, C. S., 1991, “Galaxy Formation through Hierarchical Clustering”, *Astrophys. J.*, v. 379 (set.), pp. 52. doi: 10.1086/170483.
- WHITTEN, D., PLACCO, V., BEERS, T., et al., 2019, “J-PLUS: Identification of low-metallicity stars with artificial neural networks using SPHINX”, *Astronomy & Astrophysics*, v. 622, pp. A182.
- WHITTEN, D. D., PLACCO, V. M., BEERS, T. C., et al., 2021, “The Photometric Metallicity and Carbon Distributions of the Milky Way’s Halo and Solar Neighborhood from S-PLUS Observations of SDSS Stripe 82”, *Astrophys. J.*, 912(2): 147. doi: 10.3847/1538-4357/abee7e.
- WILSON, E. B., 1927, “Probable Inference, the Law of Succession, and Statistical Inference”, *Journal of the American Statistical Association*, v. 22, n. 158, pp. 209–212. doi: 10.1080/01621459.1927.10502953. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1080/01621459.1927.10502953>>.
- WOLF, C., MEISENHEIMER, K., RIX, H. W., et al., 2003, “The COMBO-17 survey: Evolution of the galaxy luminosity function from 25 000 galaxies with $0.2 < z < 1.2$ ”, *Astron. Astrophys.*, v. 401 (abr.), pp. 73–98. doi: 10.1051/0004-6361:20021513.
- WOLF, C., ONKEN, C. A., LUVAUL, L. C., et al., 2018, “SkyMapper Southern Survey: First Data Release (DR1)”, *Publications of the Astronomical Society of Australia*, v. 35, pp. e010. doi: 10.1017/pasa.2018.5.
- WRIGHT, E. L., EISENHARDT, P. R., MAINZER, A. K., et al., 2010, “The Wide-field Infrared Survey Explorer (WISE): mission description and initial on-orbit performance”, *The Astronomical Journal*, v. 140, n. 6, pp. 1868.
- WU, Y., LUO, A.-L., LI, H.-N., et al., 2011, “Automatic determination of stellar atmospheric parameters and construction of stellar spectral templates of the

- Guoshoujing Telescope (LAMOST)”, *Research in Astronomy and Astrophysics*, v. 11, n. 8, pp. 924.
- WU, Y., DU, B., LUO, A., et al., 2014, “Automatic stellar spectral parameterization pipeline for LAMOST survey”, *Proceedings of the International Astronomical Union*, v. 10, n. S306, pp. 340–342. doi: 10.1017/S1743921314010825.
- XIANG, M., LIU, X., YUAN, H., et al., 2015, “The LAMOST stellar parameter pipeline at Peking University–LSP3”, *Monthly Notices of the Royal Astronomical Society*, v. 448, n. 1, pp. 822–854.
- XIAO, K., YUAN, H., 2022, “Validation and Improvement of the Pan-STARRS Photometric Calibration with the Stellar Color Regression Method”, *Astron. J.*, 163 (4):185. doi: 10.3847/1538-3881/ac540a.
- YANG, L., YUAN, H., ZHANG, R., et al., 2021, “Correction to the Photometric Magnitudes of the Gaia Early Data Release 3”, *Astrophys. J. Lett.*, 908(2):L24. doi: 10.3847/2041-8213/abdbae.
- YANG, LIN, YUAN, HAIBO, XIANG, MAOSHENG, et al., 2022, “J-PLUS: Stellar parameters, C, N, Mg, Ca, and [abundances for two million stars from DR1”, *A&A*, v. 659, pp. A181. doi: 10.1051/0004-6361/202142724. Disponível em: <<https://doi.org/10.1051/0004-6361/202142724>>.
- YANNY, B., ROCKOSI, C., NEWBERG, H. J., et al., 2009, “SEGUE: A Spectroscopic Survey of 240,000 Stars with $g= 14-20$ ”, *The Astronomical Journal*, v. 137, n. 5, pp. 4377.
- YORK, D. G., ADELMAN, J., ANDERSON, JR., J. E., et al., 2000, “The Sloan Digital Sky Survey: Technical Summary”, *Astron. J.*, v. 120 (set.), pp. 1579–1587. doi: 10.1086/301513.
- YOUAKIM, K., STARKENBURG, E., AGUADO, D. S., et al., 2017, “The Pristine survey - III. Spectroscopic confirmation of an efficient search for extremely metal-poor stars”, *Mon. Not. Roy. Astron. Soc.*, v. 472, n. 3 (dez.), pp. 2963–2974. doi: 10.1093/mnras/stx2005.
- YOUAKIM, K., STARKENBURG, E., MARTIN, N. F., et al., 2020, “The Pristine Survey – VIII. The metallicity distribution function of the Milky Way halo down to the extremely metal-poor regime”, *Monthly Notices of the Royal Astronomical Society*, v. 492, n. 4 (01), pp. 4986–5002. ISSN: 0035-8711. doi: 10.1093/mnras/stz3619. Disponível em: <<https://doi.org/10.1093/mnras/stz3619>>.

