



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

INPE-10047-PUD/130

ESTIMAÇÃO DE PARÂMETROS DE DADOS FÍSICOS

Clezio Marcos Denardini

Publicação vinculada ao Curso de Pós-Graduação em Geofísica Espacial.

SUMÁRIO

	Pág.
Lista de Figuras	
Estimação de Parâmetros de Dados Físicos	5
1 Características Técnicas Básicas do Radar de Retro-Espalhamento Coerente de 50 MHz (RESCO)	6
2 O Pré-Processamento dos Dados Brutos do Radar	6
2.1 Transmissão de Sinal do Radar	7
2.2 Aquisição dos Ecos Retro-Espalhados	8
2.3 Análise Espectral, Espectrogramas e Mapas RTI	9
3 Parâmetros do Modelo de Ajuste dos Dados	12
4 Determinação do Nível de Ruído	15
5 Efeito do Janelamento Sobre os Dados	18
6 Modelo de Ajuste dos Dados	20
7 Média, Mediana ou Integração dos Dados	22
8 Método de Ajuste de Espectro	24
9 Gradiente e Linearização da Função “Objetivo”	26
10 Modificação de Levenberg-Marquardt	30
11 Restrições aos Parâmetros	31
REFÊRENCIAS BIBLIOGRÁFICAS	35
Apêndice A - Fluxograma da Rotina de Determinação do Nível de Potência Espectral de Ruído	37
Apêndice B - Algoritmo da Rotina de Determinação do Nível de Potência Espectral de Ruído	39
Apêndice C - Prova Matemática da Eliminação do Fator 2 na Transformação da Equação (16) para (21)	41
Apêndice D - Rotina de Aplicação de Restrições de Parâmetros Ajustados pelo Modelo	43

LISTA DE FIGURAS

	Pág.
1 - Parâmetros de controle dos sinais transmitido e recebido pelo radar de espalhamento coerente de São Luís do Maranhão.	7
2 - Esquema de funcionamento da primeira filtragem de informação utilizando os detectores de fase e os filtros de 50 kHz. Nessa figura, A é a amplitude de uma onda genérica, δ é o coeficiente devido ao efeito Doppler e ω é a frequência angular da onda devido aos 50 MHz do sinal emitido.	7
3 - Exemplo de variação das componentes em fase e em quadratura com os pulsos recebidos e do espectro de potência obtido através da análise espectral de Fourier. Os dados foram obtidos em São Luís do Maranhão - MA em 06 de dezembro de 2000, às 14 h 07 min 06 s e correspondem à altitude de 107,8 km.	11
4- - Exemplo de espectrograma obtido pelo agrupamento de todos os espectros referentes à altura de 107,8 km (gate 12). Os dados foram obtidos em São Luís do Maranhão - MA em 06 de dezembro de 2000.	11
5 - Mapa de variação temporal de intensidade típico obtido pela integração em frequência de todos os espectrogramas. Os dados foram obtidos em São Luís do Maranhão - MA entre 23 de janeiro e 15 de fevereiro de 2001.	12
6 - Espectro de potência simulado ajustado pelo método de mínimos quadrados usando a modificação de Levenberg-Marquardt sem o uso de restrições de parâmetros. A curva preta mostra o espectro simulado. A curva azul grossa e tracejada mostra o ajuste do modelo. As linhas vermelhas tracejadas determinam o centro de distribuição de frequência das curvas gaussianas. A diferença entre as linhas vermelha e violeta tracejadas determinam os respectivos desvios padrões das curvas gaussianas. A linha verde traço-ponto determina o nível de ruído do espectro simulado. A área amarela determina a soma das áreas das duas curvas gaussianas.	13
7 - Exemplo de espectro de potência obtido pela análise dos ecos retro-espalhado pelas irregularidades de 3 metros do EEJ e recebido pelo radar RESCO.	15
8 - Espectros de potência do tipo Farley-Buneman simulados (superior) sem incluir o efeito do janelamento amostral sobre o conjunto de dados e (inferior) incluindo este efeito.	19
9 - Diagrama esquemático do modelo de ajuste por mínimos quadrados dos espectros de potência dos sinais retro-espalhados das irregularidades do EEJ.	21
10 - Espectro de potência ruidoso, obtido dos dados do radar RESCO em 1999. Sobrepostos estão os espectros suavizados pelos métodos de (vermelha) média corrida de 5 pontos e de (azul) mediana corrida de 5 pontos.	23

ESTIMAÇÃO DE PARÂMETROS DE DADOS FÍSICOS

Este trabalho versa sobre o método de análise dos dados baseado em estimação de parâmetros. São usados como exemplos os dados do radar de espalhamento coerente de 50 MHz – RESCO para ilustrar o método.

Nas seções que se seguem mostramos com são escolhidos e quais são os parâmetros dos sinais emitidos pelo radar para realização dos estudos desejados. Apresentamos, em seqüência, a forma como os ecos das irregularidades do eletrojato equatorial (EEJ) são recebidos pelo radar e armazenados adequadamente. Demonstramos a forma como os dados brutos do radar RESCO são pré-processados desde sua organização até a obtenção dos espectrogramas e mapas RTI diários. Apresentamos os parâmetros dos espectros de potências das irregularidades do EEJ (dentre os quais o nível de potência do ruído dos dados) que foram estimados pelo ajuste de duas Gaussianas. Neste contexto, discutimos os efeitos da discretização do sinal no momento de sua amostragem, o qual foi levado em consideração em toda nossa análise. Finalmente apresentamos o modelo de ajuste escolhido, o método utilizado, algumas considerações que foram verificadas durante a análise e, por ultimo, toda a abordagem matemática necessária para a utilização deste método. No nosso caso trataremos exclusivamente de Gaussianas, porém a abordagem apresenta um formato genérico permitindo que esta técnica seja utilizada com qualquer tipo de curva.

No que tange ao método de ajuste apresentado neste capítulo, vamos nos deter à teoria diretamente relacionada à estimação de parâmetros, endereçando todo o embasamento teórico e a utilização atual desta técnica aos trabalhos de Marquardt (1963), Cohen (1973), Bard (1974), Lim e Oppenheim (1988), Fukao (1989), Bevington e Robinson (1992), Press et al. (1992), Thompson (1998), Kudeki et al. (1999), Hysell e Burcham (2000), Chau et al. (2000), entre outros referenciados nestes.

1 Características Técnicas Básicas do Radar de Retro-Espalhamento Coerente de 50 MHz (RESCO)

O radar de retro-espalhamento coerente de 50 MHz, desenvolvido e construído no INPE para investigações do EEJ, está totalmente operacional em São Luís (2.33° S; 44.2° W; dip: -0.5 °) desde 1999. Observações das irregularidades de plasma de 3 metros do EEJ são conduzidas rotineiramente durante 10 dias por mês cobrindo o Calendário Geofísico Internacional ou períodos específicos, de interesse científico. As características básicas dos sistemas do radar RESCO são detalhadas na:

TABELA 1 - Características Técnicas Básicas do Radar RESCO.

PARÂMETROS	BREVE DESCRIÇÃO
Transmissão	Sistema modular consistindo de 8 transmissores com controle de fase do sinal transmitido.
Potência de Pico	120 kW.
Conjunto de Antenas	16 antenas magneticamente alinhadas norte-sul do tipo COCO (<i>collinear co-array</i>) com 48 dipolos cada uma, usadas para transmissão e recepção.
Feixe da Antena	Possui abertura de ~5° no plano leste-oeste e ~2,5° no plano norte-sul, com possibilidade de inclinação no plano leste-oeste.
Resolução de Altura	A menor resolução de altura teoricamente estimada é de 7,5 km para o feixe oblíquo com 30° de ângulo zenital;

2 O Pré-Processamento dos Dados Brutos do Radar

São considerados dados brutos do radar RESCO os dados digitalizados dos dois canais de recolhimento dos sinais em fase e em quadratura. Nas seções a seguir estão detalhados os meios como estes dados são obtidos, como é sua estrutura, qual a informação contida nestes e os processos matemáticos aplicados a eles até obtermos os mapas RTI e espectrogramas.

2.1 Transmissão de Sinal do Radar

O radar emite um sinal senoidal de 50 MHz, modulado por um trem de 256, 512 ou 1024 pulsos de onda quadrada. A quantidade de pulsos emitidos (NP) varia de acordo com o fenômeno que está sendo observado, pois aumentar a quantidade de pulsos implica em aumentar a resolução de frequência do espectro obtido pela análise de Fourier, o que é aplicável aos estudos onde pequenos efeitos Doppler são esperados. Esse sinal é emitido pelo conjunto de antenas do radar RESCO que atualmente permite transmissão de forma oblíqua (30° de ângulo de zênite) ou de forma vertical.

A FIGURA 1 ilustra alguns dos parâmetros de controle do radar RESCO para o caso de um pulso sendo transmitido. Nesta figura também podem ser observados alguns dos parâmetros de controle do sinal recebido, o qual será discutido na próxima seção.

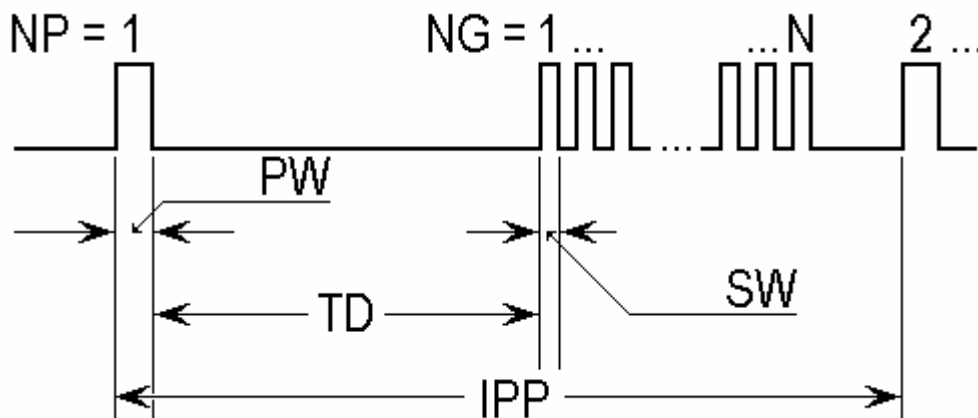


FIGURA 1 - Parâmetros de controle dos sinais transmitido e recebido pelo radar de espalhamento coerente de São Luís do Maranhão.

Os parâmetros de controle do sinal emitido pelo radar são a largura de pulso de transmissão (PW, do inglês: *pulse width*), a qual é variável de 20 μs até 100 μs e o período entre pulsos consecutivos (IPP, do inglês: *inter-pulse period*), o qual é variável de 1 ms até 20 ms. Valores típicos destes parâmetros para estudos do EEJ são: PW = 20 μs e IPP = 1 ms.

2.2 Aquisição dos Ecos Retro-Espalhados

Quando um sinal retro-espalhado pelas irregularidades de 3 metros do EEJ é recebido pelo conjunto de antenas, ele é amplificado e reconstituído de acordo com o tipo de inclinação de feixe utilizada na transmissão. Em seguida ele é dividido em dois sinais idênticos. Esses dois sinais seguem para dois detectores de fase distintos. Em um destes detectores de fase o sinal recebido é comparado com o sinal senoidal de 50 MHz gerado pelo radar e utilizado na transmissão. No outro detector, o sinal recebido é comparado com o sinal senoidal de 50 MHz do radar defasado de 90° elétricos.

O resultado da comparação do sinal recebido com o sinal de 50 MHz em fase é um sinal senoidal de 100 MHz adicionado do co-seno do efeito Doppler percebido no eco. No resultado da outra comparação observa-se o seno do efeito Doppler ao invés do co-seno. Após uma filmagem destes sinais resultantes com a utilização de filtros passa-baixa (LPF, do inglês: *low pass filter*) de 50 ou 100 kHz, o que resta são somente os seno e co-seno do efeito Doppler observáveis no eco. Em resumo, deste processamento do sinal elétrico complexo recebido resultam as componentes complexas (seno e co-seno) do efeito Doppler sofrido por esse sinal ao ser retro-espalhado pelas irregularidades de plasma do EEJ. A FIGURA 2 exemplifica matematicamente os processos eletrônicos descrito acima.

