



**Universidade Federal do Rio de Janeiro**  
**Centro de Ciências Matemáticas e da Natureza**  
**Observatório do Valongo**  
**Departamento de Astronomia**

# **Métodos de Determinação de Períodos: Testes Comparativos para a Estrela AV Cir**

**Aluna: Alessandra Casati de Almeida**

**Orientadores:**

**José Adolfo S. de Campos**

**Encarnación A. M. Gonzalez**

*(Observatório do Valongo, UFRJ)*

**Projeto de Final de Curso para a Obtenção do  
Título de Astrônomo**

*À minha mãe, Marlene.*

## **Agradecimentos:**

- Ao meu avô, Gilberto Casati (in memoriam).
- A Albano, Marlene, Aline e Gilberto, minha família, que com amor sempre me ajudaram a superar as muitas dificuldades.
- Aos meus orientadores que tornaram possível este projeto.
- Aos professores Jorge de Albuquerque, Edgar Rangel e Ronaldo Perseke (in memoriam) que nos ensinaram que além das equações matemáticas existe uma Astronomia mais humana.
- A professora Encarnación, o anjo da guarda dos alunos do Valongo, pelo seu continuado apoio e orientação em todos os momentos da minha vida acadêmica.
- Ao professor Gustavo Porto de Melo por me incentivar e acreditar em mim.
- A Leandro Martelli, meu companheiro de todos os momentos, por toda a sua valiosa ajuda.
- Aos meus amigos, Alberto Silva Betzler, Annelisie Aiex, Carlos André de Moraes, Flávia Requeijo, Flávia Pedrosa, Jorge Albuquerque, Marco Aurélio, Luciana Rios, Verônica Huziwara e outros, que partilharam comigo bons e maus momentos na universidade.
- Aos astrônomos do passado que deram à luz as verdades que hoje trazem a evolução.

*“Emocionada por esse peso de infinito que lhe pesava sobre o coração, Estela abriu a janela que dava para os castanheiros de um grande parque. O ar estava fresco e perfumado; a noite, silenciosa nesse quarteirão deserto. A Lua, em quarto crescente, flutuava, qual pequenina barca luminosa, sobre os vapores do horizonte ocidental, vagamente iluminado pelas luzes de Paris; Vênus e Júpiter brilhavam na constelação dos Gêmeos, acima de Castor e Póllux, e as quatro estrelas de Leão pareciam no seu alinhamento mostrar ao longe, a leste, a aresta da Virgem, por cima da qual brilhavam Arctúrus, a constelação e as pequenas estrelas da Coroa Boreal. As estrelas mais resplandecentes cintilavam bonançosas e a atraíam o olhar e o pensamento.*

*Com os cotovelos apoiados à janela, contemplou-as, identificando-as pelos nomes, e sua imaginação voou até elas. A beleza da noite, a calma atmosférica, os lumes do céu, a imensidade do Espaço, transportaram seu pensamento às altas regiões que acabara de visitar na sua leitura. Paris que dormia, os edifícios dos quais se notava algumas cúpulas escuras, a torre quadrada do Convento dos Pássaros, as próprias igrejas, tudo lhe parecia coisas inferiores, terrestres e humanas. O mistério do céu estrelado arrebatou sua alma, qual um sonho divino. E pela primeira vez sentiu que a verdade pairava lá, em cima; que ninguém a encontrou cá em baixo; que as religiões são tentativas incompletas, e que se uma dentre elas pretendesse confiscar o Deus - das estrelas, seria vítima de infantil puerilidade.”*

**Camille Flammarion**

**Estela**

## *Resumo*

O objetivo deste projeto é definir o método mais eficiente na determinação de periodicidades em séries temporais desigualmente espaçadas, para faixas de períodos de um a seis dias. A escassez de métodos numéricos eficientes destinados ao tratamento destas séries, é devido em grande parte, à sua complexidade, que torna difícil o seu controle e previsão. Estas séries, ao contrário das igualmente espaçadas, não podem ser modeladas pela teoria estatística convencional, exigindo técnicas particulares para a obtenção de resultados significativos. Procuramos generalizar as características das técnicas mais populares e comparar seus desempenhos através da utilização de séries temporais artificiais com diversos níveis de relação sinal-ruído, que simulam características de observações astronômicas reais. Finalmente, discutimos a aplicação dos métodos em uma série real da estrela variável *AV Cir* e suas implicações.

## *Abstract*

The aim of this project is to define the most efficient method for establishing periodicities in unevenly time series. The limited number of effective numerical methods to treat these series, is due to its complexity, which makes its control and forecasting more difficult. This kind of series can't be modelled by the conventional statistical theory, contrary to the equally spaced in time, they require specific techniques to obtain significant results. We intend to spread the general characteristics of the most popular techniques and compare its performances by using synthetic time series, with distinct S/N ratio, that simulates typical astronomical observations. Finally, we discuss the application of the methods to a real serie of the star *AV Cir* and its applications.

# Conteúdo

Introdução	1
<b>1 A Busca por Periodicidades</b>	<b>2</b>
1.1 Introdução . . . . .	2
1.2 A Análise de Períodos . . . . .	5
1.3 Definições Importantes . . . . .	5
<b>2 As Séries Temporais Utilizadas</b>	<b>13</b>
2.1 Introdução . . . . .	13
2.2 Séries Artificiais para teste . . . . .	15
2.3 Estrela Variável AV CIR . . . . .	17
<b>3 Métodos de Determinação de Períodos</b>	<b>19</b>
3.1 Introdução . . . . .	19
3.2 Método String Length . . . . .	20
3.3 Método PDM . . . . .	23
3.4 Método Fourier . . . . .	27
3.5 Algoritmo Clean . . . . .	30
<b>4 Resultados e Conclusões</b>	<b>38</b>
4.1 Introdução: . . . . .	38
4.2 Séries Artificiais para teste . . . . .	40
4.3 Série da Estrela AV Cir . . . . .	46
4.4 Comparação dos Métodos: . . . . .	47

4.5 Conclusões: . . . . . 52



# *Introdução*

A observação dos fenômenos naturais, principalmente os astronômicos, é geralmente representada sob a forma de uma série de medidas no tempo, chamada série temporal. Existe uma grande variedade de processos que produzem séries temporais periódicas. As evidências de periodicidade são extremamente reveladoras das propriedades do fenômeno e ou da estrutura do objeto estudado. Desta forma, a procura por periodicidades nos dados, é uma conseqüência natural e a escolha de um método que identifique e determine estes períodos é de fundamental importância.

Porém, devido à natureza das observações astronômicas, os dados geralmente são desigualmente espaçados e contém grandes lacunas, ao contrário do que acontece em outras áreas. Vários métodos diferentes podem ser encontrados na literatura, mas as técnicas existentes que são poderosas ferramentas para as observações igualmente espaçadas, quando aplicadas as desigualmente espaçadas apresentam diversos problemas.

Neste projeto pretendemos analisar o desempenho de todos os métodos propostos para a aplicação em séries desigualmente espaçadas típicas de estrelas variáveis Cefeidas.

Visando uma melhor compreensão das propriedades gerais destes métodos, serão feitas simulações computacionais de séries com diversos níveis de relação Sinal-Ruído. Finalmente, os métodos são testados na série de observações da estrela variável, *AV Cir* e os resultados discutidos, a partir das informações da etapa anterior.



# Capítulo 1

## A Busca por Periodicidades

### 1.1 Introdução

A Ciência estuda a natureza através de observações do mundo físico procurando explicá-las e relacioná-las criando generalizações. Em Ciência, especialmente em Astronomia, a observação possui um papel decisivo no entendimento dos diversos fenômenos.

Um sistema físico representa uma fonte de informação que muda de estado ao longo do tempo. Essa mudança de estado provoca eventos e cadeias de eventos que podem ser detectados e, por vezes, medidos. A série de medidas ordenadas no tempo segundo sua seqüência natural é chamada Série Temporal. Os estados da fonte, ou seja, suas medidas, representam a fonte em algum contexto, revelando características importantes do sistema, e até mesmo a natureza do fenômeno. Portanto, a análise de séries temporais tem desempenhado um papel importante na explicação de diversos fenômenos em ciências físicas, biológicas, sociais e etc. e ocupa um lugar importante na Metodologia Estatística, bem como nas suas interfaces com diversas áreas do conhecimento. Há dois aspectos a considerar no estudo de séries temporais: a análise e a modelagem.

O objetivo da análise é sintetizar as propriedades de uma série e determinar suas principais características. Isto pode ser feito no domínio do

tempo ou da frequência. No domínio do tempo é destacada a relação entre observações em diferentes instantes do tempo, enquanto que no domínio da frequência o interesse está na frequência com que certos eventos ocorrem em determinado período de tempo. As duas formas de análise não são alternativas mas sim complementares; a mesma informação é processada de maneiras diferentes, desse modo dando diferentes interpretações sobre a natureza da série. Podemos dizer que a análise no domínio da frequência é conveniente na procura de características determinísticas da série temporal, enquanto a do domínio do tempo é usada na análise de processos não determinísticos. Na análise de séries temporais, tentamos identificar os aspectos periódicos contidos no conjunto de medidas. A presença de periodicidade é um indicador das condições e da estrutura do objeto ou do fenômeno estudado.

Na modelagem, o objetivo é encontrar um modelo apropriado para uma série temporal, o que não é fácil em Astronomia. Um aspecto importante é a construção de um modelo que auxilie no entendimento do comportamento do fenômeno que originou a série, na determinação de periodicidades relevantes e no controle e previsão do futuro. Predizer um evento futuro é uma forma poderosa de confirmar uma teoria científica, daí a importância dos vários métodos que se preocupam em extrair informações sobre periodicidades e suas amplitudes. Porém, essa se torna uma tarefa extremamente difícil, pois relacionar uma teoria ao mundo real sempre envolve a inclusão de alguma quantidade de erros nas medidas.

Podemos ressaltar algumas peculiaridades dos problemas em séries temporais astronômicas que dificultam esta modelagem. Primeiramente, os tempos de observação são geralmente descontínuos e irregulares. Em Astronomia, as observações são diretamente afetadas, por exemplo, pelas condições do tempo, e/ou pelo tempo disponível no telescópio. Podemos apenas coletar os dados durante a noite e na época em que a estrela ou o objeto a ser estudado, está acessível à observação. Além disto, a maior parte dos objetos astronômicos varia em um período de tempo maior do que uma noite. Portanto, a nossa rotina de coleta de dados nos força a utilizar séries com

lacunas para descrever totalmente o fenômeno.

Em segundo lugar, os dados da série são freqüentemente acompanhados de ruído devido a erros de medida e flutuações atmosféricas. Neste caso, a dificuldade principal na busca de periodicidades surge do fato de que além de componentes periódicas, o sinal, os dados da série também contém componentes aleatórias, o ruído. Qualquer tentativa de se detectar periodicidades nos dados deve levar em conta a presença desta componente aleatória, e isto representa um obstáculo para as análises estatísticas.

Outra característica que podemos identificar, é que o objetivo do astrônomo não se resume apenas na previsão de valores futuros da série. Além disto, ele procura principalmente caracterizar a variação, baseado em um modelo físico. A série representa a chave para o entendimento das causas físicas da variabilidade.

Como conseqüência, não podemos aplicar os mesmos métodos usados para séries de dados igualmente espaçadas à grande maioria das séries astronômicas, daí a necessidade de estudarmos técnicas que se apliquem a séries desigualmente espaçadas.

Os métodos desenvolvidos para a aplicação em séries desigualmente espaçadas apresentam problemas na previsão do comportamento da série. Dentre as várias técnicas propostas com este objetivo, não existe uma que trate de forma totalmente eficaz com todos os tipos de séries, ou seja, não podemos definir um método ideal para o tratamento das mesmas. Diferentes métodos podem ser considerados adequados à análise de seqüências distintas, sendo que a escolha entre um deles dependerá das características da série.

A utilização de séries artificiais para o estudo dos métodos mais adequados é muito útil neste caso. Essas séries são geradas, de forma a simular uma série astronômica, e seus parâmetros são escolhidos convenientemente. Com o uso das séries artificiais, conseguimos investigar as principais qualidades e defeitos do método que estamos usando e descobrir se ele se aplica aos nossos dados.



## 1.2 A Análise de Períodos

Existem dois objetivos distintos associados com a análise de períodos. Inicialmente, queremos descobrir se existe algum sinal periódico significativo no conjunto de dados. Procuramos períodos e testamos a significância de cada possibilidade encontrada. A este procedimento chamamos de *Detecção*. O segundo objetivo da análise é, de posse de um sinal significativo, determinar os parâmetros associados a ele e as respectivas incertezas. Este segundo passo é chamado de *Estimativa*. A determinação da melhor estimativa dos parâmetros da série e de suas incertezas irá envolver propriedades estatísticas do método de determinação de períodos empregado. Esta incerteza está diretamente relacionada à *Relação Sinal-Ruído (S/R)*. A S/R é uma medida do nível de ruído comparado com um certo valor do sinal, indicando o quanto de ruído está presente nas medidas. Com frequência utilizamos a S/R como uma especificação da qualidade das observações.

Entretanto, quando estamos tratando de um objeto de propriedades já conhecidas, não é necessário comprovar a existência de periodicidade, podemos partir diretamente para o estabelecimento dos parâmetros do sinal, como período, amplitude e fase e suas incertezas associadas.