- YUAN, H.-B., LIU, X.-W., XIANG, M.-S., 2013, “Empirical extinction coefficients for the GALEX, SDSS, 2MASS and WISE passbands”, *Monthly Notices of the Royal Astronomical Society*, v. 430, n. 3, pp. 2188–2199.
- YUAN, H., LIU, X., XIANG, M., et al., 2015, “Stellar Color Regression: A Spectroscopy-based Method for Color Calibration to a Few Millimagnitude Accuracy and the Recalibration of Stripe 82”, *Astrophys. J.*, 799(2):133. doi: 10.1088/0004-637X/799/2/133.
- ZACHARIAS, N., FINCH, C. T., GIRARD, T. M., et al., 2013, “The Fourth US Naval Observatory CCD Astrograph Catalog (UCAC4)”, *Astron. J.*, 145(2):44. doi: 10.1088/0004-6256/145/2/44.
- ZEL'DOVICH, Y. B., 1970, “Gravitational instability: An approximate theory for large density perturbations.” *Astron. Astrophys.*, v. 5 (mar.), pp. 84–89.
- ZHAO, G., ZHAO, Y.-H., CHU, Y.-Q., et al., 2012, “LAMOST spectral survey—An overview”, *Research in Astronomy and Astrophysics*, v. 12, n. 7, pp. 723.
- ZINN, R., 1993, “The Dwarf Spheroidal Satellite Galaxies of the Milky Way”. In: Smith, G. H., Brodie, J. P. (Eds.), *The Globular Cluster-Galaxy Connection*, v. 48, *Astronomical Society of the Pacific Conference Series*, p. 302, jan.
- ZWICKY, F., 1935, “Remarks on the redshift from nebulae”, *Physical Review*, v. 48, pp. 802–806.
- ZWICKY, F., 1937, “On the Masses of Nebulae and of Clusters of Nebulae”, *The Astrophysical Journal*, v. 86, pp. 217–246.

Apêndice A

Como baixar dados do J-PLUS

Este apêndice tem como objetivo apresentar a query que foi utilizada para obter os dados J-PLUS DR2 para formar a base da amostra de ouro mencionada ao longo da tese. Os resultados aqui apresentados devem ser completamente reprodutíveis para qualquer leitor que use este trabalho como referência para desenvolver outras pesquisas relacionadas.

```
SELECT TILE_ID,NUMBER,ALPHA_J2000,DELTA_J2000,FWHM_WORLD,
-- J-PLUS Magnitudes
MAG_APER_6_0[jplus::uJAVA] as uJAVA,
MAG_ERR_APER_6_0[jplus::uJAVA] as err_uJAVA,
FLAGS[jplus::uJAVA] as FLAGS_uJAVA,
NORM_WMAP_VAL[jplus::uJAVA] as WMAP_uJAVA,
MASK_FLAGS[jplus::uJAVA] as MASK_uJAVA,
SINGLE_DETECT[jplus::uJAVA] as SINGLE_DETECT_uJAVA,
MAG_APER_6_0[jplus::J0378] as J0378,
MAG_ERR_APER_6_0[jplus::J0378] as err_J0378,
FLAGS[jplus::J0378] as FLAGS_J0378,
NORM_WMAP_VAL[jplus::J0378] as WMAP_J0378,
MASK_FLAGS[jplus::J0378] as MASK_J0378,
SINGLE_DETECT[jplus::J0378] as SINGLE_DETECT_J0378,
MAG_APER_6_0[jplus::J0395] as J0395,
MAG_ERR_APER_6_0[jplus::J0395] as err_J0395,
FLAGS[jplus::J0395] as FLAGS_J0395,
NORM_WMAP_VAL[jplus::J0395] as WMAP_J0395,
MASK_FLAGS[jplus::J0395] as MASK_J0395,
SINGLE_DETECT[jplus::J0395] as SINGLE_DETECT_J0395,
MAG_APER_6_0[jplus::J0410] as J0410,
MAG_ERR_APER_6_0[jplus::J0410] as err_J0410,
FLAGS[jplus::J0410] as FLAGS_J0410,
```

```
NORM_WMAP_VAL[jplus::J0410] as WMAP_J0410,  
MASK_FLAGS[jplus::J0410] as MASK_J0410,  
SINGLE_DETECT[jplus::J0410] as SINGLE_DETECT_J0410,  
MAG_APER_6_0[jplus::J0430] as J0430,  
MAG_ERR_APER_6_0[jplus::J0430] as err_J0430,  
FLAGS[jplus::J0430] as FLAGS_J0430,  
NORM_WMAP_VAL[jplus::J0430] as WMAP_J0430,  
MASK_FLAGS[jplus::J0430] as MASK_J0430,  
SINGLE_DETECT[jplus::J0430] as SINGLE_DETECT_J0430,  
MAG_APER_6_0[jplus::gSDSS] as gSDSS,  
MAG_ERR_APER_6_0[jplus::gSDSS] as err_gSDSS,  
FLAGS[jplus::gSDSS] as FLAGS_gSDSS,  
NORM_WMAP_VAL[jplus::gSDSS] as WMAP_gSDSS,  
MASK_FLAGS[jplus::gSDSS] as MASK_gSDSS,  
SINGLE_DETECT[jplus::gSDSS] as SINGLE_DETECT_gSDSS,  
MAG_APER_6_0[jplus::J0515] as J0515,  
MAG_ERR_APER_6_0[jplus::J0515] as err_J0515,  
FLAGS[jplus::J0515] as FLAGS_J0515,  
NORM_WMAP_VAL[jplus::J0515] as WMAP_J0515,  
MASK_FLAGS[jplus::J0515] as MASK_J0515,  
SINGLE_DETECT[jplus::J0515] as SINGLE_DETECT_J0515,  
MAG_APER_6_0[jplus::rSDSS] as rSDSS,  
MAG_ERR_APER_6_0[jplus::rSDSS] as err_rSDSS,  
FLAGS[jplus::rSDSS] as FLAGS_rSDSS,  
NORM_WMAP_VAL[jplus::rSDSS] as WMAP_rSDSS,  
MASK_FLAGS[jplus::rSDSS] as MASK_rSDSS,  
SINGLE_DETECT[jplus::rSDSS] as SINGLE_DETECT_rSDSS,  
MAG_APER_6_0[jplus::J0660] as J0660,  
MAG_ERR_APER_6_0[jplus::J0660] as err_J0660,  
FLAGS[jplus::J0660] as FLAGS_J0660,  
NORM_WMAP_VAL[jplus::J0660] as WMAP_J0660,  
MASK_FLAGS[jplus::J0660] as MASK_J0660,  
SINGLE_DETECT[jplus::J0660] as SINGLE_DETECT_J0660,  
MAG_APER_6_0[jplus::iSDSS] as iSDSS,  
MAG_ERR_APER_6_0[jplus::iSDSS] as err_iSDSS,  
FLAGS[jplus::iSDSS] as FLAGS_iSDSS,  
NORM_WMAP_VAL[jplus::iSDSS] as WMAP_iSDSS,  
MASK_FLAGS[jplus::iSDSS] as MASK_iSDSS,  
SINGLE_DETECT[jplus::iSDSS] as SINGLE_DETECT_iSDSS,
```

```

MAG_APER_6_0[jplus::J0861] as J0861,
MAG_ERR_APER_6_0[jplus::J0861] as err_J0861,
FLAGS[jplus::J0861] as FLAGS_J0861,
NORM_WMAP_VAL[jplus::J0861] as WMAP_J0861,
MASK_FLAGS[jplus::J0861] as MASK_J0861,
SINGLE_DETECT[jplus::J0861] as SINGLE_DETECT_J0861,
MAG_APER_6_0[jplus::zSDSS] as zSDSS,
MAG_ERR_APER_6_0[jplus::zSDSS] as err_zSDSS,
FLAGS[jplus::zSDSS] as FLAGS_zSDSS,
NORM_WMAP_VAL[jplus::zSDSS] as WMAP_zSDSS,
MASK_FLAGS[jplus::zSDSS] as MASK_zSDSS,
SINGLE_DETECT[jplus::zSDSS] as SINGLE_DETECT_zSDSS,
morph_prob_star
FROM jplus.MagABDualObj
JOIN jplus.StarGalClass USING (tile_id, number)
WHERE ARRAY_MAX_FLOAT(MAG_APER_6_0) < 21.0 AND
ARRAY_MAX_FLOAT(MAG_ERR_APER_6_0) < 0.1 AND
ARRAY_MAX_FLOAT(NORM_WMAP_VAL) > 0.8 AND
ARRAY_MAX_INT(MASK_FLAGS) = 0 AND
ARRAY_MAX_INT(FLAGS) = 0 AND
ARRAY_MAX_INT(SINGLE_DETECT) > 0 AND
morph_prob_star > 0.9

