

Identificação e classificação de variabilidade estelar em aglomerados através de análise estatística multidimensional



Autora: Maitê Mückler
Orientador: J. E. S. Costa



INTRODUÇÃO

Aglomerados globulares são sistemas estelares que podem conter até alguns milhões de estrelas. Uma fração significativa dessas são variáveis. A identificação e classificação manual de variáveis com base na inspeção visual de curvas de luz tem baixa eficácia e está sujeita a erros. Diferentes testes estatísticos têm sido propostos para detecção de variabilidade, entretanto, porém estes recursos são limitados a certos padrões de variabilidade quando usados individualmente. Um problema inerente a esta abordagem é que a capacidade de cada índice depende do padrão de variabilidade e da relação sinal-ruído da curva de luz.

OBJETIVO

O objetivo principal deste trabalho é combinar o uso de diferentes métodos estatísticos para a identificação e classificação automática de estrelas variáveis a partir de curvas de luz extraídas de séries temporais de imagens de um aglomerado estelar.

DADOS

Neste projeto, utilizou-se dados do aglomerado globular NGC 6397 (Fig.1, esq.), obtidos com o telescópio VLT (Very Large Telescope), de 8.2 metros de abertura, através do imageador FORS2 (2 CCDs de 2048x2048 pixels). As observações ocorreram nos dias 27 e 28 de julho de 2009. Na primeira noite foram obtidas 205 imagens (com tempo de exposição 60 segundos) e na segunda noite 100 imagens (com tempos de exposição de 80 e 170 segundos). (Martinazzi *et al.*, 2014). Desses dados, foram geradas cerca de 13.000 curvas de luz (Fig.1, dir.).

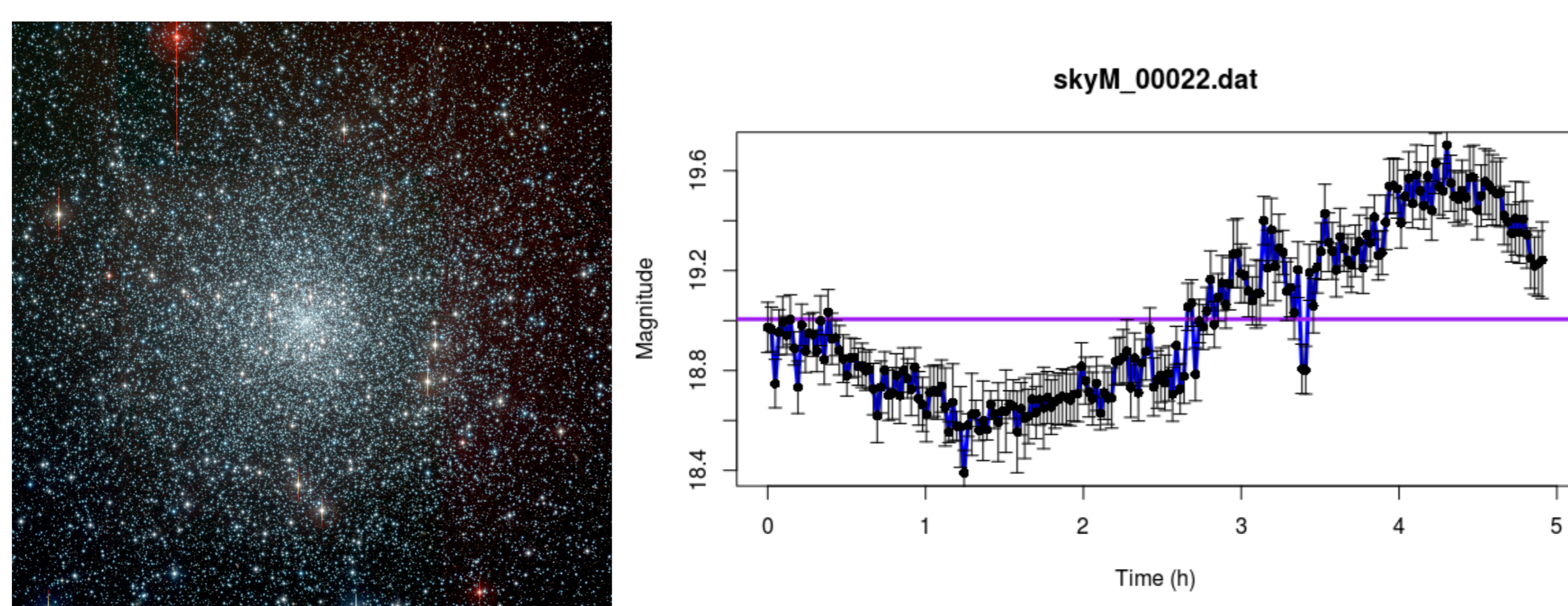


Figura 1: Esquerda: Aglomerado globular NGC 6397 (Créditos: ESO). Direita: Exemplo de curva de luz variável para uma das estrelas do aglomerado.