Finalmente, estes sinais são amostrados, digitalizados e gravados em arquivos binários. O número de amostras tomadas do eco determina o número de faixas de alturas que escolhemos para analisar individualmente, ou seja, o número de *gates* (NG). Usualmente esse número é de 16 amostras pra estudos do EEJ.

Esse processo se repete para todos os pulsos emitidos pelo radar, cujos ecos processados são armazenados no mesmo arquivo binário. Portanto, para o caso de 256 pulsos emitidos ($NP = 256$) e 16 *gates* amostrados ($NG = 16$), cada arquivo binário gerado pelo sistema terá $NP \times NG$ pares complexos, ou seja, 4096 pares complexos neste exemplo.

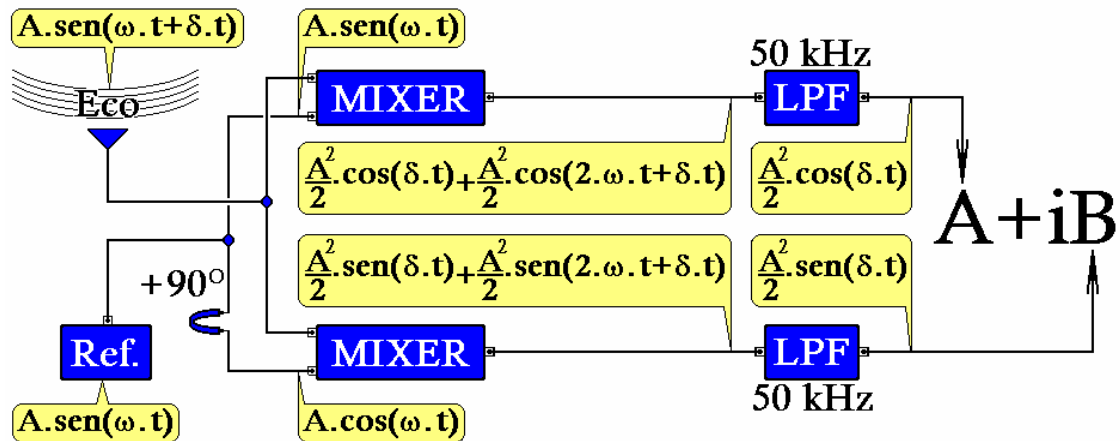


FIGURA 2 - Esquema de funcionamento da primeira filtragem de informação utilizando os detectores de fase e os filtros de 50 kHz. Nessa figura, A é a amplitude de uma onda genérica, δ é o coeficiente devido ao efeito Doppler e ω é a frequência angular da onda devido aos 50 MHz do sinal emitido.

Assim como na transmissão do sinal, é possível controlar certos parâmetros da aquisição dos dados. As variáveis de controle do sinal recebido são: a atenuação exigida pela unidade de aquisição de dados (ATT, do inglês *attenuation*); o tempo entre a emissão do sinal do radar e o início dos pulsos de amostragem (TD, do inglês: *time delay*), que determina a altura de início do estudo; e o número de amostras do sinal retro-espalhado NG, que representam as alturas de estudo (veja a FIGURA 1).

2.3 Análise Espectral, Espectrogramas e Mapas RTI

A repetição de emissão de vários conjuntos de pulsos ao longo de todo dia produz uma quantidade significativa de arquivos binários. Cada arquivo contendo a quantidade de pares complexos do modo que foi descrito nas seções anteriores. A técnica usual de processamento destes arquivos, doravante denominados dados brutos, se dá pela análise espectral de Fourier.

Contudo, antes de procedermos à análise espectral das componentes em fase e em quadratura (real e imaginária, respectivamente) do efeito Doppler, estas componentes são analisadas individualmente. O primeiro passo é verificar se não há um nível contínuo adicionado a essas componentes. Para isso, é tomada a média da variação de cada componente para o conjunto de pulsos recolhidos. Em seguida, essa média é subtraída de cada valor do conjunto. O

segundo passo é verificar se não há a presença ruídos espúrios no sinal digitalizado. Neste processo é verificado se não é observada nenhuma variação brusca no conjunto de amplitudes relativas ao conjunto de pulsos sendo processados. No caso positivo essas variações bruscas são eliminadas.

Após estas verificações, as componentes são então organizadas em pares complexos e os dados são finalmente transportados para o domínio da frequência. De cada arquivo de dados brutos são obtidos um espectro de potência para cada *gate* que foi amostrado. No caso do arquivo do exemplo anterior, onde foram tomadas amostras de 16 *gates*, são obtidos 16 espectros por arquivo.

A FIGURA 3 a seguir mostra um exemplo da variação das componentes em fase e em quadratura de acordo com a seqüência de ecos recebidos. Neste exemplo já foram eliminados os efeitos do nível contínuo e dos picos espúrios. Ainda nesta figura, é apresentado o gráfico correspondente do espectro de potência obtido pela análise de Fourier desta série complexa.

A análise espectral de todos os dados proporciona vários conjuntos de espectros. Os espectrogramas são obtidos pelo agrupamento em ordem cronológica de todos estes espectros relativos ao mesmo *gate*. Portanto, ao final da análise espectral de todo o conjunto de dados coletado durante um dia, teremos tantos espectrogramas quantos forem os *gates* utilizados na amostragem do eco recebido. No caso do corrente exemplo, teríamos obtido 16 espectrogramas. A FIGURA 4 apresenta um exemplo de espectrograma obtido pelo agrupamento de todos os espectros referentes à altura de 107,8 km do mesmo conjunto de dados da FIGURA 3. Nesta figura o código de cores determina a distribuição em frequência da potência espectral (em Watts) do eco recebido ao longo dos horários do dia.

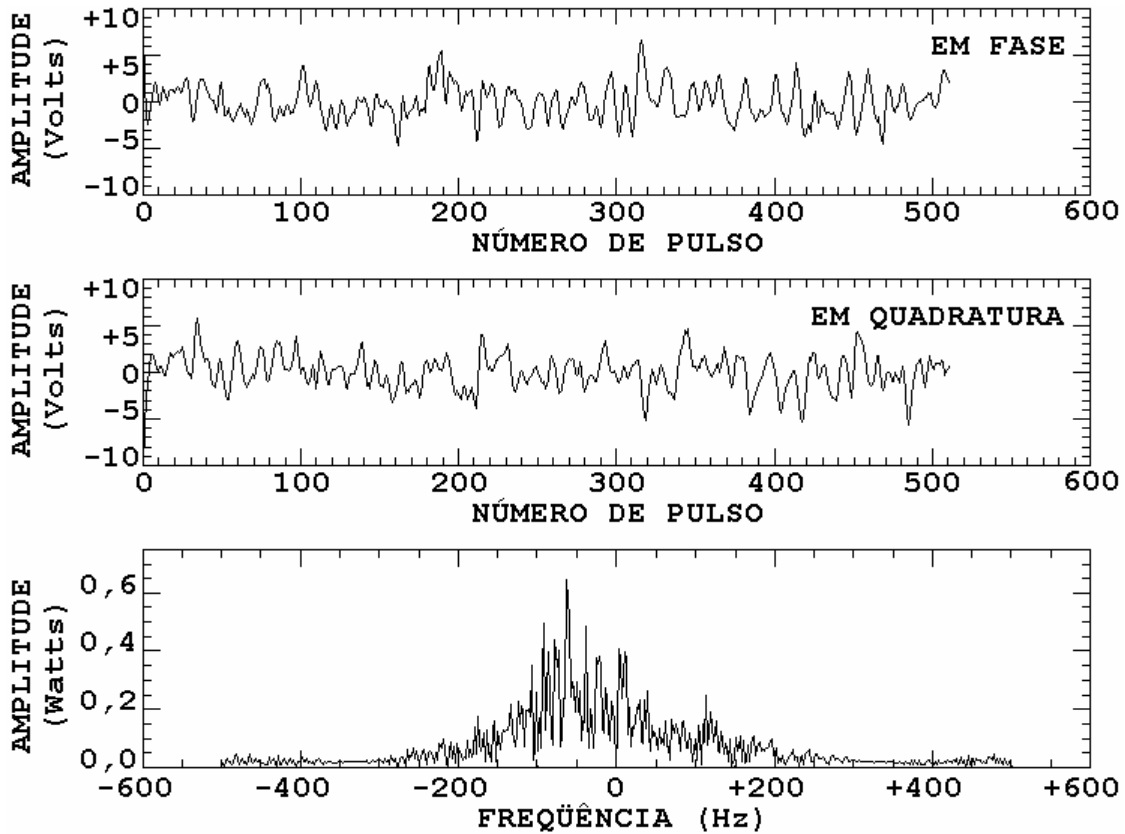


FIGURA 3 - Exemplo de variação das componentes em fase e em quadratura com os pulsos recebidos e do espectro de potência obtido através da análise espectral de Fourier. Os dados foram obtidos em São Luís do Maranhão - MA em 06 de dezembro de 2000, às 14 h 07 min 06 s e correspondem à altitude de 107,8 km.

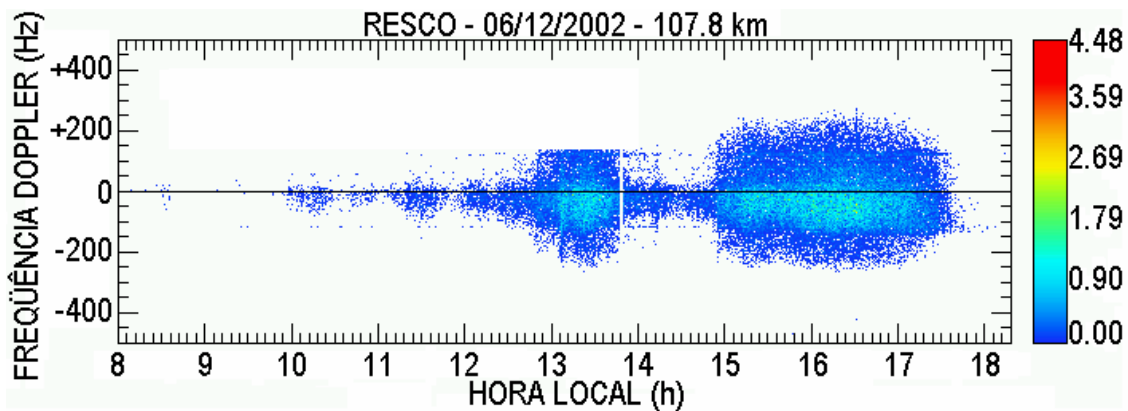


FIGURA 4 - Exemplo de espectrograma obtido pelo agrupamento de todos os espectros referentes à altura de 107,8 km (gate 12). Os dados foram obtidos em São Luís do Maranhão - MA em 06 de dezembro de 2000.

Na etapa seguinte, todos os espectrogramas são integrados em frequência e agrupados de acordo com a altura à qual correspondem. A integração em frequência de cada espectro de potência componente dos espectrogramas resulta na variação da potência total de sinal retro-espalhado ao longo do período de aquisição para a faixa de altura correspondente do espectrograma. Com isso é formado um mapa de variação de intensidade de potência (mapas RTI) para o dia de aquisição. A FIGURA 5 apresenta um exemplo de mapa RTI típico, obtido pela integração dos espectrogramas de todas alturas observadas.

