



Trabalho de Conclusão de Curso

**Astroestatística de estrelas variáveis:
identificação de variabilidade estelar em
aglomerados**

Maitê Mückler

2 de junho de 2021

Maitê Mückler

**Astroestatística de estrelas variáveis:
identificação de variabilidade estelar em aglomerados**

Trabalho de Conclusão apresentado à comissão de Graduação do Departamento de Estatística da Universidade Federal do Rio Grande do Sul, como parte dos requisitos para obtenção do título de Bacharel em Estatística.

Orientador: Prof. Dr. José Eduardo da Silveira Costa

Porto Alegre
Maio de 2021

Maitê Mückler

**Astroestatística de estrelas variáveis:
identificação de variabilidade estelar em aglomerados**

Este Trabalho foi julgado adequado para obtenção dos créditos da disciplina Trabalho de Conclusão de Curso em Estatística e aprovado em sua forma final pelo Orientador e pela Banca Examinadora.

Orientador: Prof. Dr. José Eduardo da Silveira Costa, UFRGS Doutor em Física pela Universidade Federal do Rio Grande do Sul, Rio Grande do Sul, Brasil

Banca Examinadora:

Profa. Dra. Márcia Helena Barbian, UFRGS
Doutora pela Universidade Federal de Minas Gerais - Belo Horizonte, MG

Porto Alegre
Maio de 2021

Our village life would stagnate if it were not for the unexplored forests and meadows which surround it. We need the tonic of wildness – to wade sometimes in marshes where the bittern and the meadow-hen lurk, and hear the booming of the snipe; to smell the whispering sedge where only some wilder and more solitary fowl builds her nest, and the mink crawls with its belly close to the ground.

At the same time that we are earnest to explore and learn all things, we require that all things be mysterious and unexplorable, that land and sea be infinitely wild, unsurveyed and unfathomed by us because unfathomable. We can never have enough of nature.

We must be refreshed by the sight of inexhaustible vigor, vast and titanic features, the sea-coast with its wrecks, the wilderness with its living and its decaying trees, the thunder-cloud, and the rain which lasts three weeks and produces freshets. We need to witness our own limits transgressed, and some life pasturing freely where we never wander. We are cheered when we observe the vulture feeding on the carrion which disgusts and disheartens us, and deriving health and strength from the repast.

There was a dead horse in the hollow by the path to my house, which compelled me sometimes to go out of my way, especially in the night when the air was heavy, but the assurance it gave me of the strong appetite and inviolable health of Nature was my compensation for this. I love to see that Nature is so rife with life that myriads can be afforded to be sacrificed and suffered to prey on one another; that tender organizations can be so serenely squashed out of existence like pulp – tadpoles which herons gobble up, and tortoises and toads run over in the road; and that sometimes it has rained flesh and blood!

With the liability to accident, we must see how little account is to be made of it. The impression made on a wise man is that of universal innocence. Poison is not poisonous after all, nor are any wounds fatal. Compassion is a very untenable ground. It must be expeditious. Its pleadings will not bear to be stereotyped.

Henry David Thoreau, "Walden: Or, Life in the Woods"

Para Meta e Walter Mückler.

Agradecimentos

À Luísa, iluminação de todos os meus porquês. Ao Emílio, que abastece diariamente o meu anseio pela Ciência. À Sofia, Odin e Chico (e todos os outros animais que passaram pela minha vida), que me tornam um ser humano mais indulgente e menos amargo. À Cristiane e Dolores, condutoras do meu amadurecimento e primeiras professoras.

Ao meu orientador, professor José Eduardo da Silveira Costa, que me estendeu a mão e me conduziu neste trabalho. Aos meus colegas e amigos, agentes de boa parte do meu bem-estar durante esses anos de graduação. Aos departamentos de Estatística e Astronomia da Universidade Federal do Rio Grande do Sul e seus corpos docentes por guiarem meu aprendizado. À Universidade Federal do Rio Grande do Sul, Conselho Nacional de Desenvolvimento Científico e Tecnológico e Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul por serem possibilitadores deste momento.

Muito obrigada por favorecerem e facilitarem a materialização de um sonho.

Resumo

A maioria das estrelas concentram-se em grupos, chamados aglomerados estelares. Boa parte das estrelas de um aglomerado são chamadas de estrelas variáveis, objetos cuja luminosidade varia no tempo de forma detectável em escalas de um segundo até séculos. Essa variação é detectada através da análise das suas séries temporais de brilho, que podem ter diferentes padrões dependendo do tipo de estrela observada. O estudo de estrelas variáveis fornece informações fundamentais sobre propriedades estelares, como massa, raio, luminosidade, temperatura, e uma melhor compreensão sobre a estrutura interna, composição química e evolução das estrelas. A identificação da variabilidade estelar pode ser feita por inspeção visual das séries temporais. Porém, esse método, tem eficácia limitada pois depende diretamente da interpretação do observador, além de ser demasiadamente demorado, visto o grande número de estrelas que um aglomerado pode conter. Dessa forma, este trabalho propôs desenvolver dois métodos estatísticos e automáticos, chamados *tempo de não-constância* (TNC), para detecção de estrelas variáveis em grandes conjuntos de dados, e *área de não-constância* (ANC), como índice de confirmação de variabilidade auxiliar. Este trabalho utilizou dados de 3250 estrelas do aglomerado globular NGC 6397, obtidos nos filtros U_HIGH, B_HIGH e V_HIGH, além do filtro de banda larga FILT_465_250. Foi utilizado um critério probabilístico para separar estrelas pertencentes ao aglomerado das estrelas de campo, o que resultou numa amostra de 1132 estrelas. Foi realizada inspeção visual das séries temporais e bem como aplicação dos métodos TNC e ANC, que demonstraram possuir eficiência em detectar variabilidade comparável à identificação por inspeção visual, porém numa fração do tempo.

Abstract

Most stars are concentrated in groups called star clusters and some of them are called variable stars, objects which luminosity varies over time in a detectable way, in scales from one second to centuries. This variability is detected through the analysis of its brightness time series, which has different patterns depending on the type of star observed. The study of variable stars provides fundamental informations about stellar properties such as mass, radius, luminosity, temperature, as well as a better understanding of the internal structure, chemical composition and evolution of stars. Stellar variability can be identified by time series visual inspection. However, this method has limited effectiveness due to its dependency on observer's interpretation, in addition to being too time consuming, given the large number of stars that a cluster can contain. Therefore, this work proposes to develop two statistical and automatic methods called *time of non-constancy* (TNC), for detecting variable stars in large data sets, and *area of non-constancy* (ANC), as an auxiliary index for variability confirmation. This work used data from 3250 stars from the globular cluster NGC 6397, obtained in the filters U_HIGH, B_HIGH and V_HIGH, in addition to the broadband filter FILT_465_250. A probabilistic criterion was used to separate stars belonging to the cluster to field stars resulting in a sample of 1132 stars. Visual inspection of the time series was carried out, as well as the application of the TNC and ANC methods, which demonstrated an efficiency in detecting variability comparable to identification by visual inspection, but in a fraction of the time.

Sumário

1	Introdução	14
1.1	Contexto e tema	14
1.1.1	Astroestatística	14
1.1.2	Estrelas	14
1.1.3	Aglomerados estelares	15
1.1.4	Estrelas variáveis e seus tipos	16
1.2	Problematização	17
1.2.1	<i>Surveys</i> em larga escala	18
1.2.2	Automatização da detecção e classificação de variabilidade estelar	18
1.3	Objetivo	19
2	Dados observacionais	21
2.1	Fonte de dados	21
2.2	Descrição dos dados	23
3	Metodologia de pesquisa	25
3.1	Identificação de estrelas do aglomerado globular	25
3.1.1	Diagrama cor-magnitude	26
3.2	Identificação visual de estrelas variáveis	28
3.3	Identificação automática de estrelas variáveis	29
3.3.1	Modelo empírico	30
3.3.2	Intervalos de confiança	33
3.3.3	Tempo de não-constância	34
3.3.4	Área de não-constância	36
4	Resultados	39
4.1	Distribuição de magnitudes	39
4.2	Distribuição de incertezas	41
4.3	Identificação de variáveis por inspeção visual	43
4.4	Identificação automática de variáveis	44
4.4.1	Tempo de não-constância	44
4.4.2	Área de não-constância	46
4.5	Amostra total: Comparando os resultados da identificação automática com a identificação visual	48
4.6	Identificação e descrição das estrelas do aglomerado globular NGC 6397	50
4.7	Estrelas variáveis do aglomerado globular NGC 6397	53

4.7.1	Identificação visual	53
4.7.2	Detecção automática: Tempo de não-constância	55
4.7.3	Detecção Automática: Área de não-constância	57
4.8	Amostra do aglomerado globular: Comparando os resultados da identificação automática com a identificação visual	61
5	Conclusões e perspectivas	62
	Referências bibliográficas	63

Lista de Figuras

1.1	Representação artística da formação estelar	15
1.2	Aglomerado aberto M39	16
1.3	Aglomerado globular Messier 15	16
1.4	Diagrama de uma estrela pulsante	17
1.5	Diagrama de um sistema de estrelas binário eclipsante	17
1.6	Curva de luz de uma Cefeida	17
1.7	Curva de luz de uma RR Lyrae	17
2.1	Imagem do <i>Very Large Telescope</i>	21
2.2	Aglomerado globular NGC 6397	22
2.3	CCDs do aglomerado globular NCG 6397	22
2.4	Curvas de transmissão	22
2.5	Exemplo de série temporal de magnitude	24
3.1	Exemplo de estrelas de campo do aglomerado globular NGC 6397	26
3.2	Diagrama de Hertzsprung-Russell	27
3.3	Screenshot da interface do programa VLC	29
3.4	Exemplos de séries temporais classificadas por inspeção visual . .	29
3.5	Exemplos de séries temporais suavizadas com o super suavizador de Friedman	30
3.6	Exemplos de séries temporais suavizadas com o super suavizador de Friedman e seus intervalos de confiança	33
3.7	Exemplos de curvas de luz classificadas pelo TNC como variável e não-variável	35
3.8	Exemplos de curvas de luz classificadas pelo TNC como variável suas respectivas áreas de não-constância	37
4.1	Distribuições das médias das magnitudes das 3 250 estrelas da amostra nos filtros U, B, V e FILT_465_250	40
4.2	Distribuição de magnitudes teórica, funções de completeza e distribuição de magnitudes observada	41
4.3	Distribuições das incertezas fotométricas das 3250 estrelas da amostra nos filtros U, B, V e FILT_465_250	42
4.4	Gráficos de dispersão das médias das magnitudes pelas incertezas fotométricas das 3 250 estrelas da amostra nos filtros U, B, V e FILT_465_250	42

4.5	Gráficos de barra da fração de variáveis e não-variáveis classificadas por inspeção visual contra os intervalos de magnitudes em cada filtro.	44
4.6	Gráficos de barra da fração de variáveis e não-variáveis classificadas pelo TNC contra os intervalos de magnitudes em cada filtro.	45
4.7	Gráficos de barra dos intervalos de ANC das estrelas classificadas pelo TNC como variáveis contra os intervalos de magnitudes em cada filtro.	48
4.8	Curvas de luz de estrelas cuja inspeção visual e o método TNC discordam	49
4.9	Curvas de luz de estrelas cuja inspeção visual e o método TNC concordam	50
4.10	Diagrama cor-magnitude V por (U-B)	50
4.11	Diagrama cor-magnitude V por (B-V)	51
4.12	Distribuições das médias das magnitudes das estrelas pertencentes ao aglomerado nos filtros U, B, V e FILT_465_250	51
4.13	Distribuições das incertezas fotométricas das estrelas pertencentes ao aglomerado nos filtros U, B, V e FILT_465_250	52
4.14	Distribuições das médias das magnitudes das estrelas não pertencentes ao aglomerado nos filtros U, B, V e FILT_465_250	52
4.15	Distribuições das incertezas fotométricas das estrelas não pertencentes ao aglomerado nos filtros U, B, V e FILT_465_250	53
4.16	Diagramas cor-magnitude para as estrelas do aglomerado, coloridas pela inspeção visual	54
4.17	Gráficos de barra da fração de variáveis e não-variáveis do aglomerado classificadas por inspeção visual contra os intervalos de magnitudes em cada filtro.	55
4.18	Diagrama cor-magnitude para as estrelas do aglomerado, coloridas pelo índice TNC	56
4.19	Gráficos de barra da fração de variáveis e não-variáveis do aglomerado classificadas pelo TNC contra os intervalos de magnitudes em cada filtro.	57
4.20	Gráficos de barra dos intervalos de ANC das estrelas do aglomerado classificadas pelo TNC como variáveis contra os intervalos de magnitudes em cada filtro.	59
4.21	Diagrama cor-magnitude para as estrelas do aglomerado identificadas como variáveis pelo TNC e coloridas pelo índice ANC	60

Lista de Tabelas

4.1	Classificação da variabilidade estelar na amostra de 3250 estrelas, via inspeção visual	43
4.2	Classificação da variabilidade estelar na amostra de 3250 estrelas, via tempo de não-constância	45
4.3	Quantidade de estrelas classificadas como variáveis pelo TNC por faixa de ANC	46
4.4	Tabela com os resultados do índice TNC em comparação com a inspeção visual	49
4.5	Classificação da variabilidade estelar na amostra de 1132 estrelas do aglomerado, via inspeção visual	54
4.6	Classificação da variabilidade estelar na amostra de 1132 estrelas do aglomerado, via tempo de não-constância	56
4.7	Quantidade de estrelas do aglomerado globular classificadas como variáveis pelo TNC por faixa de ANC	57
4.8	Tabela com os resultados do índice TNC em comparação com a inspeção visual, para estrelas do aglomerado	61

1 Introdução

A primeira seção deste capítulo introduz os conceitos básicos utilizados neste trabalho, bem como o contexto em que este está inserido. A segunda seção apresenta o problema de pesquisa abordado e a terceira seção traça o objetivo primário desta monografia.

1.1 Contexto e tema

Nesta seção serão apresentados os conceitos astronômicos fundamentais que envolvem os objetos de estudo desta monografia: *estrelas variáveis* e *aglomerados de estrelas*. Antes, uma breve explanação sobre a importância da área de pesquisa no qual este trabalho está inserido, a *astroestatística*.

1.1.1 Astroestatística

Áreas de pesquisa que situam-se na interseção da astronomia observacional, estatística, desenvolvimento de algoritmos e ciência de dados têm emergido rapidamente. Um dos primeiros levantamentos digitais do céu em grande escala, o *Sloan Digital Sky Survey* (SDSS), que teve seu início na virada do século, ampliou consideravelmente o interesse em estatísticas e técnicas de *machine learning* (Eadie et al., 2019). Ainda na última década, houve um aumento em cerca de 3 vezes o uso de palavras-chave “Métodos: Estatística” nos artigos científicos em astronomia, o que expõe uma demanda ampla por metodologias e técnicas estatísticas para resolver problemas astronômicos difíceis ou inviáveis de abordar com os métodos tradicionais (Siemiginowska et al., 2019).

A aplicação de qualquer novo método envolve o uso de dados, estatísticas e desenvolvimento de algoritmos e, portanto, uma colaboração interdisciplinar do conhecimento e da experiência de astrônomos, estatísticos e especialistas em computação (Siemiginowska et al., 2019). Com isso, o termo *astroestatística* se tornou popular, sendo uma referência adequada para o campo interdisciplinar entre a astronomia e a estatística. Uma das áreas da astronomia com grande interesse em técnicas estatísticas é a *astrofísica estelar*.

1.1.2 Estrelas

Estrelas são corpos gasosos, formados primordialmente de hidrogênio e hélio, que possuem formato aproximadamente esférico. No interior das estrelas ocorrem

reações termonucleares que liberam grande quantidade de energia na forma de radiação. Essa radiação, por sua vez, se propaga do centro para a superfície da estrela, passando por diferentes camadas que constituem o seu interior, até atingir o espaço em forma de ondas eletromagnéticas (de Oliveira e Saraiva, 2013).

Para estrelas se formarem, uma nuvem de gás e poeira deve sofrer um colapso gravitacional (Figura 1.1). Duas características dessa nuvem inicial são fundamentais para se entender o tipo de estrela e detalhes da sua evolução: (1) a massa da nuvem, que é a propriedade determinante, e (2) a composição química da nuvem, pois, mesmo sendo compostas principalmente de hidrogênio e hélio, as estrelas se diferem em termos da pequena fração de outros elementos químicos mais pesados que podem possuir.

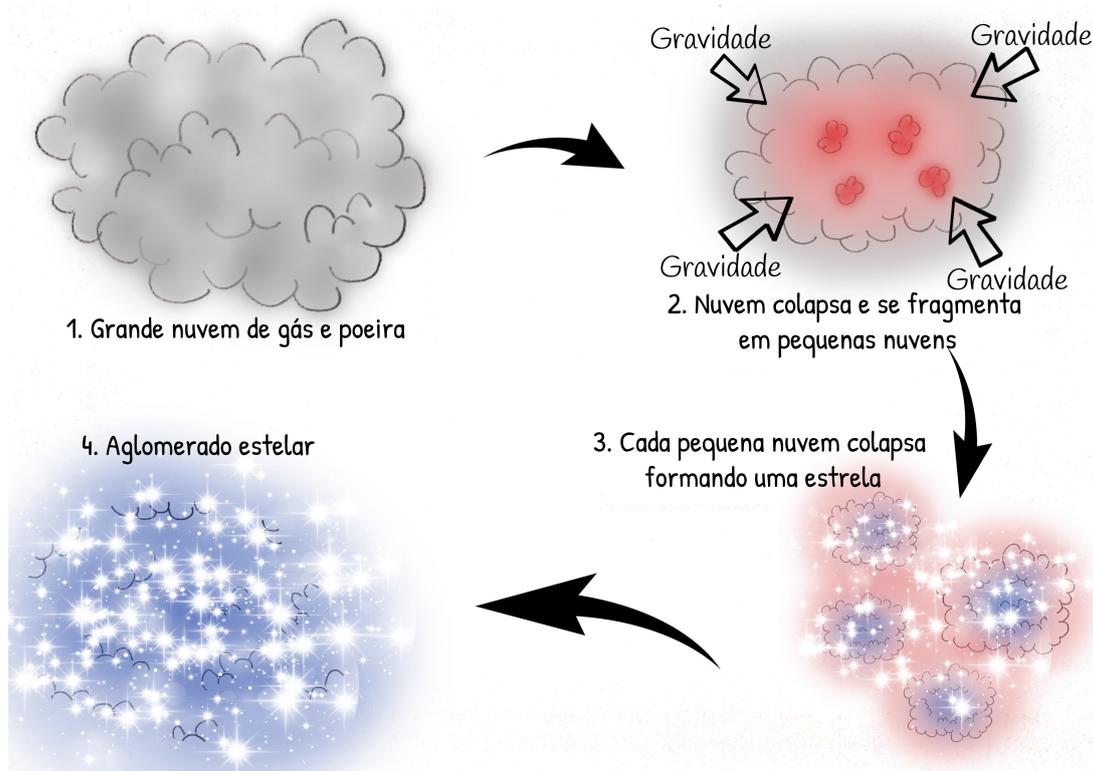


Figura 1.1: Representação artística da formação estelar.

Apesar dessa definição geral, existem vários tipos de estrelas. Algumas são mais brilhantes e outras mais fracas, seja porque estão a distâncias diferentes ou porque, de fato, emitem quantidades diferentes de luz. Elas também possuem cores distintas, algumas se parecem mais azuis, outras mais vermelhas, ou laranjas, ou amarelas, ou brancas. Além disso, há também a composição química de outros elementos além de hidrogênio e hélio, que caracterizam o tipo de estrela. Por fim, na maioria dos casos, as estrelas não nascem isoladas, mas sim em grupos, que chamamos de *aglomerados estelares* (de Oliveira e Saraiva, 2013).

1.1.3 Aglomerados estelares

Um aglomerado estelar é um grupo de estrelas que se formaram a partir da mesma nuvem e estão gravitacionalmente ligadas. As estrelas que nascem juntas em um aglomerado possuem a mesma idade e mesma metalicidade, isto é, a mesma

proporção de elementos químicos diferentes do hidrogênio e hélio. Existem dois tipos principais de aglomerados de estrelas: os *aglomerados globulares* e os *aglomerados abertos*.

Os aglomerados abertos (Figura 1.2) são muito mais jovens, sendo o local de nascimento recente de estrelas, e menores, contendo algumas dezenas a algumas centenas de estrelas (Karttunen et al., 2016). Eles são fracamente ligados gravitacionalmente, fazendo com que possam ter formatos irregulares e, apesar de suas estrelas se formarem juntas, elas não permanecem juntas ao longo do tempo e acabam se espalhando. Esses aglomerados são usados para estudar estrelas jovens e os processos de formação de estrelas.

Já os aglomerados globulares (Figura 1.3) são antigos e permaneceram em um sistema fortemente ligado gravitacionalmente, dando a eles um formato esférico. Eles podem conter de alguns milhares a um milhão de estrelas e têm idade média de 11 bilhões de anos, o que significa que contêm algumas das estrelas mais antigas da nossa galáxia.



Figura 1.2: Aglomerado aberto M39.¹



Figura 1.3: Aglomerado globular Messier 15.²

Tanto em aglomerados abertos e globulares há uma fração significativa de estrelas que possuem uma característica importante em relação ao seu brilho, sendo chamadas de *estrelas variáveis* (Clement, 2017).

1.1.4 Estrelas variáveis e seus tipos

Estrelas variáveis são aquelas cuja luminosidade varia no tempo de forma detectável em escalas de um segundo até séculos (Percy, 2007). Essa variação é detectada pela mudança em sua magnitude (brilho) ao longo do tempo, ou seja, através de uma *série temporal do brilho* da estrela, chamada comumente pelos astrônomos de *curva de luz*.

Existem dois diferentes tipos de causas para a variação do brilho da estrela: *causas intrínsecas* e *causas extrínsecas*. Estrelas variáveis intrínsecas são aquelas cuja luminosidade altera por fatores que ocorrem no interior da estrela, por exemplo, quando a estrela pulsa com pequenas expansões e contrações periódicas de seu volume (Figura 1.4). Já nas estrelas variáveis extrínsecas a mudança de

¹APOD, 2009. Disponível em: <<https://apod.nasa.gov/apod/ap090412.html>>. Acesso em: 28 de jan. de 2021

²NASA, 2017. Disponível em: <<https://www.nasa.gov/feature/goddard/2017/messier-15>>. Acesso em: 28 de jan. de 2021

brilho decorre de causas externas, por exemplo, como ocorre quando a estrela forma um sistema binário com uma estrela companheira que às vezes a eclipsa (Figura 1.5).

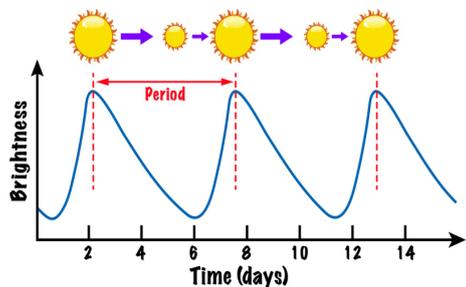


Figura 1.4: Diagrama de uma estrela pulsante.³

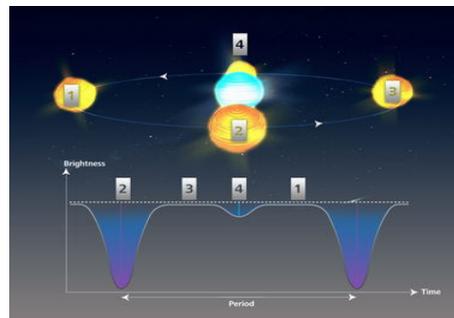


Figura 1.5: Diagrama de um sistema de estrelas binário eclipsante.⁴

Diferentes estrelas variáveis possuem diferentes padrões em suas curvas de luz (Figuras 1.6 e 1.7). O *General Catalogue of Variable Stars* (GCVS) lista mais de 100 tipos e subtipos de estrelas variáveis (Eyer e Mowlavi, 2008).