```

Existem ao menos duas formas relativamente simples de executar a query apresentada e baixar os dados numa estrutura de tabela: Executando a query diretamente no site oficial do J-PLUS ou executando a query usando o software Topcat.

A.1 Executando a query no website J-PLUS

Os passos para escrever queries e baixar dados são mostrados nas Figuras [A.1](#), [A.2](#) e [A.3](#). As Figuras são autoexplicativas e leitor pode utilizá-las como referência para escrever suas próprias queries e baixar os dados para desenvolver diferentes pesquisas.

A.2 Executando a query no Topcat

Nas Figuras [A.4](#), [A.5](#) e [A.6](#) se apresenta um procedimento similar para acessar os dados J-PLUS, porém usando o software livre Topcat¹.

¹<http://www.star.bris.ac.uk/~mbt/topcat/>

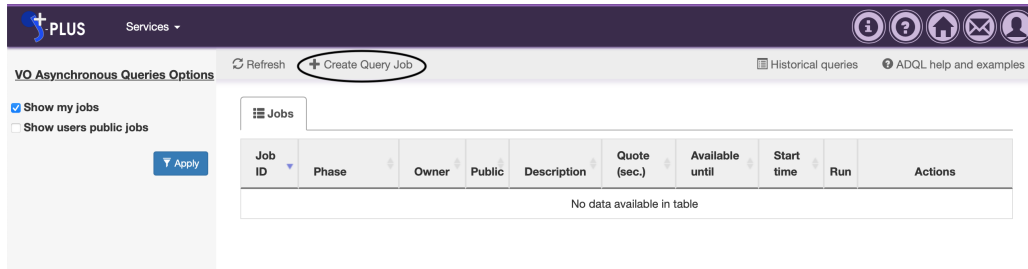


Figura A.1: Procedimento para usar VO services no portal do J-PLUS.