## 1.3 Definições Importantes

Uma série temporal é um conjunto de observações de uma variável,  $m$ , observada em intervalos de tempo igualmente espaçados ou não. Uma série de  $n$  observações será denotada por  $m_1, \dots, m_n$ . As observações vizinhas são dependentes e o estudo da série temporal (no domínio do tempo) consiste em analisar essa dependência. A maior parte dos procedimentos estatísticos foi desenvolvida para analisar observações independentes e, logo, a maioria dos resultados não é relevante para o estudo de séries temporais. Portanto, novas técnicas e resultados tem-se desenvolvido para aplicação específica em séries temporais, cujo estudo se constitui numa importante área da estatística. A série temporal é representada por um conjunto de observações

$$\{m_t, t \in T\} \quad (1.1)$$

de uma variável  $m$ , onde  $T$  é um conjunto de índices (tempo, espaço, etc...).

A variável observada  $m$  pode ser discreta ou contínua, univariada ou multivariada e o índice “tempo” pode ser discreto ou contínuo, unidimensional ou multidimensional.

Para cada  $t$ ,  $m_t$  é uma variável aleatória. Assim,  $f(m_t)$  será a densidade de probabilidade de  $m_t$  para cada  $t$ . As funções densidade poderão ser diferentes para diferentes valores de  $t$ .

A seqüência de variáveis aleatórias,

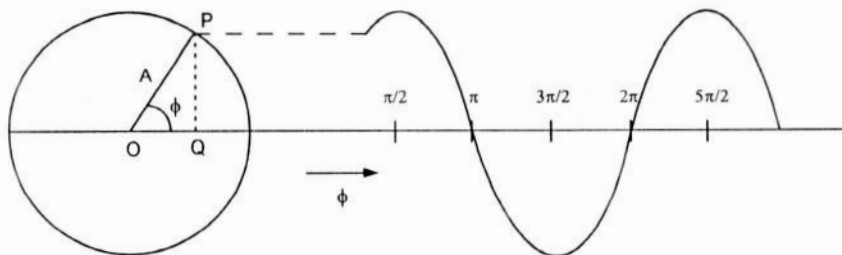
$$\{m_1, m_2, \dots\} \quad (1.2)$$

é chamada de *processo estocástico*. Cada observação num processo estocástico é uma variável aleatória e as observações evoluem no tempo de acordo com certas leis probabilísticas. Assim, um processo estocástico pode ser definido como uma coleção de variáveis aleatórias ordenadas no tempo. Em Astronomia, a variável  $m$  pode ser, por exemplo, a magnitude ou a velocidade radial de uma estrela, e como já foi dito, estas observações estão contaminadas por algum tipo de ruído aleatório.

Algumas definições são muito usadas quando estão sendo analisados períodos de estrelas variáveis, como:

- *Fase, Frequência, Período*

Observemos a figura I e consideremos que o ponto  $P$  tem um movimento circular uniforme.



**Figura I** -Parâmetros da onda senoidal.

Uma curva mostrando esta variação pode ser construída através da determinação de posições igualmente espaçadas ao longo da circunferência e, então marcando a altura destes pontos em intervalos uniformes ao longo de uma linha reta. A altura  $PQ$  pode ser expressa na fórmula:

$$PQ = A \text{ sen } \phi \quad (1.3)$$

Se o ângulo  $\phi$  é igual à  $\alpha$  em  $t = 0$  e se o ponto  $P$  leva um tempo  $T$  para fazer uma revolução completa ( $2\pi$  radianos), o ângulo em um tempo  $t$  é  $\frac{2\pi t}{T} + \alpha$  e a expressão para  $PQ$  é:

$$PQ = A \text{ sen } \left( \frac{2\pi t}{T} + \alpha \right) \quad (1.4)$$

Podemos dizer que o ponto faz  $f$  revoluções por unidade de tempo:

$$f = \frac{1}{T} \quad (1.5)$$

e a fórmula pode ser reescrita:

$$PQ = A \text{ sen } (2\pi ft + \alpha) \quad (1.6)$$

Como utilizamos nesta fórmula a função seno, a curva na figura é chamada de curva senoidal. Apenas 3 coeficientes são suficientes para descrevê-la; a amplitude  $A$ , a constante de fase  $\alpha$ , também chamada de fase inicial, e a frequência  $f$  (ou o período  $T$ , que neste trabalho será identificado por  $P$ ).

Considerando um intervalo de tempo  $\Delta t = t_i - t_{i-1}$ , podemos escrever:

$$\Delta t = \frac{\phi}{\lambda} \quad (1.7)$$



onde  $\lambda$  é o comprimento de onda.

Como  $\lambda = \frac{2\pi}{P}$ , a fase pode ser definida da seguinte forma:

$$\phi = \frac{2\pi \Delta t}{P} \quad (1.8)$$

- *Diagrama de Fase*

O diagrama de fase é uma curva que exibe o fenômeno observado em função da fase para um dado período. Dependendo do período escolhido, este gráfico pode fornecer uma representação das observações, que pode ser adequada ou não.

- *Periodograma*

Periodograma é um gráfico utilizado na obtenção de períodos onde a abcissa representa o período e a ordenada, uma função que depende do método que está sendo utilizado. Os valores da abcissa para os quais a ordenada atinge o máximo ou mínimo, dependendo do método, indicam os prováveis períodos da série em questão.

Existem algumas considerações a serem feitas a respeito da informação apresentada pela frequência(ou período) em alguma série particular. Os limites superior e inferior devem ser definidos em função das características da série. O período mais longo (frequência mínima) capaz de ser detectado no conjunto de dados analisado é dado por:

$$P_{máx} = T \quad (1.9)$$

$$f_{min} = T^{-1} \quad (1.10)$$

onde  $T$  é o intervalo de tempo entre a primeira e a última observação da série.

$$T = t_n - t_1 \quad (1.11)$$

Fica claro por 1.10 que quanto maior a duração da série, menor será a frequência capaz de ser detectada nos dados.

A frequência mais alta capaz de ser detectada é determinada pelos intervalos entre medidas sucessivas,  $\Delta t$ . Para dados igualmente espaçados, 3 pontos são o suficiente para definir um sinal senoidal, desta forma o período mínimo capaz de ser detectado é  $2\Delta t$ . A frequência máxima correspondente é conhecida como *Frequência de Nyquist*:

$$\nu_{Ny} = f_{Max} = (2\Delta t)^{-1} \quad (1.12)$$

Porém, a frequência de Nyquist não é bem definida para dados desigualmente espaçados, pois os valores de  $\Delta t$  não são constantes ao longo da série. Uma solução prática para isto é definir a *Pseudo Frequência de Nyquist* que será definida através da substituição de  $\Delta t$ , constante para dados igualmente espaçados, por um intervalo médio entre as observações,  $\langle \Delta t \rangle$ . A *Pseudo Frequência de Nyquist* é definida por:

$$\nu_{Ny} = 2 \langle \Delta t \rangle^{-1} \quad (1.13)$$

Desta forma para séries de dados desigualmente espaçadas, o menor período será dado por:

$$P_{min} = 2 \langle \Delta t \rangle \quad (1.14)$$

$$f_{máx} = (2 \langle \Delta t \rangle)^{-1} \quad (1.15)$$

Neste caso o intervalo médio poderá ser obtido através da expressão:

$$\langle \Delta t \rangle = \frac{\sum_{i=1}^{n-1} \langle \Delta t_i \rangle}{n-1} \quad (1.16)$$

ou

$$\langle \Delta t \rangle = \frac{n-1}{\sum_{i=1}^{n-1} \frac{1}{\langle \Delta t_i \rangle}} \quad (1.17)$$

onde  $\langle \Delta t_i \rangle$  é o espaçamento entre os instantes de tempo sucessivos.

$$\Delta t_i = t_{i+1} - t_i; \quad i = 1, \dots, (n-1) \quad (1.18)$$

A expressão 1.17 é uma média harmônica que exclui grandes interrupções nos dados, enquanto a expressão 1.16 leva em conta todos os intervalos. Desta forma, os períodos teste que utilizarmos devem estar contidos no intervalo, entre os limites que foram estabelecidos:

$$P_{\min} < P_T < P_{\max} \quad (1.19)$$

onde  $P_T$  é o período teste. Dentro deste intervalo, a função teste será calculada para  $np$  pontos, onde  $p$  é um incremento, separados por  $dP$ , formando uma rede de pontos. A determinação das características desta rede depende do objetivo da análise. Se estamos interessados em detectar frequências desconhecidas que podem estar presentes ou não, então é importante que os pontos da rede sejam independentes ou quase independentes uns dos outros. Esta propriedade assegura que o teste estatístico utilizado para avaliar a significância do extremo, seja válido. Porém, quando estamos apenas estimando as propriedades de um sinal conhecido, os pontos da rede não precisam ser necessariamente independentes. Desta forma, uma rede de pontos arbitrária é permitida para a estimativa mas não é para a detecção. Os períodos desta rede são dados por:

$$P_i = P_{i-1} + dP_i, \quad i = 1, \dots, np \quad (1.20)$$

ou de forma equivalente:

$$f_i = f_{i-1} + df_i, \quad i = 1, \dots, np \quad (1.21)$$

onde  $dP_i$  é o intervalo entre períodos sucessivos e  $df_i$  é o intervalo entre frequências sucessivas.

Caminhando pela grade de períodos, ao passarmos de um período teste  $P$  para o próximo,  $P + dP$ , verificamos que a fase vai se alterando ao longo da grade. A fase para um dado período  $P_i$  da grade é dada por:

$$\phi_i = FRAC \left[ \frac{(t_i - t_0)}{P_i} \right] = FRAC \left[ (t_i - t_0) f_i \right]; \quad i = 1, \dots, n. \quad (1.22)$$

onde  $t_o$  é o instante médio da série de observações determinado por  $t_0 = \frac{t_1 + t_n}{2}$ , e FRAC significa que o resultado é a parte fracionária da expressão.

Fazendo a diferenciação da expressão 1.18 e considerando que  $|t_i - t_0| \leq \frac{T}{2}$ , obtemos a diferença de fase em função do período:

$$d\phi = \frac{T}{2P^2} dP \quad (1.23)$$

Da mesma forma obtemos para a frequência:

$$d\phi = \frac{T}{2} df \quad (1.24)$$

Para que a resolução no espaço de fases seja a mesma ao longo de toda a grade de períodos, o valor de  $d\phi$  deverá ser *constante*. Portanto, ao examinarmos as expressões 1.21 e 1.22, concluímos que para que  $d\phi$  seja constante, as resoluções em frequência e período também devem ser. Porém, vemos na

expressão 1.21 que um incremento constante em período não garante que  $d\phi$  também seja , pois além de depender de  $dP$  ele também depende de  $P^2$  que é uma quantidade variável. Por outro lado, um incremento em frequência constante já representa uma condição suficiente. Isto possibilita a definição do *Fator de Resolução da Série*,  $R$ , que será igual à  $df$ :

$$R = df = \frac{2d\phi}{T} \quad (1.25)$$

A *Resolução da Série* pode ser entendida como a capacidade de distinguir na grade de frequências, pequenas estruturas de informação.

Sendo o *Periodograma*, um gráfico da *função de amostragem* versus frequências teste, então a *Resolução do Periodograma*, ou seja, a capacidade de resolver estruturas, como picos muito estreitos irá depender do valor da *Resolução da Série*. A escolha do valor de  $R$  é fundamental para uma correta estimativa das periodicidades, pois uma escolha inadequada pode resultar na ocultação do verdadeiro período da série.

A solução para este problema surge através da introdução do parâmetro  $G$ , chamado de *fator de oversampling*, que varre o espaço de frequências. Supondo que o incremento em frequência seja função da frequência mínima, definimos:

$$df = Gf_{min} \quad (1.26)$$

Uma rede de frequências com espaçamento muito pequeno é permitida durante a fase de estimativa, não sendo adequada à detecção. Com o propósito de detecção, podemos fazer  $G = 1$ . Adotando valores de  $G < 1$ , estamos fazendo uma interpolação no espaço de frequências. A interpolação é utilizada principalmente quando conhecemos somente alguns valores numéricos de uma função e é necessário calcular valores desconhecidos.



# Capítulo 2

## As Séries Temporais Utilizadas

### 2.1 Introdução

De uma forma geral uma estrela variável é aquela cujo brilho varia ao longo do tempo. Porém, neste caso uma definição mais precisa é necessária, caso contrário todas as estrelas poderiam ser consideradas como variáveis, pois todas exibem variações de brilho, conectadas com mudanças evolutivas em algum estágio do seu desenvolvimento, além disto, também são observadas estrelas com variações em uma pequena escala de magnitude e outras que são consideradas variáveis quando observadas em comprimentos de onda diferentes.

As estrelas da Sequência Principal, que são em sua maioria constantes, estão constantemente vibrando com uma amplitude muito pequena. Estas oscilações foram detectadas no Sol, e estão sendo procuradas em outras estrelas desta região, o que não é uma tarefa simples, exigindo medidas extremamente acuradas para que sejam detectadas.

Contudo, existe uma região do *Diagrama Hertzsprung-Russel* chamada de *Faixa de Instabilidade* onde estas oscilações são aumentadas. Estas estrelas possuem uma variedade de nomes; encontramos as *Variáveis Cefeidas e RR Lyrae* entre as gigantes vermelhas, as *Estrelas  $\delta$  Scuti* na sequência principal, e as *Estrelas ZZ Ceti e PG1159* entre as anãs brancas. Além disto podemos



citar as *Estrelas Variáveis Mira*, as *Estrelas Carbonadas* e as *Estrelas R Cor Bor* que apresentam instabilidades menos regulares em suas atmosferas, onde a produção variável de grãos de poeira modula a emissão da luz.