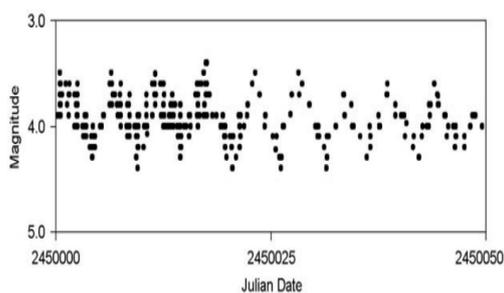


Figura 1.6: Curva de luz de uma Cefeida.
Créditos: AAVSO, Cambridge, MA

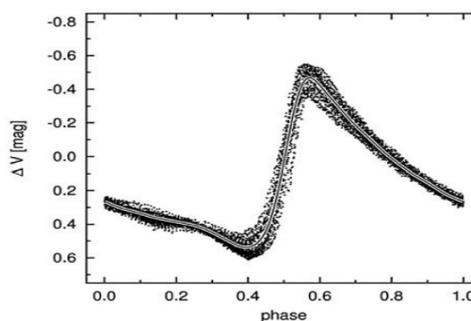


Figura 1.7: Curva de luz de uma RR Lyrae.
Créditos: AAVSO, Cambridge, MA

O estudo de estrelas variáveis fornece informações fundamentais sobre propriedades estelares, como massa, raio, luminosidade, temperatura, e uma melhor compreensão sobre a estrutura interna, composição química e evolução das estrelas (Percy, 2007).

1.2 Problematização

A identificação e classificação da variabilidade estelar em aglomerados pode ser feita com base na inspeção visual das séries temporais do brilho das estrelas. Essa abordagem, porém, tem eficácia limitada pois, além de depender diretamente da interpretação do observador, o processo pode ser demasiadamente demorado, visto o grande número de estrelas que um aglomerado estelar pode conter,

³Scioly, 2021. Disponível em: <https://scioly.org/wiki/index.php/Astronomy/Variable_Stars>. Acesso em: 18 de abril de 2021

⁴Write Science, 2015. Disponível em: <<https://writescience.wordpress.com/2015/03/27/gravity-9-the-evolving-universe/>>. Acesso em: 18 de abril de 2021

1.2.1 *Surveys* em larga escala

Nas últimas duas décadas, inúmeros *surveys* (levantamentos) em larga escala do céu vêm sendo realizados por telescópios espaciais ou por telescópios terrestres robotizados, como *Optical Gravitational Lensing Experiment* (OGLE) (Udalski et al., 2015), *Kepler Mission* (Koch et al., 2010), *Massive Compact Halo Object* (MACHO) (Alcock et al., 1997), *All Sky Automated Survey* (ASAS) (Pojmański, 2004), *The Catalina Real-Time Transient Survey* (CRTS) (Drake et al., 2017), *Northern Sky Variability Survey* (NSVS) (Wozniak et al., 2004) e outros. Todos esses *surveys* produzem grandes quantidades de dados.

A inspeção visual das séries temporais de brilho, é a forma mais direta e simples de detecção de variabilidade estelar, além de muito usual (Sokolovsky et al., 2017). A eficácia da detecção de variáveis por inspeção visual depende fortemente da quantidade (duração da série temporal) e qualidade (relação sinal/ruído) dos dados, da amplitude e do perfil da variação, além da aptidão de quem o faz.

A detecção por inspeção visual pode ser praticamente inequívoca nos casos em que a amplitude de variação é relativamente grande (relação sinal/ruído alta) e o padrão de variação característico de alguma classe já conhecida de estrelas variáveis. Entretanto, torna-se incerta, ou mesmo inviável, quando a relação sinal/ruído é baixa ou as séries temporais de brilho não são suficientemente longas (Percy, 2007).

Paralelo aos grandes *surveys*, observações de populações estelares como as de aglomerados globulares, possibilitam a fotometria simultânea de milhares ou de milhões de estrelas, produzindo uma série temporal de medidas fotométricas para cada uma delas. Logo, o grande volume de dados tornou inviável utilizar somente métodos visuais para a detecção e posterior classificação de estrelas variáveis (Hosenie et al., 2020). Portanto, é desejável procurar métodos confiáveis e automatizados, que possam ser aplicados diretamente a bancos de dados contendo grandes quantidades de séries temporais de brilho.

1.2.2 Automatização da detecção e classificação de variabilidade estelar

Uma abordagem atrativa para automatizar a detecção e classificação da variabilidade estelar é utilizar métodos que integrem, em algoritmos computacionais, técnicas astronômicas e estatísticas. Atualmente diferentes abordagens são utilizadas com este propósito, como transformadas de Fourier (Deb e Singh, 2009), índices de variabilidade estatísticos, como qui-quadrado (Martinazzi et al., 2017) e outros (Sokolovsky et al., 2017), *machine learning* (Richards et al., 2011), *Gaussian Mixture Model* (GMM) (Shin et al., 2009), modelos autoregressivos (Elorrieta et al., 2019), análise de componentes principais (Deb e Singh, 2009), entre outros.

Apesar dessas técnicas serem populares e relativamente eficazes elas também podem apresentar limitações em seus usos práticos. Ao contrário dos dados de séries temporais em outras áreas de pesquisa, é comum que dados astronômicos de séries temporais sejam amostrados irregularmente e menos homoganeamente devido ao ciclo dia/noite e à variações climáticas que afetam a transparência do céu. Essa diferença dificulta o pré-processamento do dados com um método comum, como a

análise de componentes principais (Shin et al., 2009). Além disso, os algoritmos de *machine learning* concentram-se, em sua maioria, na classificação das fontes, o que requer conjuntos de dados de treinamento bem rotulados. A fragilidade desses algoritmos se dá na identificação de eventos incomuns e nas situações em que, por exemplo, não há dados de treinamento disponíveis. (Soraisam et al., 2020).

Em alguns casos os métodos apresentados podem não ser suficientes por si próprios para selecionar automaticamente estrelas variáveis de um grande conjunto de dados, sendo dependentes ainda de validação visual dos seus resultados. Na maior parte dos casos essa necessidade está ligada à métodos que não conseguem distinguir séries temporais com alto nível de ruído que imitam a variabilidade de variáveis reais, devido a variações de noite a noite, desvio da imagem do objeto causado por um defeito cosmético e assim por diante (Sokolovsky et al., 2017).

Em Shin et al. (2009), os autores sintetizam o que seriam requisitos fundamentais para um método robusto de detecção de estrelas variáveis.

- Primeiramente, eles declaram que este deve ser um procedimento orientado aos dados, sem depender de suposições sobre os mesmos.
- Outra condição necessária para o método de detecção ser eficiente é conseguir lidar com uma ampla variedade de tipos de variabilidade estelar e de conjunto de dados, como número de dados, incerteza nos dados medidos e padrões de amostragem de tempo (Carbonell et al., 1992). Isso é importante pois pode-se adotar valores diferentes para os cortes na qualidade dos dados devido à variação dos ambientes de observação e às propriedades diferentes de cada campo de observação, como a precisão da fotometria. Tais diferenças podem resultar em uma distribuição heterogênea de pontos de dados (Shin et al., 2009).
- Além disso, o método deve ser capaz de derivar uma inferência probabilística para que se possa estimar a confiabilidade da detecção, pois, à medida que a quantidade de dados aumenta, o controle da identificação de um falso positivo se torna indispensável (Shin et al., 2009).
- Por fim, utilizar um método que leve em consideração a coerência temporal da série temporal pode ser extremamente proveitoso, porém grande parte das técnicas atualmente utilizadas para detecção de variabilidade estelar não faz uso dessa coerência temporal (Welch e Stetson, 1993).

Dado às circunstâncias citadas, torna-se desejável que o algoritmo de detecção e classificação de variabilidade estelar seja adaptável à complexidade dos dados de séries temporais de estrelas em aglomerados e eficiente computacionalmente.

1.3 Objetivo

Este trabalho têm como objetivo desenvolver um algoritmo estatístico que detecte e automaticamente estrelas variáveis em grandes populações estelares a partir de suas séries temporais de brilho. Assim, é necessário que este algoritmo seja robusto, adaptando-se aos diferentes tipos de variabilidade estelar, e computacionalmente eficiente. Para isso, este trabalho propõe um novo método totalmente orientado aos

dados, capaz de derivar uma inferência probabilística para estimar a confiabilidade da detecção e, ainda, levar em consideração a coerência temporal da curva de luz.

Esta monografia está organizada da seguinte forma: o Capítulo 2 apresenta a fonte de obtenção dos dados observacionais utilizados neste trabalho, além do tratamento inicial realizado aos dados brutos. No Capítulo 3 é apresentado a metodologia de pesquisa, começando por uma abordagem utilizando um diagrama cor-magnitude para separar estrelas de campo das estrelas do aglomerado. Posteriormente é descrito o procedimento usual de identificação visual de estrelas variáveis utilizado e, em seguida, apresenta e detalha os critérios estatísticos e numéricos utilizados na aplicação da metodologia de *tempo de não-constância* e *área de não-constância* aos dados. O Capítulo 4 resume os resultados encontrados neste trabalho e, por fim, o Capítulo 5 apresenta as discussões finais acerca dos resultados encontrados e fornece sugestões de trabalhos futuros a serem desenvolvidos.

2 Dados observacionais

Neste trabalho, utilizamos dois tipos de dados do aglomerado globular NGC 6397: (1) séries temporais com medidas das variações do brilho de estrelas ao longo do tempo e (2) medidas pontuais do brilho estelar em três diferentes bandas espectrais (filtros). Neste capítulo é apresentada uma descrição dos dados, bem como dos processos de aquisição e tratamento dos mesmos.

2.1 Fonte de dados

Os dados utilizados neste trabalho foram obtidos conforme descrito detalhadamente em [Martinazzi et al. \(2017\)](#). Em síntese, as séries temporais fotométricas foram geradas a partir de uma sequência de imagens do aglomerado globular NGC 6397 (Figura 2.2), obtidas com o telescópio de 8.4 m do *Very Large Telescope* (ESO-VLT-UT1) (Figura 2.1) do *European Southern Observatory* (ESO), localizado na cordilheira dos Andes, no Chile. As observações foram realizadas nas noites de 27 e 28 de julho de 2009, com o uso do imageador FORS2 (*Focal Reducer/low dispersion Spectrograph 2*), composto por dois CCDs (Figura 2.3).



Figura 2.1: Imagem do *Very Large Telescope*¹.

¹Créditos: ESO/Y. Beletsky. Disponível em: <<https://www.eso.org/public/images/potw1119a/>>. Acesso em: 04 de maio de 2021



Figura 2.2: Aglomerado globular NGC 6397.²

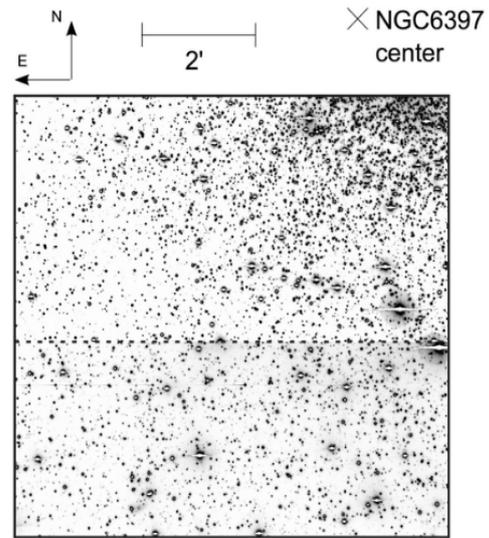


Figura 2.3: Uma das imagens do aglomerado globular NGC 6397, obtida com CCD1 (parte superior) e CCD2 (parte inferior) (Martinazzi et al., 2017).

As imagens para as séries temporais foram obtidas utilizando um filtro FILT_465_250 (comprimento de onda central $\lambda_c = 4\,650\text{ \AA}$ e $\Delta\lambda = 250\text{ \AA}$), pois este cobre a região espectral correspondente à do sistema fotométrico *UBV* (de Ultraviolet, Blue, Visual), o que corresponde a uma faixa do ultravioleta próximo até a região visível do espectro eletromagnético (Figura 2.4).

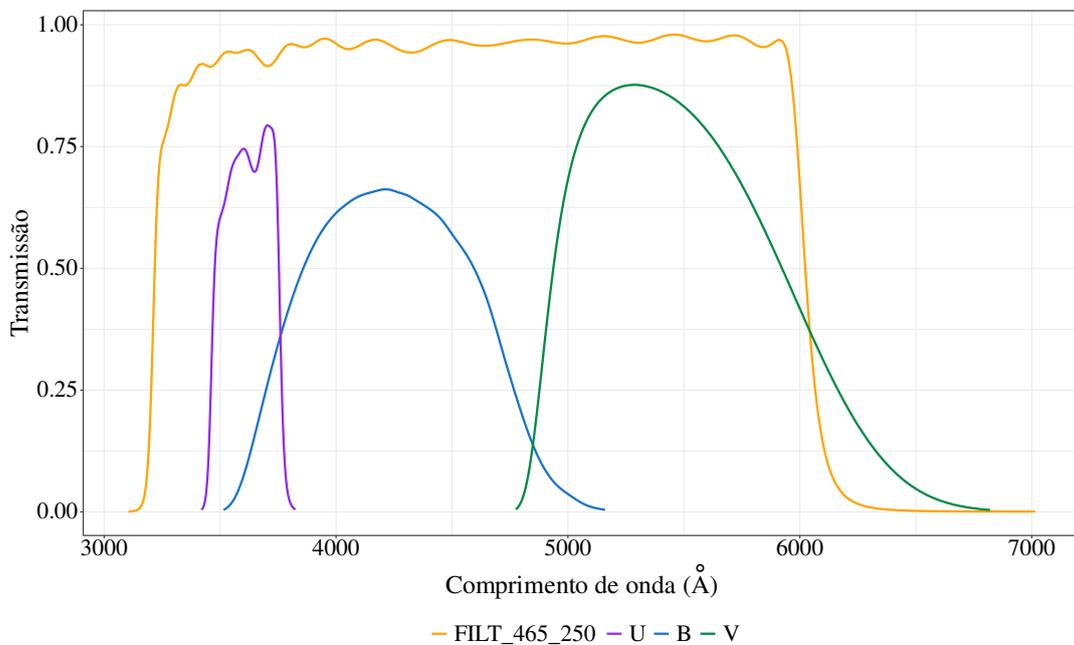


Figura 2.4: Curvas de transmissão dos filtros U, B, V e FILT_465_250, em roxo, azul, verde e laranja, respectivamente. Dados extraídos de Spanish Virtual Observatory, em 11 de março de 2021.

²ESO, 2004. Disponível em: <<https://www.eso.org/public/brazil/images/eso0425a/>>. Acesso em: 24 de jun. de 2020

Um total de 305 imagens foram obtidas: 205 com 60 segundos de tempo de integração na primeira noite, 28 com 80 segundos na primeira parte da segunda noite e 72 com 170 segundos na segunda parte da segunda noite, começando no dia juliano $JD = 2\,455\,038.507350$, resultando em 5.1 h de fotometria na primeira noite e 4.8 h na segunda noite, com separação de 19.95 h entre as duas séries temporais. As imagens foram centradas no ponto do céu com coordenadas $\alpha = 17^{\text{h}}41^{\text{m}}01.78^{\text{s}}$ e $\delta = -53^{\circ}44'47.2''$ (J2000). A extração das séries temporais a partir das imagens foi realizada por Elizandra Martinazzi durante o seu doutorado (Martinazzi, 2016), através de rotinas padrão do *software* IRAF (*Image Reduction and Analysis Facility*) (Tody, 1986) amplamente usados em astronomia.

Para construção dos diagramas cor-magnitude descritos na seção 3.1.1 foram obtidas imagens adicionais com o uso dos filtros U_HIGH, B_HIGH e V_HIGH, referenciados neste trabalho por U, B e V, respectivamente. Para os filtros B e V, as imagens foram obtidas com tempos de exposição de 1, 10, 120 e 300 segundos e para o filtro U, com tempos de exposição de 1, 300 e 600 segundos. A fotometria de estrelas fracas exige tempos de exposição mais longos, enquanto que a fotometria de estrelas mais brilhantes exigem tempos de exposição mais curtos para evitar a saturação da imagem. Por exemplo, estrelas mais brilhantes do que magnitude $V \simeq 19$ estavam saturadas em imagens com 300 segundos de tempo de exposição. O uso de diferentes tempos de exposição introduz um viés observacional importante que será discutido no capítulo 4.

2.2 Descrição dos dados

Foram encontradas 7617 estrelas no CCD1 e 3739 no CCD2, totalizando 11 356 estrelas em todo o campo com magnitudes entre 15 (mais brilhantes) e 25 (menos brilhantes)³. No entanto, escolheu-se as séries temporais das 8 391 estrelas mais brilhantes que magnitude 23, pois as menos brilhantes que magnitude 23 são altamente afetadas por variações do brilho do céu. Utilizou-se apenas séries temporais de brilho de estrelas detectadas em mais de 100 imagens, sendo eliminadas as estrelas próximas às bordas das imagens, a uma distância de 15 pixels das bordas. Destas, apenas as curvas de luz que continham o número máximo de pontos (205 pontos) foram utilizadas, afim de homogeneizar a quantidade de pontos das curvas de luz, o que resultou em uma amostra com um total de 3250 séries temporais.

Um exemplo de série temporal obtida através deste processo é mostrada na figura 2.5. Os pontos indicam as medidas fotométricas (em magnitudes) para o brilho da estrela, obtidas ao longo da primeira noite, em intervalos aproximadamente regulares da ordem de 60 segundos, com as respectivas barras de erro, geradas a partir das incertezas fotométricas. A linha vermelha indica a magnitude média para a série temporal. Em todas as séries temporais obtidas, as incertezas fotométricas seguem um padrão praticamente homocedástico, sendo aproximadamente constantes para todas as medidas de uma mesma série.

³A escala das magnitudes é decrescente.

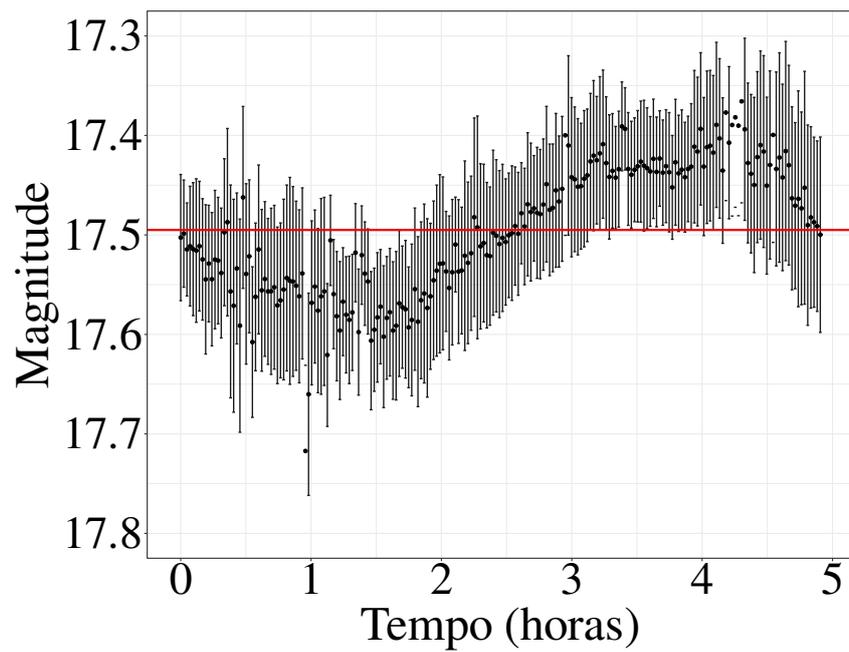


Figura 2.5: Exemplo de série temporal obtida a partir do processo descrito na seção 2.1. A linha vermelha representa a magnitude média.

3 Metodologia de pesquisa

Este trabalho tem como objetivo principal desenvolver um algoritmo para identificação automática de estrelas variáveis em uma amostra de estrelas, utilizando critérios estatísticos aplicados à séries temporais fotométricas. Espera-se que este algoritmo seja mais eficiente do que a tradicional identificação de variáveis por inspeção visual das séries temporais. O algoritmo será aplicado a séries temporais de uma amostra de estrelas com o objetivo de determinar a fração de variáveis no aglomerado globular NGC 6397. Esta amostra está contaminada por estrelas que não pertencem ao aglomerado, por estrelas binárias ou duplas, ou mesmo por estrelas em estágios evolutivos avançados do próprio aglomerado, o que exige um critério estatístico de seleção.

Este capítulo descreve detalhadamente os critérios estatísticos utilizados na identificação automática de variáveis e na seleção de estrelas para o estudo da variabilidade estelar no aglomerado globular NGC 6397. A linguagem R ([R Development Core Team, 2020](#)) foi utilizada como principal linguagem de programação na implementação dos algoritmos aqui descritos.

3.1 Identificação de estrelas do aglomerado globular

Quando observa-se uma imagem de um aglomerado de estrelas (como os exemplificados nas figuras [1.2](#) e [1.3](#)) pode-se ter a impressão de que todas as estrelas da imagem pertencem, de fato, ao aglomerado. Porém, muitas delas, são estrelas de campo (Figura [3.1](#)), ou seja, estrelas que não pertencem ao aglomerado mas estão dentro no mesmo campo observado. Essas são estrelas mais próximas da Terra do que o aglomerado e, ao observarmos o aglomerado estelar, elas ficam na frente da imagem, se confundindo com as estrelas que são, de fato, do aglomerado estelar.

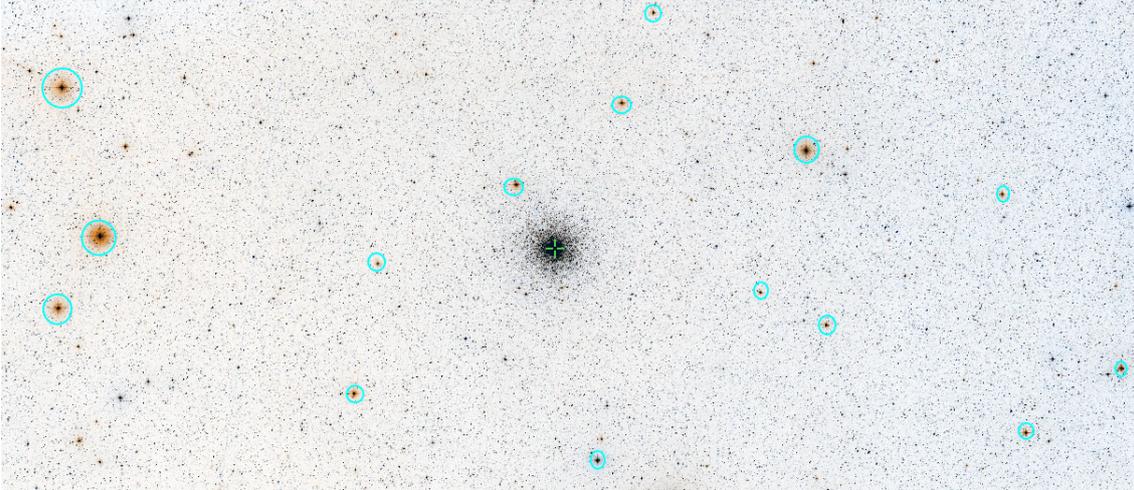


Figura 3.1: Exemplo de estrelas de campo do aglomerado globular NGC 6397. A imagem teve suas cores invertidas. Algumas das estrelas de campo estão circuladas em azul.

Para estudar a população de um aglomerado de estrelas, precisa-se utilizar um procedimento que consiga separar as estrelas que são de fato do aglomerado daquelas que são estrelas de campo.

3.1.1 Diagrama cor-magnitude

Uma abordagem para excluir prováveis objetos que não pertencem ao aglomerado baseia-se em construir um *diagrama cor-magnitude*, ou CMD (*color-magnitude diagram*). O CMD é um gráfico de dados observacionais que mostra como uma população de estrelas pode ser representada em termos de seu brilho (ou magnitude) e cor, que é a diferença de magnitude entre dois filtros, e que está associada à temperatura da superfície da estrela. Com essa temperatura pode-se traçar o tipo espectral da estrela. Um CMD é uma variante do *diagrama de Hertzsprung-Russell* (Ian Ridpath, 2012) ou, diagrama (H-R) (Figura 3.2). Enquanto o diagrama (H-R) é um resumo das temperaturas e magnitudes de estrelas individuais, um CMD dedica-se especificamente ao estudo de um aglomerados de estrelas. Abaixo, uma concepção artística do diagrama (H-R).