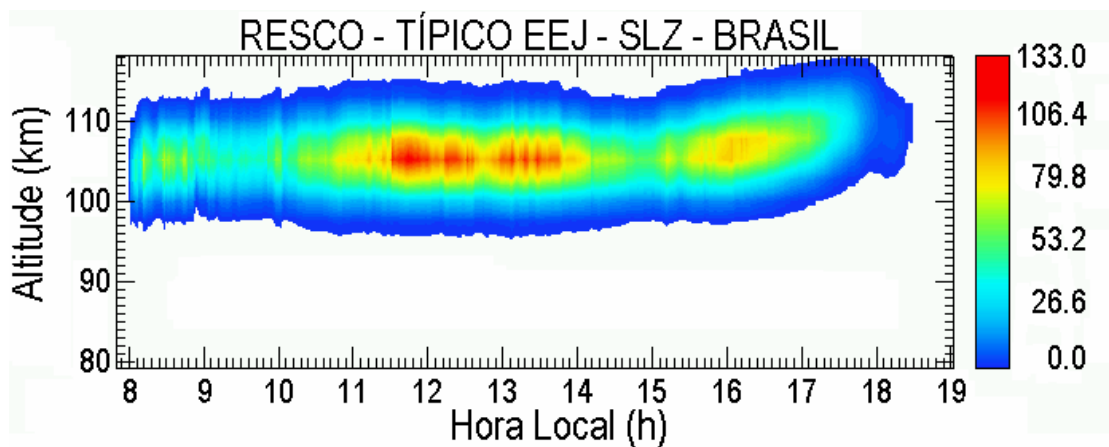


FIGURA 5 - Mapa de variação temporal de intensidade típico obtido pela integração em frequência de todos os espectrogramas. Os dados foram obtidos em São Luís do Maranhão - MA entre 23 de janeiro e 15 de fevereiro de 2001.

Neste mapa é possível visualizar a região que está retro-espelhando o sinal transmitido pelo radar em relação ao tempo e à altitude. O código de cores deste mapa representa a intensidade de potência recebida. Um comentário pertinente é que este mapa já foi corrigido com relação ao nível de ruído. Este aspecto será discutido na seção apropriada subsequente.

3 Parâmetros do Modelo de Ajuste dos Dados

A modelagem que fazemos é sobre cada espectro individualmente. Ajustamos cada espectro de potência com o somatório de duas gaussianas. Assim, ao invés de cada espectro ser representado por um conjunto de pontos obtidos experimentalmente ele será representado pelos parâmetros de duas curvas gaussianas que melhor se ajustam a este conjunto de pontos e por um nível de

potência de ruído comum às duas gaussianas. A FIGURA 6 exemplifica os parâmetros que obtemos de cada espectro.

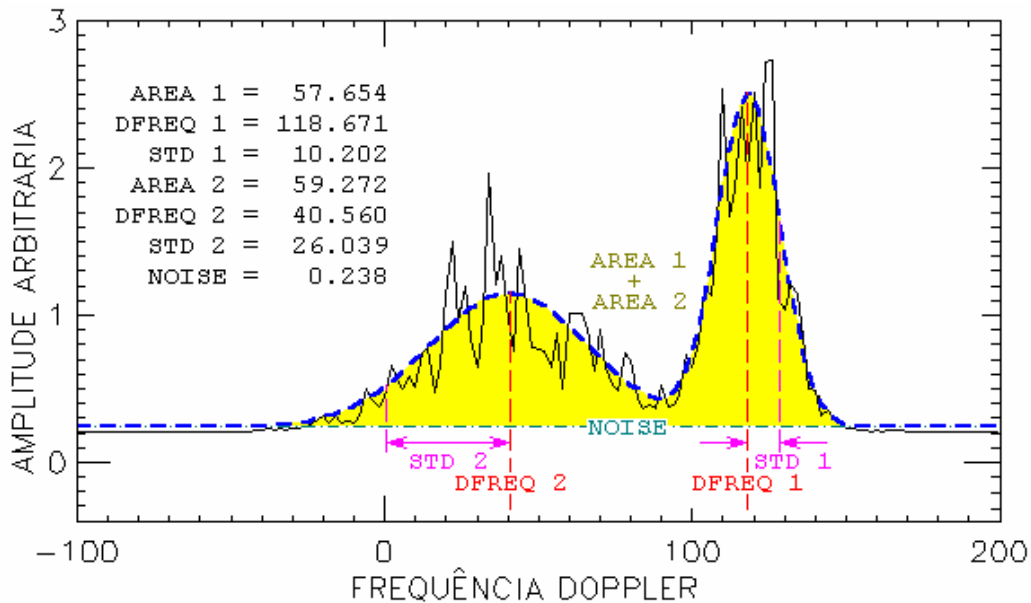


FIGURA 6 - Espectro de potência simulado ajustado pelo método de mínimos quadrados usando a modificação de Levenberg-Marquardt sem o uso de restrições de parâmetros. A curva preta mostra o espectro simulado. A curva azul grossa e tracejada mostra o ajuste do modelo. As linhas vermelhas tracejadas determinam o centro de distribuição de frequência das curvas gaussianas. A diferença entre as linhas vermelha e violeta tracejadas determinam os respectivos desvios padrões das curvas gaussianas. A linha verde traço-ponto determina o nível de ruído do espectro simulado. A área amarela determina a soma das áreas das duas curvas gaussianas.

Como estamos modelando o espectro, os dados do radar de Jicamarca e do RESCO podem ser ajustados pelo mesmo modelo e usar o mesmo algoritmo. Não importa a maneira como os espectrogramas são montados, mas sim como o modelo é alimentado. A resposta do modelo será sempre 3 parâmetros para cada gaussiana e um nível de ruído comum.

O porquê de escolhermos duas gaussianas para modelar os dados de espectros de potência de radares é uma pergunta bem pertinente. Entretanto, sua resposta não é trivial. Alguns pesquisadores os quais têm trabalhado com radares tendem a dizer que é um simples caso de semelhança entre curvas. E, desde os primeiros experimentos de retro-espalhamento nas irregularidades do

EEJ, os espectros têm mostrado sempre grande semelhança com gaussianas, sustentando essa resposta (Cohen, 1973).

Contudo, não nos contentamos com essa resposta e encontramos na literatura especializada alguns pesquisadores que discutem distribuições gaussianas de dados de radares, mas nunca dando enfoque específico ao ajuste de espectro de potência. Assim, após realizar uma busca nos artigos relacionados e baseados em nossa própria experiência, reunimos um conjunto de informações que nos leva a uma possível conclusão.

Quando obtemos o conjunto de dados de um grande somatório, como é o caso em problemas de estimação espectral, a estatística gaussiana pode ser justificada (Fukao, 1989; Kudeki et al., 1999). Esse é exatamente o caso dos espectros das irregularidades do EEJ. Somado a isso temos o fato de que, em geral, realizamos diversas integrações incoerentes para obter um espectro mais representativo e menos ruidosos. Essas integrações não são nada mais que a soma de diversas medidas de uma grandeza em um curto espaço de tempo comparado com a escala de tempo do evento. Além disso, associado a cada medida, temos as incertezas do equipamento, da amostragem e da própria natureza aleatória do evento natural. Assim, cada medida do radar, mesmo que realizada em intervalos de 6 segundos, difere de sua média subsequente por um erro aleatório.

Com isso, podemos afirmar que nossos espectros de potência representam a distribuição estatística das velocidades de propagação das irregularidades do EEJ que retro-espalham o sinal emitido pelo radar. Assim, a característica aleatória do evento fica bem caracterizada (Jenkins e Watts, 1968).

O porquê de nosso modelo espectral utilizar o somatório de duas gaussianas se deve ao fato de co-existirem no EEJ, em torno do meio dia local, dois tipos de irregularidades distintos.

4 Determinação do Nível de Ruído

Como já foi mencionado na seção anterior o nível de ruído dos espectros de potência é um dos sete parâmetros que são obtidos para caracterizá-los. Para estes parâmetros nos baseamos nas características de formação dos espectros de potência. Na FIGURA 7, apresentamos um exemplo de espectro de potência normalizado, típico do EEJ. Nela estão identificados os parâmetros do espectro usados para determinar o nível de ruído.

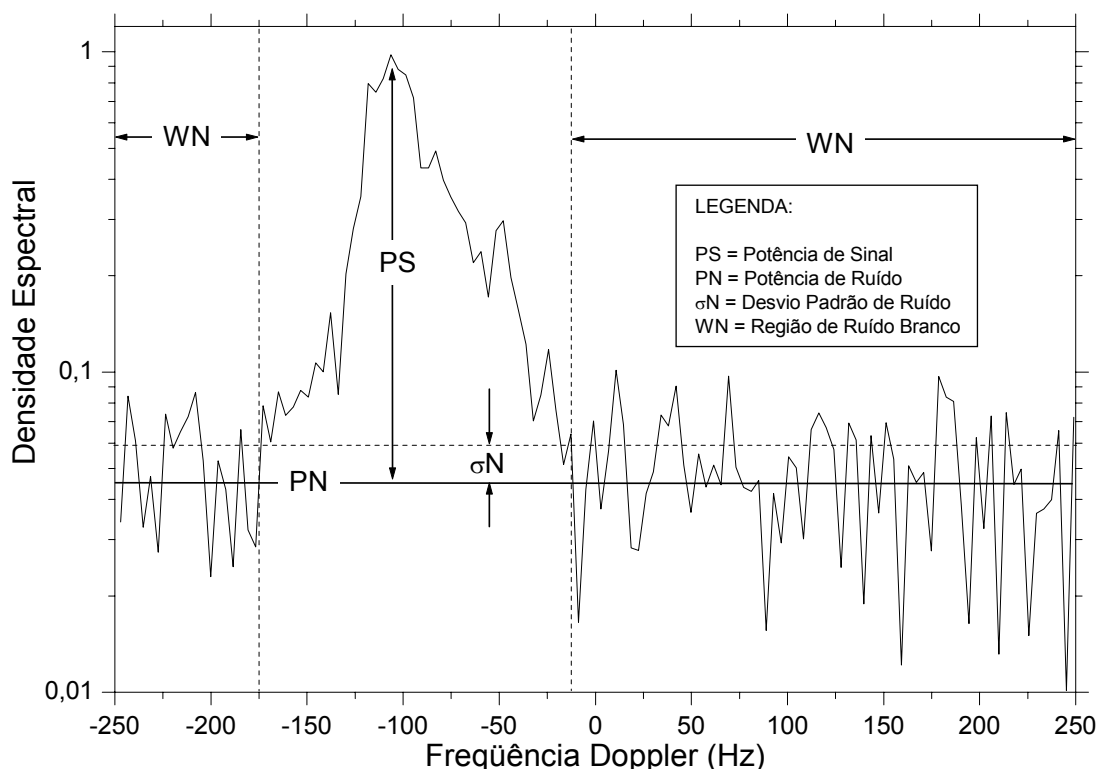


FIGURA 7 - Exemplo de espectro de potência obtido pela análise dos ecos retro-espalhado pelas irregularidades de 3 metros do EEJ e recebido pelo radar RESCO.

Nesta figura, PS é potência de densidade espectral do sinal recebido médio, PN é a potência de densidade espectral do ruído espectral médio, σN é o desvio padrão do ruído ou flutuação de nível de ruído. A faixa de frequência é determinada pelo IPP do sinal transmitido pelo radar.

Para a perfeita compreensão de como o nível de ruído é determinado, temos que introduzir o conceito de detectabilidade. Um critério de detectabilidade de

um sinal retro-espalhado por um alvo não-rígido está definido no trabalho de Gage e Balsley (1978) e Balsley e Gage (1980). Este critério é reproduzido aqui na forma da Equação (1).

$$\frac{PS}{\sigma N} \cong \frac{\bar{P}_t \cdot A_e \cdot F_1 \cdot F_2 \cdot c \cdot \tau \cdot \eta \cdot \alpha^2 \cdot \sqrt{n_i}}{16 \cdot \sqrt{2} \cdot \pi \cdot r^2 \cdot k \cdot (T_s + \alpha T_c) \cdot \delta f} \quad (1)$$

Nesta equação definimos como P_t a potência média transmitida; A_e a área efetiva da antena; F_1 a fração do sinal que passa pelo filtro de recepção; F_2 a fração de sinal que sofre processo de integração coerente; c a velocidade da luz no vácuo; τ a largura do pulso de transmissão; η a refletividade volumétrica do processo de espalhamento; α a eficiência linear de transmissão; n_i o número de integrações incoerentes; r a distância do radar à região de espalhamento; k a constante de Boltzmann; T_s a temperatura de ruído; T_c a temperatura de ruído cósmico nos terminais da antena; e finalmente, δf a faixa de frequência do espectro.