Fisicamente, a variabilidade do brilho das estrelas pode estar associada à processos explosivos ou eruptivos, à atividade magnética, à rotação de uma estrela cuja superfície seja não-uniforme ou com um formato elíptico, à órbita da estrela, à geometria de um sistema binário, ou a acreção entre suas componentes, e finalmente à variação real do tamanho da estrela através de contrações e expansões alternadas, que são as comumente chamadas de estrelas pulsantes.

Estas pulsações são consideradas periódicas quando os máximos e mínimos de brilho ocorrem em intervalos de tempo iguais. Neste caso, uma curva de luz da estrela pode ser obtida quando construímos um gráfico da magnitude versus tempo, e assim ajustamos uma curva aos pontos. A curva de luz mostra a forma de variação do brilho da estrela e fornece informações importantes a respeito de seu período e amplitude.

Periodicidades nas propriedades observadas das estrelas podem ser usadas para estabelecer a sua natureza física. Um grande número de estrelas apresenta propriedades observáveis que variam no tempo. Esta variabilidade geralmente nos permite classificá-las em classes, de acordo com suas características físicas principais e isto é de grande importância, pois desta forma alcançamos um entendimento maior da estrutura e evolução estelar e da galáxia.

Estrelas variáveis  $\delta$  *Cephei* e *RR Lyrae* são duas classes de estrelas gigantes vermelhas que, apresentam pulsações em escalas de tempo que vão de horas à dias. Elas são extremamente importantes para a Astronomia pois o período dos seus pulsos estão fortemente conectados com as suas luminosidades intrínsecas. Isto significa que medidas dos seus períodos de pulsação e brilho médio combinadas, resultam em uma *relação período- luminosidade*, que possibilita a determinação das suas distâncias e as transforma em indicadores de distância no universo.

As observações de estrelas variáveis dão origem à séries com um comportamento particular no tempo. Elas praticamente nunca são igualmente espaçadas e isto se deve à vários fatores: várias estrelas apenas podem ser visíveis em épocas específicas do ano, e a disponibilidade do telescópio é concedida por um tempo limitado ou durante alguma fase particular da Lua. Além disto, as condições meteorológicas são extremamente importantes, podendo causar uma quantidade extra de interrupções na aquisição de dados quando as condições não estão favoráveis. Todos estes fatores podem dar origem a este tipo de séries. A importância da detecção de períodos de estrelas têm motivado astrônomos a desenvolver e aplicar um grande número de algoritmos com este objetivo.

## 2.2 Séries Artificiais para teste

As propriedades dos métodos são muito difíceis de serem estabelecidas teóricamente, desta forma partimos para um estudo comparativo da atuação de cada algoritmo através de simulações numéricas.

Produzimos seis séries, onde três delas são as que simulam observações astronômicas típicas, representando os dados que coletamos das estrelas. Os seus parâmetros são escolhidos adequadamente dependendo da característica do objeto astronômico que desejamos analisar. Sendo assim, geramos seis séries com S/R distintas, três com S/R alta e outras três com S/R baixa, que apresentam curvas de luz simétrica, assimétrica e de duplo período. Com isso, podemos estudar os possíveis efeitos que os métodos exercem sobre os dados, e quais deles são os mais eficientes dependendo da série pesquisada e da amostra obtida.

Após a análise das propriedades gerais dos métodos, passamos à aplicação dos mesmos na série da estrela variável *AV Cir* e em outra série gerada artificialmente, que possui características semelhantes à da estrela para que sejam comparados os resultados. Esta série tem a forma simétrica porém é mais longa que a gerada anteriormente, e o seu tamanho é escolhido de forma

a se aproximar do tamanho da série AV Cir.

Nas tabelas abaixo podemos ver as características principais das sete séries artificiais:

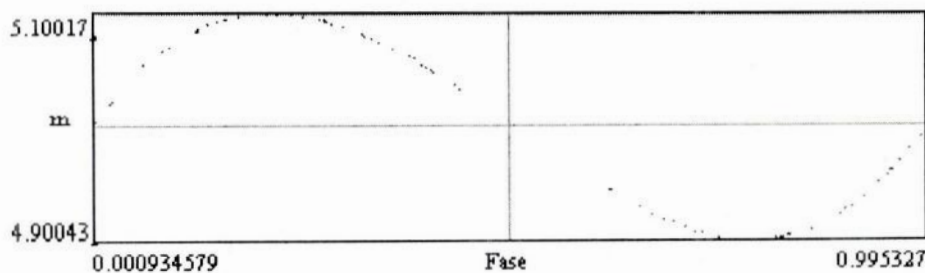
Relação S/N	Alto	Baixo
Frequência da Série	0.648088	0.648088
Número total de Pontos(N)	40	40
Intervalo de tempo total(dias)	161.1402	196.9507
Frequência Máxima(Ciclos/dia)	8.28953	9.96801
Frequência Mínima(Ciclos/dia)	0.00621	0.00508

**Tabela I** - Características das séries artificiais usadas nos testes.

Relação S/R	Alto
Frequência da Série	0.648088
Número total de Pontos(N)	200
Intervalo de tempo total(dias)	199.0911
Frequência Máxima(Ciclos/dia)	25.58862
Frequência Mínima(Ciclos/dia)	0.00502

**Tabela II** - Características da série artificial simétrica longa.

A forma da curva é um parâmetro importante na análise dos métodos, sendo também escolhida para representar dados astronômicos reais. No caso deste projeto a curva de luz pertence à uma estrela variável, desta forma evitamos curvas de formas improváveis. Vamos nos concentrar em três casos mais importantes: curva simétrica, assimétrica e com duplo período.



**Figura II** - Curva de luz simétrica.

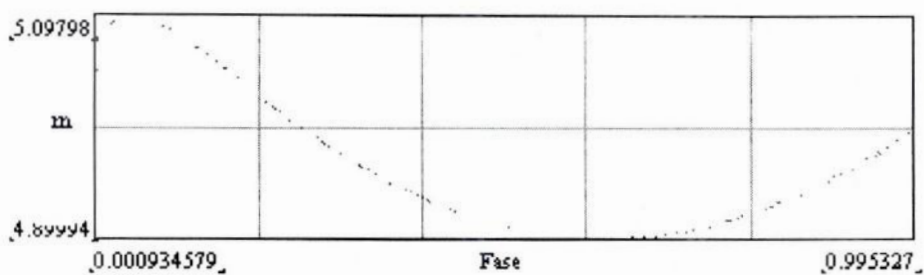


Figura III - Curva de luz assimétrica.

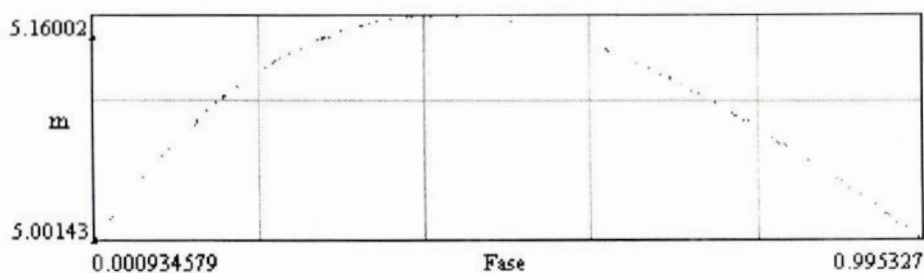


Figura IV - Curva de luz de duplo período.

## 2.3 Estrela Variável AV CIR

A série temporal que estamos analisando neste projeto pertence à estrela variável *AV Cir*, que é uma Cefeida. A série de observações fotométricas foi feita pelo satélite *Hipparcos*.

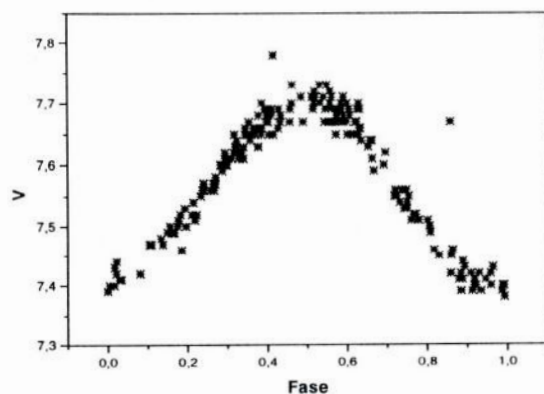


Figura V-Curva de Luz da estrela *AV Cir* para um período de 3.06524 dias.



A série da estrela *AV Cir* possui as seguintes características:

Ascensão Reta(J2000)	14:50:30.30
Declinação(J2000)	-67:29:51.41
Tipo Espectral	F7II
Período	3.0652400 dias
Época de referência	8501.8788(JD-2440000)
Período(da Literatura)	3.06510 dias
Magnitude Máxima(da Literatura)	8.00
Magnitude Mínima(da Literatura)	8.60
Época(da Literatura)	2438206.05
Número total de Pontos(N)	197
Intervalo de tempo total(dias)	1170.1999
Erro padrão associado às magnitudes	0.01
Frequência máxima(Ciclos/dia)	0.00085
Frequência mínima(Ciclos/dia)	16.35642

**Tabela III** - Características da estrela variável *AV Cir*

## Capítulo 3

# Métodos de Determinação de Períodos

### 3.1 Introdução

Como já definimos anteriormente, uma série temporal é um registro do valor de alguma quantidade variável ao longo do tempo. Assim, se desejamos explicar um comportamento variável particular é necessário construir uma descrição matemática dos dados, ou seja, um modelo que explique as características desta série. A construção destes modelos é um dos objetivos principais da modelagem de séries temporais e, se podemos obter um modelo adequado às nossas séries, isto deve fornecer uma visão ampla do mecanismo físico que gera os dados, o que pode ser usado para prever valores futuros da série.

Há uma grande necessidade de determinação de períodos com precisão, em Astronomia, e desta forma vários métodos têm sido criados com este propósito. A grande maioria dos métodos desenvolvidos é direcionada ao tratamento de dados igualmente espaçados, apresentando grandes dificuldades quando aplicados a dados obtidos em intervalos irregulares.

A construção de instrumentos astronômicos mais sensíveis e a aplicação de técnicas de observação mais refinadas têm feito com que periodicidades



sejam detectadas em um número maior de objetos, que antes eram considerados como estáveis. Além disto, o desenvolvimento de computadores modernos têm incentivado a criação de métodos mais eficientes e facilitado a sua aplicação computacional. Porém, quando estamos tratando de séries astronômicas temos que nos lembrar de que a forma clássica de aquisição de dados em Astronomia resulta, na maior parte da vezes, em séries desigualmente espaçadas e com grandes lacunas.

Os tipos de dados em que estamos interessados aqui são os desigualmente espaçados. Os problemas encontrados na análise destes dados são diferentes, e precisam de técnicas distintas. As técnicas de determinação de períodos aplicadas à estas séries estão em constante atualização, apresentando um grande número de novas propostas e de variações de uma mesma técnica. Esses métodos podem ser separados em três grandes classes, de acordo com os desenvolvimentos matemáticos usados na modelagem do problema:

- Métodos *String Length*
- Métodos *Phase Binning*
- Método de *Fourier* e *CLEAN*

Estes métodos buscam periodicidades e estimam suas propriedades. Alguns são mais eficientes do que outros dependendo das características da série que se está pesquisando. A seguir, está uma descrição do princípio de funcionamento dos principais métodos que representam cada uma destas classes.

## 3.2 Método String Length

O método String Length é uma das várias técnicas utilizadas na determinação de períodos de séries astronômicas. Este método tem como base, a minimização da soma dos comprimentos de segmentos de reta (chamados de strings), que ligam observações sucessivas no diagrama de fase, construído a

partir de um dado período teste. O período correto da série será aquele que fornecer o menor valor nessa soma.

Como na maioria dos métodos, o passo essencial é a construção de uma curva exibindo o fenômeno observado em função da fase para um período teste escolhido. A análise dos resultados fornecerá uma representação adequada das observações.

Inicialmente, calculamos as fases das medidas para um período teste,  $P$ , utilizando a definição de fase:

$$\phi = \text{FRAC} \left[ \frac{(t_i - t_0)}{P} \right], 0 \leq \phi \leq 1$$

Então, as observações são colocadas em ordem crescente de fase e assim a soma dos comprimentos dos segmentos de reta é calculada.

A estimativa do comprimento total dos segmentos de reta que unem os pontos no espaço de fase é objeto de discussão entre vários autores. *Lafler e Kinman (1965)* determinaram uma expressão que leva em consideração apenas a separação vertical entre os pontos sucessivos, *Renson (1978)* também utilizou apenas a componente vertical, mas acrescentou uma forma de “peso” às medidas, relacionado com as diferenças de fase e com os erros das observações. *Burke, Rolland e Boy (1970)* introduziram uma expressão que considerava os dois eixos, calculando assim a distância real entre as observações no diagrama de fase. Porém *Renson (1978)* alertou para o fato de, nesta expressão, as magnitudes da ordenada possuem unidades diferentes das fases da abscissa. Foi *Dworetzky (1983)* que introduzindo uma função normalizadora, a equação 3.2, obteve a expressão consistente:

$$l(P) = \sum_{i=1}^{n-1} [(m'_i - m'_{i-1})^2 + (\phi_i - \phi_{i-1})^2]^{\frac{1}{2}} + [(m'_i - m'_n)^2 + (\phi_1 - \phi_n + 1)^2]^{\frac{1}{2}} \quad (3.1)$$

sendo que:

$$m'_i = \frac{(m_i - m_{\min})}{2(m_{\max} - m_{\min})} - 0.25 \quad (3.2)$$

onde  $m$  representa a grandeza medida nas observações e  $n$  é o número de observações

Observando atentamente, podemos notar que a expressão acima nada mais é do que a aplicação repetida do teorema de Pitágoras no diagrama de fase. Sendo assim, finalmente calculamos  $l(P)$  para um certo conjunto de períodos teste, e a análise gráfica de  $l(P)$  versus os períodos de teste, determina o período verdadeiro da série, que será o correspondente ao menor valor de  $l(P)$ .