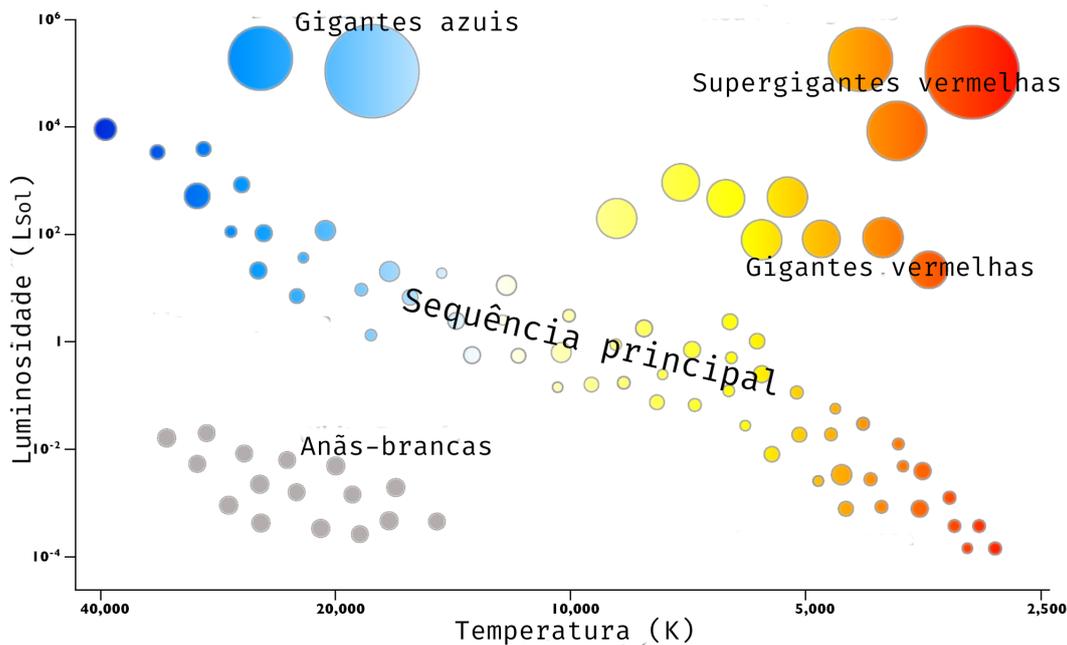


Figura 3.2: Diagrama de Hertzsprung-Russell.

As estrelas de um aglomerado que estão na fase evolutiva inicial, isto é, que ainda estão convertendo hidrogênio em hélio por meio de reações térmicas em seus núcleos, aparecem dentro de uma estreita faixa no CMD. A posição de uma estrela dentro desta faixa é determinada pela massa estelar, variando de forma contínua de um dos extremos da faixa onde estão as estrelas menos massivas (e menos brilhantes), até o outro extremo, onde estão as estrelas mais massivas (e mais brilhantes). A sequência de estrelas ao longo desta faixa é conhecida como *sequência principal*.

Modelos teóricos podem calcular a posição da sequência principal no CMD, para diferentes idades e metalicidades e a curva produzida por estes modelos é chamada de *isócrona* (do grego **iso**, *igual* + **cronos**, *tempo*), uma vez que todas as estrelas da sequência principal possuem a mesma idade. A idade e a metalicidade de um aglomerado é determinada pela isócrona que melhor se ajusta aos seus CMDs (Ian Ridpath, 2012).

3.1.1.1 Critério de seleção

As incertezas nas medidas fotométricas de U, B e V causam uma dispersão dos pontos em relação à isócrona. Entretanto, há outros fatores a considerar na análise de um CMD:

1. **A presença de estrelas de campo.** Estrelas de campo são estrelas localizadas no espaço entre a Terra e o aglomerado, que estão dentro do campo observado, mas que não pertencem ao aglomerado. Logo, a posição dessas estrelas no CMD não possui nenhum vínculo com a isócrona do aglomerado e podem aparecer como pontos distantes em relação à isócrona.
2. **A presença de estrelas evoluídas do aglomerado.** Estrelas mais massivas convertem mais rapidamente hidrogênio em hélio até o esgotamento do hidrogênio no núcleo. Quando isto ocorre, a estrela sai da sequência

principal. No CMD, estas estrelas aparecem em regiões afastadas em relação à isócrona.

3. **Estrelas binárias pertencentes ao aglomerado.** Estrelas binárias são estrelas que nasceram juntas e se mantêm muito próximas a ponto de não serem resolvidas pelo telescópio e aparecerem no CMD como uma única estrela. Como os brilhos se somam, isto afeta as magnitudes U, B e V, e a posição da estrela no CMD.
4. **Estrelas duplas pertencentes ao aglomerado.** São estrelas que não formam um sistema binário, mas estão aproximadamente alinhadas com a mesma linha de visada.

As quatro condições listadas acima tendem a aumentar a fração de estrelas mais brilhantes na amostra: (1) estrelas de campo próximas são mais brilhantes, por serem próximas; (2) estrelas evoluídas que ainda não chegaram em seus estágios finais de evolução são estrelas brilhantes; (3) e (4) estrelas binárias e estrelas duplas não resolvidas aparecem como uma única estrela nas imagens, mais brilhante do que cada componente isolada. Essa concentração de estrelas brilhantes introduz uma bimodalidade nas distribuições de magnitudes, como mostrado na figura 4.1.

Para discriminar entre estrelas com alta probabilidade de pertencer ao aglomerado globular das demais, utilizamos um critério de seleção implementado através do algoritmo descrito abaixo:

Seja x o índice de cor (B-V) ou (U-B) e y a magnitude V de uma estrela na posição (x, y) do CMD e σ_x e σ_y suas respectivas incertezas fotométricas. E, seja x' e y' o índice de cor e a magnitude V de um ponto (x', y') qualquer sobre a isócrona. Define-se a distância quadrática d^2 entre a posição (x, y) da estrela e um ponto (x', y') qualquer sobre a isócrona por:

$$d^2 = (\Delta x)^2 + (\Delta y)^2 \quad ,$$

onde $\Delta x = x - x'$ e $\Delta y = y - y'$ são os desvios nas duas direções.

Assume-se que o ponto (x'_o, y'_o) sobre a isócrona, com maior probabilidade de ser a posição real da estrela no CMD é aquele que minimiza a distância quadrática d^2 . Calcula-se os desvios normalizados pelas incertezas, $\eta_x = \Delta x / \sigma_x$ e $\eta_y = \Delta y / \sigma_y$. São excluídos os pontos com η_x e η_y maiores do que um certo limite η_o , com valor entre 3 e 4 e calculado em função do número de pontos no CMD.

3.2 Identificação visual de estrelas variáveis

A fim de comparar os resultados que serão obtidos a partir do processo de identificação automática de estrelas variáveis proposto neste trabalho com a tradicional classificação baseada em inspeção visual das séries temporais, foi realizada a classificação visual como variável ou não variável de cada uma das 3250 séries temporais utilizadas como objetos de estudo neste trabalho. Para visualizar as curvas de luz foi utilizado um script desenvolvido em linguagem R, o qual foi chamado *View Light Curves* (VLC). Um screenshot da interface do programa VLC é mostrado na figura 3.3.

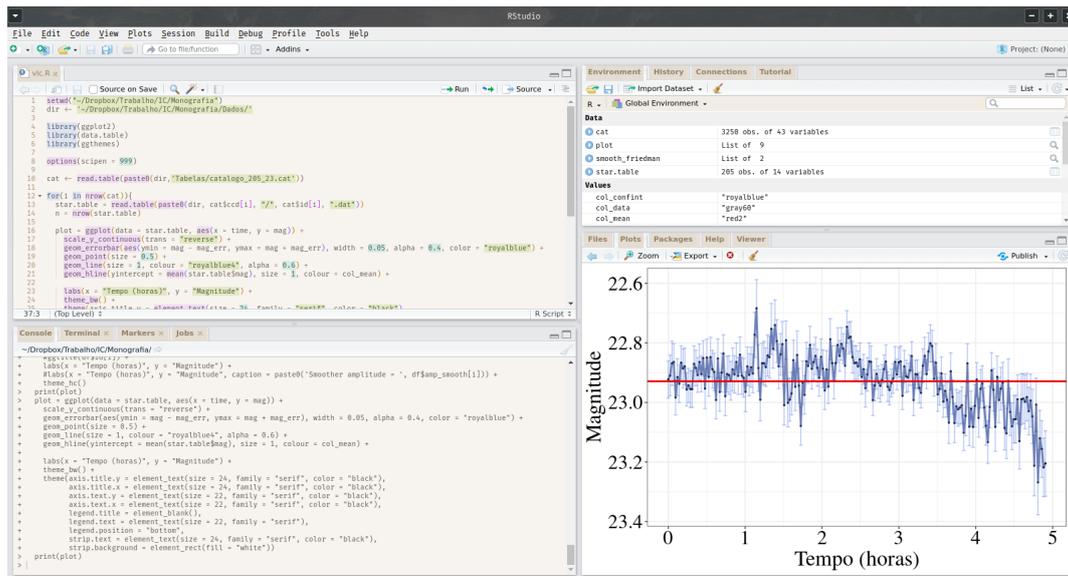


Figura 3.3: Screenshot da interface do programa VLC.

Na inspeção visual utilizou-se como referência a magnitude média, juntamente com a amplitude de variação aparente de cada curva de luz e as incertezas fotométricas associadas a cada medida de magnitude como critério visual de seleção. Dois exemplos são mostrados na figura 3.4.

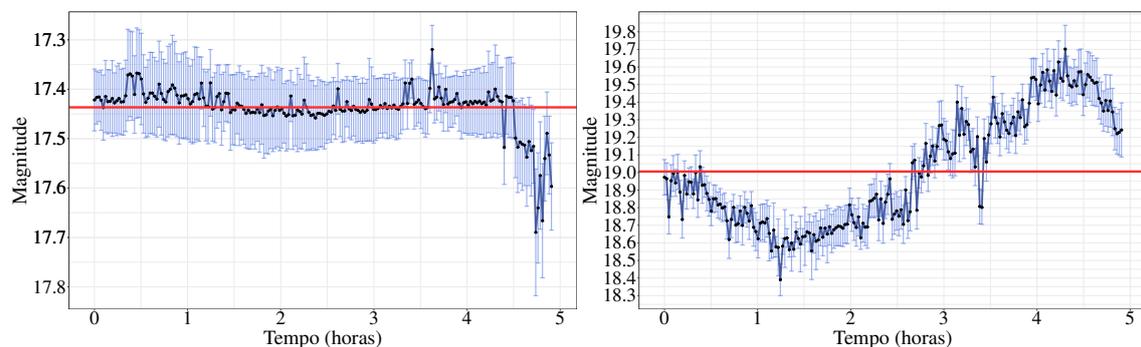


Figura 3.4: Exemplos de séries temporais classificadas por inspeção visual como não-variável (gráfico à esquerda) e como variável (gráfico à direita). A curva sólida representa a série temporal e as barras ds incertezas fotométricas de cada medida são representadas em azul claro. A linha horizontal vermelha indica o brilho médio.

3.3 Identificação automática de estrelas variáveis

O processo de identificação automática proposto neste trabalho consiste em, basicamente, avaliar a suposição de que o brilho da estrela seja constante e aproximadamente igual à magnitude média da série temporal. Caso essa suposição não seja atendida, pode-se dizer que o brilho da estrela varia no tempo, conforme descrito por um modelo empírico calculado a partir da série temporal. Dado que o brilho da estrela varia no tempo, então ela pode ser classificada como candidata a estrela variável. O procedimento se dá da seguinte forma:

1. Para cada série temporal calcula-se um modelo empírico a partir do *super*

suavizador de Friedman (Friedman, 1984) e o seu respectivo intervalo de confiança;

2. Calcula-se a magnitude média da série temporal e o respectivo intervalo de confiança;
3. Utiliza-se um índice estatístico, fundamentado na seção 3.3.3, como avaliador da suposição de brilho constante e igual à magnitude média da série temporal para classificar a estrela como variável ou não-variável;
4. Um índice estatístico adicional, fundamentado na seção 3.3.4, é utilizado, apenas para as estrelas classificadas como variáveis pelo índice anterior, como indicador da amplitude de variabilidade;

3.3.1 Modelo empírico

Ao longo de uma noite de observação, variações observáveis no brilho de uma estrela, sejam decorrentes de causas intrínsecas ou extrínsecas, ou mesmo causadas por mudanças na transparência do céu, ocorrem em escalas de tempo da ordem de poucos minutos ou horas e, em geral, são relativamente suaves. A fim de modelar as variações observadas no brilho e contrastá-las com o ruído inerente ao processo de medida, aplicou-se uma suavização às séries temporais de brilho utilizando o super suavizador de Friedman, por meio da função `supsmu()` do pacote `stats` do R. Este suavizador não-paramétrico utiliza regressão linear local de amplitude variável, com larguras de banda adaptáveis. Exemplos da suavização das curvas de luz encontram-se na figura 3.5. Flutuações em relação à curva suavizada são interpretadas como ruído.

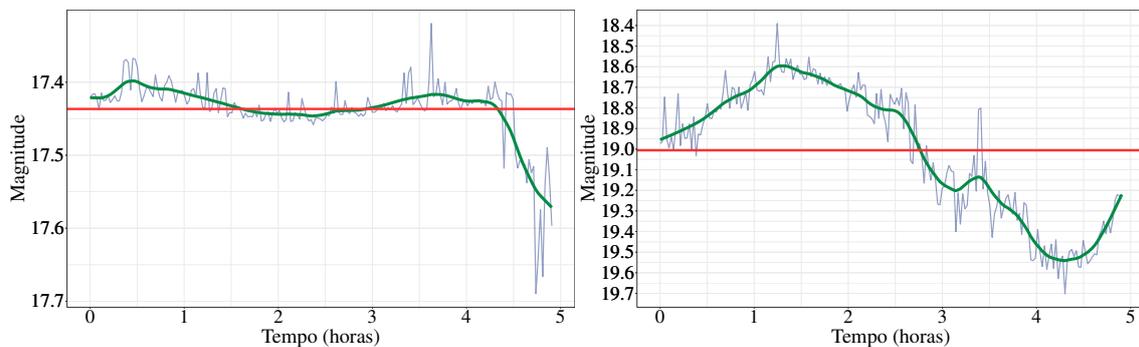


Figura 3.5: Exemplos de séries temporais suavizadas com o super suavizador de Friedman. As curvas azuis são os dados temporais, as linhas verdes são a suavização e as linhas vermelhas são os brilhos médios.

3.3.1.1 Super suavizador de Friedman

O algoritmo no qual o super suavizador de Friedman se baseia é descrito precisamente em Luedicke (2015) e resumido a seguir:

1. Estima-se, por regressão linear local, um número de suavizações com largura de banda fixa.

2. Os resíduos da validação cruzada *leave-one-out* (Efron, 1982) de cada uma dessas estimativas iniciais são então suavizados usando uma largura de banda constante.
3. Com base nos resíduos suavizados, as melhores larguras de banda das estimativas iniciais são selecionadas em cada ponto de dados ao longo do intervalo da variável preditora.
4. Essas larguras de banda locais são então suavizadas com uma largura de banda constante e as duas estimativas das estimativas iniciais com valores de largura de banda mais próximas às larguras de banda suavizadas são selecionadas.
5. Os resultados suavizados do passo anterior são interpolados linearmente.
6. Os pontos interpolados são então suavizados novamente com uma largura de banda fixa, resultando na estimativa final.

Formalmente, o processo pode ser definido como especificado a seguir:

Sejam x_1, \dots, x_n e y_1, \dots, y_n n amostras aleatórias da distribuição conjunta $P(X, Y)$ gerada pelo processo

$$Y = f(X) + \epsilon \quad , \quad (3.1)$$

onde f é uma função arbitrária de X e ϵ são erros aleatórios i.i.d. (independentes e identicamente distribuídos) com valor esperado igual a zero. O objetivo é estimar a esperança condicional $E[Y|X = x]$, de forma que o valor esperado da diferença quadrática $E[Y - f(X)]^2$ seja minimizado, onde $f(X)$ é a verdadeira função subjacente. Para isso precisamos, então, encontrar a estimativa $\hat{f}(x)$ em

$$y_i = \hat{f}(x_i) + \epsilon_i \quad . \quad (3.2)$$

Regressão linear local

Uma forma de estimar $\hat{f}(x_i)$ é ajustar localmente regressões de mínimos quadrados lineares da forma

$$E[Y|x_i] = \hat{\alpha} + \hat{\beta}x_j \quad , \quad x_j \in N_i \quad (3.3)$$

onde N define a vizinhança local em torno de x_i e ϵ é o parâmetro de ajuste, também conhecido como largura de banda, que controla o *tradeoff* entre viés e variância. No caso da largura de banda fixa, o tamanho da janela N é constante, enquanto no caso da largura de banda adaptável pode variar ao longo do intervalo da variável preditora.

Dada uma largura de banda fixa J , onde J é o número de observações em uma janela, podemos escrever o estimador linear local como

$$\hat{y}_k = \hat{\alpha} + \hat{\beta}x_k \quad , \quad k = 1, \dots, n \quad (3.4)$$

onde $\hat{\alpha}$ e $\hat{\beta}$ são obtidos a partir de ajustes locais para os pontos de dados $i_{-J/2}, \dots, i_{+J/2}$, com $x_i \leq x_{i+1}$ para $i = J/2, \dots, n - J/2$.

Uma largura de banda ideal que minimiza o erro quadrático esperado,

$$e^2(J) = E[Y - f(X|J)]^2 \quad , \quad (3.5)$$

pode ser obtida estimando-se $e^2(J)$ via validação cruzada *leave-one-out*:

$$\hat{e}^2(J) = \frac{1}{n} \sum_{i=1}^n [y_i - \hat{f}_{(-i)}(x_i|J)]^2 \quad . \quad (3.6)$$

Minimizar $\hat{e}^2(J)$ produz então a largura de banda ótima com validação cruzada

$$\hat{e}^2(J_{cv}) = \min_{0 < J \leq n} \hat{e}^2(J) \quad . \quad (3.7)$$

Os quadrados dos resíduos do *leave-one-out* podem ser calculados analiticamente:

$$\hat{e}^2(J) = \frac{1}{n} \sum_{i=1}^n [y_i - \hat{f}(x_i|J)]^2 / \left[1 - \frac{1}{J} - \frac{(x_i - \bar{x}_J)^2}{V_J} \right]^2 \quad , \quad (3.8)$$

onde $\sum_{j=i-J/2}^{i+J/2} (x_j - \bar{x}_J)^2$ e $\bar{x}_J = \frac{1}{J} \sum_{j=i-J/2}^{i+J/2} x_j$.

Larguras de banda adaptáveis

A fim de obter um estimador com larguras de banda localmente adaptáveis, [Friedman \(1984\)](#) propôs minimizar a estimativa de

$$e^2(f, J) = E[Y - f(X|J(X))]^2 \quad , \quad (3.9)$$

com respeito à $f(x)$ e $J(x)$. Para minimizar e^2 em (3.9) estima-se primeiro \hat{y}_k em (3.4) por regressão linear local sobre uma grade de valores para J . Enquanto [Friedman \(1984\)](#) propôs originalmente usar $J = 0,05n$, $J = 0,2n$ e $J = 0,5n$, a função `supsmu()` generaliza essa condição permitindo a especificação de qualquer número de larguras de banda no intervalo $0 < J < n$ para permitir um espaço de largura de banda de granulação mais fina. Em seguida, calcula-se os resíduos da validação cruzada para cada uma dessas estimativas iniciais de largura de banda constante:

$$r_{(i)(J)} = [y_i - \hat{f}(x_i|J)] / \left[1 - \frac{1}{J} - \frac{(x_i - \bar{x}_J)^2}{V_J} \right] \quad (3.10)$$

e suavizamos $r_{(i)(J)}$ contra x_i com largura de banda $J = 0,2n$ para estimar $\hat{e}(f, J|x_i)$ que usamos para encontrar a largura de banda ótima em cada ponto:

$$\hat{e}(f, J_{cv}(x_i)|x_i) = \min_J \hat{e}(f, J|x_i) \quad . \quad (3.11)$$

As larguras de banda ótimas $J_{cv}(x_i)$ são então suavizadas novamente ($J = 0,2n$) contra x_i e as duas estimativas iniciais com larguras de banda mais próximas são selecionadas, sujeitas a

$$J_1 \leq J_{cv}(x_i) \leq J_2 \quad . \quad (3.12)$$

A penúltima suavização é então calculada por interpolação linear entre essas duas estimativas iniciais com relação a $J_{cv}(x_i)$. Finalmente, o resultado da interpolação é então suavizado novamente com largura de banda $J = 0,05n$.

3.3.2 Intervalos de confiança

A curva suavizada é usada como modelo empírico para variações suaves no brilho da estrela e a dispersão dos pontos na série temporal em relação à curva suavizada como caracterização do ruído inerente às observações. O modelo empírico obtido a partir dos dados não necessariamente descreve as variações reais no brilho da estrela. Mas, pode-se calcular um intervalo ao redor do modelo, dentro do qual existe uma certa probabilidade da curva real estar. Este intervalo é o *intervalo de confiança* para o modelo e a probabilidade da curva real nele estar contida é definida através do *nível de confiança*.

Uma suposição a ser verificada é de que o brilho da estrela é constante ao longo da noite e que quaisquer desvios fazem parte do ruído inerente às observações. Neste caso, a melhor estimativa para o brilho da estrela a partir da série temporal é a magnitude média, para a qual também pode-se estimar um intervalo de confiança.

Foram calculados os intervalos de confiança para a suavização, bem como para a magnitude média, de cada série temporal, utilizando um nível de confiança de 99,9% (exemplos na figura 3.6).

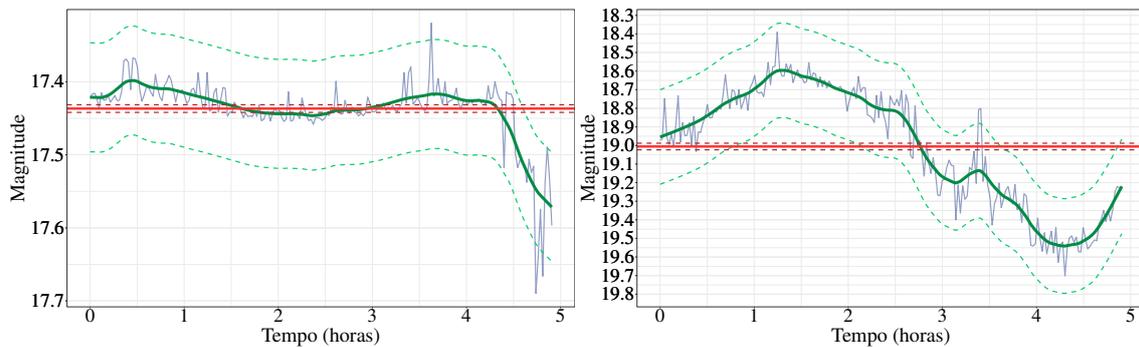


Figura 3.6: Exemplos de séries temporais suavizadas com o super suavizador de Friedman e seus intervalos de confiança de 99,9%. As curvas azuis são as séries temporais, a linha sólida verde é a suavização, as linhas tracejadas verdes são os limites dos intervalos de confiança de 99% para a suavização, a linha sólida vermelha é a magnitude média e as linhas tracejadas vermelhas são os limites dos intervalos de confiança para a média.

Formalmente, o intervalo de confiança de um parâmetro θ , é definido pelo intervalo de valores inferior e superior $[l(X), u(X)]$ que dependem das variáveis X definidas de modo que

$$P[l(X) < \theta < u(X)] = 1 - \alpha \quad , \quad (3.13)$$

onde $0 < \alpha < 1$ é geralmente um valor pequeno como $\alpha = 0,05$ ou $0,01$, chamado de *nível de significância*. Ou seja, se θ for o parâmetro verdadeiro, então a probabilidade de o intervalo $[l(X), u(X)]$ conter θ é de $(1 - \alpha)$. Para $\alpha = 0,05$, o intervalo de confiança estimado de 95% de um estimador de algum parâmetro θ é um intervalo (l, u) de modo $P(l < \hat{\theta} < u) = 0,95$. Se o experimento foi repetido 100 vezes, uma média de 95 intervalos obtidos conterão o valor do parâmetro θ . Para este trabalho foram construídos intervalos de confiança para a suavização das séries temporais através do super suavizador de Friedman e para a magnitude média de cada série temporal. Para a suavização toma-se cada medida de magnitude suavizada $s(t_i)$ no tempo t_i , $i = 1, \dots, n$ para fazer

$$U_s(t_i) = s(t_i) + z_{\alpha/2} * \sigma_r \quad e \quad L_s(t_i) = s(t_i) - z_{\alpha/2} * \sigma_r \quad , \quad (3.14)$$

os respectivos intervalos de confiança superior e inferior, onde $z_{\alpha/2}$ é o valor do $(1 - \alpha * 100)$ percentil da distribuição Normal padrão e σ_r é o desvio-padrão dos resíduos da suavização.

Para o intervalo de confiança da média superior e inferior é feito, respectivamente,

$$U_m(t_i) = \bar{m} + z_{\alpha/2} * \frac{\sigma_r}{\sqrt{n-1}} = U_m \quad e \quad L_m(t_i) = \bar{m} - z_{\alpha/2} * \frac{\sigma_r}{\sqrt{n-1}} = L_m, \quad \forall i \quad , \quad (3.15)$$

onde \bar{m} é a magnitude média da série temporal.