METODOLOGIA

Na fase inicial do projeto, elaboramos um código para visualização de curvas de luz, escrito em R (R Core Team, 2016), uma linguagem de programação dedicada à análise estatística e visualização de dados. Testamos a aplicação de diferentes métodos de suavização das curvas de luz, como o super suavizador de Friedman, suavização por regressão local e suavização por *kernel*. O código foi utilizado na visualização de aproximadamente 13.000 estrelas do aglomerado globular NGC 6397 (Martinazzi *et al.*, 2015 e 2016), permitindo uma classificação visual, preliminar, de cada estrela como variável (*variable=1*) ou não-variável (*variable=0*), a fim criar uma amostra de controle para testes posteriores.

Na fase atual, implementamos o cálculo de 11 índices de variabilidade (Fig. 2) propostos na literatura (Sokolowski *et al.*, 2016), listados a seguir:

- Teste χ^2
- Desvio padrão, σ
- Desvio padrão ponderado, σ_w
- Desvio absoluto da mediana (MAD)
- Intervalo interquartil (IQR)
- Mediana estatística robusta (RoMS)
- Excesso de variância normalizada, σ_{NXS}^2
- Variabilidade pico-a-pico, ν
- Autocorrelação Lag-1, l_1
- Razão de von Neumann, η
- Excesso no valor Abbe, ϵ_A

Cada índice é tratado como uma dimensão de um espaço multidimensional. Espera-se que estrelas variáveis concentrem-se em certas regiões deste espaço. Há, pelo menos, três problemas que esperamos enfrentar. Primeiro, alguns índices estão fortemente correlacionados entre si (pela própria definição), o que introduz um certo grau de redundância em seus usos. O conjunto dos índices não forma, necessariamente, a base mais adequada para a definição deste espaço multidimensional. Por fim, o grande número de dimensões pode dificultar o entendimento da distribuição dos dados dentro este espaço multidimensional. Para lidar com isto, vamos utilizar Análise de Componentes Principais (PCA).

star	variable	chl2red	stdev	wstdev	mad	lqr	roms	nxs	peak	il	ratioeta	abbe	
27	skyM_00027.dat	0	2.87582604	0.394642827	0.441569788	0.28910	0.594500	1.2892807	1.216690e-04	3.206533e-02	2.398015332	5.116537e-01	9.772234e-01
28	skyM_00028.dat	0	1.94781940	0.358014417	0.459514388	0.33290	0.524800	0.9914408	-3.718777e-05	2.610074e-02	-0.151382650	7.543287e-01	6.628410e-01
29	skyM_00029.dat	0	0.71734809	0.274466451	0.290038756	0.03260	0.366400	0.6410176	-1.114785e-04	1.242714e-02	12.875253265	7.091110e-03	7.051082e+01
30	skyM_00030.dat	0	1.80034699	0.292608890	0.287887721	0.13800	0.387900	1.0763504	1.056300e-04	3.296764e-02	3.142644292	2.178210e-01	2.295462e+00
31	skyM_00031.dat	0	0.20305538	0.023231522	0.024023510	0.04410	0.036350	0.3807329	-7.325420e-06	-6.484300e-05	-0.120203806	7.969038e-01	6.274283e-01
32	skyM_00032.dat	0	1.36443171	0.098893694	0.091503441	0.08640	0.102400	0.8842131	5.436263e-06	9.138887e-03	3.270622146	1.939591e-01	2.577862e+00
33	skyM_00033.dat	0	4.66964412	0.604564904	0.725036379	0.12250	0.815825	1.4278251	4.623978e-04	7.650523e-02	-2.557207426	7.902811e-02	6.326862e+00
34	skyM_00034.dat	1	0.90068820	0.063007716	0.062784993	0.04510	0.091400	0.7963377	-1.359262e-06	5.504244e-03	0.885228252	7.591542e+01	6.586277e-03

Figura 2: Tabela com valores dos 11 índices para algumas estrelas.

RESULTADOS PRELIMINARES

Na análise exploratória dos resultados, observamos três padrões de simetria na distribuição dos índices (Fig.3). Verifica-se que alguns índices estão fortemente correlacionados enquanto outros não (Fig. 4).