Uma qualidade importante do método é a presença de um teste estatístico que considera o efeito de erros aleatórios nas observações, que poderiam dar origem à períodos falsos. Este teste rejeita as periodicidades que foram introduzidas nas medidas, através de erros.

Assim, o procedimento computacional do método se inicia com a normalização da escala dos dados através da expressão 3.2. Então, calculamos  $l(P)$  para vários valores de período, utilizando a expressão 3.1, sendo que antes de cada procedimento, os dados devem ser ordenados em fase para o período que está sendo avaliado. Construimos um gráfico de  $l(P)$  versus os períodos teste e o valor mínimo de  $l(P)$  é tomado como sendo o mais provável.

Para nos certificarmos de que este período é significativo, Dworetsky(1983), sugere que analisemos se o valor mínimo:

- É menor do que  $1.6 + 1.2\delta l$ , onde:

$$\delta l = 0.34\left(\epsilon - \frac{\epsilon^2}{2}\right)\left(N_{obs} - \sqrt{\frac{10}{\epsilon}}\right) \quad (3.3)$$

e  $\epsilon$  é dado por:

$$\epsilon = \frac{\varepsilon}{2(m_{max} - m_{min})} \quad (3.4)$$

onde  $\varepsilon$  é o erro interno das observações.

- ou é menor do que  $0.1054n$ , sendo  $n$  o número de observações, desde que este limite seja maior que o limite estabelecido para  $L$ , que varia

entre  $1.4 < L < 1.8$ , dependendo da forma da oscilação.

No caso de não encontrarmos nenhum pico significativo, a procura deve ser estendida à outros intervalos de frequência. Já, se encontramos mais de um pico significativo, isso indica que mais observações serão necessárias.

A grande vantagem desse método é a simplicidade dos conceitos envolvidos. Além disso, ele é intuitivo e particularmente eficiente na determinação de periodicidades em séries com uma pequena quantidade de dados  $N_{obs} \gtrsim 20$ , sendo o seu uso fortemente recomendado neste caso. Os resultados não são afetados pela forma da curva de luz do objeto, o que amplia a sua atuação. Porém, a necessidade dos dados serem ordenados em fase, para cada período teste, resulta em um longo tempo de execução do programa.

### 3.3 Método PDM

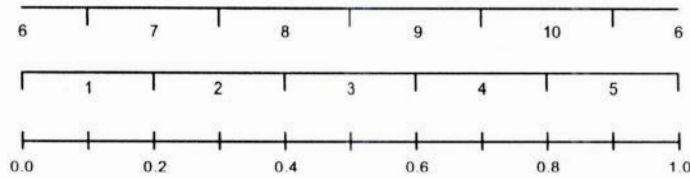
O PDM é uma técnica bastante popular, que procura reduzir ao mínimo a dispersão dos dados em fases constantes, vindo daí o nome *Minimização da dispersão de fase*. Ele pertence à classe de métodos *Phase Binning*, cuja característica básica é a separação dos dados observacionais em “caixas” de acordo com as suas fases, calculadas a partir de um certo período teste. No método PDM, uma estrutura padrão de “caixas” é definida e nelas agrupamos os dados da série de acordo com suas fases, que são calculadas através da expressão já conhecida para a fase:

$$\phi = FRAC\left[\frac{(t_i - t_0)}{P}\right], 0 \leq \phi \leq 1 \quad (3.5)$$

O intervalo unitário de fase é dividido em  $N_b$  caixas, todas de tamanho igual a  $\frac{1}{N_b}$ . Cada caixa está relacionada a um intervalo de fase. Geralmente, também se considera a superposição de um certo número  $N_c$  de “coberturas” defasadas em fase da cobertura anterior. Assim obtemos  $M = N_b N_c$  caixas cujos pontos médios são uniformemente espaçados ao longo do intervalo unitário.



Uma estrutura típica é caracterizada por 5 caixas e 2 coberturas,  $(N_b, N_c) = (5, 2)$ , porém esses valores podem ser modificados de acordo com o nosso interesse.



**Figura VI**-Esquema da estrutura de caixa (5, 2) do método PDM. O espaço de fase é dividido em 5 caixas iguais, cada uma contendo um intervalo de 0,2 em fase. Existem dois conjuntos de coberturas, cada uma deslocada de 0,1 em fase da outra. As 10 caixas resultantes estão numeradas. Podemos notar que cada ponto cai em duas caixas.

Considerando a série temporal com  $N$  pontos, podemos calcular a variância desses dados através da expressão:

$$\sigma^2 = \frac{\sum (m_i - \bar{m})^2}{N - 1} \quad (3.6)$$

onde  $\bar{m} = \sum \frac{m_i}{N}$  é a média das observações e  $m_i$  é a observação obtida no tempo  $t_i$ . Para um período teste escolhido, as variâncias das amostras de cada uma das caixas  $S_j^2$  são calculadas e somadas, fornecendo uma medida do espalhamento em torno da curva de luz média, definida em termos da média de  $m_i$  em cada amostra, considerada como função de  $\phi$ . A variância total para todas as amostras é dada por:

$$S^2 = \frac{\sum (n_j - 1) S_j^2}{\sum n_j - M} \quad (3.7)$$

A soma das variâncias das caixas é comparada com a variância global dos

dados de entrada através da definição da estatística:

$$\theta = \frac{S^2}{\sigma^2} \quad (3.8)$$

Se  $P$  não é o período verdadeiro da série, então os dados estarão aleatoriamente distribuídos no espaço de fase, o valor de  $S^2$  será aproximadamente igual ao de  $\sigma^2$  e a dispersão será alta, resultando em um  $\theta \approx 1$ . Porém, no caso de um período teste que corresponda ao verdadeiro período da série, a dispersão das amostras das caixas sofrerá uma redução drástica ( $S^2 \rightarrow 0$ ), indicando uma dispersão mínima entre os pontos e a curva média. Neste caso,  $\theta$  também se aproxima de zero, alcançando um valor mínimo, se comparado com os períodos vizinhos.

Analisando graficamente, podemos construir um periodograma, ou seja, um gráfico de  $\theta$  versus período, onde a periodicidade verdadeira será indicada por um mínimo profundo. A profundidade é uma indicação da significância do ajuste e revela o quanto o espalhamento em torno da curva de luz média foi reduzido.

Porém, quando estamos procurando periodicidades em fenômenos naturais, tal como a variação de brilho nas estrelas, devemos nos certificar de que o comportamento periódico que encontramos é real e não causado pela ação conjunta do fenômeno e do método de obtenção dos dados, gerando ruídos. A determinação do período correto da série não se resume a encontrar um valor extremo no periodograma, temos que testar a significância estatística deste extremo encontrado, para sabermos o quanto podemos confiar nesse resultado, especialmente no caso de dados desigualmente espaçados.

Observemos que para determinar  $\theta$ , tomamos a razão das variâncias nos dois subconjuntos de  $m$ , ou seja, a soma das variâncias das caixas é comparada com a variância global das observações. Como consequência disso,  $\theta$  possuirá uma densidade de probabilidade dada por uma distribuição  $F$  com graus de liberdade  $\sum n_j - M$  e  $N - 1$ , considerando que as observações  $m$  possuam uma distribuição Normal. A distribuição  $F$  desempenha um papel



importante em muitas aplicações estatísticas e está relacionada com a idéia de comparação de variâncias. É conveniente definir  $F$  como um número maior do que 1 (Hoel, 1971), sendo assim  $F \equiv \theta^{-1}$ . A probabilidade de que o valor de  $\Theta$  encontrado não seja devido à flutuações aleatórias tende à 1 à medida que a variância amostral se aproxima da variância da população. Para uma significância  $P$  :

$$F_{(\frac{P}{2}, N_{1f}, N_{2f})} = \frac{1}{\theta}, \quad N_{1f} = N - 1, N_{2f} = \sum n_j - M \quad (3.9)$$

Se o número de observações é grande,  $N > 100$ , o teste  $\chi^2$  pode ser usado.

Uma variação deste teste de significância é proposta por Davies(1990). Ele sugere um novo critério de significância, que é estatisticamente mais completo para amostras de todos os tamanhos e para qualquer taxa S/R, além de ser mais simples e mais sensível à periodicidade do que os testes comumente usados. Aqui, a procura por períodos pode ser efetuada usando-se o método PDM convencional para uma estrutura de caixa (5,1), porém, segundo *Davies*, a significância correta do pico deve ser calculada através da estatística  $L$ , ao invés de  $\theta$ . A transformação do estimador  $\theta$  no estimador  $L$  é dada pela fórmula:

$$L = \frac{(N - 1) - (N - M)\theta}{(M - 1)\theta} \quad (3.10)$$

A probabilidade associada à estatística  $L$ , é obtida diretamente da distribuição  $F$ :

$$F = L \quad (3.11)$$

O valor da Estatística  $L$  é determinado para cada período teste. Isto é, o valor de  $L$  não segue a mesma distribuição  $F$  dos outros períodos, ele precisa ser calculado para cada P.

Da mesma forma como nos métodos *String Length*, as técnicas de *Phase Binning* não se preocupam com a forma da variação que está sendo estudada.

Além disso, os métodos *Phase Binning* não requerem computação tão intensa quanto os métodos *String Length*, pois os dados não precisam ser ordenados no interior das caixas.

Este método é adequado a casos de variações não-senoidais compostas de poucas observações desigualmente espaçadas. A computação é bem direta, permitindo uma automatização completa da procura por períodos. Esse método é uma generalização do método proposto por *Lafler e Kinman*. Esta generalização aumenta muito a utilidade do algoritmo, além de permitir um grau arbitrário de suavização e fornecer uma informação estatística completa.

### 3.4 Método Fourier

O método Fourier é uma das técnicas de determinação de períodos mais populares devido à sua base matemática clássica e aos conceitos elegantes que ele emprega. Desta forma, apresenta um grande número de variações, e publicações. É um método equivalente ao conhecido método de ajuste de mínimos quadrados, porém neste caso estamos ajustando ondas senoidais aos pontos da série.

A transformada de Fourier de uma função é uma especificação das amplitudes e fases de ondas senoidais que, quando adicionadas, reproduzem a função. Considerando uma função  $X(t_i)$ ,  $i = 1, 2, \dots, N_0$  que representa os dados obtidos no tempos  $t_j$ , escrevemos:

$$FT_x(\omega) = \int X(t_j)e^{-i\omega t_j} dx \quad (3.12)$$

A função  $FT_x$  é a transformada de Fourier de  $X(t_i)$  e  $\omega$  é a frequência angular  $\omega = 2\pi\nu = 2\pi(P)^{-1}$ . A expressão acima é usada se a função é contínua e infinita, porém quando estamos analisando dados discretos e finitos no tempo, a fórmula da transformada de Fourier discreta deve ser utilizada:

$$FT_x(\omega) = \sum_{j=1}^{N_0} X(t_j) e^{(-i\omega t_j)} \quad (3.13)$$

O termo  $e^{(-i\omega t_j)}$  pode ser expandido usando a fórmula de *Euler*:  $e^{(-i\omega t_j)} = \cos \omega t - i \sin \omega t$ . Matematicamente  $X(t_i)$  poderia ser uma função complexa, porém como estamos estudando um fenômeno físico,  $X(t_i)$  representa a magnitude de uma estrela, e possui apenas componente real. Com isso  $FT_x$  será dado por:

$$FT_x(\omega) = \sum_{j=1}^{N_0} \cos \omega t - i \sum_{j=1}^{N_0} \sin \omega t \quad (3.14)$$

Essa expressão é de grande importância, pois é através dela que determinamos o chamado Periodograma Clássico. O periodograma é uma função sensível à periodicidades presentes nos dados e serve como uma estimativa, podendo ser definido da seguinte forma:

$$\begin{aligned} P_x(\omega) &= \frac{1}{N_0} |FT_x(\omega)|^2 = \frac{1}{N_0} \left| \sum_{j=1}^{N_0} X(t_j) e^{(-i\omega t_j)} \right|^2 \\ &= \frac{1}{N_0} \left[ \left( \sum_j X_j \cos \omega t_j \right)^2 + \left( \sum_j X_j \sin \omega t_j \right)^2 \right] \end{aligned} \quad (3.15)$$

Para entendermos a utilidade do periodograma, consideremos que os dados da série possuam um sinal senoidal de frequência angular  $\omega_0$ . Quando  $\omega \simeq \omega_0$  os termos  $X(t)$  e  $e^{(-i\omega t)}$  estão aproximadamente em fase e se adicionam de modo que  $P(\omega)$  alcança um valor alto. Porém quando o valor da frequência teste  $\omega$  não se aproxima de  $\omega_0$ , os termos  $\cos \omega t_j$  e  $\sin \omega t_j$  ficam fora de fase, apresentando sinais aleatórios em relação um ao outro e o cancelamento resultante nos fornece um valor baixo na soma, ou seja um valor baixo para

$P(\omega)$ . Sendo assim, a presença de um sinal periódico será revelada no periodograma e indicada por um pico distinto próximo à uma certa frequência  $\omega_0$ .

Ao contrário dos métodos anteriores, o método Fourier trabalha no domínio das frequências. Para cada frequência  $\omega$  temos sempre um valor de  $P(\omega)$  correspondente e quando repetimos este procedimento para vários valores de frequências podemos construir um gráfico de  $P(\omega)$  versus  $\omega$ . Outra peculiaridade deste método é a presença de periodicidade ser indicada no gráfico por um máximo, ao invés de mínimos.