Apesar de aparentemente complexa, a Equação (1) determina uma relação direta entre a potência espectral do sinal, o desvio padrão do ruído e o número de integrações coerentes. Se definirmos D como a detectabilidade, dada pela equação:

$$D \equiv \frac{\bar{P}_t \cdot A_e \cdot F_1 \cdot F_2 \cdot c \cdot \tau \cdot \eta \cdot \alpha^2}{16 \cdot \sqrt{2} \cdot \pi \cdot r^2 \cdot k \cdot (T_s + \alpha T_c) \cdot \delta f} \quad , \quad (2)$$

podemos re-escrever a Equação (1) da forma:

$$\frac{PS}{\sigma N} \cong D \cdot \sqrt{n_i} \quad (3)$$

e representar a detectabilidade por:

$$D \cong \frac{PS}{\sigma N \cdot \sqrt{n_i}} \quad (4)$$

Em observações de radar de espalhamento, é usual postular que é possível realizar a estimação de parâmetros espectrais se a detectabilidade for superior a três (Fukao, 1989).

Uma vez definida a detectabilidade de um radar, dada pela Equação (4), temos que lembrar que no caso do ruído branco (região definida como WN na FIGURA 7) a potência espectral do ruído PN é da ordem de grandeza ou maior que o desvio padrão do ruído σN (Fukao, 1989). Em outras palavras, a relação entre potência espectral do ruído PN e o desvio padrão do próprio ruído σN é, estatisticamente, maior ou igual a um.

Agora temos as ferramentas para determinar o nível de ruído em distribuições espectrais de potência. Se calcularmos a detectabilidade dentro da região de ruído, estamos usando a potência espectral do ruído PN ao invés da potência de densidade espectral do sinal recebido médio PS. Usando também a relação recém mencionada entre a potência espectral do ruído PN e desvio padrão do ruído σN , vamos chegar a Inequação (5).

$$\frac{PN}{\sigma N \cdot \sqrt{n_i}} \geq 1 \Rightarrow \sigma N \leq \frac{PN}{\sqrt{n_i}} \quad (5)$$

Finalmente, temos uma função de teste para determinar a potência de ruído em espectros de potência. É necessário apenas ordenar os pontos do espectro em ordem crescente de amplitude de modo que o ruído esteja localizado na parte inferior do vetor de amplitudes. Em seguida, é necessário percorrer o vetor de amplitudes verificando a relação da Inequação (5), aumentando o número de pontos tomados gradativamente. No ponto em que essa relação não for mais satisfeita, denominado k-ésimo ponto, a potência espectral de ruído está determinada. O nível de ruído do espectro estará entre o k-ésimo elemento e

seu antecessor. Por razões de simplicidade costuma-se assumir que a potência de ruído é a potência espectral do elemento imediatamente anterior ao k-ésimo elemento.

Neste algoritmo temos de um lado o cálculo do desvio padrão da amplitude de um conjunto de pontos tomados e o do outro o somatório das amplitudes desses pontos, levando em conta o número de integrações incoerentes realizadas para obter o espectro. Como o somatório envolve uma grande quantidade de pontos ruidosos e alguns poucos pontos com sinal, quando o desvio padrão das amplitudes for superior ao outro lado da desigualdade significa que estamos entrando na região de sinal. Em outras palavras, houve uma variação abrupta de amplitude espectral comparada com a amplitude média corrente. No Apêndice A apresentamos o fluxograma deste método de determinação de ruído e no Apêndice B, o algoritmo do método para seu melhor entendimento.

5 Efeito do Janelamento Sobre os Dados

Imaginemos um evento que é, por natureza, contínuo no tempo. Se tomarmos amostras de uma das características desse evento para construir uma série histórica dessa característica, estamos discretizando algo que tem natureza contínua. Matematicamente, estamos discretizando uma função contínua. Em outras palavras estamos multiplicando a função original por uma função pulso ou trem de pulsos. Isso é o que chamamos aqui de efeito do janelamento sobre os dados, ou efeito da amostragem do sinal.

Nesta seção pretendemos discutir o efeito da amostragem do sinal retro-espalhados das irregularidades do EEJ porque esse efeito foi considerado no nosso ajuste e, portanto, merece comentários específicos. Especialmente, porque estamos ajustando um espectro de potência de uma função contínua no tempo que foi discretizada.

A seguir apresentamos a FIGURA 8 que exemplifica o efeito do janelamento sobre o conjunto de dados. Os dois gráficos superiores desta figura

apresentam a mesma curva Gaussiana sem considerar o efeito do janelamento do sinal. No gráfico à direita a escala vertical é logarítmica e a escala horizontal cobre todo a faixa de freqüência limitada pela análise espectral. No gráfico à esquerda a escala vertical é linear e a escala horizontal cobre somente a faixa de freqüência em torno do máximo da Gaussiana. Os dois gráficos abaixo apresentam a mesma Gaussiana dos gráficos superiores, porém neles foi adicionado o efeito do janelamento. A diferença entre os gráficos inferiores é a mesma apresentada para os gráficos superiores.

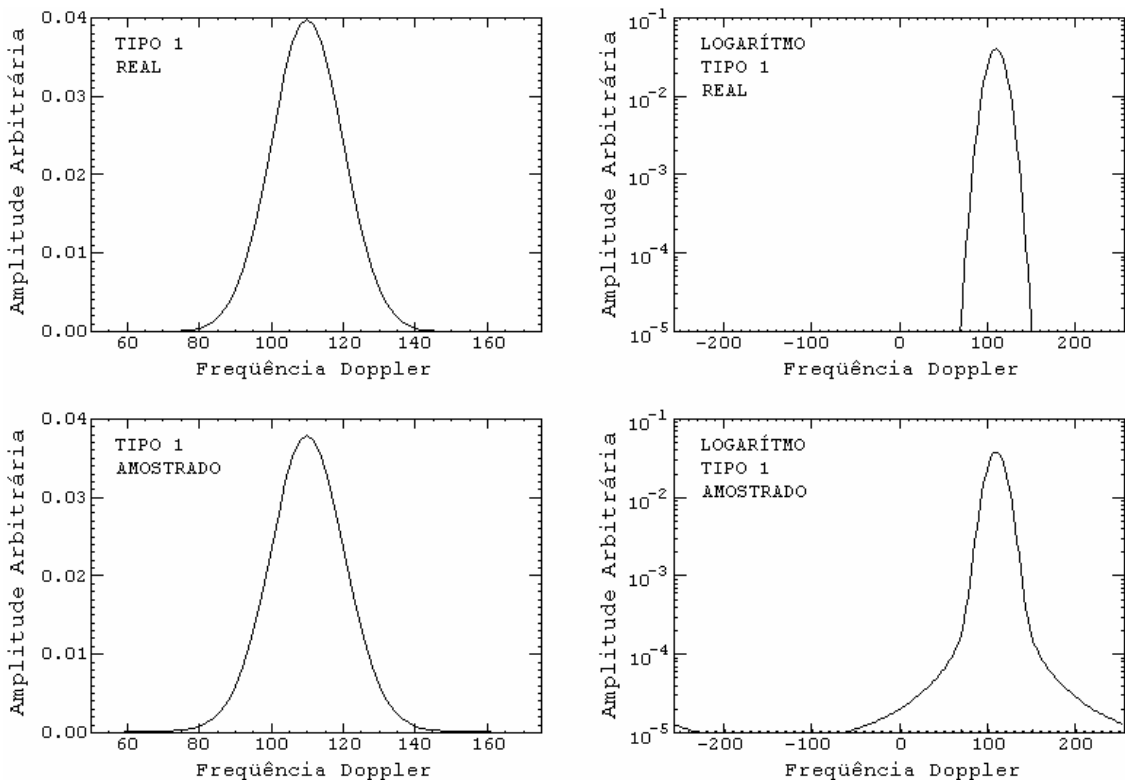


FIGURA 8 - Espectros de potência do tipo Farley-Buneman simulados (superior) sem incluir o efeito do janelamento amostral sobre o conjunto de dados e (inferior) incluindo este efeito.

Acreditamos que com esse conjunto de gráficos fica claro que o resultado obtido na construção de uma Gaussiana tomando uma série histórica discreta difere em alguns aspectos do resultado obtido pela aplicação da função Gaussiana contínua. Comparando os gráficos à direita é possível notar um alargamento na base da Gaussiana, ao passo que comparando os gráficos à esquerda nota-se uma diminuição da amplitude máxima da Gaussiana.

O motivo matemático dessas diferenças é, obviamente, o recém comentado efeito da multiplicação da janela unitária (função pulso) sobre o sinal recebido, ou simplesmente, efeito do janelamento do sinal recebido. Contudo, é importante mencionar que estes efeitos não são independentes, ou seja, a diminuição da amplitude máxima da Gaussiana está intimamente relacionada com o alargamento de sua base. O que ocorre é que o desvio padrão do conjunto de dados foi aumentado pelo efeito dessa multiplicação enquanto a área total da Gaussiana amostrada permaneceu inalterada comparada com a Gaussiana real.

Também é possível pensar em termos da convolução da função espectral da janela unitária pela Gaussiana real. Ou então, como preferimos tratar em nosso trabalho, é possível pensarmos em termos da multiplicação da convolução da janela unitária pela anti-transformada de Fourier da Gaussiana real que estamos tentando ajustar. Usamos a convolução da janela unitária porque a anti-transformada de Fourier da Gaussiana é a convolução do sinal recebido.

Assim, podemos dizer que a Gaussiana resultante do sinal amostrado é o resultado da multiplicação, no domínio do tempo, da anti-transformada da Gaussiana real pela convolução a janela unitária. Com isso, temos a relação que necessitamos para incluir no nosso modelo o efeito do janelamento sobre os dados.

6 Modelo de Ajuste dos Dados

Na FIGURA 9 está esquematizado o modelo de ajuste que utilizamos para obter os parâmetros dos espectros de potência do radar RESCO. Segundo este modelo, o sinal recebido pela antena é amostrado e sofre a transformada para o domínio da frequência. Desse espectro resultante são retirados os parâmetros iniciais que vão gerar a primeira Gaussiana ajustada. Em seguida esta função Gaussiana é transportada para o domínio do tempo e multiplicada pela função de correlação da janela unitária para incluir o efeito do janelamento. Na seqüência, o resultado dessa multiplicação é novamente transportado para o domínio da frequência e comparado com o espectro

original. Se os erros médios quadráticos forem satisfatórios o modelo então retorna os últimos parâmetros utilizados na geração da Gaussiana de ajuste. Caso contrário, os parâmetros sofrem modificações segundo uma função “objetivo” até que o ajuste obtenha a convergência ou que um número máximo de iterações seja atingido.