3.3.2.1 Relação Sinal-Ruído

A largura do intervalo de confiança depende da *razão sinal-ruído* (S/N) ou SNR (do inglês, *Signal-to-Noise Ratio*) da série temporal, isto é, da razão entre a intensidade do brilho da estrela o nível de ruído associado às medidas. Numa primeira aproximação, S/N pode ser calculado como a razão entre a magnitude média \bar{m} da série temporal e o desvio padrão do ruído. Para dados fotométricos (de Oliveira e Saraiva, 2013),

$$S/N \propto 10^{-\bar{m}/5} \quad .$$

Quanto menos brilhante é a estrela (menor \bar{m}), menor será a razão sinal-ruído S/N e maior será a largura do intervalo de confiança. Isto é ilustrado na figura 3.6 onde, no caso à esquerda, para uma série temporal com magnitude média $\bar{m} \simeq 17.42$ (mais brilhante), a largura do intervalo de confiança de 99,9% é de $\simeq 0.1$ mag, enquanto que no caso à direita, para uma série temporal com $\bar{m} \simeq 19.0$ (menos brilhante), a largura do intervalo de confiança é de $\simeq 0.6$ mag.

3.3.3 Tempo de não-constância

Estamos considerando duas suposições mutuamente exclusivas sobre o comportamento do brilho da estrela: para o intervalo de observação (a) o brilho da estrela é constante e aproximadamente igual à magnitude média da série temporal (estrela não-variável) e (b) o brilho da estrela não é constante, apresentando variações ao longo do tempo (estrela variável).

Para avaliar essas suposições, aplicou-se um índice estatístico desenvolvido por nós durante o projeto de Iniciação Científica, ao qual denominamos *tempo de não-constância*. Este índice mede o tempo em que o intervalo de confiança para a curva suavizada, $[U_s(t_i); L_s(t_i)]$, não contém o intervalo de confiança para a média do brilho da estrela, $[U_m(t_i); L_m(t_i)]$. Se $[U_s(t_i); L_s(t_i)]$ conter $[U_m(t_i); L_m(t_i)]$ ao longo de toda a série temporal, o tempo de não-constância é zero e a suposição (a) é aceita. Neste caso, a estrela é classificada como não-variável, caso contrário, como variável (Figura 3.7). O tempo de não constância pode ser usado não somente para avaliar a variabilidade da estrela como uma variável dicotômica, mas também serve como um indicador do grau de evidência da suposição de variabilidade.

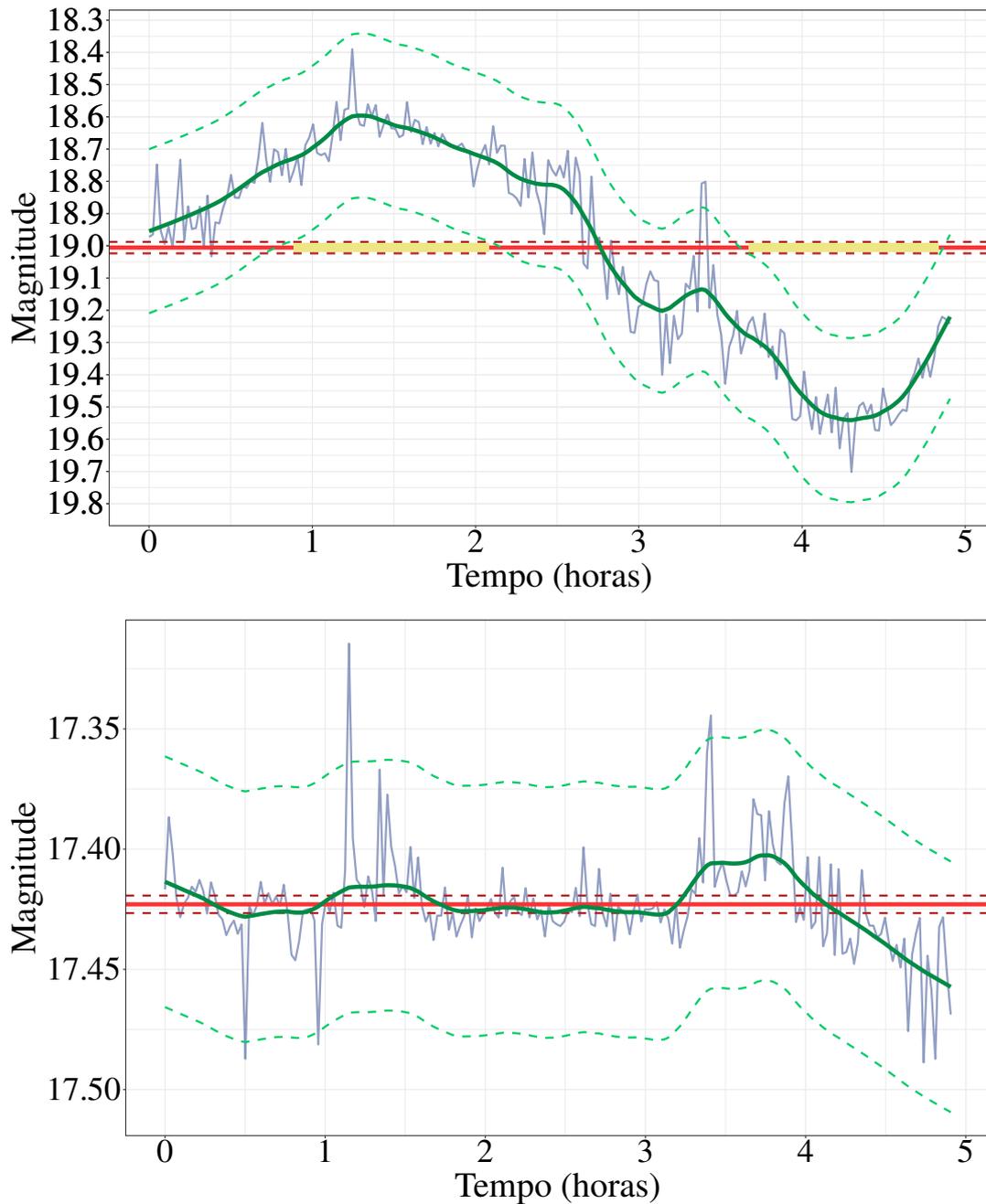


Figura 3.7: Exemplos de curvas detectadas pelo TNC como *variável* (acima) e como *não-variável* (abaixo). A linha sólida verde é a suavização e as linhas tracejadas verdes são os respectivos intervalos de confiança de 99,9% para a suavização. A linha sólida vermelha é o brilho médio e as linhas tracejadas vermelhas são os respectivos intervalos de confiança de 99,9% para a média. Na figura acima, a parte em amarelo indica o tempo em que o intervalo de confiança para a média não está contido no intervalo de confiança para a suavização.

Formalmente, calcula-se a função indicadora do *tempo de não-constância superior* tomando a diferença entre o intervalo de confiança inferior para a suavização e o intervalo de confiança superior para a magnitude média da curva de luz, de modo que

$$I_{TNC_U}(t_i) = \begin{cases} 1, & \text{se } L_s(t_i) - U_m > 0 \\ 0, & \text{caso contrário} \end{cases} \quad (3.16)$$

e, analogamente, a função indicadora do *tempo de não-constância inferior* utiliza a diferença entre o intervalo de confiança superior para a suavização e o intervalo de confiança inferior para a magnitude média da curva de luz, tal que

$$I_{TNC_L}(t_i) = \begin{cases} 1, & \text{se } U_s(t_i) - L_m < 0 \\ 0, & \text{caso contrário} \end{cases} \quad (3.17)$$

Para um dado tempo t_i , se $I_{TNC_U}(t_i) = 1$, então neste tempo t_i existe um *ponto de variabilidade superior* na série temporal e se $I_{TNC_L}(t_i) = 1$, então neste tempo t_i há um *ponto de variabilidade inferior*.

Seja T_U o conjunto de todos os t_i tais que $I_{TNC_U}(t_i) = 1$ e T_L o conjunto de todos os t_i tais que $I_{TNC_L}(t_i) = 1$. Ainda, seja $T = T_U \cup T_L$. A soma de todos os t_i em T de forma que

$$TNC = \sum_{i \in T} t_{i+1} - t_i \quad (3.18)$$

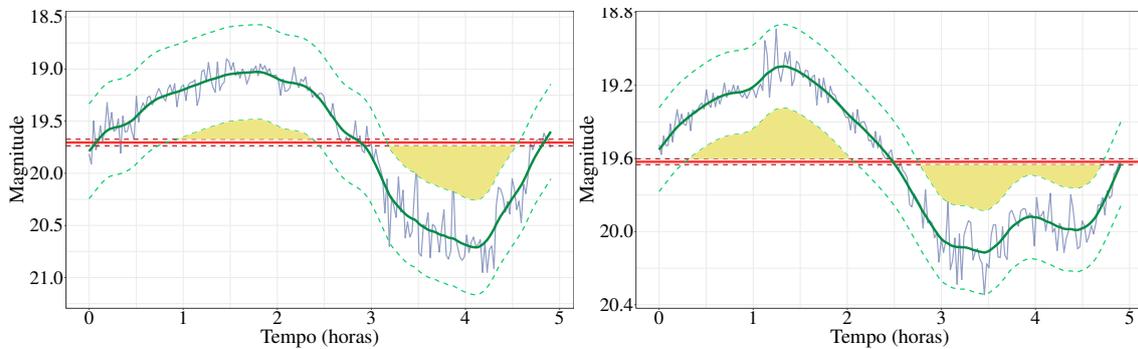
é o *tempo de não-constância total* da série temporal.

3.3.4 Área de não-constância

Quanto maior é a amplitude da variabilidade estelar, maior é o desvio da curva suavizada em relação à uma curva constante e igual à magnitude média e, conseqüentemente, maior é a área entre as duas curvas. Portanto, a área entre as duas curvas pode ser utilizada como um indicador da amplitude da variabilidade na série temporal. Neste trabalho, utilizou-se um indicador mais conservativo que leva em conta os intervalos de confiança das duas curvas, denominados *área de não-constância*, definido formalmente por:

$$\tau(c) \equiv \int_{t_1}^{t_2} \delta(\tilde{m}^2(t) - 2\bar{m}\tilde{m}(t) + \bar{m}^2 - \sigma^2(c))dt \quad (3.19)$$

onde \tilde{m} é a curva suavizada, \bar{m} é a magnitude média, t é o tempo, $\sigma(c)$ é o semi-intervalo de confiança e c é o nível de confiança. Exemplos de séries temporais de estrelas classificadas como variáveis pelo tempo de não-constância com suas respectivas áreas de não-constância são encontrados na figura 3.8.



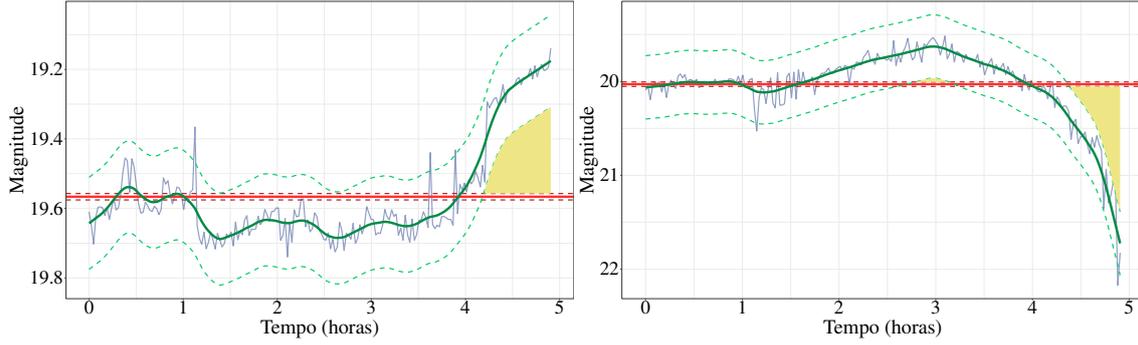


Figura 3.8: Exemplos de curvas de luz classificadas pelo TNC como variável suas respectivas áreas de não-constância. As áreas em amarelo indicam a área de não-constância. A linha sólida verde é a suavização e as linhas tracejadas verdes são os respectivos intervalos de confiança de 99,9% para a suavização. A linha sólida vermelha é o brilho médio e as linhas tracejadas vermelhas são os respectivos intervalos de confiança de 99,9% para a média.

3.3.4.1 Caracterização numérica da área de não-constância

Numericamente, a área de não-constância pode ser calculada através de uma soma de Riemann, onde toma-se a soma das áreas dos trapézios referentes ao tempo de não-constância superior, definidos como

$$A_{\text{trapézio}_U}(t_i) = \frac{(L_s(t_i) - U_m) * (t_{i+1} - t_i)}{2}, \quad \forall t_i \in T_U, \quad (3.20)$$

e a soma das áreas dos trapézios referentes ao tempo de não-constância inferior, definidos como

$$A_{\text{trapézio}_L}(t_i) = \frac{(U_s(t_i) - L_m) * (t_{i+1} - t_i)}{2}, \quad \forall t_i \in T_L. \quad (3.21)$$

Dessa forma, define-se a *área de não constância superior* como

$$ANC_U = \sum_{i \in T_U} A_{\text{trapézio}_U}(t_i) \quad (3.22)$$

e, analogamente, a *área de não constância inferior* como

$$ANC_L = \sum_{i \in T_L} A_{\text{trapézio}_L}(t_i). \quad (3.23)$$

Finalmente, a soma da área de não constância superior e da área de não constância inferior,

$$ANC = ANC_U + ANC_L \quad (3.24)$$

resultam na *área de não constância total*.

Nota-se que a área de não-constância é equivalente à uma área retangular, cuja base é o tempo de não-constância e cuja altura está relacionada com a amplitude média da variabilidade da série temporal, penalizada pelos intervalos de confiança das duas curvas. A área de não-constância permite comparar séries temporais com o mesmo tempo de não-constância.

Neste ponto, temos um indicador estatístico, o tempo de não-constância, que pode ser usado como avaliador da variabilidade na análise de cada série temporal para classificar a estrela como variável ou não-variável. Além disso, temos um segundo indicador estatístico, a área de não-constância, para a amplitude da variabilidade estelar. Os dois indicadores podem ser aplicados a todas as séries temporais.

Em resumo, este capítulo descreveu a metodologia adotada para a seleção automática de estrelas variáveis do aglomerado globular NCG 6397. Essa metodologia se baseia em 3 procedimentos básicos,

1. identificar estrelas com alta probabilidade de pertencer ao aglomerado globular;
2. classificar essas estrelas como variáveis ou não-variáveis a partir da inspeção visual das séries temporais;
3. e, por fim, a partir do cálculo do tempo de não-constância e da área de não-constância, classificar automaticamente as estrelas como variáveis ou não-variáveis.

Os resultados obtidos são discutidos no próximo capítulo.

4 Resultados

Analisamos um conjunto de séries temporais com medidas fotométricas da luz de estrelas obtidas ao longo de uma noite de observação com o objetivo de classificar cada estrela como variável ou não. A classificação foi feita tanto por inspeção visual da série temporal, quanto automaticamente, utilizando dois critérios estatísticos propostos neste trabalho: o tempo de não-constância e a área de não-constância.

Além disso, analisamos medidas de magnitude estelar em três bandas espectrais para separar estrelas com alta probabilidade de pertencer à sequência principal do aglomerado globular NGC 6397 das demais e assim ter uma amostra com alto grau de pureza para analisar a variabilidade estelar no aglomerado.

Os resultados são apresentados neste capítulo.

4.1 Distribuição de magnitudes

Uma primeira informação de grande interesse físico obtida em nosso estudo foi a distribuição de magnitudes para a amostra de 3 250 estrelas. Magnitude é uma medida do brilho estelar para uma certa banda do espectro eletromagnético definida pelo filtro utilizado nas observações. Os dados incluem medidas pontuais para magnitudes nos filtros U, B e V, indicadas por estas mesmas letras. Além disso, há as magnitudes médias no filtro `FILT_465_250`, calculadas a partir das magnitudes médias das séries temporais.

Para as estrelas do NGC 6397, as magnitudes nos quatro filtros variam de ≈ 15 para as mais brilhantes a ≈ 23 para as mais fracas. Como mencionado anteriormente, a escala de magnitudes é decrescente. As distribuições para os filtros U, B, V e `FILT_465_250` são mostradas na figura 4.1.

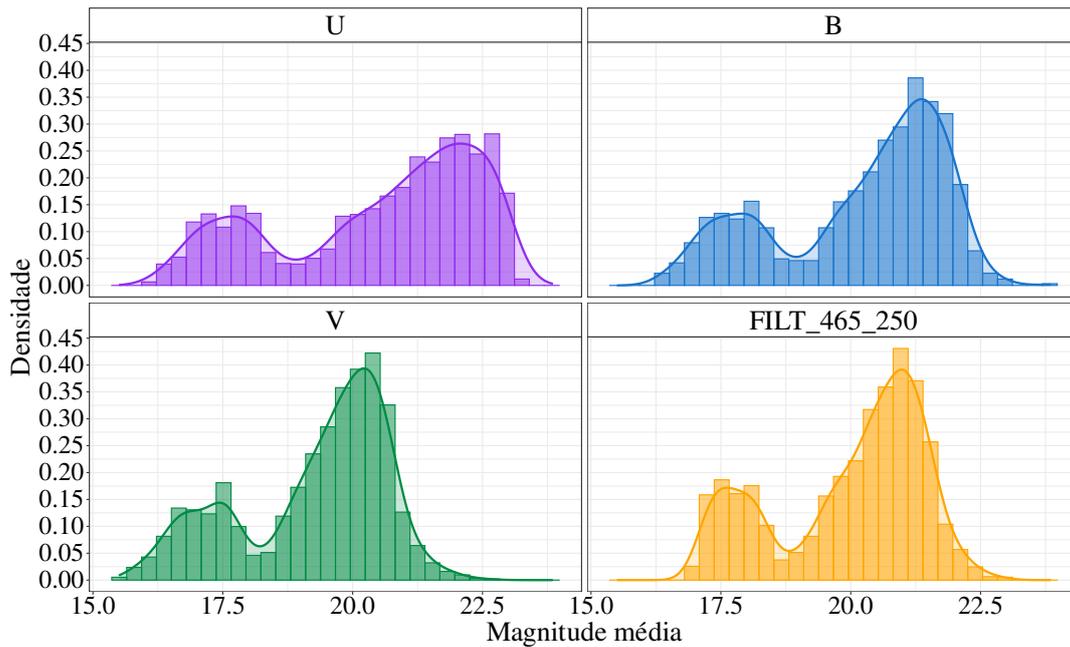


Figura 4.1: Distribuições das médias das magnitudes das 3 250 estrelas da amostra nos filtros U, B, V e FILT_465_250.

As distribuições obtidas são fortemente marcadas por um viés observacional. Numa população de estrelas, a distribuição de brilho segue uma lei de potência muito bem conhecida pelos astrônomos (Salpeter, 1955), onde a densidade de estrelas decresce com o aumento do brilho. A razão disso é que a natureza produz muito mais estrelas fracas do que estrelas mais brilhantes. A figura 4.2 (a) mostra a distribuição teórica de magnitudes para o aglomerado globular NGC 6397, calculada para o filtro FILT_465_250, porém similar a dos outros filtros.

A fração de estrelas detectáveis (completeza da amostra) depende de dois fatores: a magnitude da estrela e o tempo de exposição. Estrelas mais fracas são mais difíceis de serem observadas. Com tempo de exposição de 1 segundo consegue-se observar quase 100% das estrelas mais brilhantes, mas esta fração diminui progressivamente a medida que o brilho diminui e abaixo de um certo limite de magnitude, nenhuma estrela pode ser observada. Isto é esboçado pela curva preta na figura 4.2 (b). Uma maneira de se observar estrelas mais fracas é aumentando o tempo de exposição. Mas, neste caso, a observação de estrelas mais brilhantes é impedida pela saturação. Isto é esboçado pela curva vermelha na figura 4.2 (b) para um tempo de exposição de 300 segundos.

Quando se multiplica as funções de completeza para os dois tempos de exposição pela distribuição real de magnitudes, obtém-se uma distribuição bimodal, conforme esboçado na figura 4.2 (c). Esta é explicação para a bimodalidade na distribuição de magnitudes obtida (Figura 4.1) neste trabalho, não sendo esta, representativa da distribuição real.

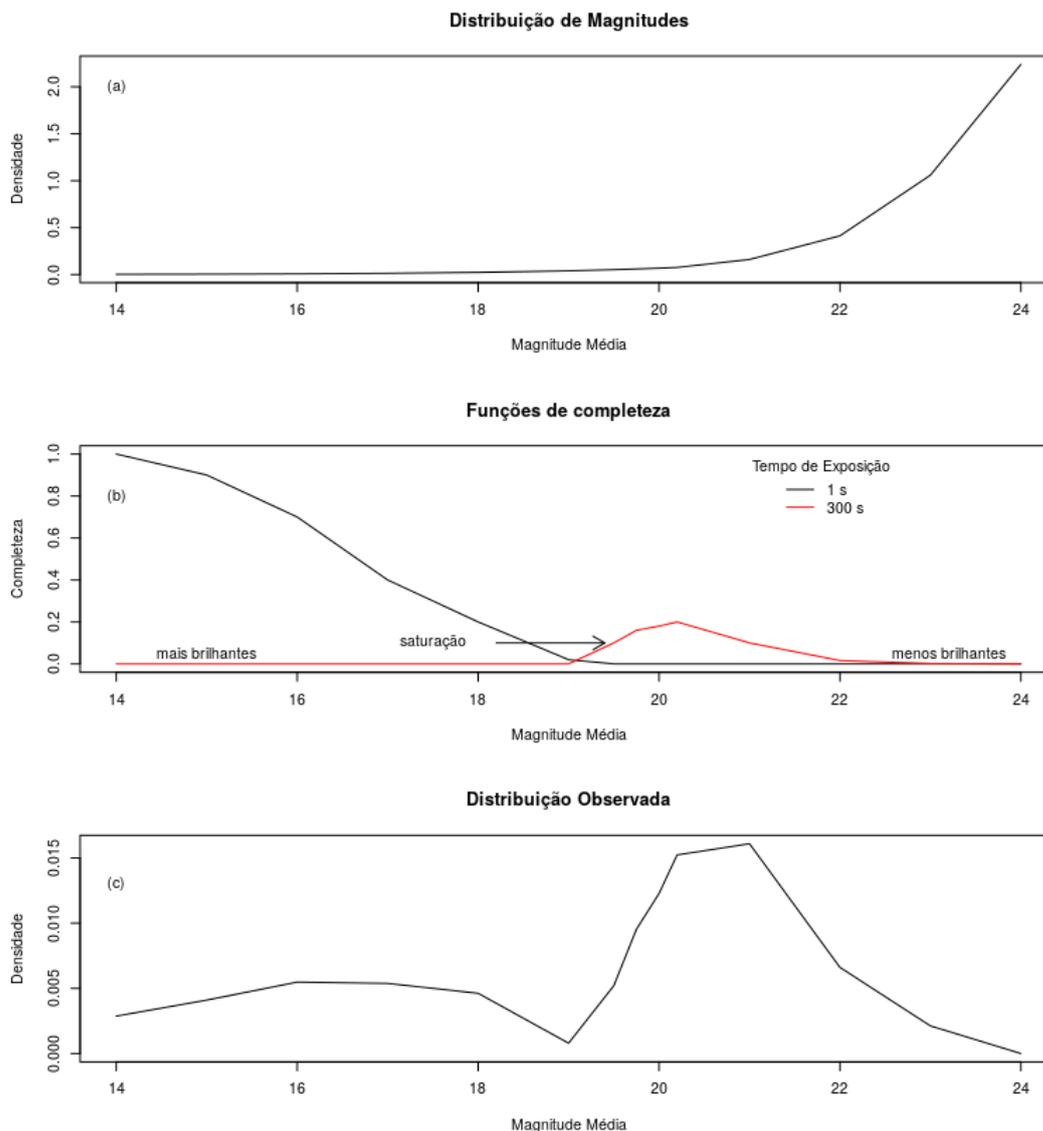


Figura 4.2: No painel (a) é mostrada a distribuição teórica de magnitudes para o NGC 6397, seguindo uma lei de potência. Em (b) são esboçadas as funções de completiza para tempos de exposição de 1 e 300 segundos. No painel (c), a distribuição observada.