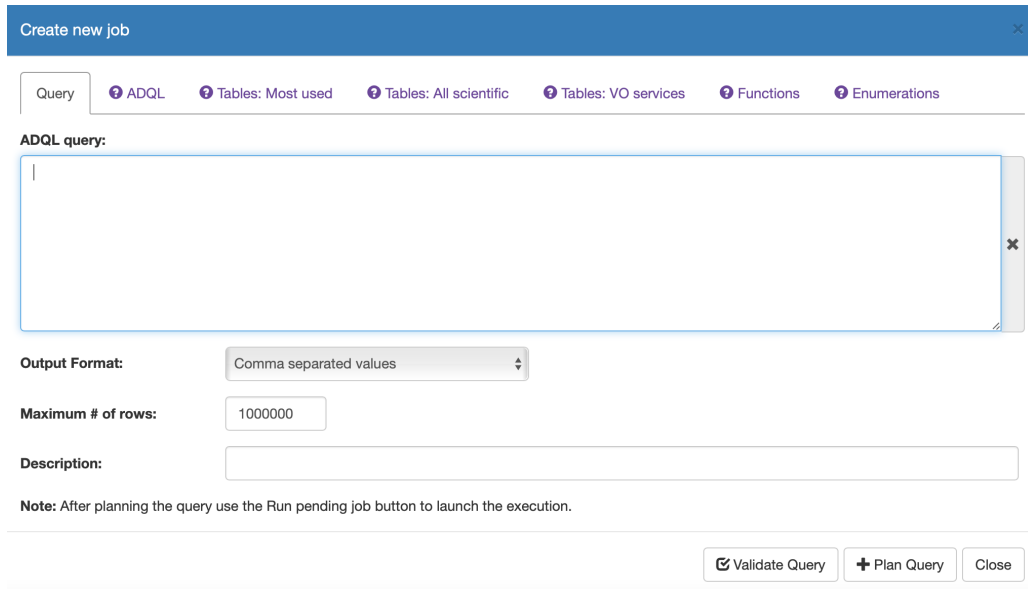


Figura A.2: Plataforma para escrever queries e gerar catalogos de observações feitas com J-PLUS.

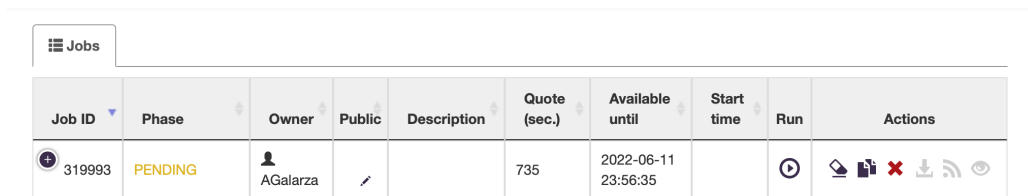


Figura A.3: Processo de compilação de queries e descarga de dados para uso em diferentes projetos de pesquisa.

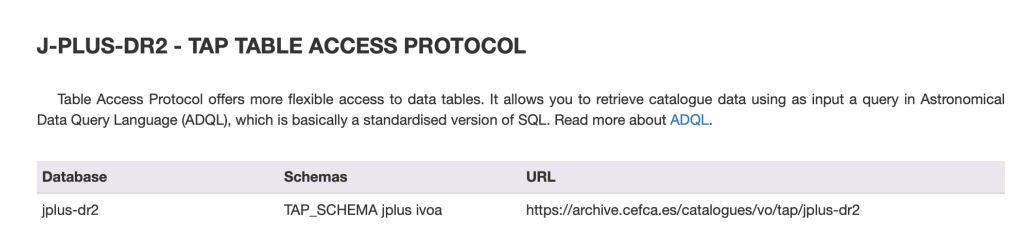


Figura A.4: Protocolo para acessar aos dados J-PLUS com Topcat.

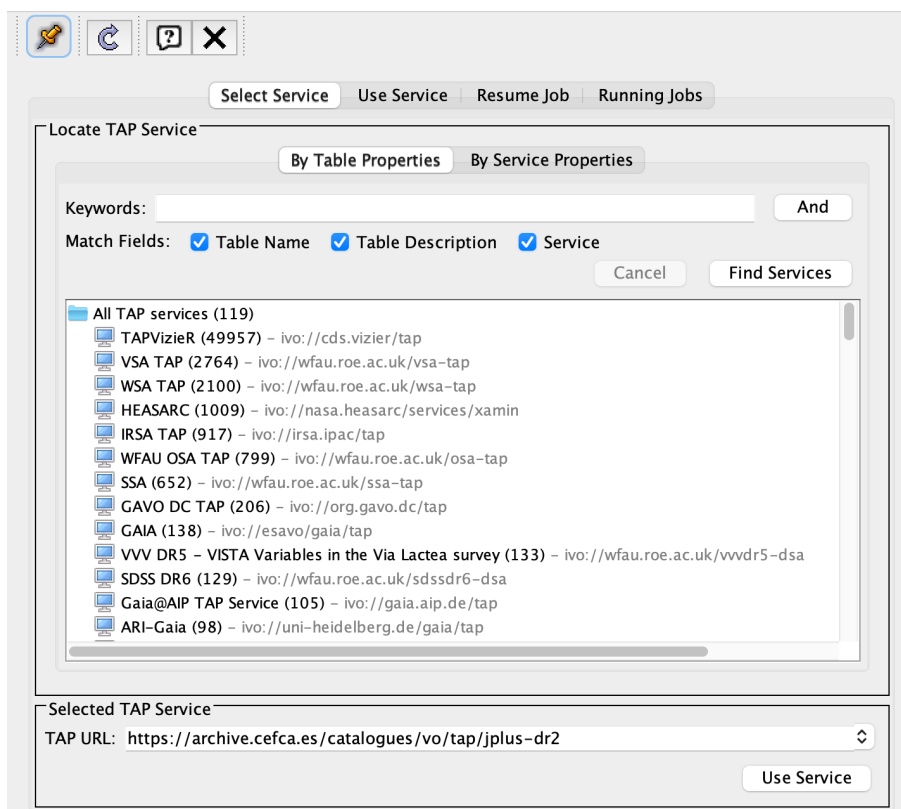


Figura A.5: Seleção do serviço para acessar aos dados J-PLUS com Topcat.