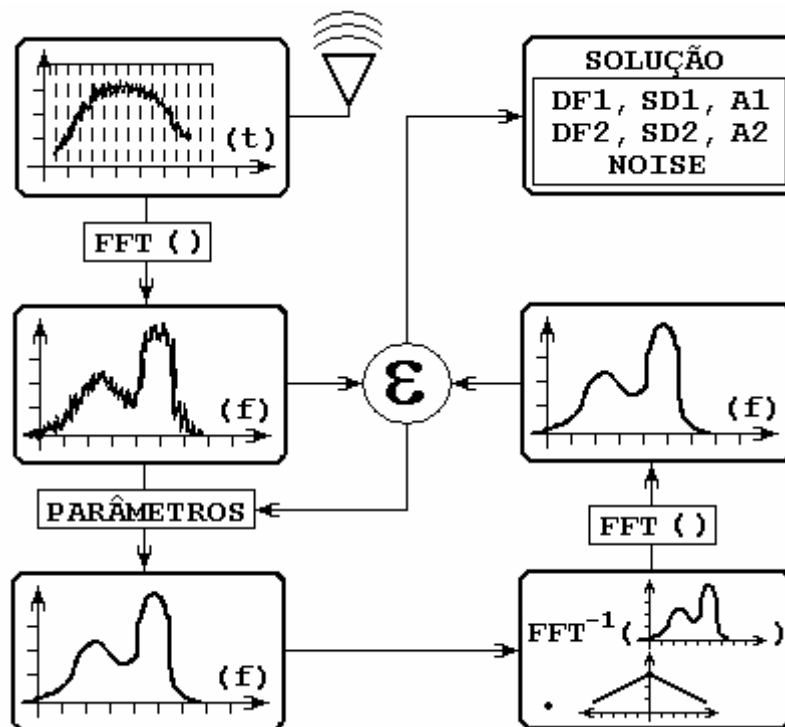


FIGURA 9 - Diagrama esquemático do modelo de ajuste por mínimos quadrados dos espectros de potência dos sinais retro-espalhados das irregularidades do EEJ.

O método pelo qual os parâmetros da Gaussiana de ajuste são obtidos é, na verdade, o cerne desse modelo. Ele envolve a função “objetivo” que minimiza o processo de escolha dos parâmetros e toda a matemática computacional necessária para isto. As demais etapas do modelo, apesar de necessárias, são simples testes da eficiência do método determinação dos parâmetros. A função principal das etapas periféricas é, em realidade, fornecer respostas e/ou estímulos que definem o quão bem o ajuste está sendo realizado. Com base nesses estímulos o método se ajusta e/ou se adapta para convergir para a resposta solicitada.

Nas seções que se seguem serão detalhados o método e a função “objetivo utilizados”. Para o momento, é importante mencionar que este método foi testado com dados simulados e reais. Em todos os casos ele apresentou excelentes resultados.

7 Média, Mediana ou Integração dos Dados

Em alguns casos específicos, quando o espectro se mostrou muito ruidoso, foi necessário utilizarmos artifícios matemáticos para melhorar o desempenho do algoritmo. Entre eles o que apresentou o melhor resultado foi o método de suavização do espectro de potência. Foram testados dois métodos de suavização: (a) a diminuição do número de pontos através de integração incoerente em frequência, reduzindo a resolução espectral; e (b) a aplicação da média e/ou da mediana corrida, com o manutenção do número de pontos e, por consequência, da resolução espectral.

O método de integração em frequência não foi exaustivamente testado porque, em geral, os métodos de média e/ou mediana corrida apresentaram excelente resposta ao problema. Porém, cabe uma pequena discussão entre estes dois últimos.

Ambos os métodos de média e de mediana corrida apresentaram uma muito boa resposta na redução das incertezas do algoritmo, fazendo com que esse convergisse para a resposta na maioria dos casos. Além disso, ambas respostas sempre foram muito aproximadas. Porém, acreditamos que a utilização da mediana corrida seja mais representativa na maioria dos casos. Isso se deve ao fato de, em alguns poucos espectros, termos observado alguns picos espúrios que, no caso da média corrida, influenciaram a forma final da Gaussiana.

Veja, por exemplo o caso apresentado na FIGURA 10. Para este espectro de potência específico, na região de principal interesse (de -150 a 0 Hz), a curva de suavização por média corrida de 5 pontos ficou quase sempre acima da curva de mediana corrida de 5 pontos. Possivelmente, essa média foi

influenciada pelo número excessivo de picos agudos desse espectro. E, aparentemente, neste caso a mediana é mais apropriada. Entretanto, o objetivo deste trabalho não é investigar qual o melhor tipo de suavização dos dados. Mesmo porque o número de casos em que foi necessário decidir entre um método de suavização ou outro foi insignificante, comparado com o número de casos onde o algoritmo resolveu o ajuste sem necessidade de nenhum alisamento.

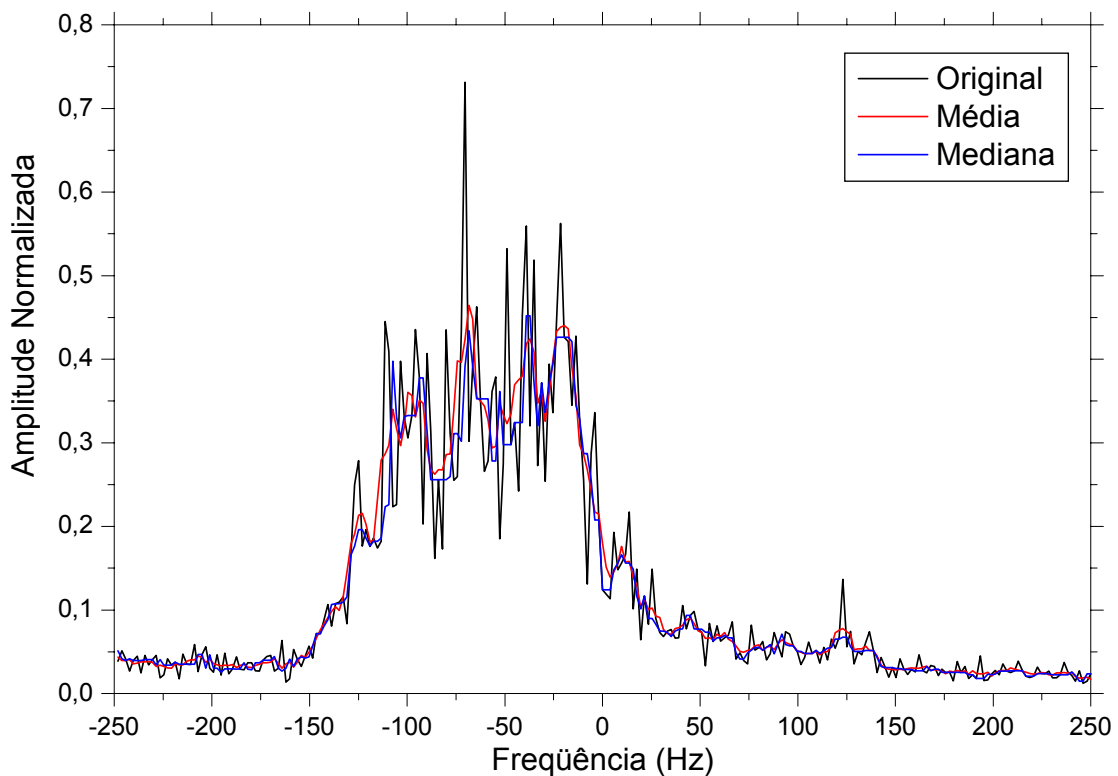


FIGURA 10 - Espectro de potência ruidoso, obtido dos dados do radar RESCO em 1999. Sobrepostos estão os espectros suavizados pelos métodos de (vermelha) média corrida de 5 pontos e de (azul) mediana corrida de 5 pontos.

A escolha pela mediana na maioria dos casos não é conclusiva ou definitiva. A escolha deve ser tomada analisando cada caso em separado. Essa atitude deve ser tomada porque não podemos afirmar com absoluta certeza que os picos dentro do espectro são realmente espúrios ou fazem parte do sinal retro-espalhado. No caso em que os picos façam parte do sinal a média corrida seria mais significativa, caso contrário deveríamos optar pela mediana corrida. A única conclusão possível é que a decisão pelo melhor método de suavização

deve ser tomada com base no conhecimento do sistema de radar com o qual os dados foram coletados.

8 Método de Ajuste de Espectro

Um modelo é um sumário do conjunto de dados que queremos representar. Em muitos casos o modelo é simplesmente uma classe de funções, como um conjunto de polinômios ou uma Gaussiana, que se ajusta ao conjunto de dados com os coeficientes ou parâmetros apropriados. Em outros casos o modelo está baseado em uma teoria que determina o comportamento básico do evento, o qual limita o conjunto de respostas. Enfim, existem muitas variações de modelagem para um conjunto de dados, como por exemplo, interpolações e extrapolações a partir de um conjunto conhecido de respostas, entre outras técnicas (Press et al., 1992). Porém, em todos os casos a abordagem é sempre a mesma. Uma função “objetivo” mede a concordância entre o conjunto de dados e o modelo escolhido baseado na escolha de um particular conjunto de parâmetros que determinam e/ou modificam as características do modelo. Essa função “objetivo” é escolhida de modo que o seu máximo resultado represente a melhor aproximação entre o modelo e os dados. Por fim, o conjunto de parâmetros é ajustado para que a função “objetivo” apresente como resposta seu máximo valor.

Se o conjunto de dados foi obtido de forma aleatória uma técnica usual de ajuste é a estimativa da máxima probabilidade (MLE, do inglês *Maximum Likelihood Estimate*) (Lim e Oppenheim, 1988). A estimativa de máxima probabilidade dos parâmetros de um modelo consiste em escolher o conjunto de parâmetros \mathbf{a} ($a_1, a_2, a_3, \dots, a_k$) que maximizam todas as probabilidades individuais P_i ($P\{y_1 \dots y_n\} | \mathbf{a}, P\{y_1 \dots y_n\} | \mathbf{a}, \dots, P\{y_1 \dots y_n\} | \mathbf{a}$) de obtermos todos os valores y_i ($y_1 \dots y_n$).

Se o conjunto de pontos y_i possui uma distribuição normal em torno de um valor médio, como no caso de grandes somatórios (Press et al., 1992), cada ponto tem associado um erro puramente aleatório em torno de seu valor “verdadeiro”. Neste caso as variâncias σ_i^2 de todos os pontos são idênticas

entre si ($\sigma_1 = \sigma_2 = \dots \sigma_n = \sigma$) e a distribuição é chamada Gaussiana. Porém se cada ponto y_i possui uma variância σ_i^2 independente das demais, então a distribuição é chamada Chi-Quadrada e um certo número de variáveis independentes deve ser considerado. Neste último caso, as variâncias independentes σ_i^2 são obtidas pela repetição no número de medidas ao obter-se o conjunto de dados y_i .

Finalmente, a probabilidade de ajustarmos o ponto y_i , que possui uma variância associada σ_i^2 , com o modelo que resulta um valor $y(x_i, \mathbf{a})$ quando utiliza o conjunto de parâmetros \mathbf{a} é dada por (Bevington e Robinson, 1992):

$$P_i = \frac{1}{\sigma_i \cdot \sqrt{2\pi}} \exp\left[-\frac{(y_i - y(x_i, \mathbf{a}))^2}{2 \cdot \sigma_i^2}\right] \quad (6)$$

Portanto, a probabilidade do modelo ajustar todos os pontos y_i é dada pelo produto das probabilidades individuais de cada ponto.

$$P = \prod_{i=1}^n P_i = \frac{1}{(2\pi)^{\frac{n}{2}} \cdot \prod_{i=1}^n \sigma_i} \exp\left[-\frac{1}{2} \cdot \sum_{i=1}^n \frac{(y_i - y(x_i, \mathbf{a}))^2}{\sigma_i^2}\right] \quad (7)$$

Mas estamos procurando os parâmetros \mathbf{a} , para o qual temos a máxima probabilidade do modelo ajustar todos os pontos y_i com respeito a esse conjunto de parâmetros. Assim, torna-se evidente que procuramos a solução para a qual a derivada do produto das probabilidades P em relação ao conjunto \mathbf{a} de parâmetros é igual a zero.