Scargle (1982) propôs no seu artigo uma nova definição de periodograma adequada ao caso de séries desigualmente espaçadas, através da modificação do periodograma clássico. Essa modificação é necessária para que o comportamento estatístico simples do caso de série igualmente espaçadas se mantenha. Desta forma, o periodograma modificado é baseado no método de ajuste de mínimos quadrados, e é dado por:

$$P_x(\omega) = \frac{1}{2} \left\{ \frac{\left[ \sum_j X_j \cos \omega(t_j - \tau) \right]^2}{\sum_j \cos^2 \omega(t_j - \tau)} + \frac{\left[ \sum_j X_j \sin \omega(t_j - \tau) \right]^2}{\sum_j \sin^2 \omega(t_j - \tau)} \right\} \quad (3.16)$$

onde  $\tau$  é definido por:

$$\tan(2\omega\tau) = \frac{(\sum_j \sin 2\omega t_j)}{(\sum_j \cos 2\omega t_j)} \quad (3.17)$$

Esta nova definição, mais complexa que a anterior, possui propriedades estatísticas desejáveis que fazem o seu uso preferível. O comportamento estatístico deste periodograma é análogo ao do periodograma para dados igualmente espaçadas.

O cálculo do periodograma de Scargle é equivalente à ajustar ondas senoidais aos dados pelo método de ajuste de mínimos quadrados. Para cada frequência teste  $\omega$ , uma combinação linear de funções seno e cosseno com argumento  $(\omega t - j)$  é ajustada aos dados através do ajuste de mínimos quadrados. Construindo um gráfico de  $P_x(\omega)$  versus frequência obtemos um



periodograma, onde os picos indicam os períodos e as alturas dos mesmos indicam a significância.

É muito importante analisar a significância estatística do pico que encontramos, para verificar qual é a probabilidade dele ter sido produzido por flutuações aleatórias. Para isso é necessário encontrar um limite  $z_0$  para a potência, que é tal que se o sinal é detectado apenas quando a potência excede este nível, então estaremos errando na detecção somente por um fator  $p_0$  do tempo. Este limite de detecção é definido por:

$$z_0 = -\ln[1 - (1 - p_0)^{\frac{1}{N}}] \quad (3.18)$$

onde  $N$  é o número de freqüências que foram analisadas na procura pelo valor máximo e  $p_0$  é a probabilidade de falso alarme, um pequeno número que é fixado de acordo com a nossa escolha. É de nosso interesse que  $p_0$  seja um número muito pequeno, esta condição modifica a expressão acima resultando em:

$$z_0 \approx \ln\left(\frac{N}{p_0}\right) = 4.6 + \ln N \quad \{\text{para } p_0 = 0.01\}. \quad (3.19)$$

No método de Fourier, os resultados dependem da forma de variação da curva de luz do objeto, sendo assim, ele assume que o sinal possui uma forma senoidal definida. Porém, isto não diminui a sua utilidade para os casos de variações não-senoidais, pois o *Teorema da Decomposição* de Fourier estabelece que qualquer sinal periódico pode ser decomposto em uma soma de ondas seno, com uma freqüência múltiplo inteiro da freqüência fundamental. Sendo assim, o método pode ser aplicado a formas de sinais arbitrárias.

### 3.5 Algoritmo Clean

O método CLEAN foi originalmente desenvolvido por Högbom (1974) sendo aplicado no tratamento de mapas em Radioastronomia, que são produzidos através da transformada de Fourier das medidas de interferômetros. O objetivo inicial era a resolução de problemas relacionados com irregularidades



nos interferômetros, o que resultava em mapas incompletos e difíceis de interpretar.

Neste método identificamos e eliminamos a contribuição esperada para uma fonte pontual localizada em um máximo do mapa e continuamos a iteração no restante do mapa até que reste apenas ruído. Cada subtração remove não só a contribuição da fonte mas também seus lóbulos secundários que se estendem ao redor. O conjunto de fontes pontuais então obtido, as componentes "CLEAN", são usadas como um modelo para a fonte e contém as posições e amplitudes reais características desta fonte, mas não incluem características falsas introduzidas pelos lóbulos secundários. Então, fazemos uma convolução das componentes CLEAN com um "cleanbeam". O "cleanbeam" é uma estrutura com a resolução original, porém sem lóbulos secundários. Finalmente, para encontrar o *mapa CLEAN* final adicionamos os resíduos ao resultado da convolução.

Durante as subtrações, não é permitido remover o valor total do máximo do mapa, mas somente uma fração  $g$  do pico. Esta fração  $g$  é chamada de *ganho* do CLEAN, e apresenta um valor típico de  $0.1 < g < 1.0$ . Esta remoção gradativa, é necessária pois se existe mais de um pico no mapa e o mesmo apresenta ruído, cada subtração apresentará um pequeno erro. O uso do ganho permite que estes erros sejam minimizados.

O método CLEAN tem sido usado em vários estudos de variabilidades astronômicas. Porém, a escolha e identificação do pico mais forte (valor máximo) ainda representam um problema. A análise da significância e a análise de erros ainda estão incompletas, e muitas vezes as propriedades do periodograma de Fourier são utilizadas para compensar esta falha.

Posteriormente, o método CLEAN foi adaptado para o caso unidimensional por *Roberts, Lehar and Dreher (1987)* e utilizado na procura por periodicidades em séries temporais. Esta aplicação é particularmente útil para o caso de séries desigualmente espaçadas, que é o nosso interesse principal neste trabalho. Inicialmente, consideremos uma série temporal discreta, composta de um número finito  $N$  de observações:

$$\{f_r, t_r\} \equiv \{f(t_r), t_r\}, \quad r = 1, 2, \dots, N. \quad (3.20)$$

Esta série pode ser representada por uma função contínua  $f$ , se supomos que os  $N$  valores discretos da série são valores desta função conhecidos apenas em certos pontos  $t_r$ . Desta forma, estes valores devem ser considerados como tendo sido obtidos de  $f$  através da multiplicação com uma função amostragem  $s(t)$  definida por:

$$s(t) = C \frac{\sum_{r=1}^N \omega_r \delta(t - t_r)}{\sum_{r=1}^N \omega_r} \quad (3.21)$$

onde a  $r$ -ésima observação é um fator peso  $\omega_r$ . Esta função consiste de uma soma de  $N$  funções Delta de Dirac multiplicadas pelo peso  $\omega_r$  para cada ponto.

Considerando  $C = 1$  e  $\omega_r = 1$ , o sinal da amostra é dado por:

$$\begin{aligned} f_s \equiv f(t)s(t) &= \frac{1}{N} \sum_{r=1}^N f(t) \delta(t - t_r) \\ &= \frac{1}{N} \sum_{r=1}^N f_r \delta(t - t_r) \end{aligned} \quad (3.22)$$

O teorema da convolução de Fourier estabelece que a transformada de Fourier de um sinal de uma amostra é a convolução do espectro,  $FT[f] = F(\nu)$  com a transformada de Fourier da função de amostragem. Em análise de séries temporais, a última transformada é a janela espectral  $W(\nu) \equiv FT[s]$  e a convolução é chamada de 'Espectro Sujo',  $D(\nu)$  ou *Dirty Spectrum*:

$$D(\nu) = F(\nu) \otimes W(\nu) \equiv \int_{-\infty}^{+\infty} F(\nu') W(\nu - \nu') d\nu' \quad (3.23)$$

As duas funções  $D(\nu)$  e  $W(\nu)$  são observáveis e podem ser diretamente calculadas dos dados da série, resultando em transformadas de Fourier discretas:

$$D(\nu) = \int_{-\infty}^{+\infty} f_s(t) e^{-2\pi i \nu t} dt = \frac{1}{N} \sum_{r=1}^N f_r e^{-2\pi i \nu t_r} \quad (3.24)$$

$$W(\nu) = \int_{-\infty}^{+\infty} s(t) e^{-2\pi i \nu t} dt = \frac{1}{N} \sum_{r=1}^N e^{-2\pi i \nu t_r} \quad (3.25)$$

Consideramos que  $D(\nu)$ ,  $W(\nu)$  e  $F(\nu)$  são reais, e desta forma satisfazem as seguintes simetrias:

$$F(-\nu) = F^*(\nu) \quad (3.26)$$

$$D(-\nu) = D^*(\nu) \quad (3.27)$$

$$W(-\nu) = W^*(\nu) \quad (3.28)$$

Como consequência de uma escolha de normalização, a componente de  $D$  relacionada à frequência zero é o valor médio dos dados:

$$D(0) = \frac{1}{N} \sum_{r=1}^N f_r = \bar{f}_r \quad (3.29)$$

e a função janela é igual a um na origem.

$$W(0) = 1 \quad (3.30)$$

Uma convolução direta, como  $F(\nu) = FT[f] = FT[\frac{f_s}{s}] = FT[f_s] \otimes FT[\frac{1}{s}]$  é impossível uma vez que a solução não é única pois a função de amostragem  $S(t)$  é aproximadamente zero em todo lugar.

É importante ressaltar que o objetivo deste desenvolvimento é estimar  $F$ , através do conhecimento de  $D$  e  $W$ , a fim de desfazer o prejuízo causado pelo

conhecimento incompleto da função  $f$ , tanto quanto possível. Assim sendo, vamos aplicar o método CLEAN, e ele irá realizar uma deconvolução na equação (4). Porém inicialmente temos que considerar que a função contínua  $f$  é dada por uma componente harmônica, ou seja, por uma onda cossenoidal de amplitude harmônica  $A$ , frequência  $\hat{\nu}$  e constante de fase  $\phi$ :

$$f(t) = A \cos(2\pi\hat{\nu}t + \phi) \quad (3.31)$$

O espectro de  $f$  é:

$$F(\nu) = a\delta(\nu - \hat{\nu}) + a\delta(\nu + \hat{\nu}) \quad (3.32)$$

onde:  $a = \frac{A}{2}e^{+i\phi}$

Este é um espectro complexo, que contém um par de componentes espectrais em  $\pm\hat{\nu}$ , de amplitudes espectrais iguais à  $\frac{A}{2}$ . As fases das componentes positiva e negativa são  $+\phi$  e  $-\phi$ , respectivamente.

O *espectro sujo* pode ser determinado através dos dados da série :

$$D(\nu) = aW(\nu - \hat{\nu}) + *aW(\nu\hat{\nu}) \quad (3.33)$$

A janela espectral contém um pico central em  $\nu = 0$ , e picos secundários de menor intensidade dispostos nas laterais do pico central, os chamados *lóbulos secundários*. O pico central é o que possui maior intensidade, contudo os lóbulos secundários maiores podem apresentar uma fração considerável da intensidade do pico central. Desta forma, é de se esperar que no *espectro sujo*, encontremos não só picos de frequência positiva e negativa, mas também picos falsos devido à lóbulos secundários dos mesmos. Os picos falsos podem alcançar uma amplitude substancial e fazer com que a interpretação do *Espectro Sujo* seja muito confusa.

Estamos interessados em subtrair funções do tipo janela espectral do *espectro sujo*. Uma fonte pontual simples produz um máximo no *mapa sujo*, já no *espectro sujo* essa mesma fonte em uma frequência  $\hat{\nu}$  produz dois máximos,



em  $\pm\hat{\nu}$ . Esses picos são contaminados pelos lóbulos secundários do outro, o que faz com que não possamos determinar a amplitude e frequência da componente apenas através da amplitude e frequência do máximo no *espectro sujo*. No caso de uma componente harmônica simples com frequência  $\hat{\nu}$ , o *espectro sujo* é:

$$D(\hat{\nu}) = aW(0) + a^*W(2\hat{\nu}) \quad (3.34)$$

Podemos encontrar a amplitude espectral desta componente em termos da frequência através da expressão:

$$a = \frac{D(\hat{\nu}) - D^*(\hat{\nu})w(2\hat{\nu})}{1 - |W(2\hat{\nu})|^2} \quad (3.35)$$

Dada a frequência de uma componente clean estimada de  $D$ , a sua amplitude complexa é definida como:

$$\alpha(D; \nu) = \frac{D(\nu) - D^*(\nu)w(2\nu)}{1 - |W(2\nu)|^2} \quad (3.36)$$

onde  $a \simeq \alpha(D; \nu)$ .

A aplicação do algoritmo CLEAN pode ser sintetizada de uma forma simples nos seguintes passos:

- 1- No *espectro sujo*  $D$  encontramos a frequência  $\nu^i$  da  $i$ -ésima componente clean do máximo de  $|R^{i-1}(\nu)|$ . Partindo de  $i=1$  e usando a notação  $D = R^0$ , a amplitude da componente é dada por:

$$c^i = g\alpha(R^{i-1}; \nu^i) \quad (3.37)$$

onde  $g$  é o ganho.