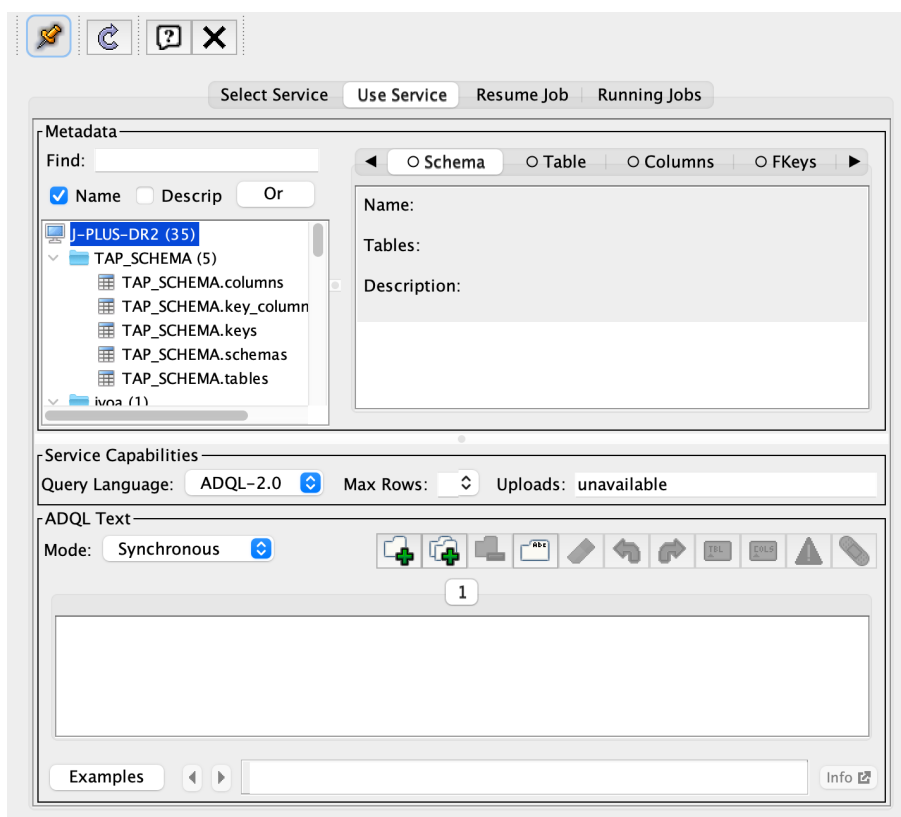


Figura A.6: Plataforma para escrever queries e gerar catalogos de observações feitas com J-PLUS no Topcat.

Apêndice B

Cruzamento com outras fontes

Este apêndice tem como objetivo apresentar um tutorial curto de como fazer cruzamentos (usando o software Topcat) entre diferentes levantamentos astronômicos para definir a amostra de treinamento usada para treinar o modelo SPEEM mencionado ao longo da tese. Os resultados aqui apresentados devem ser completamente reproduzíveis para qualquer leitor que use este trabalho como referência para desenvolver outras pesquisas relacionadas.

B.1 Carregar os dados e cruzar com bancos de dados públicos

O software Topcat permite carregar dados em diferentes formatos (fits, csv,...) para criar tabelas internas que possam ser relacionadas com outras fontes disponíveis na internet usando o icone X sinalizado na Figura B.1. Para fontes internas se usa o icone mostrado na Figura B.2, e janela para carregar as bases de dados aparece na Figura B.3.

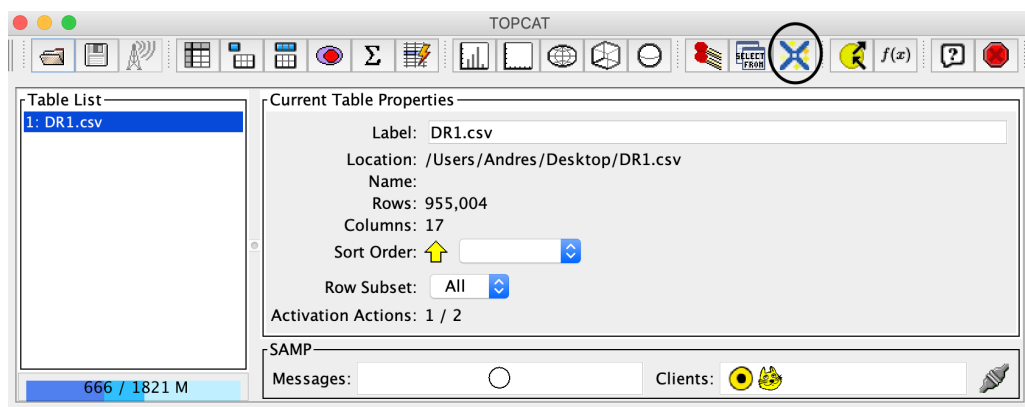


Figura B.1: Como fazer cruzamentos entre diferentes fontes de dados externos com Topcat.