Normalmente ao invés de minimizar a equação (7), o método mais usual é minimizar somente o argumento de sua exponencial. Isso é possível porque, ao se minimizar esse argumento, estamos tornando o produto das probabilidades P próximo da unidade, o que também significa maximizar o somatório das probabilidades.

$$\frac{\partial P}{\partial \mathbf{a}} = 0 \Leftrightarrow \frac{\partial}{\partial \mathbf{a}} \left[-\frac{1}{2} \cdot \sum_{i=1}^n \frac{(y_i - y(x_i, \mathbf{a}))^2}{\sigma_i^2} \right] = 0 \quad (8)$$

Por fim, o problema da aplicação do método MLE se reduziu a um problema de minimização dos erros médios quadráticos.

$$\frac{\partial}{\partial \mathbf{a}} \sum_{i=1}^n \frac{(y_i - y(x_i, \mathbf{a}))^2}{\sigma_i^2} = 0 \quad (9)$$

Esse método tem se mostrado bastante eficiente mesmo quando o conjunto de dados não é normalmente distribuído. Porém, nestes casos a estimativa dos parâmetros \mathbf{a} , necessariamente, não convergirá para a resposta ótima com respeito à maximização das probabilidades.

9 Gradiente e Linearização da Função “Objetivo”

Quando a função de ajuste é uma função linear é possível encontrar o conjunto dos parâmetros \mathbf{a} que minimizam os erros médios quadráticos em um único passo. Entretanto, quando a relação é não-linear, o que ocorre na maioria dos casos, incluindo o nosso, deve-se procurar iterativamente pelo mínimo no espaço de parâmetros \mathbf{a} . A busca por esse mínimo requer um ponto de partida e uma direção para o primeiro passo. As escolhas do valor inicial e do tamanho do passo é o que, na verdade, diferenciam os métodos de busca de soluções em problemas não-lineares (Bard, 1974).

No caso da função “objetivo” ser a função dos erros médios quadráticos, apresentada na Equação (10), pode-se optar por um método que direcione os passos no sentido negativo do gradiente da função.

$$\chi^2 \equiv \sum_{i=1}^n \frac{(y_i - y(x_i, \mathbf{a}))^2}{\sigma_i^2} \quad (10)$$

O método do gradiente consiste em ordenarmos os passos iterativos do processo sempre no sentido da máxima variação da função “objetivo”, até atingirmos seu mínimo.

$$\nabla\chi^2 = \frac{\partial}{\partial\mathbf{a}} \sum_{i=1}^n \frac{(y_i - y(x_i, \mathbf{a}))^2}{\sigma_i^2} \quad (11)$$

Entretanto, sua utilização não é recomendada na região próxima do mínimo da função. Pois a determinação do gradiente que define o sentido e a dimensão do passo, torna-se bastante imprecisa neste caso. Porém esse método torna-se efetivo nas regiões próximas do mínimo quando combinado com outras técnicas. Uma das técnicas é aproximar a função “objetivo” por uma função linearizada equivalente.

Segundo Bard (1974), a superfície que melhor aproxima a superfície χ^2 em torno de um mínimo local \mathbf{a}_{j+1} é a quadrática obtida da expansão em série de Taylor da função “objetivo”.

$$\chi^2(\mathbf{a}_{j+1}) \approx \chi_j^2 + \nabla\chi_j^2 \cdot \delta\mathbf{a} + \frac{1}{2} \cdot \delta\mathbf{a} \cdot \mathbf{H}_j \cdot \delta\mathbf{a} \quad (12)$$

Na Equação (12), as grandezas com índice j são os valores obtidos com os parâmetros correntes, $\delta\mathbf{a}$ representa a variação nos parâmetros \mathbf{a} , dada por:

$$\delta\mathbf{a} = \mathbf{a}_{j+1} - \mathbf{a}_j \quad (13)$$

e \mathbf{H} é a matriz contendo as segundas derivadas parciais da função “objetivo”, também conhecida como matriz Hessiana. Esta matriz é definida da forma:

$$\mathbf{H} = \frac{\partial^2 \chi^2}{\partial \mathbf{a}_l \partial \mathbf{a}_m} . \quad (14)$$

Com a equação expandida podemos facilmente aplicar o método do gradiente. Assim, igualando a zero a derivada da Equação (12) em relação aos parâmetros \mathbf{a} , obtemos:

$$\nabla \chi_j^2 + \mathbf{H}_j \cdot \delta \mathbf{a} = 0 \quad (15)$$

A mudança nos parâmetros $\delta \mathbf{a}$ pode então ser obtida com uma iteração, se a matriz Hessiana for não-singular. Para isso, tomamos a inversa da matriz Hessiana multiplicada pelo gradiente da função “objetivo”.

$$\delta \mathbf{a} = -\mathbf{H}_j^{-1} \cdot \nabla \chi_j^2 \quad (16)$$

Com isso, o problema não linear de obter os parâmetros \mathbf{a} , torna-se um problema linear de calcular a variação $\delta \mathbf{a}$ destes mesmos parâmetros. Mas isso ocorre às custas de introduzirmos o cálculo de uma matriz de segundas derivadas da função “objetivo”. Se esta função for relativamente simples e o número de parâmetros reduzido, então não teremos causado maiores complicações. Contudo, se a função não for simples ou se o conjunto de parâmetros forem inter-relacionados podemos ter muita dificuldade no cálculo desta matriz.

Procedemos então à explicitação dos elementos matemáticos necessários para o cálculo de $\delta \mathbf{a}$. Iniciamos com o gradiente da função “objetivo” apresentado na Equação (11).

$$\nabla \chi_j^2 = \frac{\partial \chi_j^2}{\partial \mathbf{a}_1} = -2 \cdot \sum_{i=1}^n \frac{(y_i - y(x_i, \mathbf{a}_j))}{\sigma_i^2} \cdot \frac{\partial y(x_i, \mathbf{a}_j)}{\partial \mathbf{a}_1} \quad (17)$$

A matriz Hessiana, definida na Equação (14), é explicitada na forma da Equação (18).

$$\mathbf{H}_j = \frac{\partial^2 \chi_j^2}{\partial \mathbf{a}_1 \partial \mathbf{a}_m} = 2 \cdot \sum_{i=1}^n \frac{1}{\sigma_i^2} \cdot \frac{\partial y(x_i, \mathbf{a}_j)}{\partial \mathbf{a}_1} \cdot \frac{\partial y(x_i, \mathbf{a}_j)}{\partial \mathbf{a}_m} - 2 \cdot \sum_{i=1}^n \frac{(y_i - y(x_i, \mathbf{a}_j))}{\sigma_i^2} \cdot \frac{\partial^2 y(x_i, \mathbf{a}_j)}{\partial \mathbf{a}_1 \partial \mathbf{a}_m} \quad (18)$$

Na prática, podemos aproximar a matriz Hessiana por:

$$\mathbf{H}_j = 2 \cdot \sum_{i=1}^n \frac{1}{\sigma_i^2} \cdot \frac{\partial y(x_i, \mathbf{a}_j)}{\partial \mathbf{a}_1} \cdot \frac{\partial y(x_i, \mathbf{a}_j)}{\partial \mathbf{a}_m} \quad (19)$$

O segundo termo do lado direito da Equação (18) contém termos de derivadas segundas dos dados do modelo com relação aos parâmetros \mathbf{a} . Com isso, este termo é muito pequeno comparado com o primeiro, o qual envolve as derivadas primeiras (Press et al., 1992). Além disso, o segundo termo contém o fator $(y_i - y(x_i, \mathbf{a}_j))$, o qual tende a zero no limite de $y(x_i, \mathbf{a}_j)$ tendendo a y_i para um modelo de sucesso.

Para efeitos de computação é convencional remover o fator 2 das matrizes Hessiana \mathbf{H} e derivada do gradiente $\nabla \chi^2$. Para isso, é necessário organizar os fatores da Equação (16) em função de duas novas matrizes correspondentes, denominadas matriz de curvatura α e matriz β . A prova matemática do desaparecimento do fator 2 pela utilização destas novas matrizes é apresentada no Apêndice C. A definição destas novas matrizes é apresentada na forma da Equação (20) a seguir:

$$\begin{aligned}
\boldsymbol{\alpha} &\equiv \frac{\mathbf{H}}{2} = \frac{1}{2} \cdot \frac{\partial^2 \chi^2}{\partial \mathbf{a}_l \partial \mathbf{a}_m} = \sum_{i=1}^n \frac{1}{\sigma_i^2} \cdot \frac{\partial y(x_i, \mathbf{a})}{\partial \mathbf{a}_l} \cdot \frac{\partial y(x_i, \mathbf{a})}{\partial \mathbf{a}_m} \\
\boldsymbol{\beta} &\equiv -\frac{\nabla \chi^2}{2} = -\frac{1}{2} \cdot \frac{\partial \chi^2}{\partial \mathbf{a}_l} = \sum_{i=1}^n \frac{(y_i - y(x_i, \mathbf{a}_j))}{\sigma_i^2} \cdot \frac{\partial y(x_i, \mathbf{a}_j)}{\partial \mathbf{a}_l} .
\end{aligned} \tag{20}$$

Usando agora as novas matrizes $\boldsymbol{\alpha}$ e $\boldsymbol{\beta}$, e a inter-relação entre a Equação (16) e a Equação (13), obtemos:

$$\mathbf{a}_{j+1} = \mathbf{a}_j + \boldsymbol{\alpha}_j^{-1} \cdot \boldsymbol{\beta}_j . \tag{21}$$

Finalmente, temos condição de proceder a um cálculo iterativo até reduzirmos o $\delta \mathbf{a}$ a valores considerados satisfatórios. Outra alternativa é verificar se a variação na função “objetivo” não é significativa em passos subseqüentes.

Ao final dessa seção é importante lembrarmos que os métodos de expansão funcionam muito bem quando a busca pelo conjunto de parâmetros que melhor ajustem a função aos dados ocorre dentro da região válida de aproximação parabólica (Bevington e Robinson, 1992). Quando o ponto de partida está longe desta região estes métodos tendem a falhar. Uma maneira de contornar esse problema foi apresentada por Marquardt (1963). Ele sugeriu a introdução de uma variação no tamanho do passo do gradiente a fim de torná-lo maior fora da região válida, facilitando a convergência do método.

10 Modificação de Levenberg-Marquardt

A essência do método de Marquardt está em interferir nos valores da diagonal principal da matriz de curvatura $\boldsymbol{\alpha}$. Esses elementos são modificados pela sua multiplicação pelo fator $(1+\lambda)$. Dessa forma, quando o algoritmo está buscando soluções fora da área de validade da aproximação parabólica, ou seja, longe da melhor solução, o termo λ deve apresentar um valor elevado tornando a matriz $\boldsymbol{\alpha}$ predominantemente diagonal. Isso faz com que o passo de busca dos parâmetros \mathbf{a} seja aumentado. Ao contrário, se o algoritmo aproxima-se da

melhor solução, o termo λ deve apresentar um valor baixo de modo a deixar o método do gradiente atuar eficientemente.

Portanto, a matriz α é substituída pela matriz α modificada (α') da forma:

$$\alpha' = \begin{cases} \alpha_{ij} \cdot (1 + \lambda), & i = j \\ \alpha_{ij} & , i \neq j \end{cases} . \quad (22)$$

No algoritmo modificado de Marquardt (1963) ele sugere que λ tenha um valor inicial relativamente baixo, aumentando ou diminuindo em passos sucessivos, dependendo da variação observada no valor da função “objetivo”. Esse método ficou conhecido e é referenciado hoje como método Levenberg-Marquardt. Isso se deu porque foi Marquardt (1963) quem apontou para um fator multiplicador $(1+\lambda)$ na diagonal principal da matriz de curvatura α como fator acelerador da convergência do problema. Entretanto, foi Levenberg (1944) quem primeiro foi levado a sugerir a adição de tal quantidade à diagonal da matriz α . O trabalho desenvolvido por Curry (1944) também teve influência sobre o trabalho de Marquardt (1963) uma vez que trata da validade da utilização do método do gradiente em métodos de minimização usando mínimos quadrados.

11 Restrições aos Parâmetros

Apesar de bastante eficiente nosso método não é infalível. Como todos métodos computacionais ele apresenta algumas restrições, às quais podem provocar a não convergência do modelo, ou a convergência para um mínimo local nos casos de problemas não lineares. Quando os parâmetros que descrevem o modelo têm um significado físico ou quando conhecemos o conjunto possível de respostas é possível restringir a busca de soluções no espaço de parâmetros \mathbf{a} e, com isso, aumentar as chances do sucesso nas nossas buscas.