- 2- Depois de identificado, removemos a contribuição deste componente, incluindo seus lóbulos secundários, subtraindo de  $D$  uma componente



da forma da equação (1.34) e assim formando o primeiro espectro residual. Este procedimento é repetido. Quando chegamos ao o  $i$ -ésimo passo deste procedimento iterativo, temos um conjunto de  $i$  componentes espectrais CLEAN e um espectro residual  $R^i$ :

$$R^i(\nu) = R^{i-1}(\nu) - (c^i W(\nu - \nu^i) + (c^i * W(\nu + \nu^i))) \quad (3.38)$$

- 3- A iteração deve ser encerrada por algum critério de parada, que neste caso pode ser: quando o máximo de  $R^i$  é menor que algum valor de ruído pré-determinado, o número de iterações alcançou um limite razoável, ou a soma dos módulos das componentes clean não está crescendo de maneira significativa. Desta forma, examinamos  $R^i$  e se a condição de parada ainda não é satisfeita continuamos a iteração.
- 4- Após todas as componentes clean terem sido calculadas, segue-se a convolução das componentes com o *feixe clean*  $B(\nu)$ , que restaura a resolução original do espectro. Ajustamos  $B(\nu)$  à função janela.
- 5- Com o objetivo de manter o nível de ruído e acrescentar quaisquer características que não tenham sido bem representadas pelas componentes espectrais clean, o espectro residual é adicionado para formar o espectro clean. Sendo assim, o espectro clean  $S(\nu)$  é formado pela convolução de  $K$  componentes clean com  $B(\nu)$  e adicionando o espectro residual final:

$$S(\nu) = \sum_{i=1}^K (c^i W(\nu - \nu^i) + (c^i * W(\nu + \nu^i))) + R^K(\nu) \quad (3.39)$$

O método CLEAN é um método iterativo, o que significa que ele consiste de uma seqüência de instruções que são executadas e repetidas em ciclos.

Cada iteração utiliza resultados das iterações anteriores, sempre verificando se foi atingido um resultado próximo o suficiente do esperado. Desta forma, o CLEAN permite que se faça um refinamento, melhorando as aproximações iniciais até que se obtenha uma dentro de uma precisão prefixada.

O algoritmo CLEAN faz iterações no *Espectro sujo* para formar uma versão “clean”, primeiramente identificando o pico mais forte e então limpando numericamente todas as características associadas à ele através do uso da *janela espectral*.

# Capítulo 4

## Resultados e Conclusões

### 4.1 Introdução:

Para definirmos o comportamento periódico da estrela variável *AV Cir*, aplicamos os quatro métodos já discutidos anteriormente aos dados da série. Utilizamos para isso programas que implementam os algoritmos apresentados pelos respectivos artigos, todos desenvolvidos na linguagem FORTRAN. Porém, na aplicação do método PDM utilizamos o programa PDM-Win, que é uma versão windows desenvolvida a partir do método, para calcular o espectro de frequências.

Estes programas também foram aplicados nas séries artificiais com níveis diferentes de relação sinal-ruído, de forma a reproduzir as várias condições apresentadas pelos dados de uma série real.

À princípio, consideramos que os programas não introduzem erros significativos nos resultados. Eles reproduzem o algoritmo proposto pelos artigos nas suas características básicas, reduzindo assim as operações de entrada e saída de informação ao mínimo, de forma a limitar o tempo total gasto na sua execução. A única fonte de erros que levamos em conta é o erro inerente às observações de magnitude da estrela.

A utilização dos programas envolve o uso de certos parâmetros, que devem ser estimados com base na série observada e que estão intimamente

ligados à resolução do periodograma resultante. O intervalo de frequências para o qual analisaremos os dados das séries é um deles, e a sua escolha deve basear-se nas características da estrela variável que está sendo pesquisada, como tipo de variabilidade, períodos conhecidos e outras características físicas importantes.

Os dados observacionais podem ser exibidos em um gráfico em que representamos na ordenada a magnitude da estrela e na abcissa, a fase ou a data juliana heliocêntrica. A fase é geralmente utilizada no caso de estrelas variáveis periódicas enquanto a data juliana heliocêntrica é usada quando a estrela não possui um período regular, apresentando variações aleatórias de luz.

Através da curva de luz, podemos examinar o tipo de variação da estrela. As variáveis pulsantes apresentam curvas de luz de formas senoidais que se repetem regularmente, as variáveis geométricas possuem curvas de luz com características singulares que também se repetem com regularidade. Por outro lado, as variáveis eruptivas apresentam variações aleatórias de brilho que se repetem em intervalos irregulares.

As séries de dados foram processadas por todos os quatro algoritmos para um intervalo de frequência de 0.17 a 0.9 ciclos por dia (o que equivale à um período de 1.1 a 6 dias). Cada algoritmo produziu um gráfico, o periodograma, que foi analisado levando-se em conta o período conhecido das séries artificiais e o período encontrado na literatura para a estrela *AV Cir*. A frequência da série é indicada por um valor extremo, podendo ser um máximo ou mínimo, de acordo com o método usado.

Os resultados de todos os programas de determinação de períodos estão sujeitos à uma série de dificuldades que surgem da imprecisão inerente às observações, das características da amostra de dados e do ruído gerado pelo algoritmo.

As características da amostragem das observações no tempo dão origem à dificuldades conhecidas como *aliasing* e *pseudo-aliasing*. O *aliasing* inibe e até mesmo impede a identificação da frequência correta do sinal, fazendo



com que, no lugar desta, se detecte uma frequência falsa que está relacionada com a verdadeira. Este fenômeno surge quando as observações são feitas em intervalos igualmente espaçados,  $\Delta t$ . Esta regularidade temporal permite que ondas seno de frequência  $f_0 = n(\Delta t)^{-1}$  se ajustem aos dados; nesta expressão  $f_0$  é a frequência real do sinal e  $n$  é um número inteiro. Porém, como já foi visto, as observações astronômicas raramente são igualmente espaçadas no tempo, e desta forma o aliasing não se torna um problema para as séries astronômicas.

Por outro lado, um fenômeno similar denominado pseudo-aliasing está quase sempre presente nas observações astronômicas. O pseudo-aliasing age da mesma forma que o aliasing, causando uma confusão na identificação da frequência real, porém surge devido à presença de uma regularidade na obtenção da seqüência de observações.

As propriedades estatísticas dos programas de determinação de períodos são extremamente importantes e representam a única proteção contra a detecção de períodos falsos. Dentre os quatro programas usados neste projeto, o referente ao método CLEAN é o único que não apresenta um critério de significância estatístico no seu algoritmo, possuindo ao invés disto, um critério de parada das iterações que garante que o resultado está suficientemente próximo do valor real.

## 4.2 Séries Artificiais para teste

Aplicamos os métodos às séries artificiais com relação sinal-ruído distintas. A S/R dos dados da série têm um papel fundamental na detecção da frequência da periodicidade e das características associadas, como amplitude e fase. Quanto maior a S/R, melhor a resolução. A inclusão do ruído nestas séries é importante, para que seja verificada a capacidade de discriminação de ruído dos métodos propostos. É comum os dados serem tão ruidosos que não se consegue nem mesmo afirmar a existência de periodicidades na série.

Os valores de frequência encontrados através da aplicação dos vários



métodos à cada série artificial, podem ser vistos nas tabelas abaixo:

Série	String Length	PDM	Fourier	CLEAN
Simétrica	0.6478448	0.6479762	*	0.6480447
Assimétrica	0.6478448	0.648596	0.6472241	0.6480447
Duplo período	0.6478448	0.647976	0.64908592	0.6480447
Simétrica longa	0.6481731	0.6482932	0.64666622	0.6479156

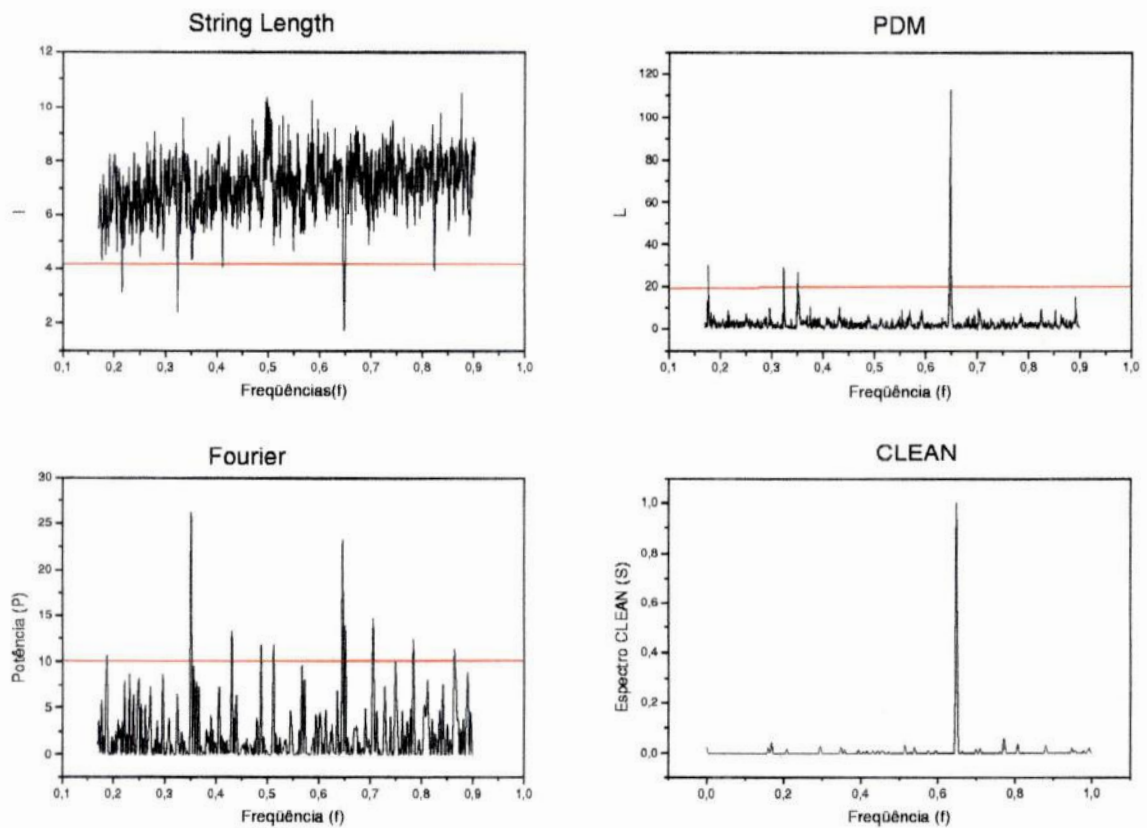
**Tabela IV** - Resultados para as séries artificiais com relação sinal-ruído alta. ciclos/dia.

Série	String Length	PDM	Fourier	CLEAN
Simétrica	*	**	0.6477845	
Assimétrica	*	*	*	*
Duplo período	*	**	*	*

**Tabela VI** - Resultados para as séries artificial com relação sinal-ruído baixa.

O símbolo \* informa que no caso em questão a frequência do pico mais significativo encontrado difere muito da frequência real da série, e o símbolo \*\* indica que as frequências encontradas não são significativas.

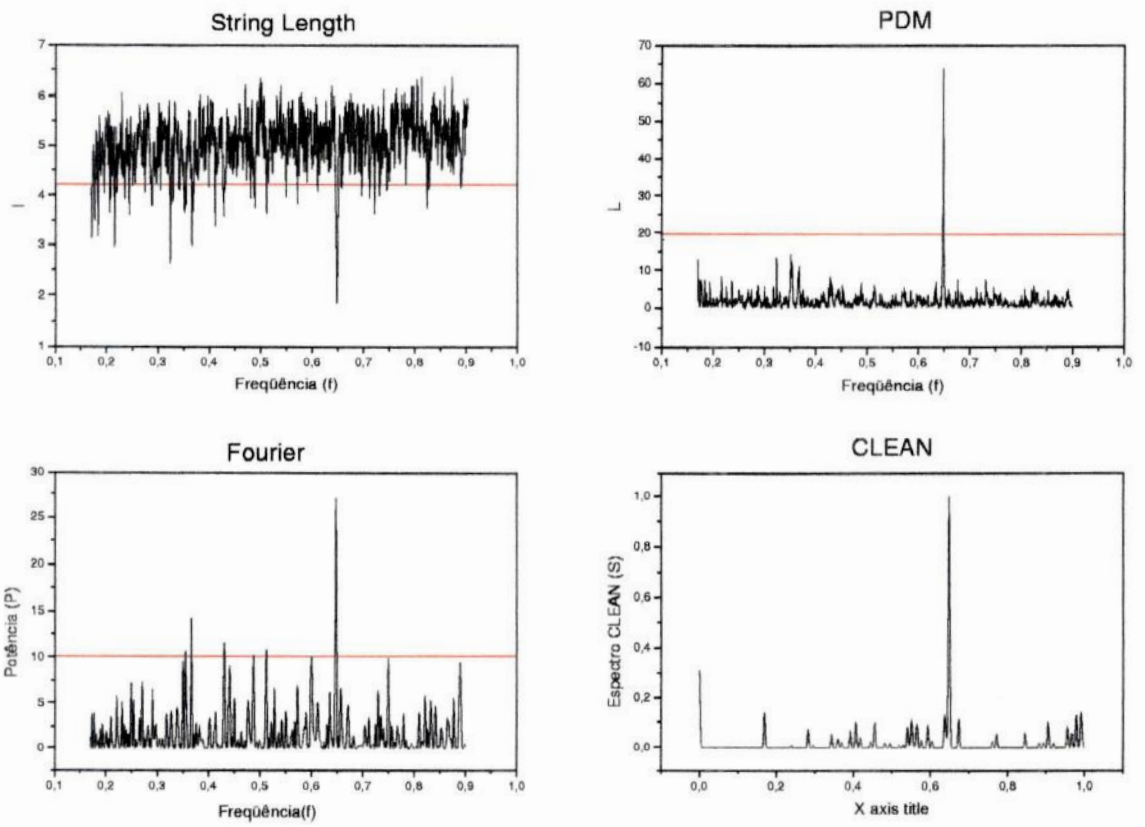
Abaixo são apresentados os periodogramas de cada método para as séries simétrica, assimétrica e com duplo período, todas com alta S/N. Podemos notar que a frequência significativa é representada no gráfico por um extremo, que no caso do método String Length é um mínimo. A barra vermelha representa o limite de significância calculado.