4.2 Distribuição de incertezas

A figura 4.3 mostra a distribuição de incertezas para as magnitudes médias nos 4 filtros. Na quase totalidade dos casos, o erro é menor do que 0.2 magnitudes e nunca excedendo 0.6 magnitudes. As incertezas são relativamente baixas, indicando séries temporais com uma alta razão sinal-ruído.

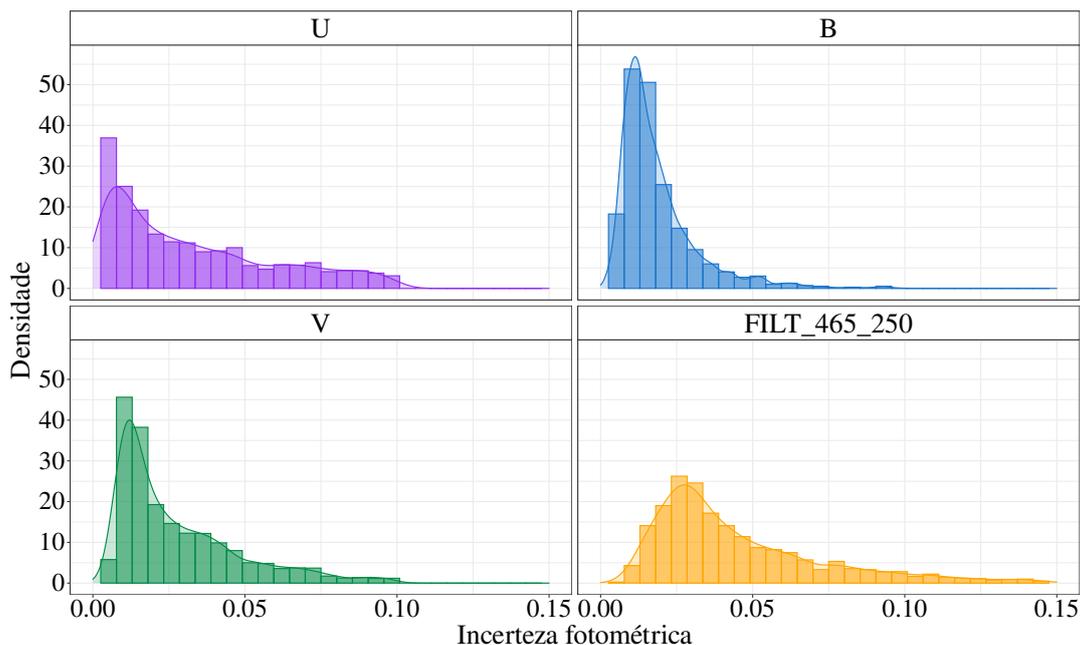


Figura 4.3: Distribuições das incertezas fotométricas das estrelas 3250 estrelas da amostra nos filtros U, B, V e FILT_465_250.

Na figura 4.4, os gráficos de dispersão das médias das magnitudes pelas incertezas fotométricas nos 4 filtros.

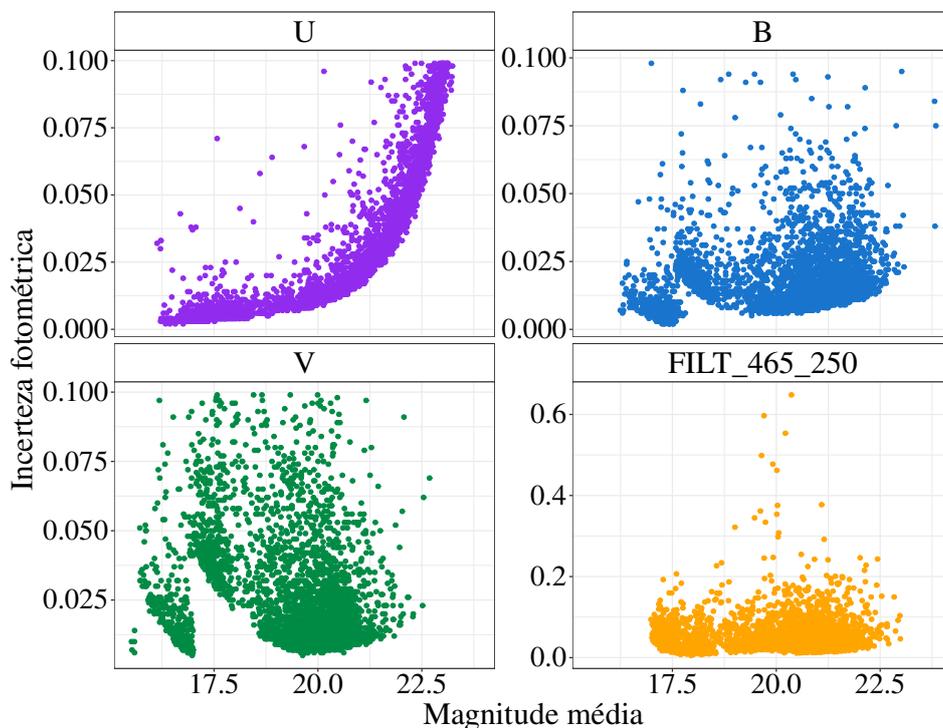


Figura 4.4: Gráficos de dispersão das médias das magnitudes pelas incertezas fotométricas das 3250 estrelas da amostra nos filtros U, B, V e FILT_465_250.

Para o filtro U, observamos que a incerteza fotométrica cresce exponencialmente conforme as magnitudes são maiores, ou seja, conforme as estrelas são mais fracas, o que é esperado, como já discutido anteriormente na seção 2.1. Os filtros B e

V possuem um padrão diferente do filtro U devido às observações realizadas com tempo de exposição diferente, como também comentado na seção 2.1. Já para o filtro FILT_465_250, tanto as magnitudes quanto as incertezas fotométricas apresentadas no gráfico são geradas através das médias das medidas temporais de magnitude e incerteza fotométrica de cada estrela. Logo, faz sentido pensar que as medidas de incertezas serão maiores, pois, ao contrário dos filtros U, B e V, em FILT_465_250 estamos observando a magnitude variando no tempo para cada ponto da imagem do CCD (Figura 2.3).

4.3 Identificação de variáveis por inspeção visual

Cada uma das 3250 séries temporais foi inspecionada visualmente com auxílio do programa VLC, por nós desenvolvido durante a Iniciação Científica. No programa, consegue-se visualizar a série temporal juntamente com as barras de erro das incertezas fotométricas e a linha da magnitude média, como mostrado nas figuras 3.4. Dessa forma, ao examinar visualmente uma série temporal, pode-se levar em consideração o tamanho das incertezas, a amplitude aparente de variação em relação à média e os padrões de oscilação de cada série temporal. Ao analisar com cuidado esses fatores, levou-se aproximadamente 30 segundos inspecionando visualmente cada série temporal, o que resulta em um tempo estimado de inspeção de $\approx 8,5$ horas a cada 1 000 séries temporais e um tempo total para avaliar as 3250 séries temporais de ≈ 27 horas. Ainda, ao inspecionar visualmente a série temporal, há situações onde a classificação como variável ou não-variável não é trivial. Por exemplo, as oscilações em torno da média podem ter uma amplitude baixa, o que pode se confundir com as oscilações devido as incertezas fotométricas, não ficando claro, então, se a série temporal trata-se de uma variável ou está variando devido ao ruído.

Das 3250 séries temporais inspecionadas, 652 dessas foram selecionadas como candidatas à variáveis ($\approx 20\%$ da amostra) através da inspeção visual, restando 2598 séries temporais classificadas como candidatas à não-variáveis (tabela 4.1).

Inspeção visual		
<i>Amostra</i>	<i>Variáveis</i>	<i>Não-variáveis</i>
3 250	652 (20%)	2 598 (80%)

Tabela 4.1: Classificação da variabilidade estelar na amostra de 3250 estrelas, via inspeção visual.

Na figura 4.5 são apresentados os gráficos de barras da fração de variáveis e não-variáveis classificadas por inspeção visual contra os intervalos de magnitudes em cada filtro.

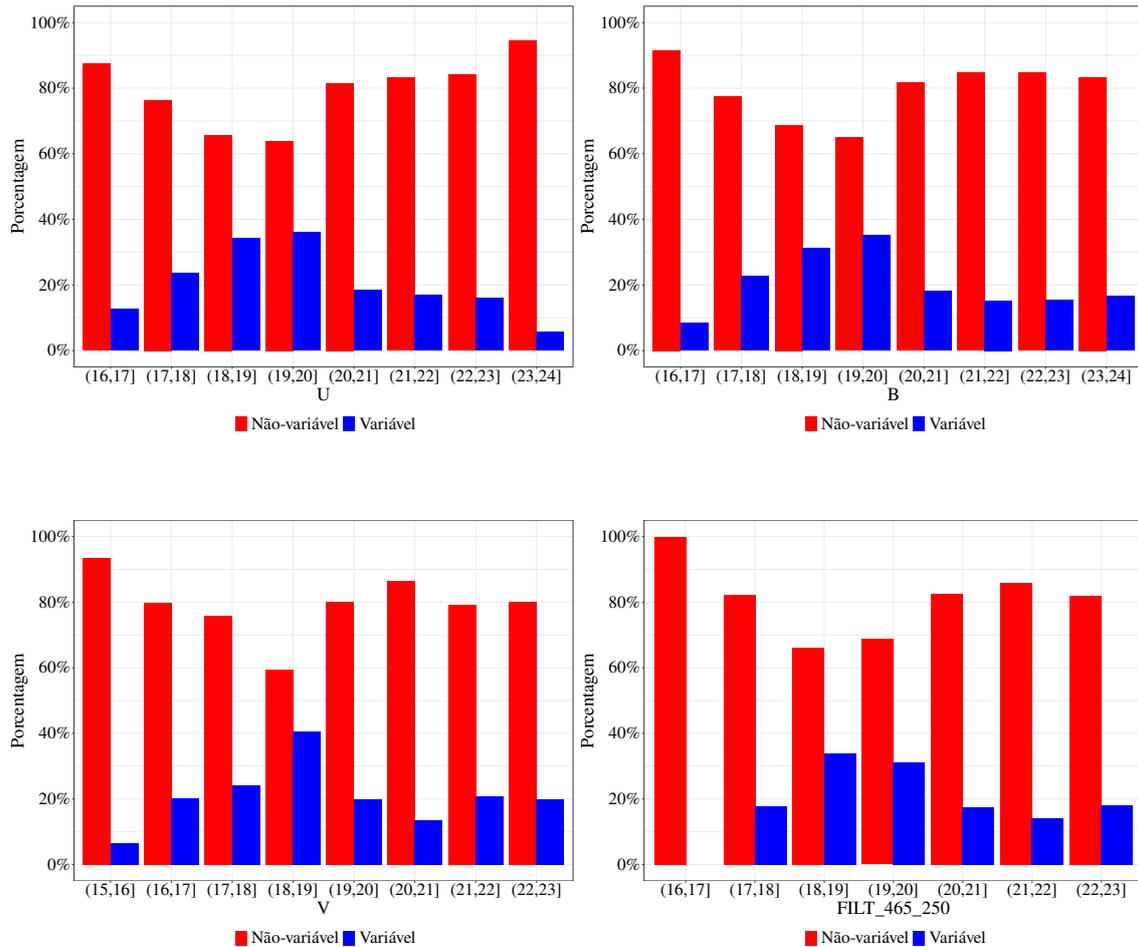


Figura 4.5: Gráficos de barra da fração de variáveis (em azul) e não-variáveis (em vermelho) classificadas por inspeção visual contra os intervalos de magnitudes em cada filtro. Os bins foram normalizados pela quantidade de estrelas da amostra de 3 250 em cada intervalo de magnitude. As porcentagens, por tanto, são relativas a quantidade total de estrelas em cada intervalo de magnitude.

Pode-se observar através dos gráficos de barra acima que, via inspeção visual, a classificação de variáveis se concentra em torno das magnitudes de 18 a 20, em todos os filtros. Nota-se, também, que a distribuição de variáveis possui uma cauda mais alongada para magnitudes maiores, ou seja, estrelas mais fracas. Isso pode ser justificado pela dificuldade de classificar uma estrela mais fraca como variável devido à incertezas fotométricas maiores nessa faixa de magnitudes.

Estes foram os resultados obtidos por inspeção visual para todas as séries temporais. Em seguida, serão apresentados os resultados obtidos por identificação automática.

4.4 Identificação automática de variáveis

4.4.1 Tempo de não-constância

Foi aplicada a metodologia do índice estatístico tempo de não-constância (TNC) às 3250 séries temporais, utilizando um nível de confiança de 99,9% em ambos os intervalos de confiança para o suavizador e para a média. Em contrapartida à

inspeção visual, o tempo médio para o TNC classificar 1000 séries temporais é $\approx 1,2$ segundos, totalizando ≈ 4 segundos para classificar a amostra total de 3250 séries temporais.

O índice indicou 1338 séries temporais como candidatas a estrelas variáveis, totalizando $\approx 41\%$ da amostra. Complementarmente, 1912 séries temporais foram classificadas como de estrelas não-variáveis (tabela 4.2).

Tempo de não-constância

<i>Amostra</i>	<i>Variáveis</i>	<i>Não-variáveis</i>
3 250	1 338 (41%)	1 912 (59%)

Tabela 4.2: Classificação da variabilidade estelar na amostra de 3250 estrelas, via tempo de não-constância.

Em seguida (Figura 4.6) são apresentados os gráficos de barras da fração de variáveis e não-variáveis classificadas pelo TNC contra os intervalos de magnitudes em cada filtro.

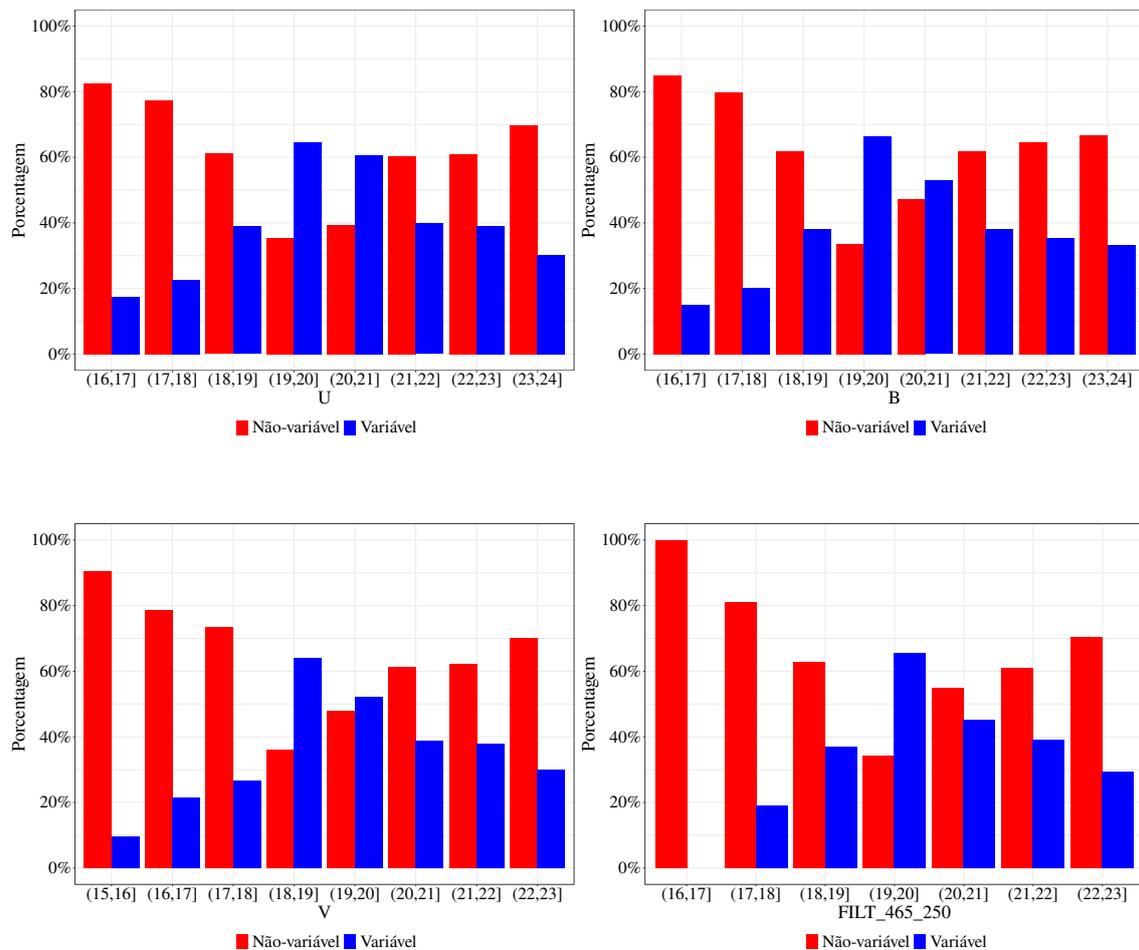


Figura 4.6: Gráficos de barra da fração de variáveis (em azul) e não-variáveis (em vermelho) classificadas pelo TNC contra os intervalos de magnitudes em cada filtro. Os bins foram normalizados pela quantidade de estrelas da amostra de 3 250 em cada intervalo de magnitude. As porcentagens, por tanto, são relativas a quantidade total de estrelas em cada intervalo de magnitude.

Os gráficos de barra da fração estrelas classificadas como variáveis e não-variáveis via TNC são similares aos obtidos via inspeção visual, com maior detecção de variáveis entre as magnitudes 18 a 20. Isso demonstra que eficiência do método é similar a da inspeção visual, com o benefício de ser um processo muito mais rápido - levando 0.004% do tempo da inspeção visual. Além disso, observa-se que a classificação como variável para estrelas mais fracas é maior no TNC do que para a inspeção visual: isso pode ser devido ao TNC não utilizar as incertezas fotométricas ao avaliar a variabilidade estelar.

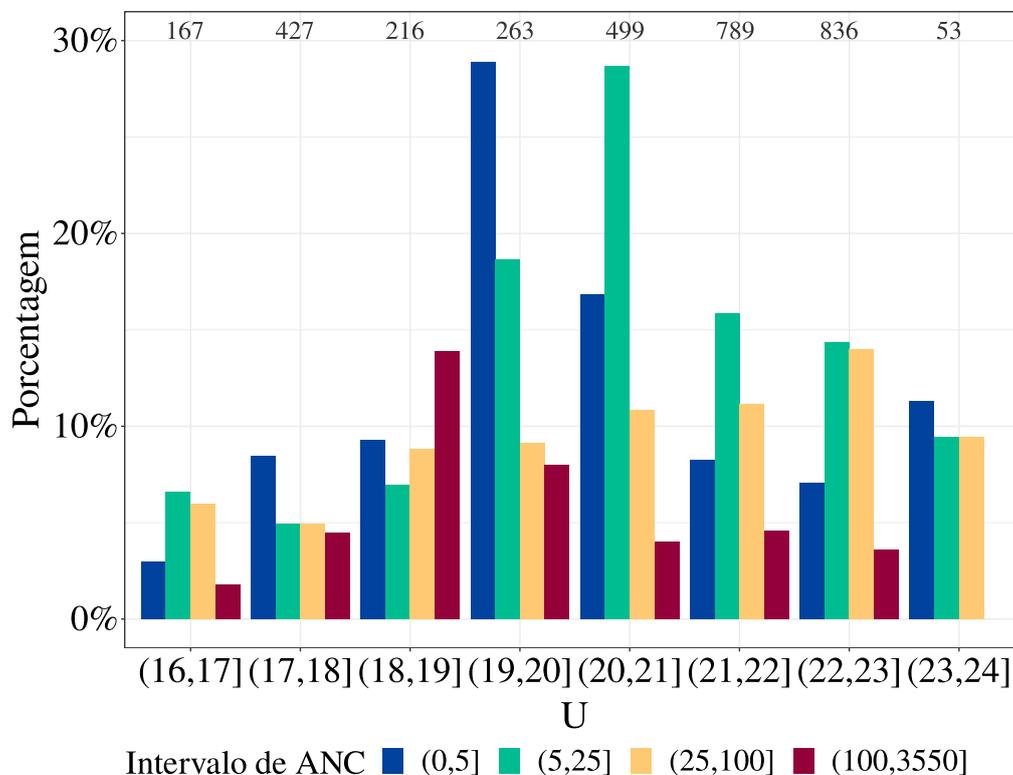
4.4.2 Área de não-constância

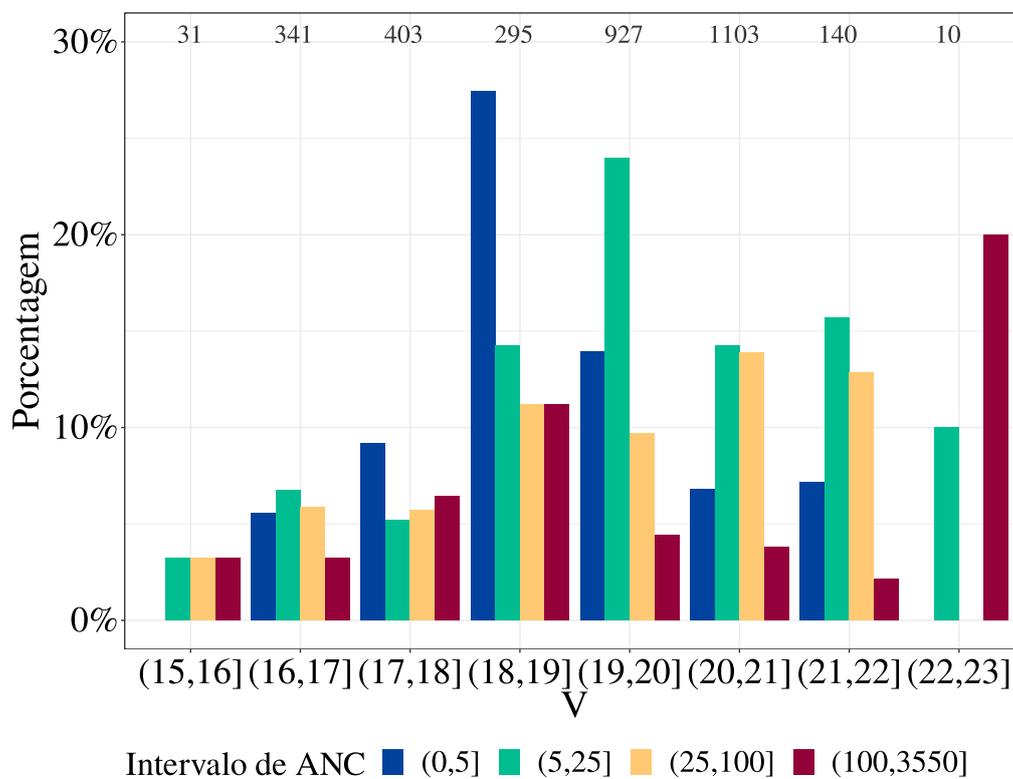
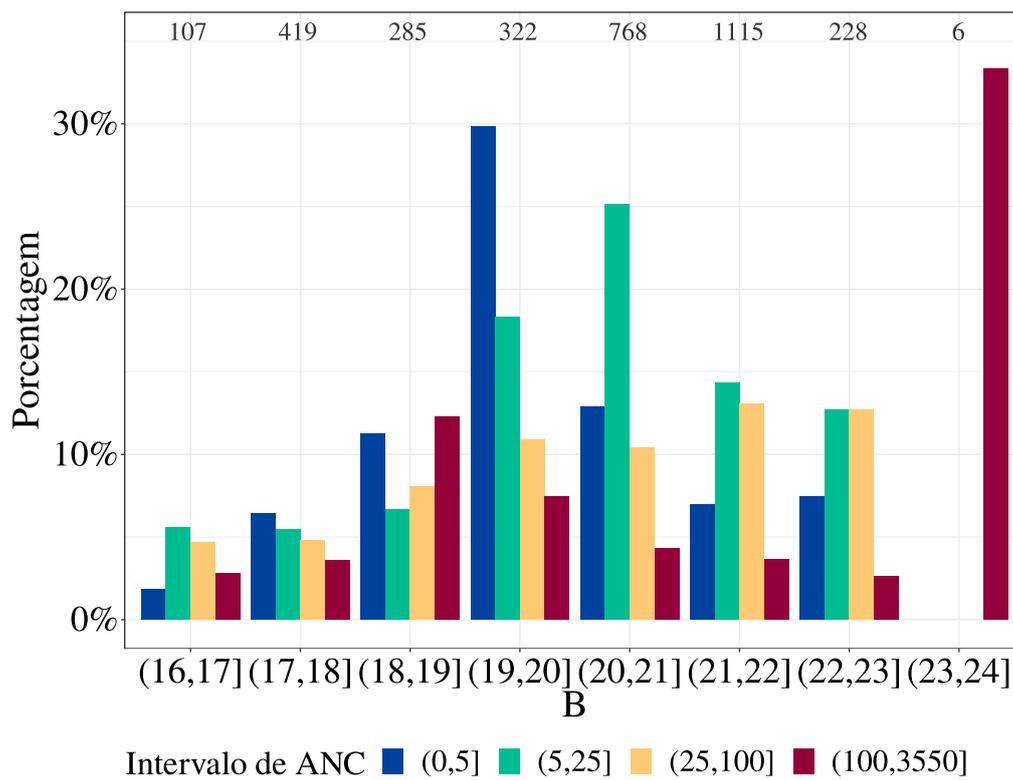
Para as 1338 estrelas - da amostra de 3250 - classificadas como variáveis pelo TNC, pode-se avaliar a área de não-constância (ANC) como índice de quão forte é a suposição de variabilidade para uma dada estrela classificada como variável pelo TNC. Aqui, as ANC variam de ≈ 0.01 a ≈ 3550 (tabela 4.3).