Essa restrição pode ser feita não aceitando que os parâmetros cruzarem as fronteiras permitidas e solicitando uma nova busca com outros valores iniciais.

Apesar de eficaz, essa não é a solução eficiente e nem a mais elegante. Pelo contrário, é uma solução robusta. Ela aumenta o tempo de processamento tantas vezes quantas forem as tentativas fracassadas. Uma solução elegante é mais eficiente foi apresentada por Bard (1974). Ele atribuiu penalidades à função “objetivo” a medida que os parâmetros se aproximam do limite das regiões proibidas. Essas penalidades são elaboradas de modo a não surtir efeito sobre a busca se os parâmetros estiverem dentro do domínio aceitável, mas devem aumentar rapidamente seu efeito se os parâmetros estiverem convergindo para o limite imposto.

As penalidades podem ser construídas usando equações ou inequações de restrições da forma apresentada a seguir:

$$h(\mathbf{a}) \geq 0 \quad , \quad (23)$$

na qual $h(\mathbf{a})$ é uma função dos parâmetros que é positiva na região permitida, mas que decresce rapidamente a medida que a busca se aproxima da região proibida e é zero exatamente na borda entre as regiões. Por exemplo, para restringir o k -ésimo parâmetro a_k acima de um limite inferior b_1 podemos usar a inequação:

$$h_j(a_k) \equiv a_k - b_1 \geq 0 \quad . \quad (24)$$

Da mesma maneira, para restringir o k -ésimo parâmetro a_k abaixo de um limite superior b_2 podemos usar a inequação:

$$h_j(a_k) \equiv b_2 - a_k \geq 0 \quad . \quad (25)$$

Por conseqüente, a restrição do parâmetro a_k dentro dos limites inferior e superior ($b_1 < a_k < b_2$) é determinada pela combinação das inequações acima.

Uma das maneiras de criar funções de penalidades ζ com base nesse tipo de restrições é usando pesos recíprocos às funções $h(\mathbf{a})$ da forma:

$$\zeta_j = \frac{\alpha_j}{h_j(\mathbf{a})} \quad (26)$$

Estas funções ζ devem ser adicionadas à função “objetivo” χ^2 para formar a função “objetivo” modificada:

$$\chi'^2 = \sum_{i=1}^n \frac{(y_i - y(x_i, \mathbf{a}))^2}{\sigma_i^2} + \sum_j \frac{\alpha_j}{h_j(\mathbf{a})} \quad (27)$$

a qual cresce rapidamente à medida que qualquer dos parâmetros se aproxima das restrições impostas, mas é dominada pela função original dos erros médio quadráticos dentro do domínio permitido.

Nosso problema agora é minimizar a função “objetivo” modificada. Isto é feito da mesma forma que anteriormente. Porém, alguns cálculos adicionais devem ser realizados a cada iteração. É necessário computar as derivadas primeira e segunda da função “objetivo” modificada e das funções de penalidades, além das próprias funções de penalidades. Como já temos as derivadas primeira e segunda da função “objetivo” necessitamos somente as derivadas das funções de penalidades que foram adicionadas à função “objetivo”.

$$\begin{aligned} \frac{\partial \zeta_j}{\partial \mathbf{a}} &= \frac{\partial}{\partial \mathbf{a}} \sum_j \frac{\alpha_j}{h_j(\mathbf{a})} = - \sum_j \frac{\alpha_j}{h_j^2(\mathbf{a})} \cdot \frac{\partial h_j(\mathbf{a})}{\partial \mathbf{a}} \\ \frac{\partial^2 \zeta_j}{\partial \mathbf{a}_1 \partial \mathbf{a}_m} &= \frac{\partial^2}{\partial \mathbf{a}_1 \partial \mathbf{a}_m} \sum_j \frac{\alpha_j}{h_j(\mathbf{a})} = \sum_j \frac{\alpha_j}{h_j^3(\mathbf{a})} \cdot \left[2 \cdot \frac{\partial h_j(\mathbf{a})}{\partial \mathbf{a}_1} \cdot \frac{\partial h_j(\mathbf{a})}{\partial \mathbf{a}_m} - \right. \\ &\quad \left. - h_j(\mathbf{a}) \cdot \frac{\partial^2 h_j(\mathbf{a})}{\partial \mathbf{a}_1 \partial \mathbf{a}_m} \right] \end{aligned} \quad (28)$$

Como pode ser visto, estas duas equações são análogas às equações do gradiente (Equação (17)) e da matriz Hessiana (Equação (18)). Note que estas derivadas devem ser muito pequenas dentro da região permitida dos parâmetros \mathbf{a} , uma vez que ambas têm, pelo menos, uma relação direta com fator $\alpha / h^2(\mathbf{a})$. Portanto, essas derivadas somente apresentam influencia no resultado nas proximidades dos limites impostos aos parâmetros. Em adição, o segundo termo dentro dos colchetes do lado direito da derivada segunda da função de penalidades tem uma dependência da segunda derivada da função de restrições dos parâmetros $h(\mathbf{a})$. Similarmente à Equação (18), esse termo deve ser muito pequeno a medida que nos aproximamos da solução que melhor ajusta o modelo. Conseqüentemente, esse termo pode ser ignorado (Bard, 1974) como foi feito na computação da matriz Hessiana.

Em resumo, o que modificamos no método original dos mínimos quadrados com Levenberg-Marquardt foram os cálculos da função “objetivo” e das matrizes α e β . Além da introdução do cálculo da inequações e de um laço externo para reduzir o peso α_j da função penalidade a medida que o algoritmo converge para a melhor solução.

Os algoritmos com as modificações no cálculo da matriz α , da matriz β e da função objetivo χ^2 , e também as modificações no laço do algoritmo para incluir o cálculo de α_j não serão apresentados neste trabalho porque envolvem uma grande quantidade de linhas de código. Porém a rotina que desenvolvemos para restrições nos parâmetros é apresentada no Apêndice D. Esta rotina, aparentemente simples, tem como principal característica grande versatilidade e abrangência. Através dela é possível de se introduzir restrições com limites superiores e inferiores em todos os parâmetros do modelo bastando apenas escolher o α_j apropriado.

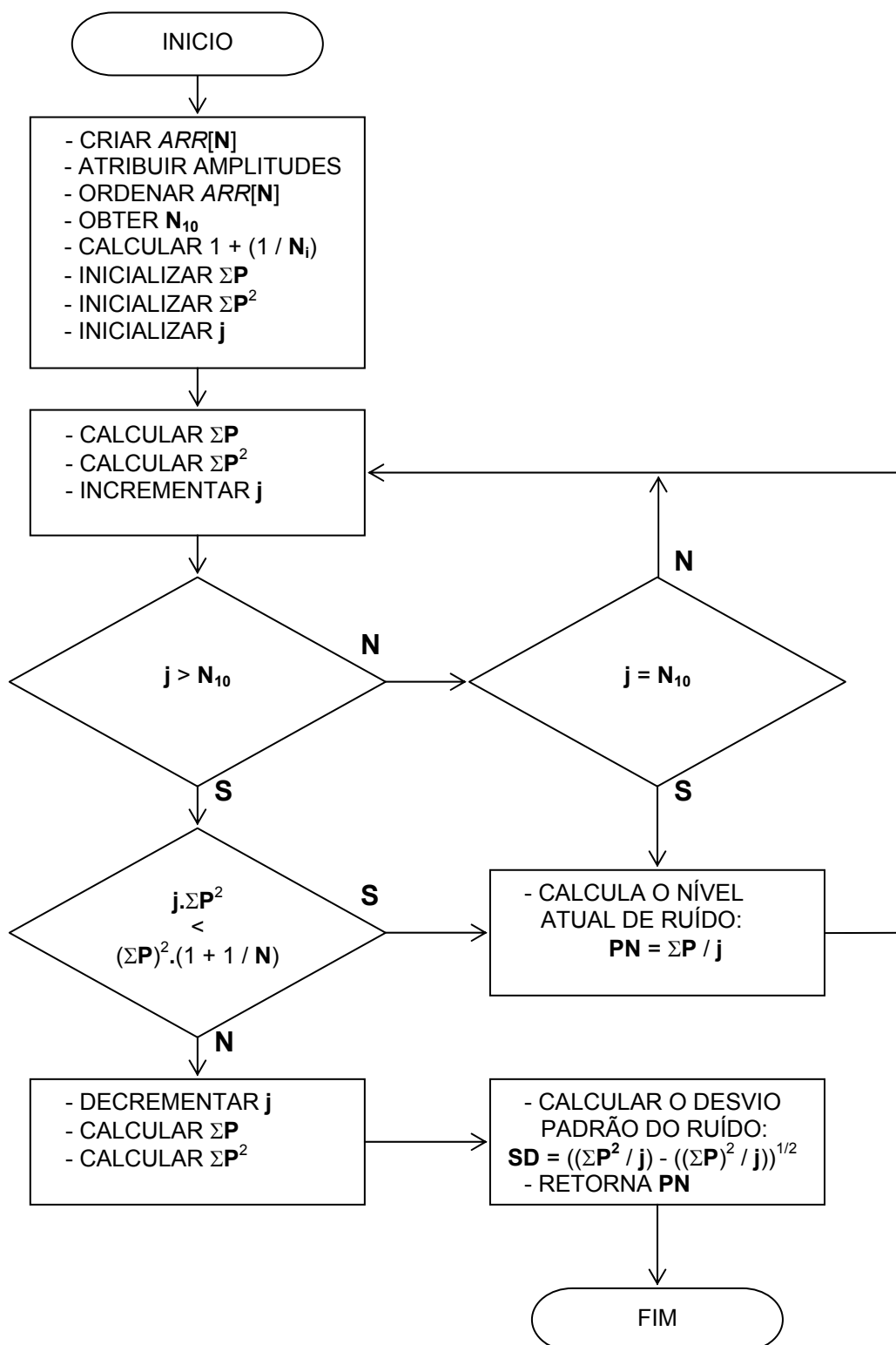
REFERÊNCIAS BIBLIOGRÁFICAS

- Balsley, B. B., Gage, K. S. *The MST radar technique: potential for middle atmosphere studies*. **Pure and Applied Geophysics**, v. 118, p. 452-493, 1980.
- Bard, Y. ***Non linear parameter estimation***. Academic Press, 1974.
- Bevington, P. R., Robinson, D. K. ***Data reduction and error analysis for the physical science***. McGraw-Hill, 1992.
- Chau, J. L., Farley, D. T., Balsley, B. B., *East-West asymmetry in type 1 electrojet echoes at Jicamarca*. In: The Tenth International Symposium on Equatorial Aeronomy 10, Antalya (Turkey), **Abstract Notebook**, May 2000.
- Cohen, R. *Phase velocities of irregularities in the equatorial electrojet*. **Journal of Geophysical Research**, v. 78, n. A13, p. 2222-2231, May 1973.
- Curry, H. B. *The method of steepest descent for non-linear minimization problem*. **Quarterly Applied Mathematics**, v. 2, n.1, p. 258-261, Apr. 1944.
- Fukao, S. **Middle atmosphere program – Handbook for map: International school on atmospheric radar**, v. 30, Urbana (IL): SCOSTEP Secretariat, Oct. 1989.
- Gage, K. S., Balsley, B. B. *Doppler radar probing of the clear atmosphere*. **Bulletin of the American Meteorological Society**, v. 59, n. 9, p. 1074-1093, Sep. 1978.
- Hysell, D. L., Burcham, J. D. *Ionospheric electric field estimates from radar observations of the equatorial electrojet*. **Journal of Geophysical Research**, v. 105, n. A2, p. 2443-2460, Feb. 2000.
- Jenkins, G. M., Watts, D. G. ***Spectral analysis and its applications***. San Francisco (CA): Holden-Day, 1968.