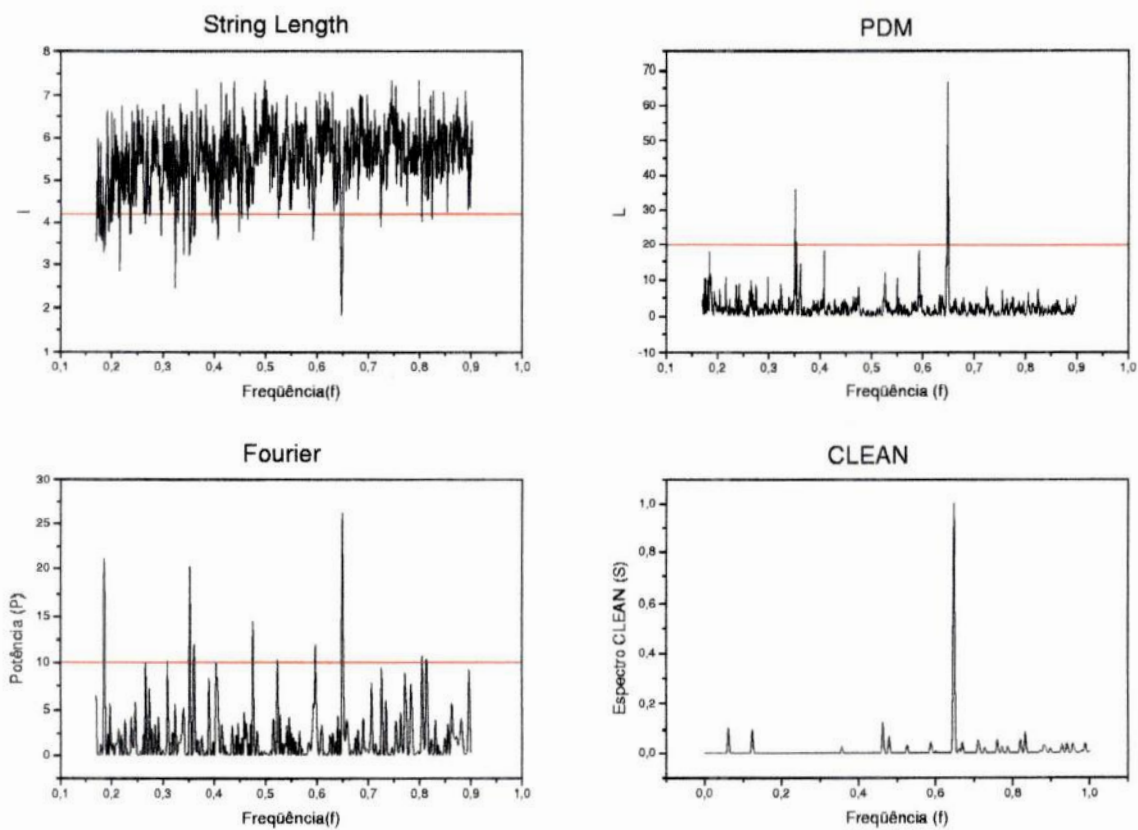
**Figura VII** - Periodogramas resultantes dos quatro métodos para a série simétrica.

Na maior parte das figuras, podemos notar a presença de outros extremos além da frequência principal, que também ultrapassam o limite de significância traçado. Estes extremos são componentes periódicas usualmente chamadas de sub-harmônicos e aliases e apresentam amplitudes que são frações da frequência principal.

Analisando o gráfico do método Fourier na figura VII, é possível verificar que o resultado obtido pelo método Fourier, não corresponde à frequência real da série, pois esta ficou representada por um pico de amplitude menor.

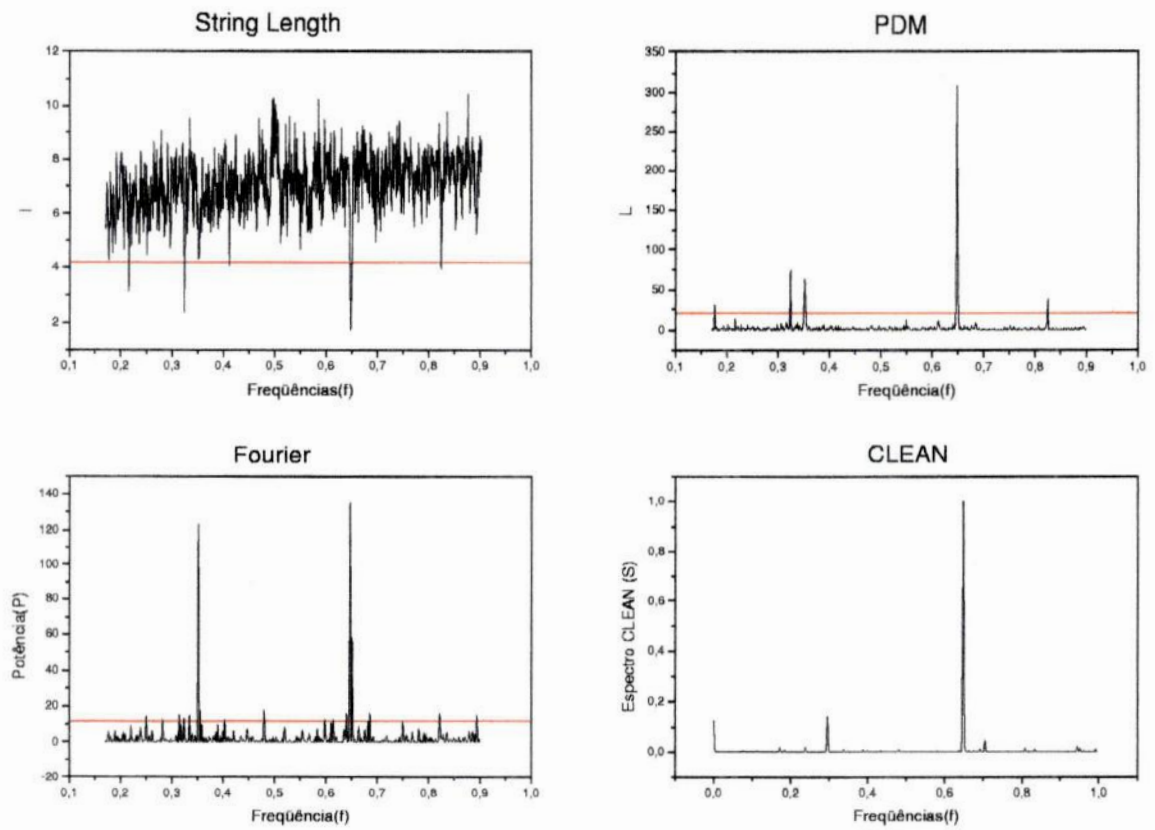


**Figura VIII** - Periodogramas resultantes dos quatro métodos para a série assimétrica.



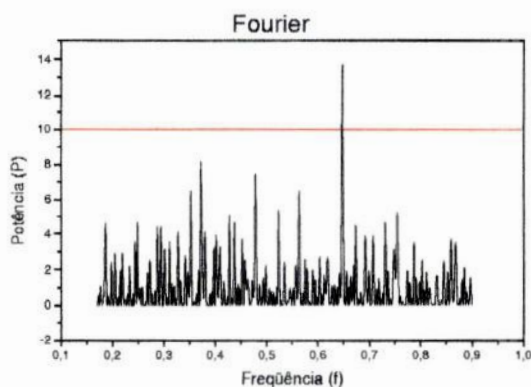
**Figura IX** - Periodogramas resultantes dos quatro métodos para a série com duplo período.





**Figura X** - Periodogramas resultantes dos quatro métodos para a série simétrica longa.

Para as séries com S/N baixo, notamos que todos os programas, com exceção do Fourier, não encontraram uma frequência significativa próxima da frequência real.



**Figura XI** - Periodograma resultante do método Fourier para a série simétrica com S/R baixa.

### 4.3 Série da Estrela AV Cir

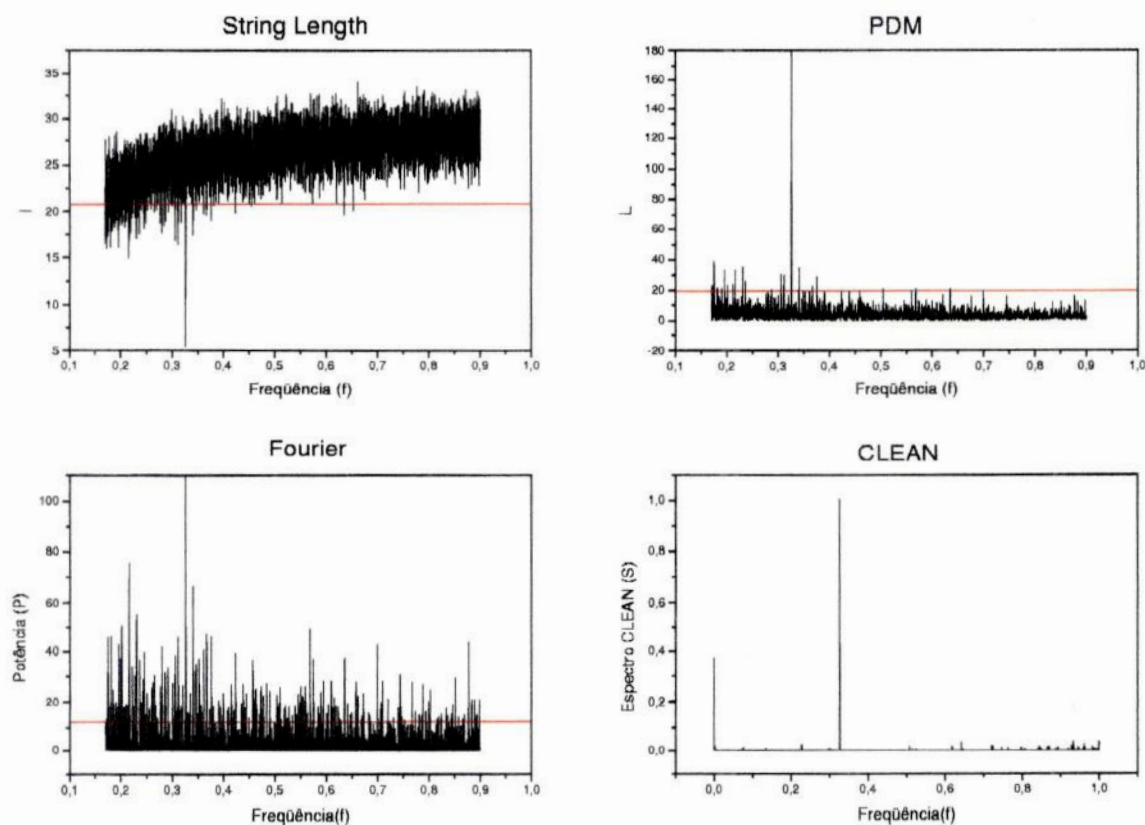
A estrela *AV Cir* é uma Cefeida, e seu período já foi calculado anteriormente como sendo de 1,58 dias ( $f = 0,633\text{ciclos/dia}$ ), desta forma podemos inicialmente examinar o intervalo de 1,1 a 6 dias, o que equivale em frequência a 0,17 e 0,9 ciclos por dia respectivamente.

Abaixo encontramos uma tabela com os resultados obtidos para a estrela *AV Cir*. Nela observa-se os valores de frequência encontrados para o pico mais significativo, nos diferentes métodos.

Método	Frequência
String Length	0.326212
PDM-Davies	0.326221
Fourier	0.325956
CLEAN	0.326269

**Tabela IV** - Estrela AV Cir

Analisando os periodogramas da figura XII, podemos observar em todos um extremo bem definido indicando a frequência da série.



**Figura XII** -Periodogramas resultantes dos quatro métodos para a estrela *AV Cir*

#### 4.4 Comparação dos Métodos:

Devido à dificuldades em se estabelecer teóricamente as propriedades estatísticas das técnicas de determinação de períodos, iniciamos um estudo comparativo dos seus desempenhos através de simulações numéricas. A escolha entre os vários algoritmos depende da estrutura do sinal, pois não existe nenhuma ferramenta válida universalmente para a análise de séries temporais astronômicas.

Desta forma, fazemos uso de séries artificiais que simulam condições observacionais diferentes através da variação da *Relação Sinal-Ruído*. Selecio-

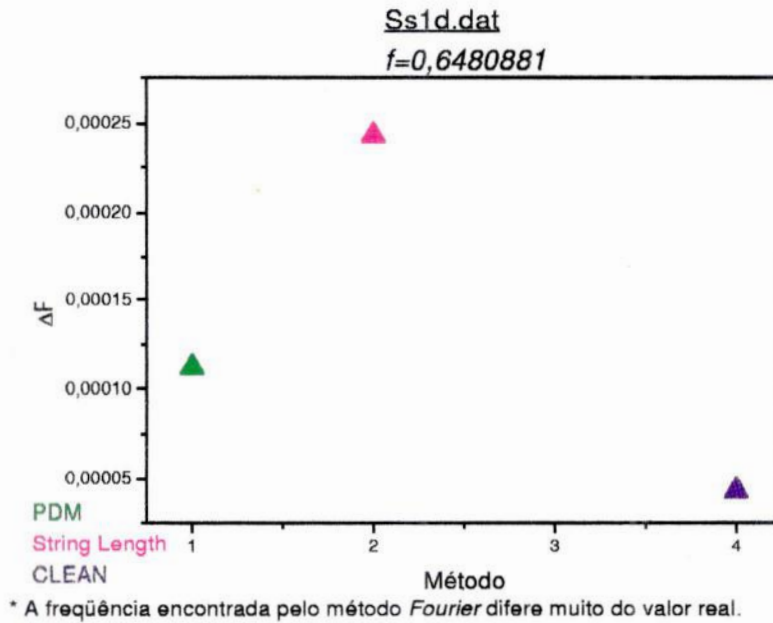
namos quatro algoritmos bastante utilizados, e testamos a sua utilidade para as diferentes séries.