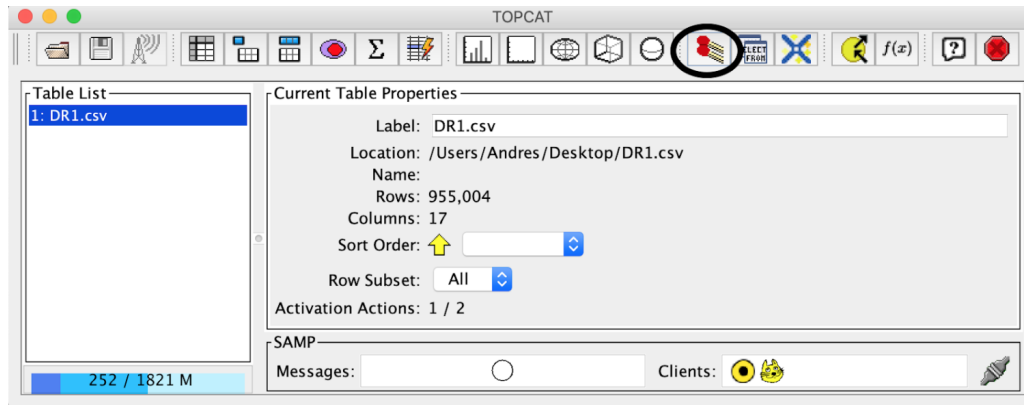


Figura B.2: Como fazer cruzamentos entre diferentes fontes de dados internos com Topcat.

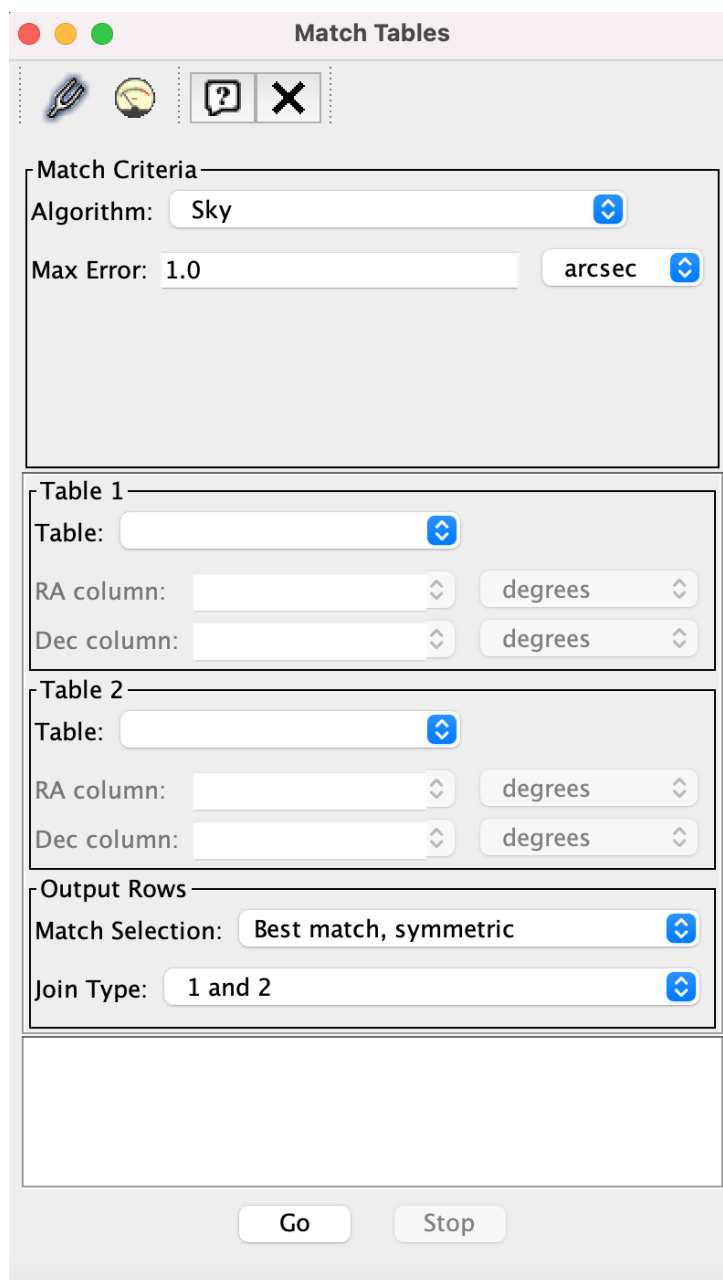


Figura B.3: Janela de seleção de dados para fazer cruzamentos de dados.

Apêndice C

Código Python para treinar o SPEEM

Este apêndice tem como objetivo apresentar o código que foi escrito para desenvolver o modelo SPEEM mencionado ao longo da tese. O objetivo é que as linhas de código aqui apresentadas possam ser adaptadas e reproduzíveis para qualquer leitor que use este trabalho como referência para desenvolver outras pesquisas relacionadas dando os créditos respectivos.

```
1 from sklearn import set_config
2 import AstroTools
3 import pandas as pd
4 import numpy as np
5 import sklearn
6
7 print("sklearn: {}".format(sklearn.__version__))
8 print("Pandas Version: {}".format(pd.__version__))
9 set_config(display="diagram")
10
11 path = AstroTools.Path()
12 AstroTools.ChangeFolder(path , "/Path_to_files/J-PLUS/Datasets/DR2/")
```