- Kudeki, E., Bhattacharyya, S., Woodman, R. F. *A new approach in incoherent scatter F region E x B drift measurements at Jicamarca.* **Journal of Geophysical Research**, v. 104, n. A12, p. 28145-28162, Dec. 1999.
- Levenberg, K. *A method for the solution of certain non-linear problem in least square.* **Quarterly Applied Mathematics**, v. 2, n.1, p. 164-168, Apr. 1944.
- Lim, J. S., Oppenheim, A. V. ***Advanced topics in signal processing.*** Prentice Hall, 1988.
- Marquardt, D. W. *An algorithm for last square estimation of non linear parameters.* **Journal of the Society for Industrial and Applied Mathematics**, v. 2, n. 2, p. 431-441, 1963.
- Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. **Numerical recipes in C: the art of scientific computing.** 2. ed. Cambridge: Cambridge University Press, 1992.
- Thomson, D. ***Física Computacional.*** Apostila (Notas de Aulas) - Pontificia Universidad Católica del Perú, Lima, 296p.1998.

APÊNDICE A

FLUXOGRAMA DA ROTINA DE DETERMINAÇÃO DO NÍVEL DE POTÊNCIA ESPECTRAL DE RUÍDO



APÊNDICE B

ALGORITMO DA ROTINA DE DETERMINAÇÃO DO NÍVEL DE POTÊNCIA ESPECTRAL DE RUÍDO

- 1) Determinar o número de pontos **N** do espectro a ser processado;
- 2) Criar um vetor com N pontos para ser manipulado nesta função;
- 3) Atribuir os valores das amplitudes ao vetor e ordenar seus valores em ordem crescente;
- 4) Determinar um número mínimo **N₁₀** de 10% do total de pontos **N** que devem ser utilizados na primeira avaliação;
- 5) Calcular o valor constante $(1 + 1 / N_i)$, onde **N_i** é o número de integrações incoerentes que foram realizadas para obter o espectro;
- 6) Preparar as variáveis acumuladoras, a variável índice dos dados **j** e o *flag* de parada;
- 7) Enquanto o número de pontos **j** for menor que o número mínimo **N₁₀** estipulado em 4 e o *flag* de parada não for sinalizado, permanecer dentro do laço de repetição:
 - 7.1) Calcular o somatório de todas amplitudes ΣP , o somatório do quadrado das amplitudes ΣP^2 e incrementar o índice dos dados **j**;
 - 7.2) Verificar se o índice dos pontos **j** é maior que o número mínimo de pontos iniciais **N₁₀**;
 - 7.2.1) Caso afirmativo verificar se a inequação $j \cdot \Sigma P^2 < (\Sigma P)^2 \cdot (1 + 1 / N)$ é verdadeira usando o número atual de pontos;
 - 7.2.1.1) Caso afirmativo ainda está na região de ruído, portanto:
 - calcular o nível atual de ruído $P_N = \Sigma P / j$ e retornar a 7.1;
 - 7.2.1.2) Caso negativo, foi adicionado o primeiro ponto de sinal, portanto:
 - decrementar o índice dos dados **j**;
 - recalcular os somatórios ΣP e ΣP^2 ; e
 - sinalizar o *flag* de parada para sair do laço de repetição;
 - 7.2.2) Caso negativo verificar se o número de pontos **j** é igual ao número mínimo de pontos iniciais **N₁₀** calculados em 4;
 - 7.2.2.1) Caso afirmativo calcular o nível de ruído $P_N = \Sigma P / j$ e retornar a 7.1;
- 8) Calcular o desvio padrão dos dados da região de ruído $SD = ((\Sigma P^2 / j) - ((\Sigma P)^2 / j))^{1/2}$
- 9) Retornar o último nível de ruído calculado em 7.2.1.1. ou em 7.2.2.1.

APÊNDICE C

PROVA MATEMÁTICA DA ELIMINAÇÃO DO FATOR 2 NA

TRANSFORMAÇÃO DA EQUAÇÃO (16) PARA (21)

A seguir é apresentada a prova matemática de como é eliminado o fator 2 na transformação da Equação (16) para a Equação (21) pela utilização das matrizes α e β definidas na Equação (20).

Seja uma matriz $\mathbf{H}_{N \times N}$ tal que:

$$\mathbf{H} = \lambda \cdot \alpha \quad .$$

Se o determinante de α ($\det \alpha$) é dado por:

$$\det \alpha = \varepsilon_{abc\dots p} \cdot \alpha_{1a} \cdot \alpha_{2b} \cdot \alpha_{3c} \dots \alpha_{Np} \quad ,$$

onde ε é o fator de anti-simétrico de Levi-Civita, então o determinante de \mathbf{H} é:

$$\begin{aligned} \det \mathbf{H} &= \varepsilon_{abc\dots p} \cdot H_{1a} \cdot H_{2b} \cdot H_{3c} \dots H_{Np} \\ \det \mathbf{H} &= \varepsilon_{abc\dots p} \cdot (\lambda \cdot \alpha_{1a}) \cdot (\lambda \cdot \alpha_{2b}) \cdot (\lambda \cdot \alpha_{3c}) \dots (\lambda \cdot \alpha_{Np}) \\ \det \mathbf{H} &= \varepsilon_{abc\dots p} \cdot \lambda^N \cdot \alpha_{1a} \cdot \alpha_{2b} \cdot \alpha_{3c} \dots \alpha_{Np} \\ \det \mathbf{H} &= \lambda^N \cdot \det \alpha \end{aligned} \quad .$$

Agora, o cofator do elemento $J \times j$ da matriz α ($\text{cof } \alpha_{Jj}$) é determinado por:

$$\text{cof } \alpha_{Jj} = (-1)^{J+j} \cdot \varepsilon_{abc\dots ik\dots p} \cdot \alpha_{1a} \cdot \alpha_{2b} \cdot \alpha_{3c} \dots \alpha_{li} \cdot \alpha_{kk} \dots \alpha_{Np} \quad .$$

Assim o cofator do mesmo elemento $J \times j$ da matriz \mathbf{H} fica:

$$\begin{aligned} \text{cof}(H_{Jj}) &= (-1)^{J+j} \cdot \varepsilon_{abc\dots ik\dots p} \cdot H_{1a} \cdot H_{2b} \cdot H_{3c} \dots H_{li} \cdot H_{kk} \dots H_{Np} \\ \text{cof}(H_{Jj}) &= (-1)^{J+j} \cdot \varepsilon_{abc\dots ik\dots p} \cdot (\lambda \cdot \alpha_{1a}) \cdot (\lambda \cdot \alpha_{2b}) \cdot (\lambda \cdot \alpha_{3c}) \dots (\lambda \cdot \alpha_{li}) \cdot (\lambda \cdot \alpha_{kk}) \dots (\lambda \cdot \alpha_{Np}) \\ \text{cof}(H_{Jj}) &= (-1)^{J+j} \cdot \varepsilon_{abc\dots ik\dots p} \cdot \lambda^{N-1} \cdot \alpha_{1a} \cdot \alpha_{2b} \cdot \alpha_{3c} \dots \alpha_{li} \cdot \alpha_{kk} \dots \alpha_{Np} \\ \text{cof}(H_{Jj}) &= \lambda^{N-1} \cdot \text{cof}(\alpha_{Jj}) \end{aligned} \quad .$$

Portanto, a matriz cofatora de \mathbf{H} pode ser escrita da forma:

$$\text{cof}\mathbf{H} = \lambda^{N-1} \cdot \text{cof}\boldsymbol{\alpha} \ .$$

E a cofatora das matrizes transpostas fica:

$$\text{cof}\tilde{\mathbf{H}} = \lambda^{N-1} \cdot \text{cof}\tilde{\boldsymbol{\alpha}} \ .$$

Finalmente, a matriz inversa de \mathbf{H} pode ser escrita em termos da inversa de $\boldsymbol{\alpha}$:

$$\mathbf{H}^{-1} = \frac{\text{cof}\tilde{\mathbf{H}}}{\det\mathbf{H}} = \frac{\lambda^{N-1} \cdot \text{cof}\tilde{\boldsymbol{\alpha}}}{\lambda^N \cdot \det\boldsymbol{\alpha}} = \frac{1}{\lambda} \cdot \boldsymbol{\alpha}^{-1} \ .$$

APÊNDICE D

ROTINA DE APLICAÇÃO DE RESTRIÇÕES AOS PARÂMETROS

AJUSTADOS PELO MODELO

```
PRO GAUSSCONST, Apar=Apar, Alpha_C=Alpha_C, LoConst=LoConst, HiConst=HiConst,$
  hCont=hCont, alpha_Mtx=alpha_Mtx, beta_Mtx=beta_Mtx
; This procedure calculates the penalty parameters "h", their 1st and 2nd
; derivates with respect to the parameters "A". To avoid any constraint just
; set its respective "Alpha_C" element to zero.
; - "Apar" are the Gaussian parameter.
; - "Alpha_C" is the alpha vector that decreases as well as "Apar" don't
; change to much from the good fit.The order of "Alpha_C" is supposed to
; be the double of the order of "Apar".
; - "LoConst" and "HiConst" are the constraint limits to the gaussians
; Both vectors "LoConst" and "HiConst" must have the order of "Apar" and
; their constraint must be order in the same order. See for example:
;   Apar[0] ; snr type 1
;   Apar[1] ; standard deviation type 1
;   Apar[2] ; Doppler freq. type 1
;   Apar[3] ; snr type 2
;   Apar[4] ; standard deviation type 2
;   Apar[5] ; Doppler freq. type 2
;   Apar[6] ; noise level
;   LoConst[0] = 10.^(SNRTH/10.) ; lower limite of snr 1
;   LoConst[1] = 2.0 ; lower limite of sd 1
;   LoConst[2] = 100.0 ; lower limite of freq. 1
;   LoConst[3] = 10.^(SNRTH/10.) ; lower limite of snr 2
;   LoConst[4] = 10.0 ; lower limite of sd 2
;   LoConst[5] = 0.0 ; lower limite of freq. 2
;   LoConst[6] = 0.0 ; lower limite of noise
;   HiConst[0] = 10.^(SNRTH) ; upper limite of snr 1
;   HiConst[1] = 20.0 ; upper limite of sd 1
;   HiConst[2] = 150.0 ; upper limite of freq. 1
;   HiConst[3] = 10.^(SNRTH) ; upper limite of snr 2
;   HiConst[4] = 50.0 ; upper limite of sd 2
;   HiConst[5] = 125.0 ; upper limite of freq. 2
;   HiConst[6] = 10.^(SNRTH/10.) ; upper limite of noise
;-----
; By : Clezio M. Denardini - clezio@dae.inpe.br (clezio@jro.igp.gob.pe)
; Date : May 7, 2002
; Modified By :
; Modified On :
; Mofification :
;-----

; determine the order of the matrices and the number of constraints
Mtx_Order = N_ELEMENTS(Apar)

; create the matrices alpha, beta and the vector h
beta_Mtx = FLTARR(Mtx_Order)
alpha_Mtx = FLTARR(Mtx_Order,Mtx_Order)
hCont = FLTARR(2*Mtx_Order)

;Apar = ABS(Apar)

FOR j=0 , Mtx_Order-1, 1 DO BEGIN
; Calculate h vector elements
hCont[j] = Apar[j] - LoConst[j] ; lower limits h
hCont[Mtx_Order+j] = HiConst[j] - Apar[j] ; upper limits h
; Calculate the beta matrix elements
beta_Mtx[j] = (Alpha_C[Mtx_Order+j] / (hCont[Mtx_Order+j]^2)) - $
(Alpha_C[j] / (hCont[j]^2))
; Calculate the alpha matrix diagonal elements (the other are zero)
alpha_Mtx[j,j] = 2 * ((Alpha_C[Mtx_Order+j] / (hCont[Mtx_Order+j]^3)) + $
(Alpha_C[j] / (hCont[j]^3)))
ENDFOR

END
```