Variáveis	(0,5]	(5,25]	(25,100]	(100,3550]
1 338	351	489	338	159

Tabela 4.3: Quantidade de estrelas classificadas como variáveis pelo TNC por faixa de ANC.

Os gráficos de barra abaixo mostram a fração de estrelas em cada intervalo de ANC contra os intervalos de magnitudes em cada filtro.





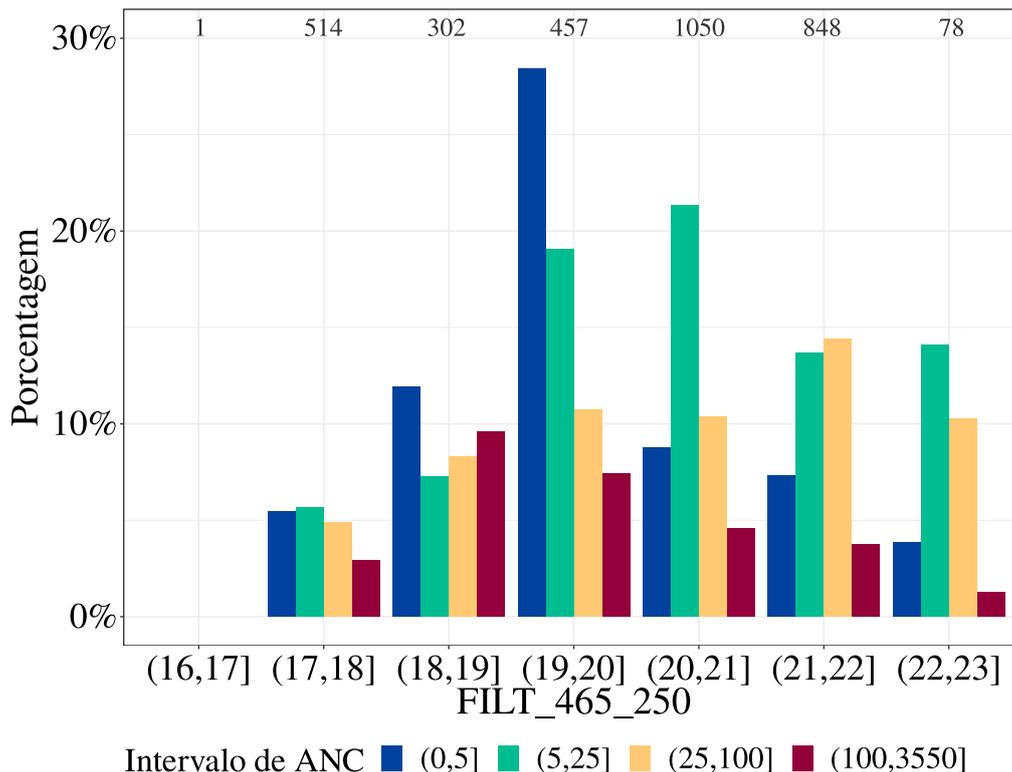


Figura 4.7: Gráficos de barra dos intervalos de ANC das estrelas classificadas pelo TNC como variáveis contra os intervalos de magnitudes em cada filtro. Os bins foram normalizados pela quantidade de estrelas da amostra de 3 250 em cada intervalo de magnitude (valores na parte superior dos gráficos, em cinza). As porcentagens são relativas a quantidade total de estrelas em cada intervalo de magnitude. Importante lembrar que a ANC é calculada somente para as estrelas indicadas como variáveis pelo TNC. Logo o eixo y das porcentagens não se prolonga até 100%.

Observa-se através dos gráficos de barra anteriores que a maior parte das estrelas com alto ANC (barras vermelhas) estão concentradas também na região de maior variabilidade detectada, ou seja, magnitudes entre 18 e 20. Note que o número de estrelas da amostra é diferente em cada bin e para cada filtro.

4.5 Amostra total: Comparando os resultados da identificação automática com a identificação visual

Nesta seção será comparado os dois métodos utilizados para identificação de variabilidade estelar na amostra de 3250 estrelas: a inspeção visual e a identificação automática através do tempo de não-constância. A detecção por inspeção visual resultou em 652 estrelas classificadas como variáveis (20% da amostra) e a detecção automática indicou 1338 estrelas como variáveis (41% da amostra). Dessas, os métodos concordam que 382 estrelas são de variáveis (12% da amostra) e 1642 são não-variáveis (50% da amostra), porém, discordam a respeito da variabilidade de 1226 estrelas (38% da amostra). A tabela 4.4 sumariza a quantidade de estrelas classificadas como variável e não-variável pelos dois métodos.

		Tempo de não-constância		Total
		Não-variável	Variável	
Inspeção visual	Não-variável	1642(50%)	956(30%)	2598(80%)
	Variável	270(8%)	382(12%)	652(20%)
Total		1912(59%)	1338(41%)	3250(100%)

Tabela 4.4: Tabela com os resultados do índice TNC em comparação com a inspeção visual.

A figura seguinte (Figura 4.8) mostra exemplos de séries temporais cujo os métodos discordaram sobre a classificação. A figura à esquerda foi classificada como variável pela inspeção visual mas não pelo TNC. A razão disso vem do fato dessa série temporal ter uma amplitude baixa e pouca frequência de variação, o que acaba fazendo com que o intervalo de confiança de 99,9% para a suavização acabe contando o intervalo de confiança para a magnitude média durante todo tempo avaliado. Já a figura à direita, apesar de não ter sido classificada como variável pela inspeção visual, foi classificada como variável pelo TNC. O motivo é o curto período em torno do tempo 4,5 horas em que o intervalo de confiança para o suavizador não conteve o intervalo de confiança para a magnitude média. Algumas estrelas da amostra apresentaram esses padrões sistemáticos de variabilidade nos extremos de suas séries temporais, que não necessariamente indicam variabilidade da estrela, mas podem ser devido à algum problema ou artefato no início ou final da observação.

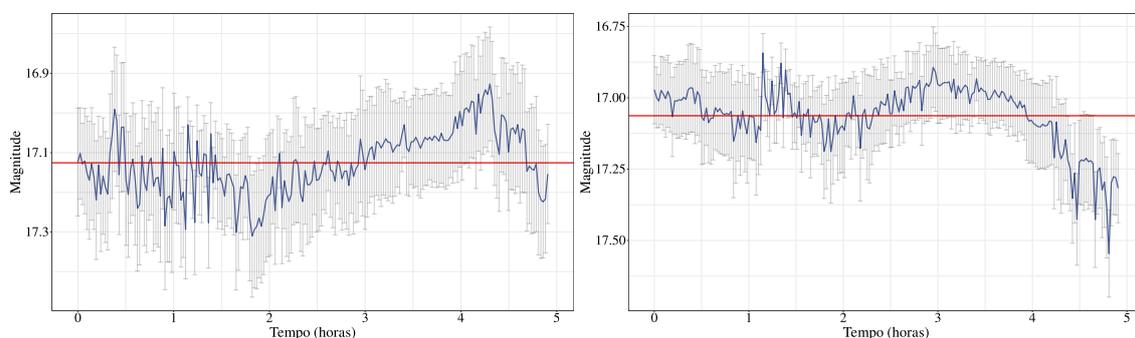


Figura 4.8: Curvas de luz de estrelas cuja inspeção visual e o método TNC discordam. No painel à esquerda, a série temporal de uma estrela classificada como variável pela inspeção visual, mas não pelo TNC. No painel à direita, a série temporal de uma estrela classificada como variável pelo TNC, mas não pela inspeção visual.

Já na figura seguinte (Figura 4.9) há exemplos de curvas de luz cujo os métodos concordam sobre suas classificações. Percebe-se que, na figura à esquerda, o padrão de variabilidade é notável, com amplitudes altas e bem definidas de variação, além de incertezas fotométricas relativamente baixas. Além disso, figura à direita, o padrão constante da magnitude da estrela prevalece, sendo fácil a classificação por ambos os métodos como não-variável.

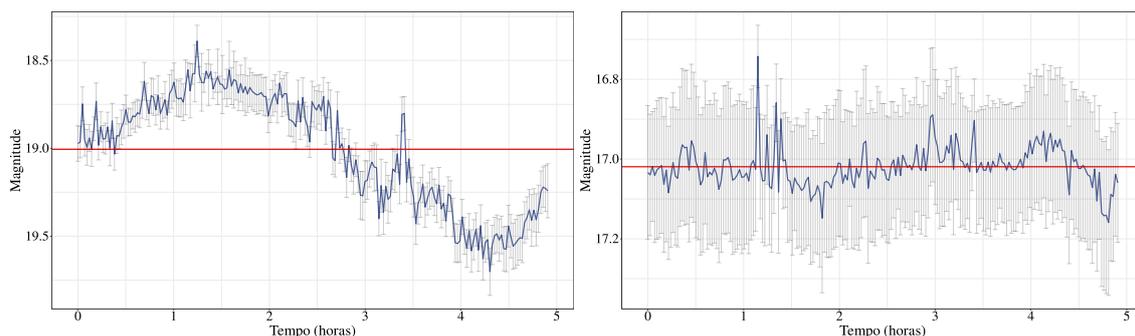


Figura 4.9: Curvas de luz de estrelas cuja inspeção visual e o método TNC concordam. No painel à direita, a série temporal de uma estrela classificada como variável pelo TNC e inspeção visual. No painel à esquerda, a série temporal de uma estrela classificada como não-variável pelo TNC e inspeção visual.

4.6 Identificação e descrição das estrelas do aglomerado globular NGC 6397

Uma estrela pode ser classificada como variável ou não independentemente de pertencer ou não ao aglomerado globular NGC 6397. Mas, para fins de estudo da variabilidade estelar neste aglomerado, é necessário identificar as estrelas com maior probabilidade de pertencer ao mesmo.

Para isso, foram produzidos dois diagramas cor-magnitude para as estrelas da amostra, um com a cor (U-B) contra a magnitude em V e outro com a cor (B-V) contra a magnitude em V. Foi aplicado o procedimento de seleção de estrelas do aglomerado descrito na seção 3.1.1.1 e as figuras 4.10 e 4.11 mostram as estrelas selecionadas como sendo do aglomerado em azul e as não pertencentes ao aglomerado em vermelho, para (U-B) por V e (B-V) por V, respectivamente.

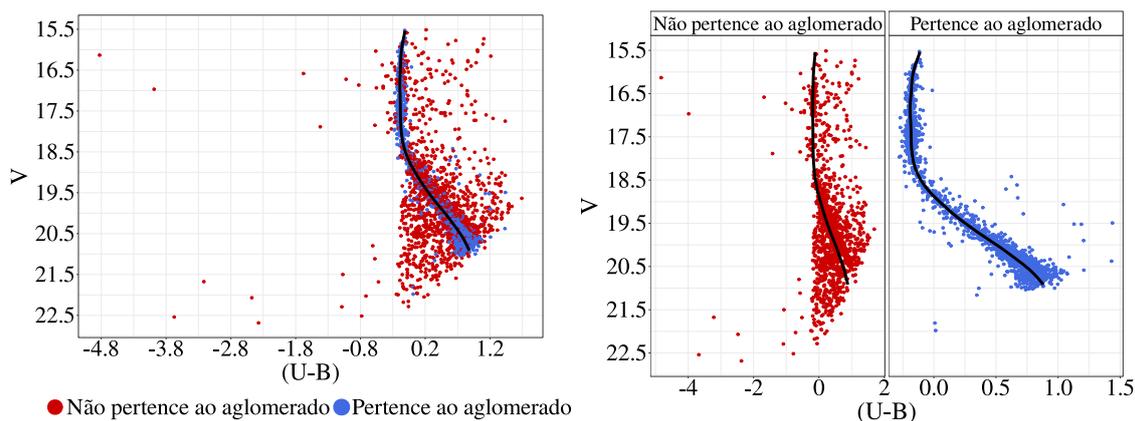


Figura 4.10: Diagrama cor-magnitude V por (U-B). Os círculos azuis são as estrelas que possuem alta probabilidade de pertencer ao aglomerado e os círculos vermelhos são as estrelas que possuem baixa probabilidade de pertencer ao aglomerado. A curva preta é a isócrona.

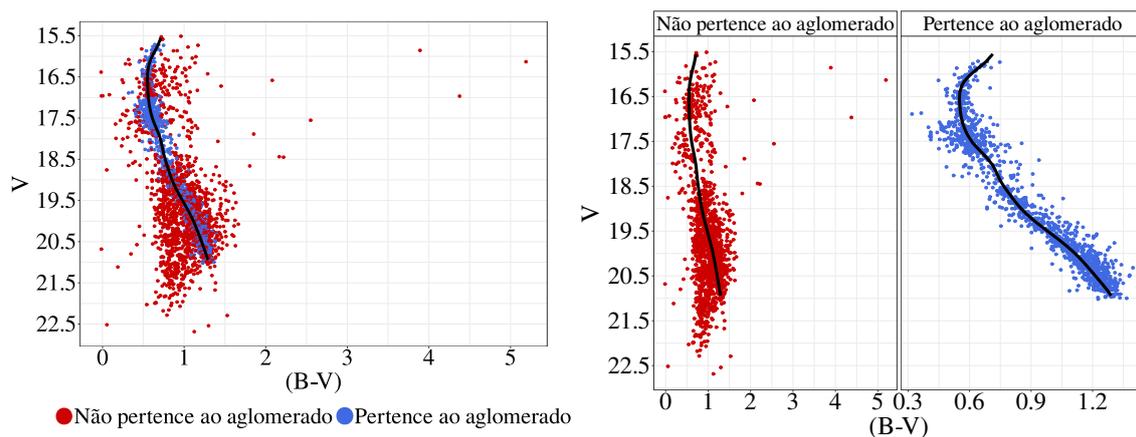


Figura 4.11: Diagrama cor-magnitude V por (B-V). Os círculos azuis são as estrelas que possuem alta probabilidade de pertencer ao aglomerado e os círculos vermelhos são as estrelas que possuem baixa probabilidade de pertencer ao aglomerado. A curva preta é a isócrona.

A amostra reduziu de 3250 para 1132 estrelas, ou seja, para cerca de 1/3 da amostra inicial. As estrelas excluídas caem em um dos quatro casos mencionados na seção 3.1.1.1 ou possuem incertezas fotométricas excessivamente grandes.

As distribuições das magnitudes médias para as estrelas pertencentes ao aglomerado nos filtros U, B, V e FILT_465_250 (Figura 4.12), bem como as incertezas fotométricas associadas (Figura 4.13) são mostradas a seguir.

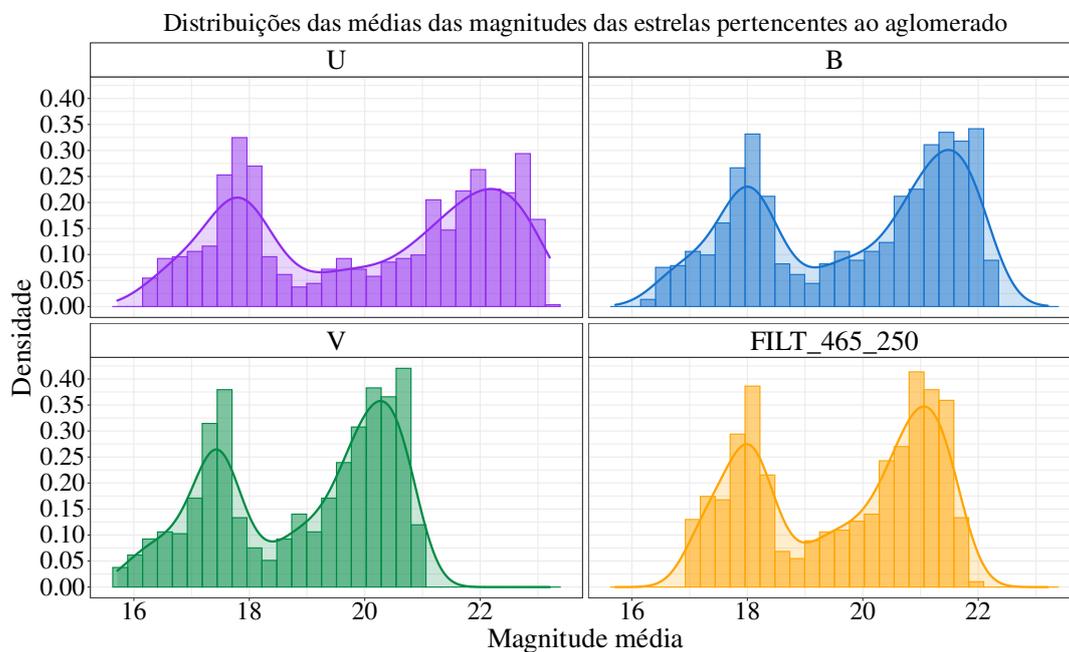


Figura 4.12: Distribuições das médias das magnitudes das estrelas pertencentes ao aglomerado nos filtros U, B, V e FILT_465_250.

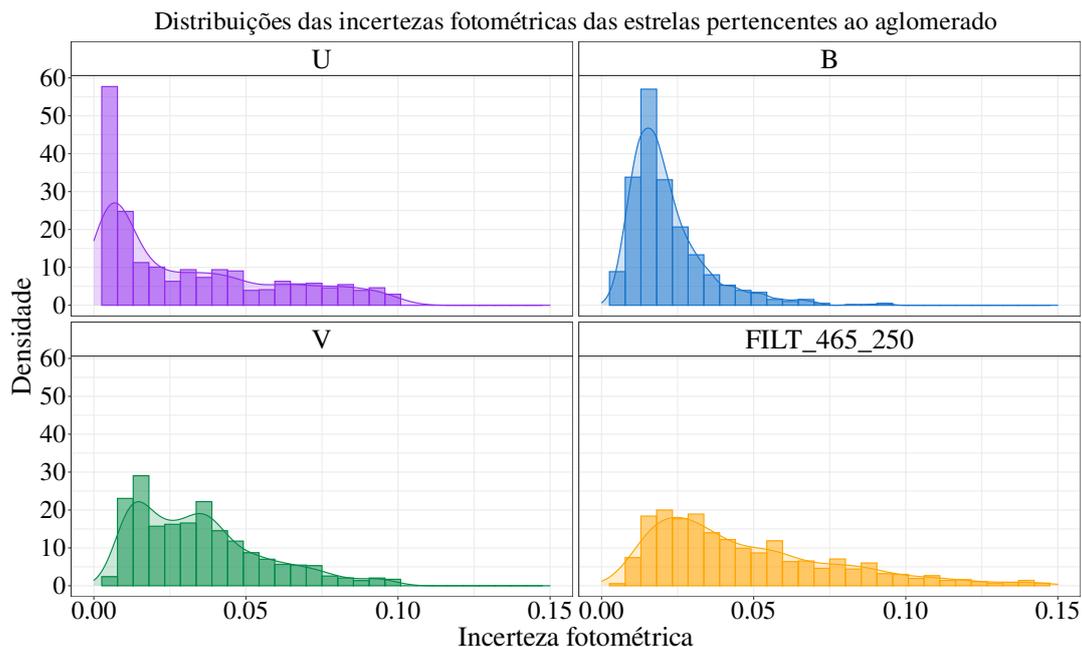


Figura 4.13: Distribuições das incertezas fotométricas das estrelas pertencentes ao aglomerado nos filtros U, B, V e FILT_465_250. As incertezas fotométricas se estendem até 0.6.

De forma similar, as distribuições das magnitudes médias para as estrelas não pertencentes ao aglomerado nos filtros U, B, V e FILT_465_250 (Figura 4.14), bem como as incertezas fotométricas associadas (Figura 4.15) são mostradas a seguir.

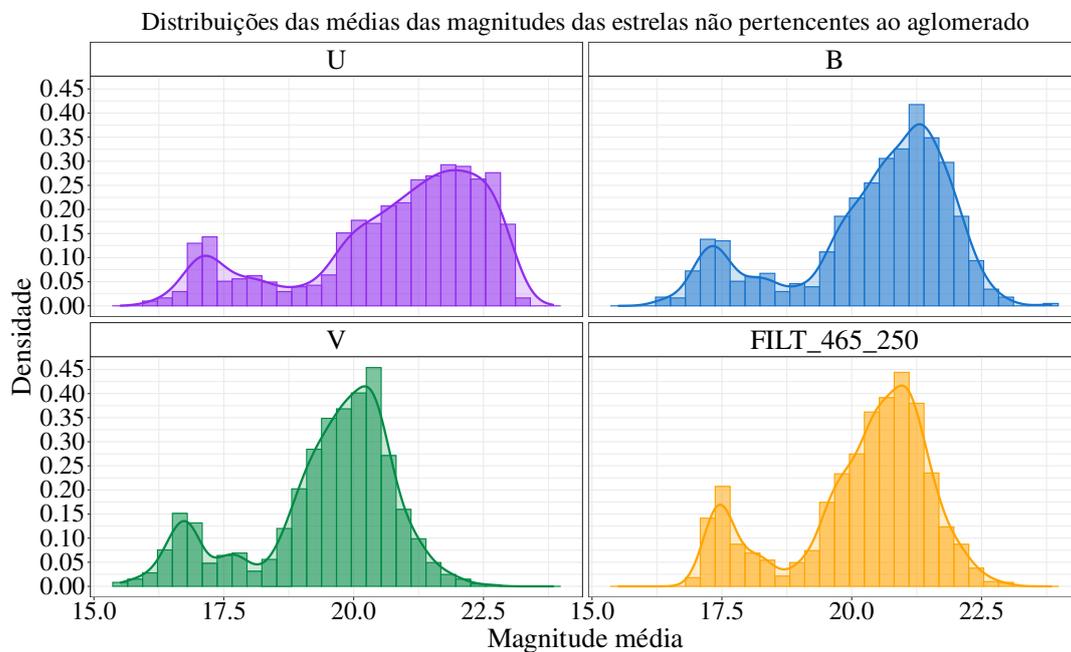


Figura 4.14: Distribuições das médias das magnitudes das estrelas não pertencentes ao aglomerado nos filtros U, B, V e FILT_465_250.

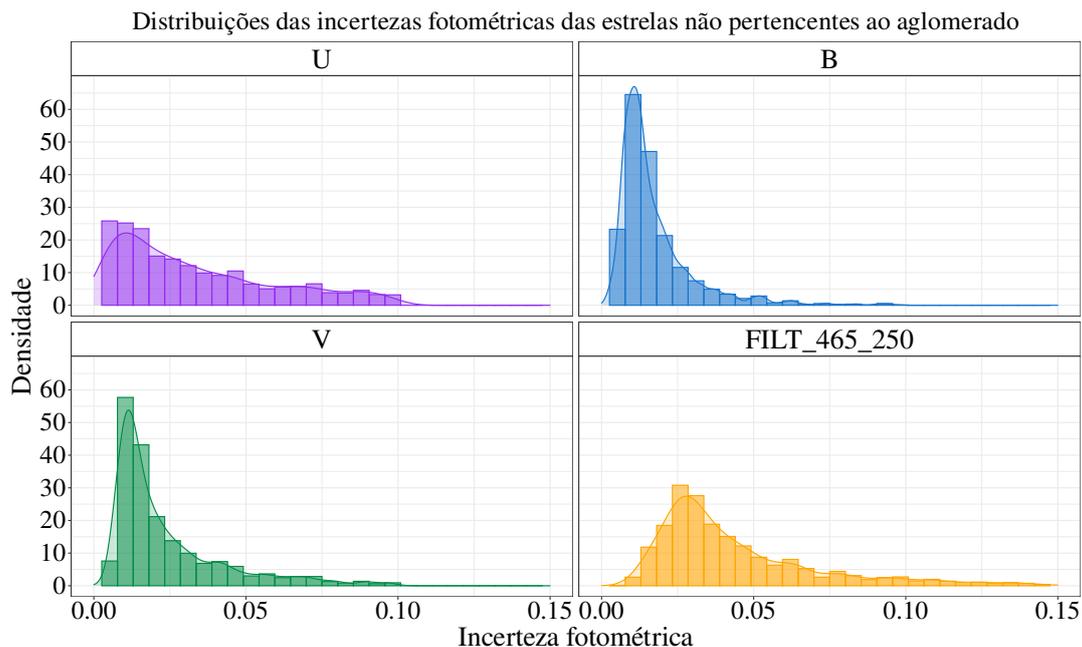


Figura 4.15: Distribuições das incertezas fotométricas das estrelas não pertencentes ao aglomerado nos filtros U, B, V e FILT_465_250. As incertezas fotométricas se estendem até 0.6.