O pacote AstroTools corresponde a um arquivo de funções escritas pelo autor com o objetivo de simplificar o código e facilitar o processamento de dados. Caso seja necessário esse arquivo pode ser solicitado ao autor.

```
1 import pickle
2
3 tiles, acor = pickle.load(open("jplus_dr2_6total.pkl", "rb"))
4
5 Aper_corrs = {"TILE_ID" : tiles, "u_ApCor" : acor[:,0],
6               "J0378_ApCor" : acor[:,1], "J0395_ApCor" : acor[:,2],
7               "J0410_ApCor" : acor[:,3], "J0430_ApCor" : acor[:,4],
8               "g_ApCor" : acor[:,5], "J0515_ApCor" : acor[:,6],
9               "r_ApCor" : acor[:,7], "J0660_ApCor" : acor[:,8],
10              "i_ApCor" : acor[:,9], "J0861_ApCor" : acor[:,10],
11              "z_ApCor" : acor[:,11]}
```

```

12
13 acor_6 = pd.DataFrame(Aper_corr)

1
2 stellar = ["FLAG", "SNR", "SPECTYPE_SUBCLASS", "Teff_GAIA",
3           "TEFF_ADOP", "LOGG_ADOP", "FEH_ADOP", "FEH_ANRR",
4           "TEFF_SPEC", "LOGG_SPEC", "FEH_SPEC"]
5
6 Filters = list(AstroTools.Filters["JPLUS"])
7
8 def correctMags(df):
9     '''
10    This function corrects the J-PLUS magnitudes by aperture correction
11    and reddening
12    '''
13
14    Filters = list(AstroTools.Filters["JPLUS"])
15    Correction = list(AstroTools.Filters["Ax"])
16    A_Correction = list(acor_6.columns[1:])
17
18    dfcp = pd.concat([df_result[a].sub(df_result[b]).add(df_result[c])
19                    for a, b, c in zip(Filters, Correction, A_Correction)], axis=1, keys=
20                    Filters).round(3)
21
22    return dfcp
23
24 def add_id_columns(df1, df2, columns=AstroTools.Reference["Full"]):
25
26    df = df1.join(df2[columns])
27
28    return df
29
30 def order_columns(df, reference=AstroTools.Reference["Full"],
31                  filters=AstroTools.Filters["JPLUS"]):
32
33    return df[reference + filters]
34
35 def add_gaia_filters(df1, df2, columns=AstroTools.Filters["GAIA"]):
36
37    df = df1.join(df2[columns])
38
39    return df
40
41 def add_wise_filters(df1, df2, columns=AstroTools.Filters["WISE"]):
42
43    df = df1.join(df2[columns])

```

```

42     return df
43
44 def add_stellar_params(df1, df2, columns=stellar):
45
46     df = df1.join(df2[columns])
47
48     return df
49
50
51
52 def remove_contaminants(df):
53
54     dfcp = df[df.SPECTYPE_SUBCLASS.str.contains("R|S|W") == False].copy()
55
56     return dfcp
57
58 def remove_bad_values(df, Column=["TEFF_ADOP", "LOGG_ADOP", "FEH_ADOP"]):
59
60     cond1 = df[Column[0]].gt(-9999.0)
61     cond2 = df[Column[1]].gt(-9999.0)
62     cond3 = df[Column[2]].gt(-9999.0)
63     allcond = cond1 & cond2 & cond3
64     return df[allcond].copy()
65
66 def remove_spec_subclasses(df):
67
68     # Replace Original Column
69     df = df.assign(SPECTYPE_SUBCLASS = df.SPECTYPE_SUBCLASS.str.replace(
70         r"(\d)", "", regex=True))
71
72     return df
73
74 def remove_bad_flags(df):
75
76     dfcp = df[df.FLAG.str.contains("nnnnn") == True].copy()
77
78     return dfcp

```

```

1 df = AstroTools.Read_Dataset("JPLUS_DR2_Training_Dataset.csv")
2 df_result = pd.merge(df, acor_6, how="inner", on="TILE_ID")
3
4     # MW Correction
5 Working_df_SSPP = (df_result.pipe(correctMags)
6     # Add TILE_ID, NUMBER, RA, DEC
7     .pipe(add_id_columns, df_result)
8     # Reorder columns

```

```

9         .pipe(order_columns)
10        # Add GAIA Filters
11        .pipe(add_gaia_filters, df_result)
12        # Add WISE and 2MASS Filters
13        .pipe(add_wise_filters, df_result)
14        # Add TEFF_ADOP, FEH_ADOP, LOGG_ADOP to Dataframe
15        .pipe(add_stellar_params, df_result)
16        # Remove QSOs, BROADLINES, STARFORMING
17        .pipe(remove_contaminants)
18        # Remove -9999.0 values
19        .pipe(remove_bad_values)
20        .pipe(remove_spec_subclasses)
21    )
22
23 Working_df_SSPP.dropna(inplace=True)
24 Working_df_SSPP.isnull().sum()

```

```

1
2 from sklearn.base import TransformerMixin
3
4 class FeaturesSelection(TransformerMixin):
5     # Useful class to create custom transformer for features selection
6
7     def __init__(self, cols):
8         self.cols = cols
9
10    def fit(self, X, y=None):
11        return self
12
13    def transform(self, X):
14        Xcols = X[self.cols]
15        return Xcols
16
17 # Create custom transformer for build JPLUS colors
18
19 class MakeColors():
20     def __init__(self, func):
21         self.func = func
22
23     def fit(self, X, y=None, **fit_params):
24         return self
25
26     def transform(self, X, **transform_params):
27         return self.func(X)
28
29 # Create custom transformer for build Pristine indicator
30 class MakePristine():