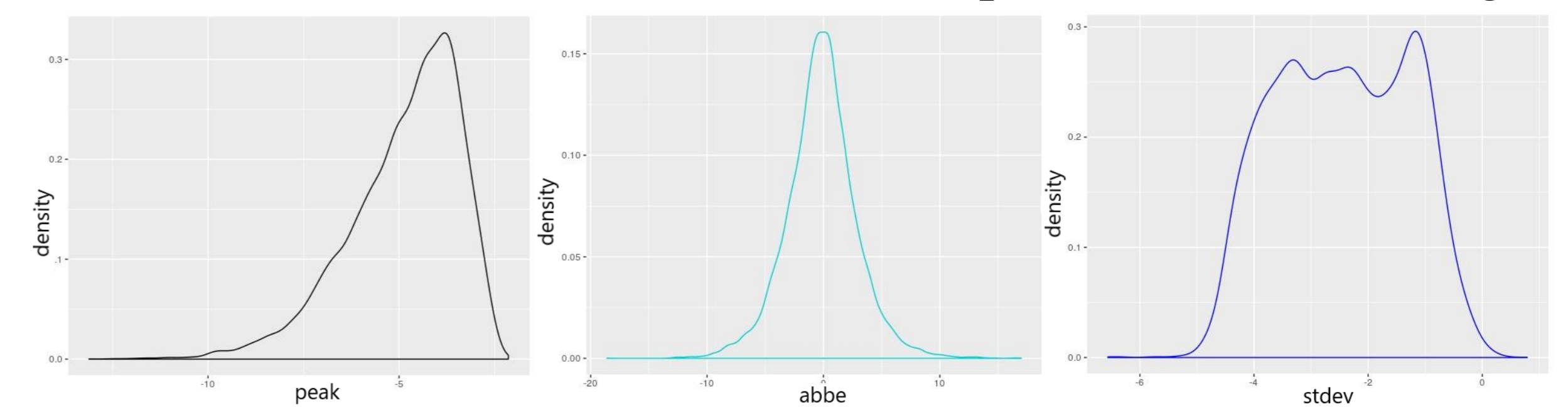


Figura 3: Distribuição de densidade dos índices variabilidade pico-a-pico (esquerda), excesso no valor Abbe (centro) e desvio padrão (direita).

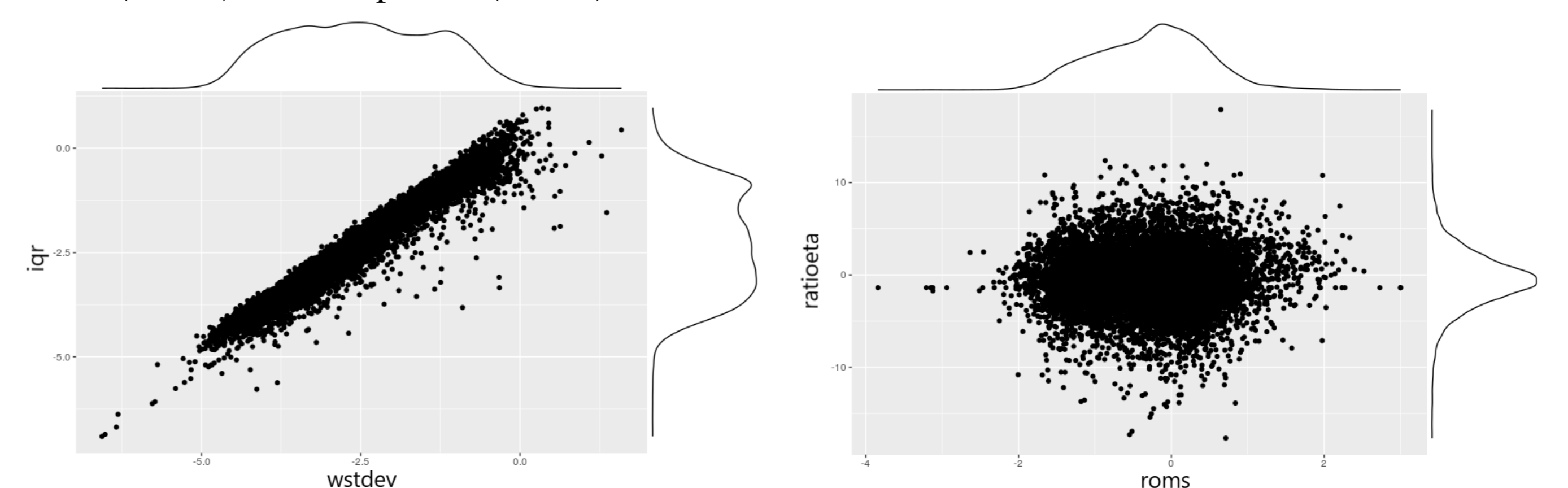


Figura 4: Os índices desvio padrão ponderado e IQR (esquerda) têm uma alta correlação, $R = 0.976994$, quando comparado com a correlação entre RoMS e razão de von Neumann, $R = 0.02612739$

O passo seguinte, será aplicar a Análise de Componentes Principais (PCA) com o objetivo de (1) definir uma base mais adequada para representação dos dados neste espaço, onde cada nova dimensão é uma combinação linear dos 11 índices e (2) identificar as componentes principais de modo a reduzir a dimensionalidade deste espaço. Em seguida, tentaremos identificar as regiões onde se concentram as estrelas variáveis.

REFERÊNCIAS

1. R Core Team, 2016, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
2. Martinazzi, E. et al. 2014, MNRAS, 442, 3105.
3. Martinazzi, E. et al. 2015, MNRAS, 447, 2235.
4. Martinazzi, E. 2016, Tese de Doutorado, IF-UFRGS.
5. K.V. Sokolowski, P. Gavras, A. Karamelas, 2016, MNRAS, 464, 274.