Tanto nas distribuições de magnitudes das estrelas identificadas como pertencentes ao aglomerado (Figura 4.12) quando nas identificadas como não pertencentes (Figura 4.14) pode-se observar o padrão bimodal explicado na seção 4.1, com uma moda em ≈ 17.5 (estrelas mais brilhantes) e outra em ≈ 21.5 (estrelas menos brilhantes). Entretanto, a densidade de estrelas mais brilhantes o grupo do aglomerado é proporcionalmente maior do que para o segundo grupo, conforme esperado, porque populações de aglomerados globulares concentram um número maior de estrelas mais brilhantes do que populações estelares mais difusas.

As distribuições de incertezas fotométricas mostradas nas figuras 4.13 e 4.15 é aproximadamente similar para os dois grupos para os filtros B e FILT_465_250. Para o filtro U (ultravioleta), há uma densidade maior de medidas com incertezas < 0.01 para o grupo de estrelas do aglomerado do que para o segundo grupo. Este resultado é esperado, uma vez que no grupo de estrelas do aglomerado há um maior número de estrelas mais brilhantes que, por sua vez, são mais quentes e emitem mais em ultravioleta.

4.7 Estrelas variáveis do aglomerado globular NGC 6397

Nesta seção será discutido os resultados da análise de variabilidade para as estrelas que possuem alta probabilidade de pertencer ao aglomerado globular NGC 6397.

4.7.1 Identificação visual

Das 1132 séries temporais de brilho da amostra do aglomerado globular NGC 6397, 284 foram selecionadas visualmente como candidatas a estrelas variáveis,

totalizando cerca de 25% da amostra do aglomerado (tabela 4.5).

Inspeção visual		
<i>Amostra</i>	<i>Variáveis</i>	<i>Não-variáveis</i>
1 132	284 (25%)	848 (75%)

Tabela 4.5: Classificação da variabilidade estelar na amostra de 1132 estrelas do aglomerado, via inspeção visual.

A figura abaixo (4.16) mostra os CMDs para as as estrelas do pertencentes ao aglomerado globular, coloridos pela classificação através da inspeção visual.

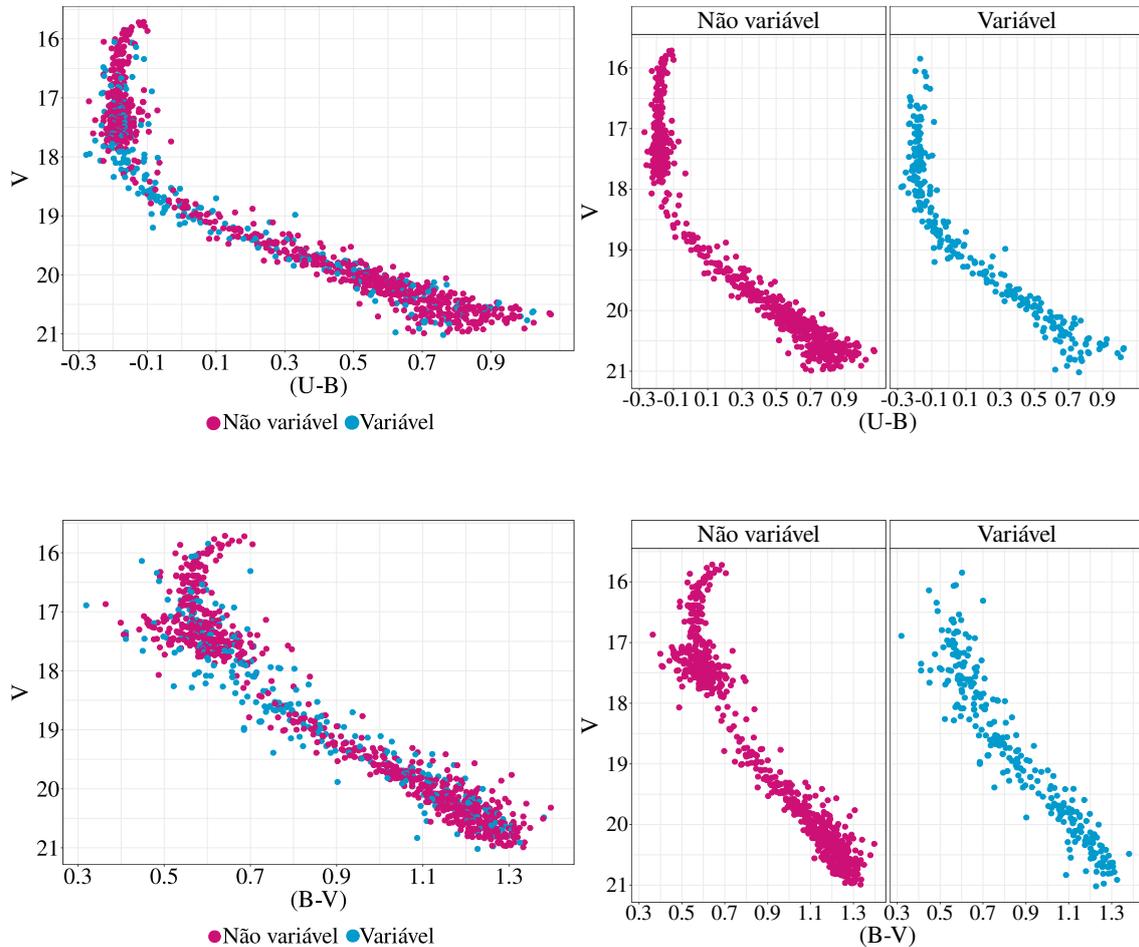


Figura 4.16: Diagramas cor-magnitude para as estrelas do aglomerado, coloridas pela inspeção visual. Nos painéis acima o CMD de V por (U-B) e nos painéis abaixo o CMD de V por (B-V). Os círculos azuis são as estrelas variáveis e os círculos magenta são as estrelas não variáveis.

Pode-se observar que a localização das estrelas variáveis e não variáveis segundo a inspeção visual é dispersa ao longo dos CMDs mostrados anteriormente (Figura 4.16). Essa homogeneidade indica que não há uma concentração de variáveis em regiões da sequência principal, segundo a inspeção visual.

Abaixo (Figura 4.17) são apresentados os gráficos de barras da fração de variáveis e não-variáveis do aglomerado globular classificadas por inspeção visual contra os intervalos de magnitudes em cada filtro.

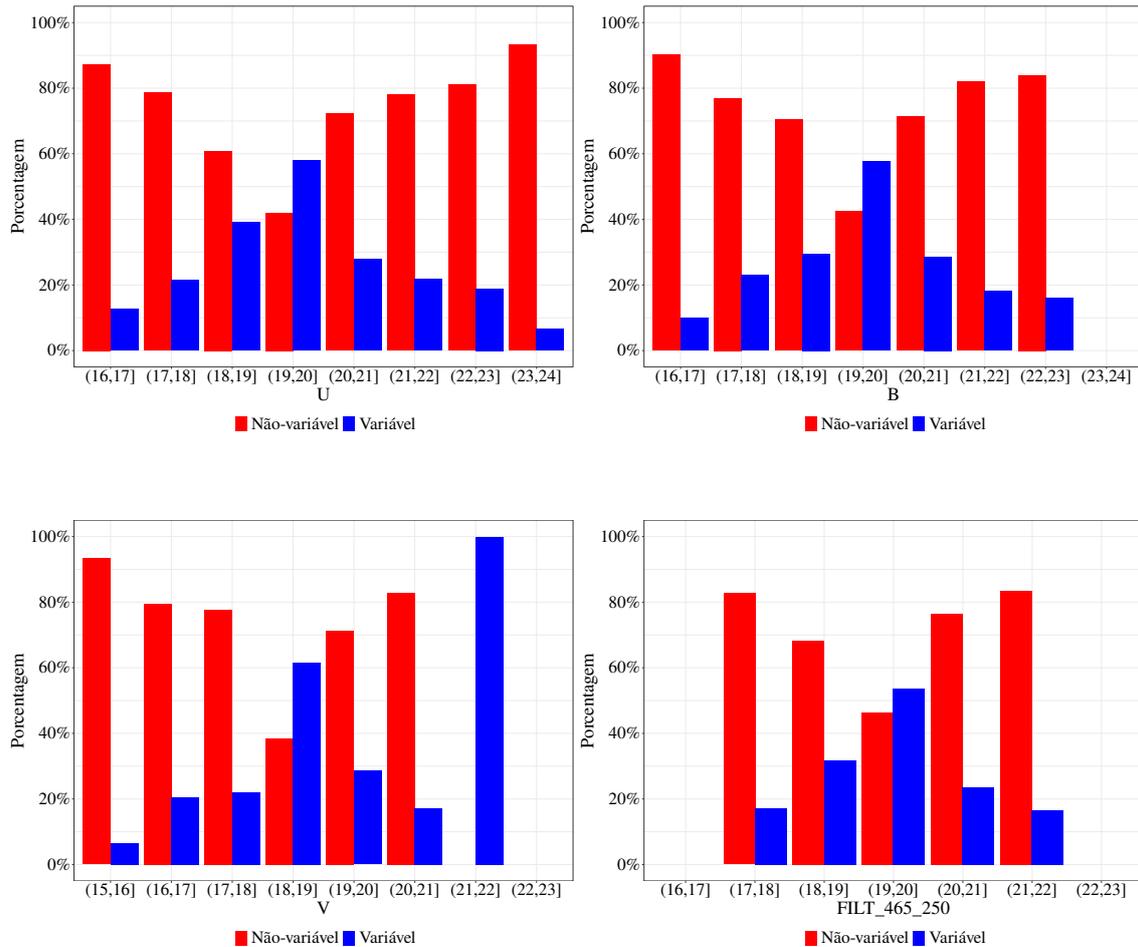


Figura 4.17: Gráficos de barra da fração de variáveis (em azul) e não-variáveis (em vermelho) do aglomerado classificadas por inspeção visual contra os intervalos de magnitudes em cada filtro. Os bins foram normalizados pela quantidade de estrelas do aglomerado (1 132) em cada intervalo de magnitude. As porcentagens, por tanto, são relativas a quantidade total de estrelas do aglomerado em cada intervalo de magnitude.

Em comparação com os gráficos de barra da figura 4.5 para a inspeção visual de todas as estrelas da amostra com a inspeção visual apenas para as estrelas do aglomerado globular (Figura 4.17), percebe-se que, removendo-se as estrelas com maior incerteza fotométrica (concentradas nas estrelas mais fracas), o resultado da variabilidade concentra-se mais ainda, como esperado, nas estrelas com magnitudes entre 18 e 20.

4.7.2 Detecção automática: Tempo de não-constância

Através do índice estatístico TNC, utilizando um nível de confiança de 99,9% em ambos os intervalos de confiança para o suavizador e para a média, foram identificadas 396 estrelas do aglomerado globular NGC 6397 como candidatas à variáveis, totalizando cerca de 35% da amostra do aglomerado (tabela 4.6).

Tempo de não-constância

<i>Amostra</i>	<i>Variáveis</i>	<i>Não-variáveis</i>
1 132	396 (35%)	736 (65%)

Tabela 4.6: Classificação da variabilidade estelar na amostra de 1132 estrelas do aglomerado, via tempo de não-constância.

A figura abaixo (4.18) mostra os CMDs para as as estrelas do pertencentes ao aglomerado globular, coloridos pela classificação através do índice TNC.

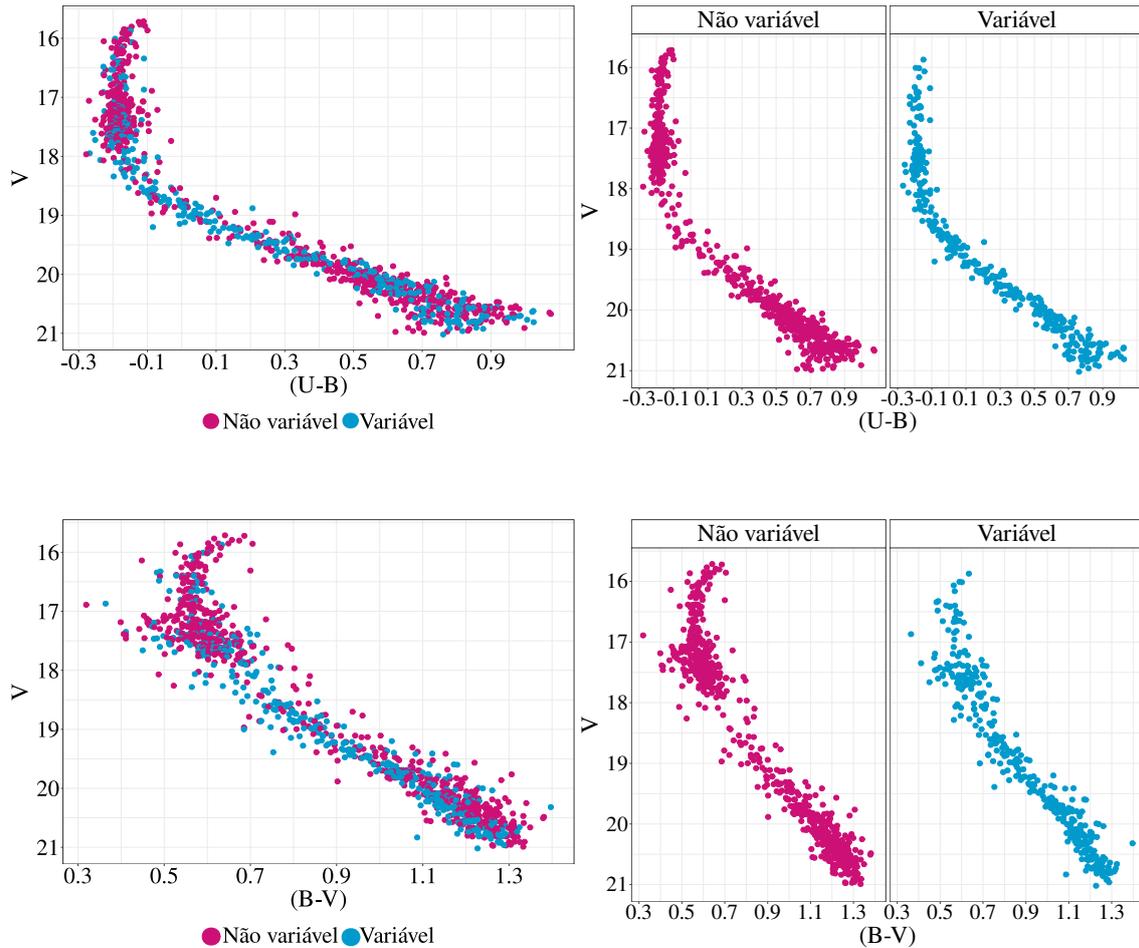


Figura 4.18: Diagrama cor-magnitude para as estrelas do aglomerado, coloridas pelo índice TNC. Nos painéis acima o CMD de V por (U-B) e nos painéis abaixo o CMD de V por (B-V). Os círculos azuis são as estrelas variáveis e os círculos magenta são as estrelas não variáveis.

Nota-se que a distribuição no CMD das variáveis detectadas pelo TNC é similar a distribuição das mesmas por inspeção visual, não contendo nenhuma preferência por uma região específica. Isso novamente demonstra que a eficiência do TNC é similar à inspeção visual, mas também indica que o TNC não introduz nenhum viés observacional referente a tipos específicos de variabilidade associados à regiões do CMD.

Abaixo (Figura 4.19) são apresentados os gráficos de barras da fração de variáveis e não-variáveis do aglomerado globular classificadas pelo TNC contra os intervalos de magnitudes em cada filtro.

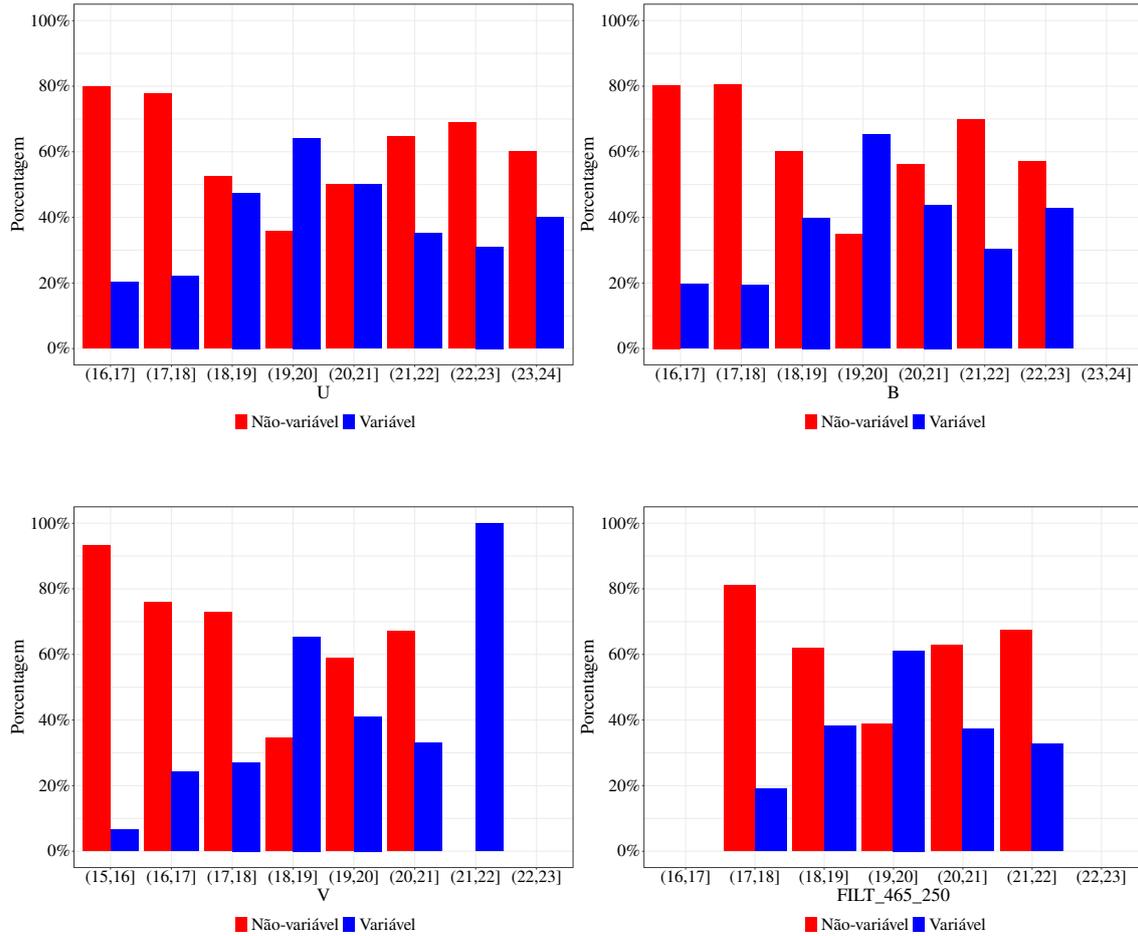


Figura 4.19: Gráficos de barra da fração de variáveis (em azul) e não-variáveis (em vermelho) do aglomerado classificadas pelo TNC contra os intervalos de magnitudes em cada filtro. Os bins foram normalizados pela quantidade de estrelas do aglomerado (1 132) em cada intervalo de magnitude. As porcentagens, por tanto, são relativas a quantidade total de estrelas do aglomerado em cada intervalo de magnitude.

Na figura 4.19 nota-se que a proporção de variabilidade detectada entre os bins de magnitude é mais homogênea que o previamente detectado com a amostra contendo todas as estrelas. Interpreta-se este resultado como consequência da remoção de estrelas com alta incerteza fotométrica. E, novamente, podemos observar que o TNC possui o mesmo padrão de detecção da inspeção visual, comparando-se com a figura 4.17.

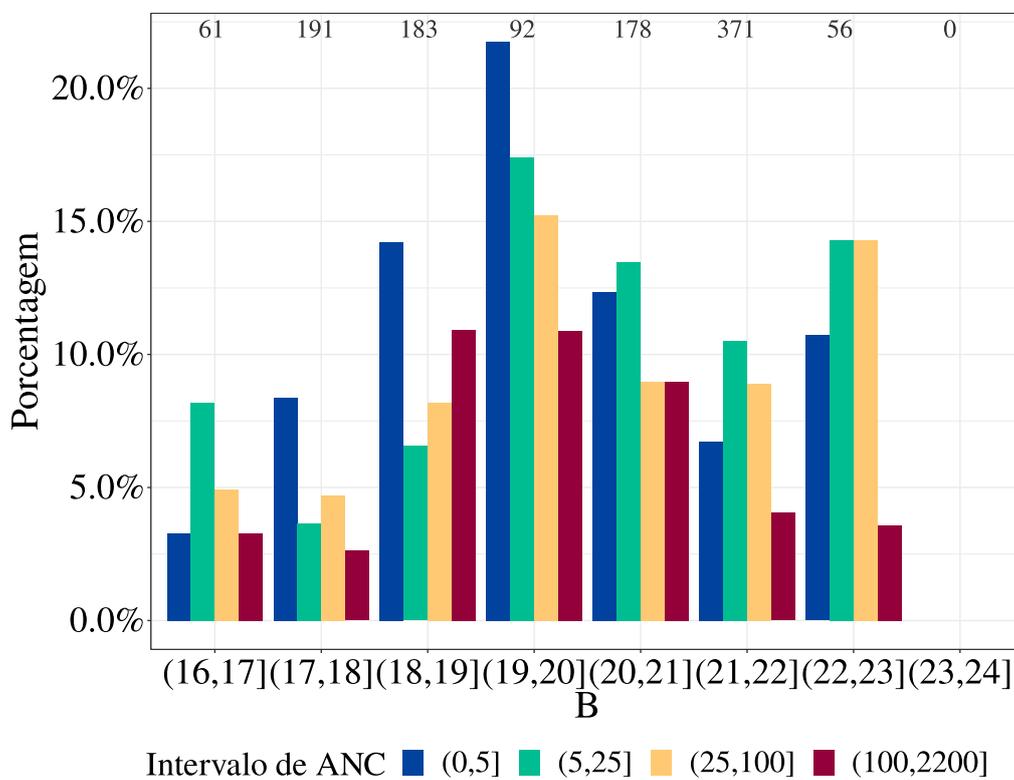
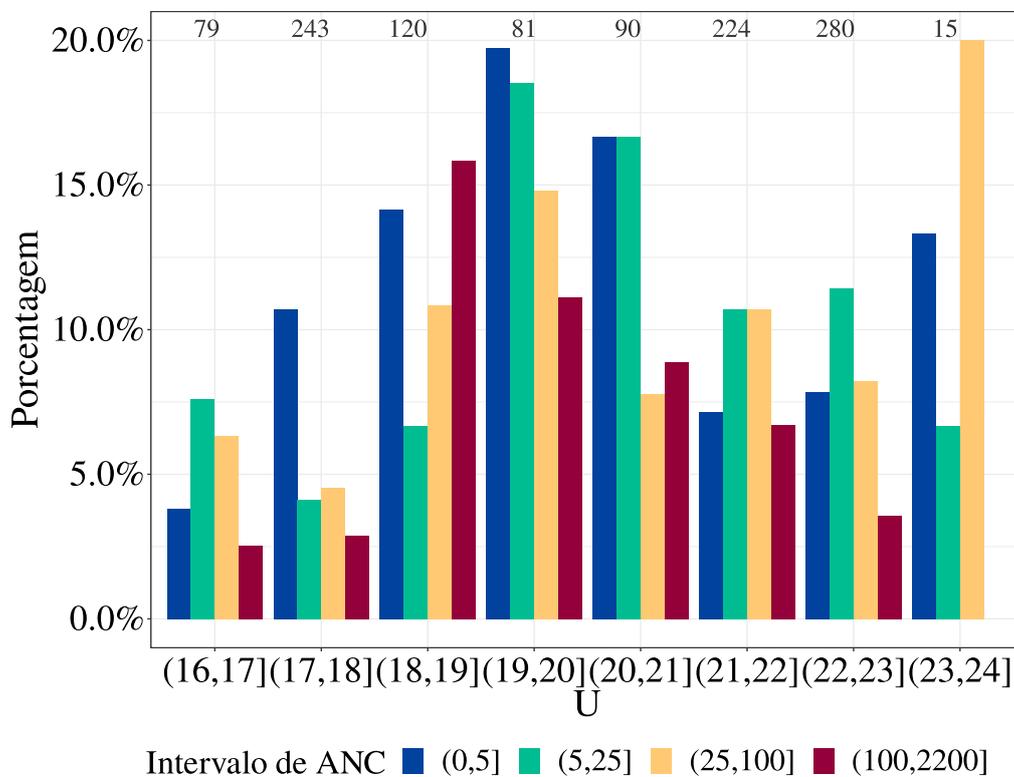
4.7.3 Detecção Automática: Área de não-constância

Foi medida, também, a área de não-constância (ANC) do para as 396 estrelas do aglomerado globular indicadas como variáveis pelo TNC. Utilizou-se, igualmente, um nível de confiança de 99,9% em ambos os intervalos de confiança para o suavizador e para a média. Aqui, as ANC variam de ≈ 0.04 a $\approx 2\ 200$.