Os periodogramas correspondentes à cada método foram analisados com o objetivo de se determinar a sua atuação no processo de determinação de períodos ou frequências. A frequência verdadeira da série corresponde ao extremo mais profundo que encontramos no gráfico, que pode ser um mínimo ou um máximo de acordo com o método. Porém existem certos casos em que o extremo não representa a frequência real da série que, neste caso, pode estar encoberta por ruído.

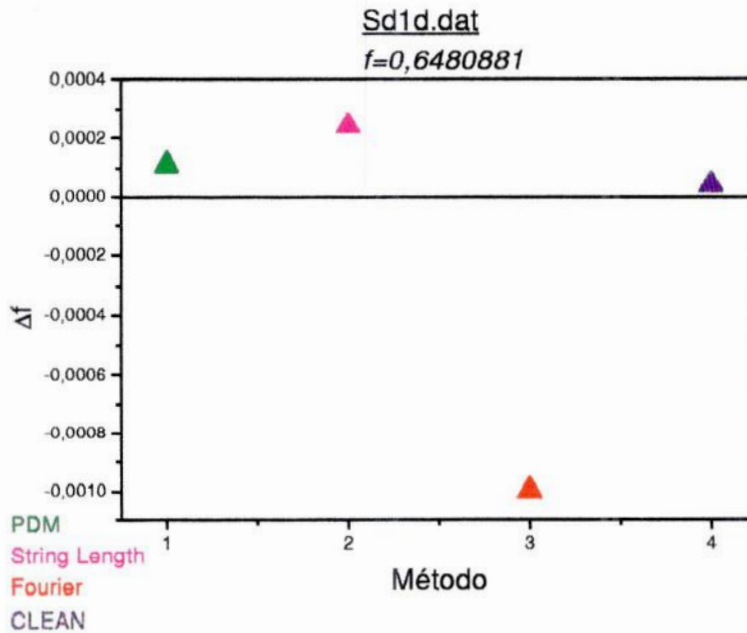
Inicialmente podemos observar os gráficos abaixo, com resultados obtidos para as séries artificiais, que indicam o quanto a frequência encontrada pelos métodos difere da frequência real da série:

- Séries com S/N Alto

O gráfico XIII corresponde à série simétrica, nele podemos notar que o método *Fourier* não foi capaz de determinar a frequência real da série. Observando as figuras XIII, XIV e XV, pode-se dizer que de uma forma geral, o método *CLEAN* forneceu resultados mais próximos dos reais enquanto o método *Fourier* sempre indicava valores mais distantes do esperado. O segundo resultado mais próximo do real foi obtido pelo método PDM, quando aplicado às séries simétrica e com duplo período e pelo String Length, quando aplicado à série assimétrica.

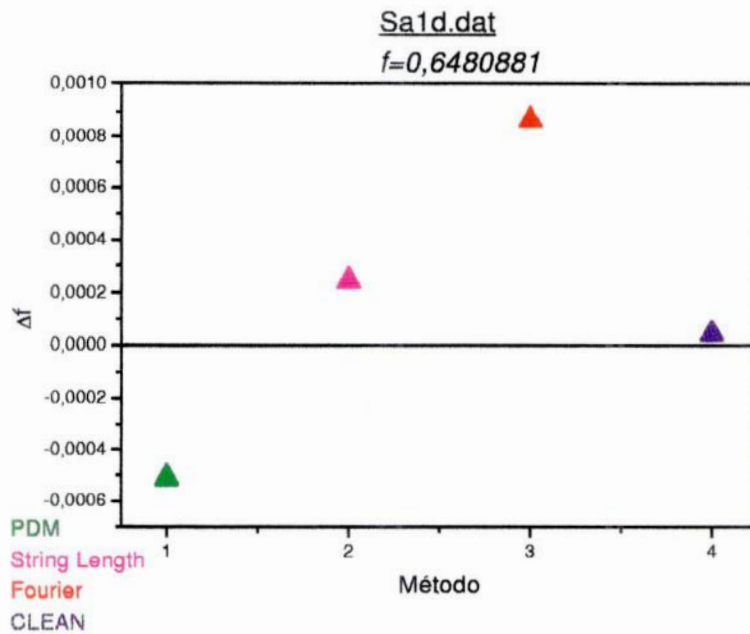


**Figura XIII** -Gráfico da diferença entre a frequência real e a encontrada por cada método para a série assimétrica.

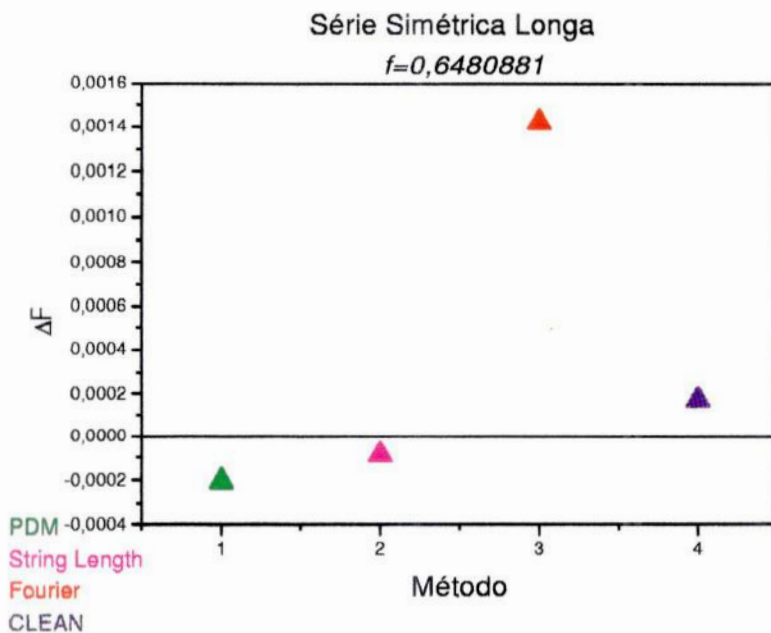


**Figura XIV** -Gráficos da diferença entre a frequência real e a encontrada por cada método para a série com duplo período.





**Figura XV** -Gráficos da diferença entre a frequência real e a encontrada por cada método para a série simétrica.



**Figura XVI** -Gráficos da diferença entre a frequência real e a encontrada por cada método para a série simétrica.

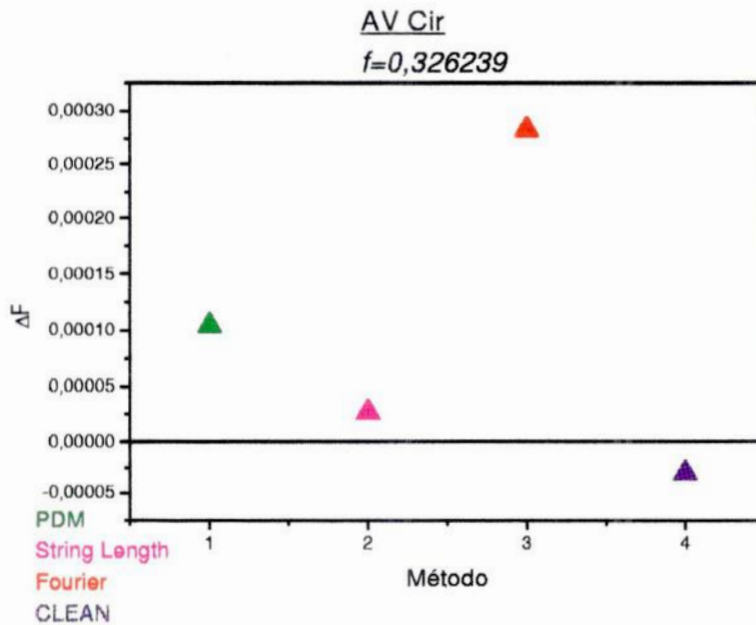
- Séries com S/N baixa

Para as séries com S/N baixa a comparação foi impossível de ser feita, já que apenas o método Fourier conseguiu determinar uma frequência adequada, porém somente para a série simétrica.

Todos os métodos mostraram dificuldades de identificar a frequência real de uma série ruidosa.

- Série da estrela AV Cir

Quando aplicamos os programas à série *AV Cir*, que possui um número de observações maior, os métodos, com exceção do *Fourier*, forneceram valores bastante próximos do ditado na literatura. Aqui o comportamento encontrado nas aplicações em séries sintéticas se repete, com o CLEAN antecedendo o PDM e com o Fourier produzindo resultados distantes do esperado.



**Figura XVII** -Gráficos da diferença entre a frequência real e a encontrada por cada método para a estrela A V Cir.

## 4.5 Conclusões:

Considerando a série da estrela *AV Cir*, observamos que ela apresenta :

- uma alta relação sinal-ruído,
- uma curva de luz simétrica(figura II).

Desta forma, é possível fazer uma comparação com os resultados obtidos para uma série artificial que partilhe as mesmas características e que tenha aproximadamente o mesmo tamanho. A série em questão é a série simétrica longa. Com base nos gráficos anteriores, ficou claro que a atuação dos métodos foi semelhante para as duas séries.

O método String Length forneceu o resultado que mais se aproximou da frequência que gerou a série artificial, e também foi o que mais se aproximou da frequência tabelada para a estrela *AV Cir*. Estes valores mais próximos obtidos pelo método, ao contrário do que foi visto para as outras séries de teste, pode estar relacionadas ao aumento do número de observações das mesmas.

Os resultados obtidos pelo método CLEAN também são razoáveis embora não tão bons quanto os obtidos pelo método String Length. A seguir vem o método PDM com resultados também próximos do esperado. O método Fourier foi o pior dos quatro.

Como na literatura (Dworetzky, 1983), observamos que o tempo de execução computacional do método String Length é muito grande em comparação com os outros. Os métodos CLEAN e o de Fourier foram os mais rápidos.

Assim, concluimos que o método CLEAN é o recomendado para séries senoidais com alta S/R por apresentar bons resultados e pequeno tempo de execução.

## Referências

- Akerlof et al.:1994, *Ap.J* **436**, 787.
- Barber,N.F.M.Sc.:1961, *Experimental Correlograms and Fourier Transforms*, (Pergamon Press), New York.
- Barthés, D.: 1995, *A&A Suppl. Ser.* **111**, 373.
- Bloomfield, P.: 1976, *Fourier Analysis of Time Series: an Introduction*, John Wiley, New York.
- Bossi, M., La Franceschina, L.: 1995, *A&A suppl. Ser.* **113**, 387.
- Burki, G., Maeder, A., & Rufener, F.:1978, *Astr. Ap.* **65**, 363.
- Cavaliere, A.G. et al.: 1998, *New Astronomy Reviews*(*Time Series, neural networks and the future of the sun*). **42**(5), 343
- Davies, S.R.: 1990, *M.N.R.A.S* **244**, 93 .
- Davies, S.R.: 1991, *M.N.R.A.S* **251**, 64 .
- Dworetzky, M.M.:1983, *M.N.R.A.S* **203**, 917 .
- Foster, G.: 1995, *A.J.* **109**, 1889 .
- Fullerton, A.W.: 1986, *The study of Variable Stars Using Small Telescopes*, (ed. J.R.Percy ), Cambridge University Press), New York, p.201.
- Gray, D.F.: 1976, *The Observation and Analysis of Stellar Photospheres*, Heck, A.W., Manfroid, J., Mersch, G.: 1985, *A&A Suppl. Ser.* **59**, 63 .
- Hoel, P.G.: , *Introduction to Mathematical Statistics*, (ed. Jon Wiley & Sons, Inc. ), New York
- Hoffmeister, C., Richter, G.,Wenzel,W.: 1984, *Variable Stars*, Springer-Verlag, Berlin, Heidelberg, New York, London.
- Högbom, J.A.: 1974, *A&A Suppl.* **15**, 417 .
- Horne, J.H.,& Baliunas, S.L.:1986, *Ap.J* **302**, 757 .
- Jurkevich, L.: 1971, *Ap. Space Sci.* **13**, 154 .
- Lafler, J., and Kinman, T.D.: 1965, *Ap.J Suppl. Ser.* **11**, 216 .
- Lohse Resende, C.E.:1996, *Proposta de Dissertação de Mestrado* , Instituto Nacional de Pesquisas Espaciais-INPE .



- Nemec, A.F.L. & Nemec, J.M.:1991, *M.N.R.A.S.* **251**, 356 .
- Perez de la Blanca, N. & Garrido, R.:1981, *Workshop on Pulsating Stars* ,  
eds. G.E.V.O.N. & C. Sterken, p.285 Nice: Nice Observatory.
- Renson, P.: 1978, *Astr. Ap.* **65**, 125 .
- Roberts, D.H., Lehar, J., Dreher, J.W.: 1987, *A.J.* **93**, 968 .
- Scargle, J.D.: 1982, *Ap.J* **263**, 835 .
- Subba Rao, T., Priestley, M.B., Lessi, O.:1997, *Applications of Time Series  
Analysis in Astronomy and Meteorology*, Chapman & Hall, London.
- Stellingwerf, R.F.: 1978, *Ap.J* **224**, 953 .