```

```

31     def __init__(self, func):
32         self.func = func
33
34     def fit(self, X, y=None, **fit_params):
35         return self
36
37     def transform(self, X, **transform_params):
38         return self.func(X)
39
40
41 # Create Pristine Pseudo-color
42
43 def Pristine(df):
44
45     df = \
46     df.assign(J0395gSDSSgSDSSiSDSS=df.J0395-df.gSDSS-df.gSDSS+df.iSDSS)
47
48     return df

```

```

1 from sklearn.model_selection import train_test_split, cross_validate,
   RandomizedSearchCV, GridSearchCV
2
3
4 Working_df_SSPP["FeH_Class"] = \
5 Working_df_SSPP.FEH_ADOP.apply(lambda x: 0 if (x > -1.0)\
6                                 else (1 if (x > -2.0)\
7                                     else (2 if (x > -3.0)\
8                                             else 3)))
9
10 Filters = AstroTools.Filters["JPLUS"]
11
12 X, y = \
13 [Working_df_SSPP.drop(columns=["TEFF_ADOP", "FEH_ADOP", "LOGG_ADOP"]),\
14 Working_df_SSPP[["TEFF_ADOP", "FEH_ADOP", "LOGG_ADOP"]]]
15
16 X_train, X_test, y_train, y_test = \
17 train_test_split(X, y, test_size=0.1,\
18                 stratify=Working_df_SSPP.FeH_Class,\
19                 random_state=42)
20
21 print(X_train.shape)
22 print(X_test.shape)

```

```

1 from sklearn.pipeline import Pipeline, FeatureUnion
2
3         # Teff indicators narrow-band
4 JPLUS_Colors2 = ["J0410J0861", "J0430J0660", "J0430J0861", "J0515J0861",
5                 # Teff indicators broad-band

```

```

6         "gSDSSiSDSS", "gSDSSzSDSS", "gSDSSrSDSS",
7         # [Fe/H] & Logg indicators
8         "uJAVAJ0378", "uJAVAgSDSS",
9         # [Fe/H] indicators
10        "uJAVAJ0430", "uJAVAJ0410",
11        "J0378J0395", "J0378J0410", "J0378J0430",
12        "J0395J0410", "J0395J0430",
13        # Logg indicators
14        "uJAVAJ0395", "J0660J0861"]
15
16        # Teff indicators
17 JPLUS_Colors4 = ["uJAVAJ0395J0410zSDSS", "J0378J0395J0410J0861",
18                 "J0430iSDSSJ0861zSDSS"
19                 # [Fe/H] indicators
20                 "J0378J0410gSDSSiSDSS", "J0378J0410J0515J0861",
21                 "J0378J0430gSDSSiSDSS", "J0378J0430gSDSSzSDSS",
22                 "J0378J0430gSDSSJ0861", "J0395J0410J0660J0861",
23                 "J0395J0430J0660J0861", "J0395J0430J0660zSDSS",
24                 "J0395J0430rSDSSJ0861", "J0395gSDSSJ0515J0861",
25                 # Logg indicators
26                 "uJAVAJ0378J0861zSDSS", "J0378J0410J0430J0861"]
27
28 Pristine_Indicator = ["J0395gSDSSgSDSSiSDSS"]
29
30 best_jplus_colors2 = \
31 Pipeline([("filters", FeaturesSelection(Filters)),\
32           ("colors_1", MakeColors(AstroTools.createColors2)),\
33           ("features_select", FeaturesSelection(JPLUS_Colors2))])
34
35 best_jplus_colors4 = \
36 Pipeline([("colors_2", MakeColors(AstroTools.colorsComb)),\
37           ("features_select", FeaturesSelection(JPLUS_Colors4))])
38
39 pristine_indicator = \
40 Pipeline([("pristine", MakePristine(Pristine)),\
41           ("features_select", FeaturesSelection(Pristine_Indicator))])
42
43 union = FeatureUnion([("Colors2", best_jplus_colors2),\
44                       ("Colors4", best_jplus_colors4),\
45                       ("Pristine", pristine_indicator)])

```

```

1 from sklearn.preprocessing import StandardScaler
2 from sklearn.compose import TransformedTargetRegressor
3 from scipy.special import exp10
4 from sklearn.multioutput import MultiOutputRegressor
5 from sklearn.ensemble import RandomForestRegressor
6 from xgboost import XGBRegressor

```

```
7 from sklearn.metrics import (mean_absolute_error, median_absolute_error,
8                               r2_score, max_error, mean_squared_error,
9                               explained_variance_score)
10
11 def Log10(X, column=0):
12     X[:, column] = np.log10(X[:, column])
13     return X
14
15
16 def Exp10(X, column=0):
17
18     X[:, column] = exp10(X[:, column])
19     return X
20
21 reg_rf = RandomForestRegressor(n_estimators=350, max_features=15, \
22                               bootstrap=True, random_state=42)
23
24 reg_xgb = XGBRegressor(booster="gbtree", subsample=0.5, \
25                       scale_pos_weight=90, num_parallel_tree=10, \
26                       max_delta_step=10, n_estimators = 100)
27
28 Model_RF = Pipeline([("features", union), \
29                      ("std_scaler", StandardScaler()), \
30                      ("rgr", MultiOutputRegressor(reg_rf))])
31
32 Model_XGB = Pipeline([("features", union), \
33                       ("std_scaler", StandardScaler()), \
34                       ("rgr", MultiOutputRegressor(reg_xgb))])
35
36 Model_FINAL = TransformedTargetRegressor(Model_XGB, func=Log10, \
37                                           inverse_func=Exp10)
38 Model_FINAL.fit(X_train, y_train)
```