Variáveis	(0,5]	(5,25]	(25,100]	(100,2200]
	396	117	111	98
				70

Tabela 4.7: Quantidade de estrelas do aglomerado globular classificadas como variáveis pelo TNC por faixa de ANC.

Os gráficos de barra abaixo mostram a fração de estrelas do aglomerado globular em cada intervalo de ANC contra os intervalos de magnitudes em cada filtro.



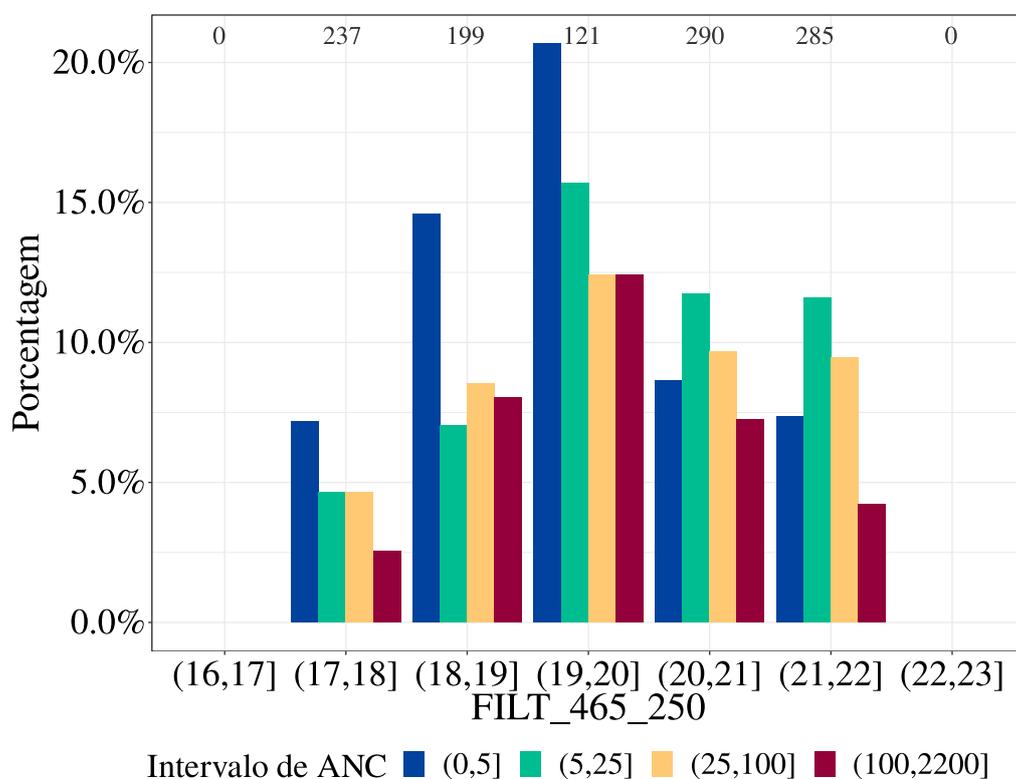
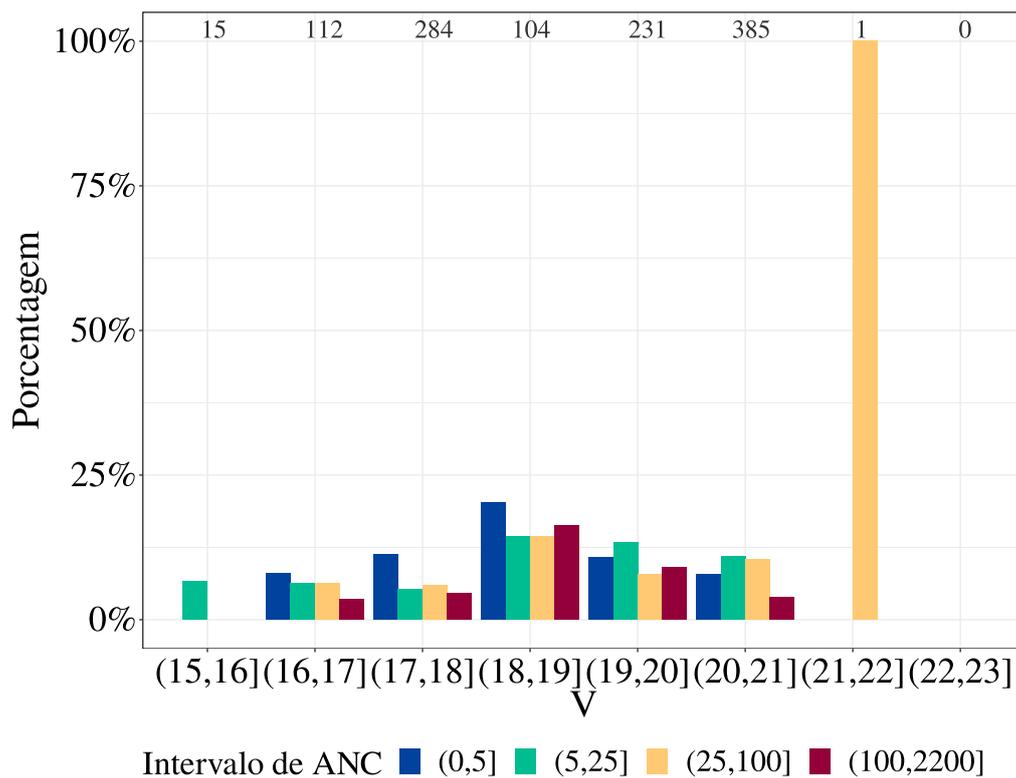


Figura 4.20: Gráficos de barra dos intervalos de ANC das estrelas do aglomerado classificadas pelo TNC como variáveis contra os intervalos de magnitudes em cada filtro. Os bins foram normalizados pela quantidade de estrelas da amostra de 1 132 em cada intervalo de magnitude (valores na parte superior dos gráficos, em cinza). As porcentagens são relativas a quantidade total de estrelas em cada intervalo de magnitude. Importante lembrar que a ANC é calculada somente para as estrelas indicadas como variáveis pelo TNC. Logo o eixo y das porcentagens não se prolonga até 100%.

Observa-se através dos gráficos de barra anteriores que a intensidade de variabilidade, demonstrada pelo índice ANC, varia conforme os bins de magnitude. De forma que as estrelas com variabilidade mais intensa (barras vermelhas, alto ANC) estão localizadas predominantemente entre magnitudes 19 e 21, em todas os filtros. Conforme segue-se para bins de magnitude mais brilhantes (ou seja, mais propensos a efeitos de saturação) ou mais fracos (com incertezas fotométricas relativamente maiores, mesmo após o corte em incerteza fotométrica), a variabilidade detectada pelo ANC é menos intensa, em média. Interpreta-se este resultado como indicativo de como o ANC é capaz de compensar vieses observacionais ao quantificar a intensidade da variabilidade detectada pelo TNC.

Na figura 4.21 pode-se avaliar a população de estrelas, em função da ANC, ao longo da sequência principal através dos CMDs coloridos pelos intervalos de área de não-constância.

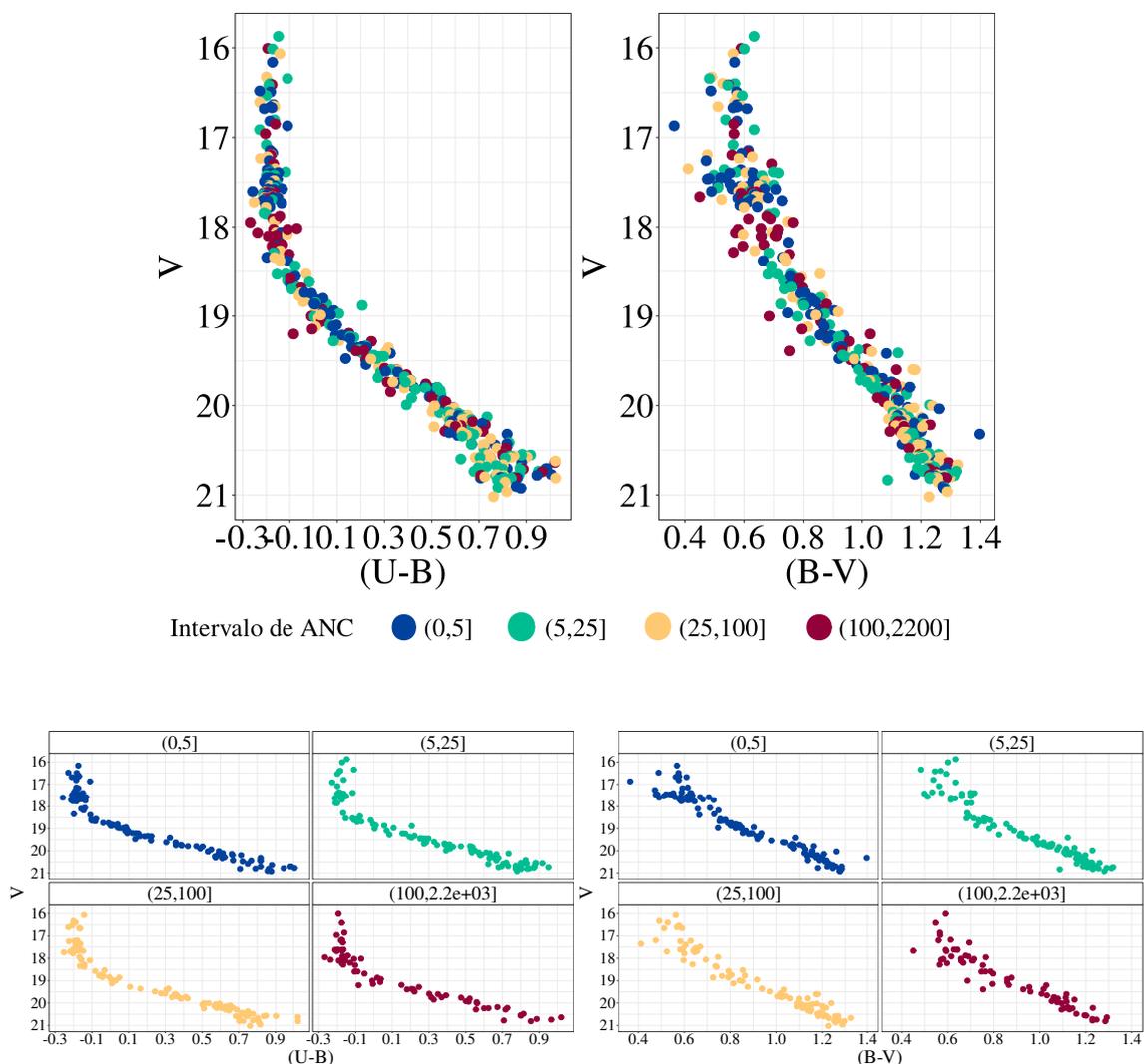


Figura 4.21: Diagrama cor-magnitude para as estrelas do aglomerado identificadas como variáveis pelo TNC e coloridas pelo índice ANC.

Percebe-se que os intervalos de ANC distribuem-se homogeneamente através dos CMDs, não havendo uma concentração maiores em regiões específicas do gráfico, o que é esperado já que a ao longo da sequência principal estrelas são mais semelhantes

umas das outras. Logo, a uniformidade dos intervalos de ANC ao longo dos CMDs garante um resultado não viesado por parte do método, caracterizando de forma verossímil a população de estrelas da sequência principal do aglomerado globular NGC 6397.

4.8 Amostra do aglomerado globular: Comparando os resultados da identificação automática com a identificação visual

Comparando os resultados do TNC com a inspeção visual para as estrelas do aglomerado globular NGC 6397, obteve-se os resultados apresentados na tabela 4.8. Pode-se observar que as proporções para as classificações em ambos os métodos seguem similares àquelas descritas na tabela 4.4, que descreve os resultados para a amostra total de 3250 estrelas.

		Tempo de não-constância		Total
		Não-variável	Variável	
Inspeção visual	Não-variável	662(55%)	226(20%)	848(75%)
	Variável	114(10%)	170(15%)	284(25%)
Total		736(65%)	396(35%)	1132(100%)

Tabela 4.8: Tabela com os resultados do índice TNC em comparação com a inspeção visual, para estrelas do aglomerado.

5 Conclusões e perspectivas

Neste projeto foi analisado um conjunto de 3250 séries temporais com medidas fotométricas de estrelas, das quais 1132 pertencem ao aglomerado globular NGC 6397 e adjacências, com o objetivo principal de estimar a fração de variáveis no aglomerado. A seguir, resume-se as principais conclusões obtidas neste trabalho.

1. Apresentamos dois índices estatísticos, o tempo de não-constância (TNC) e a área de não-constância (ANC). Seus objetivos são auxiliar na detecção e caracterização, respectivamente, de estrelas variáveis a partir de curvas de luz - isto é, séries temporais de magnitude.
2. Foram analisadas quanto à variabilidade amostras de estrelas do aglomerado NGC 6397. Primeiramente, analisou-se a amostra geral de estrelas e posteriormente, analisando-se o diagrama cor-magnitude (CMD), conclui-se que $\approx 35\%$ da amostra tem alta probabilidade de pertencer à sequência principal do aglomerado. A inspeção dos pontos excluídos sugere que o critério utilizado é altamente conservativo, tendo alto grau de pureza, mas baixo grau de completeza; adequada entretanto para fins de análise da dependência da variabilidade em função da magnitude estelar.
3. Para as estrelas com alta probabilidade de pertencer à sequência principal, encontrou-se 25% de variáveis via inspeção visual e 35% via tempo de não-constância. Ambos os métodos concordam sobre a variabilidade de cerca de 15% dessas estrelas.
4. Os índices estatísticos apresentados demonstram ter eficiência em detectar variabilidade comparável à identificação por inspeção visual, como usualmente praticado na área, porém numa fração do tempo.
5. Quanto às limitações do método automático nota-se que o mesmo tem relevante nível de detecção de falsos positivos para curvas de luz possuindo padrões de variabilidade artificiais no início ou final da série temporal, provavelmente relacionados a efeitos observacionais no início ou final da noite.

Com base nos resultados apresentados neste trabalho, para trabalhos futuros propõe-se:

- Um refinamento do método acerca de suas limitações no que se refere à natureza observacional dos dados;

- Um aprimoramento do método para caracterização de diferentes padrões de séries temporais referentes a diferentes tipos de estrelas, não somente estrelas de sequência principal.
- Extensão da análise aqui apresentada para estrelas de outros aglomerados, e portanto com diferentes características observacionais, afim não só de detecção de variáveis nestes, mas também de analisar a possível influência de outros vieses observacionais.

Referências Bibliográficas

- Alcock, C., Allsman, R. A., Alves, D., Axelrod, T. S., Becker, A. C., Bennett, D. P., Cook, K. H., Freeman, K. C., Griest, K., Guern, J., Lehner, M. J., Marshall, S. L., Peterson, B. A., Pratt, M. R., Quinn, P. J., Rodgers, A. W., Stubbs, C. W., Sutherland, W., e Welch, D. L. (1997). The MACHO Project Large Magellanic Cloud Microlensing Results from the First Two Years and the Nature of the Galactic Dark Halo. *The Astrophysical Journal*, 486(2):697–726.
- Carbonell, M., Oliver, R., Ballester, J. L., Carbonell, M., Oliver, R., e Ballester, J. L. (1992). Power spectra of gapped time series - A comparison of several methods. Technical Report 1.
- Clement, C. (2017). Catalogue of variable stars in Milky Way globular clusters. *Wide-Field Variability Surveys: A 21st Century Perspective - 22nd Los Alamos Stellar Pulsation - Conference Series Meeting, San Pedro de Atacama, Chile, Edited by Catelan, M.; Gieren, W.; EPJ Web of Conferences, Volume 152, id.01021*, 152:01021.
- de Oliveira, K. e Saraiva, M. d. F. (2013). *Astronomia & Astrofísica*. Livraria da Física, São Paulo.
- Deb, S. e Singh, H. P. (2009). Light curve analysis of variable stars using Fourier decomposition and principal component analysis. *Astronomy and Astrophysics*, 507(3):1729–1737.
- Drake, A. J., Djorgovski, S. G., Catelan, M., Graham, M. J., Mahabal, A. A., Larson, S., Christensen, E., Torrealba, G., Beshore, E., Mcnaught, R. H., Garradd, G., Belokurov, V., e Koposov, S. E. (2017). The Catalina Surveys Southern periodic variable star catalogue. *Monthly Notices of the Royal Astronomical Society*, 469:3688–3712.
- Eadie, G., Lored, T. J., Mahabal, A. A., Siemiginowska, A., Feigelson, E., Ford, E. B., Djorgovski, S. G., Graham, M., Ivezić, Z., Borne, K., Cisewski-Kehe, J., Peek, J. E. G., Schafer, C., Yanamandra-Fisher, P. A., e Young, C. A. (2019). Realizing the potential of astrostatistics and astroinformatics. *Jessi Cisewski-Kehe*, 16:19.
- Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*.
- Elorrieta, F., Eyheramendy, S., e Palma, W. (2019). Discrete-time autoregressive model for unequally spaced time-series observations. *Astronomy and Astrophysics*, 627.

- Eyer, L. e Mowlavi, N. (2008). Variable stars across the observational HR diagram. *Journal of Physics: Conference Series*, 118(1).
- Friedman, J. H. (1984). A variable span smoother. Technical report.
- Hosenie, Z., Lyon, R., Stappers, B., Mootoovaloo, A., e McBride, V. (2020). Imbalance Learning for Variable Star Classification. *Monthly Notices of the Royal Astronomical Society*, 493(4):6050–6059.
- Ian Ridpath (2012). *A Dictionary of Astronomy*. Oxford University Press.
- Karttunen, H., Kröger, P., Oja, H., Poutanen, M., e Donner, K. J. (2016). *Fundamental astronomy*.
- Koch, D. G., Borucki, W. J., Basri, G., Batalha, N. M., Brown, T. M., Caldwell, D., Christensen-Dalsgaard, J., Cochran, W. D., DeVore, E., Dunham, E. W., Gautier, T. N., Geary, J. C., Gilliland, R. L., Gould, A., Jenkins, J., Kondo, Y., Latham, D. W., Lissauer, J. J., Marcy, G., Monet, D., Sasselov, D., Boss, A., Brownlee, D., Caldwell, J., Dupree, A. K., Howell, S. B., Kjeldsen, H., Meibom, S., Morrison, D., Owen, T., Reitsema, H., Tarter, J., Bryson, S. T., Dotson, J. L., Gazis, P., Haas, M. R., Kolodziejczak, J., Rowe, J. F., Van Cleve, J. E., Allen, C., Chandrasekaran, H., Clarke, B. D., Li, J., Quintana, E. V., Tenenbaum, P., Twicken, J. D., e Wu, H. (2010). Kepler Mission Design, Realized Photometric Performance, and Early Science. *ApJL*, 713(2):L79–L86.
- Luedicke, J. (2015). Friedman’s Super Smoother. Technical report.
- Martinazzi, E. (2016). Estudo fotométrico e descoberta de novas estrelas variáveis nos aglomerados globulares NGC6397 e NGC288. Technical report.
- Martinazzi, E., Kepler, S. O., e Costa, J. E. (2017). Low-mass variable stars in the globular cluster NGC 6397. *Monthly Notices of the Royal Astronomical Society*, 468(3):2816–2821.
- Percy, J. R. (2007). *Understanding variable stars*, volume 9780521232. Cambridge University Press.
- Pojmański, G. (2004). The all sky automated survey.
- R Development Core Team, R. (2020). *R: A Language and Environment for Statistical Computing*.
- Richards, J. W., Starr, D. L., Butler, N. R., Bloom, J. S., Brewer, J. M., Crellin-Quick, A., Higgins, J., Kennedy, R., e Rischard, M. (2011). On Machine-Learned Classification of Variable Stars with Sparse and Noisy Time-Series Data. *The Astrophysical Journal*, 733(10):10.
- Salpeter, E. E. (1955). The Luminosity Function and Stellar Evolution. *The Astrophysical Journal*, 121:161.
- Shin, M.-S., Sekora, M., e Byun, Y.-I. (2009). Detecting Variability in Massive Astronomical Time-Series Data I: application of an infinite Gaussian mixture model. *Monthly Notices of the Royal Astronomical Society*, 400(4):1897–1910.

- Siemiginowska, A., Eadie, G., Czekala, I., Feigelson, E., Ford, E. B., Kashyap, V., Kuhn, M., Loredó, T., Ntampaka, M., Stevens, A., Avelino, A., Borne, K., Budavari, T., Burkhart, B., Cisewski-Kehe, J., Civano, F., Chilingarian, I., van Dyk, D. A., Fabbiano, G., Finkbeiner, D. P., Foreman-Mackey, D., Freeman, P., Fruscione, A., Goodman, A. A., Graham, M., Guenther, H. M., Hakkila, J., Hernquist, L., Huppenkothen, D., James, D. J., Law, C., Lazio, J., Lee, T., López-Morales, M., Mahabal, A. A., Mandel, K., Meng, X. L., Moustakas, J., Muna, D., Peek, J. E. G., Richards, G., Portillo, S. K. N., Scargle, J., de Souza, R. S., Speagle, J. S., Stassun, K. G., Stenning, D. C., Taylor, S. R., Tremblay, G. R., Trimble, V., Yanamandra-Fisher, P. A., e Young, C. A. (2019). Astro2020 Science White Paper: The Next Decade of Astrominformatics and Astrostatistics.
- Sokolovsky, K. V., Gavras, P., Karamelas, A., Antipin, S. V., Bellas-Velidis, I., Benni, P., Bonanos, A. Z., Burdanov, A. Y., Derlopa, S., Hatzidimitriou, D., Khokhryakova, A. D., Kolesnikova, D. M., Korotkiy, S. A., Lapukhin, E. G., Moretti, M. I., Popov, A. A., Poulialis, E., Samus, N. N., Spetsieri, Z., Veselkov, S. A., Volkov, K. V., Yang, M., e Zubareva, A. M. (2017). Comparative performance of selected variability detection techniques in photometric time series data. *Monthly Notices of the Royal Astronomical Society*, 464(1):274–292.
- Soraisam, M. D., Saha, A., Matheson, T., Lee, C.-H., Narayan, G., Vivas, A. K., Scheidegger, C., Oppermann, N., Olszewski, E. W., Sinha, S., e DeSantis, S. R. (2020). A classification algorithm for time-domain novelties in preparation for LSST alerts: Application to variable stars and transients detected with DECam in the Galactic Bulge. *Monthly Notices of the Royal Astronomical Society*.
- Tody, D. (1986). The Iraf Data Reduction And Analysis System. In Crawford, D. L., editor, *Instrumentation in Astronomy VI*, volume 0627, page 733.
- Udalski, A., Szymański, M. K., e Szymański, G. (2015). OGLE-IV: Fourth Phase of the Optical Gravitational Lensing Experiment. *Acta Astronomica*, 65(1):1–38.
- Welch, D. L. e Stetson, P. B. (1993). Robust variable star detection techniques suitable for automated searches - New results for NGC 1866. *The Astronomical Journal*, 105(5):1813.
- Wozniak, P. R., Vestrand, W. T., Akerlof, C. W., Balsano, R., Bloch, J., Casperson, D., Fletcher, S., Gisler, G., Kehoe, R., Kinemuchi, K., Lee, B. C., Marshall, S., McGowan, K. E., McKay, T. A., Rykoff, E. S., Smith, D. A., Szymanski, J., e Wren, J. (2004). Northern Sky Variability Survey (NSVS): Public data release. *The Astronomical Journal*, 127(4):2436–2449.