

UNIVERSIDADE DE SÃO PAULO

Instituto de Física.

# Métodos Bayesianos e a Priori Multiescala em fMRI. †

Selene da Rocha Amaral

Dissertação realizada sob a orientação do Prof. Nestor Felipe Caticha Alfonso apresentada à Comissão de Pós-Graduação em Física como parte dos requisitos para a obtenção do grau de Mestre em Ciências.

**Orientador:**

Nestor Felipe Caticha Alfonso

**Banca Examinadora:**

Nestor Felipe Caticha Alfonso (IF/USP)

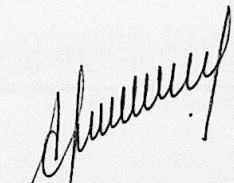
Dráulio Barros de Araújo (FFCLRP/USP)

Carlos Eugênio Imbassahy Carneiro (IF/USP)



São Paulo

2003



Prof. Armando Corbani Ferraz  
Presidente da Comissão de Pós Graduação

---

†Trabalho financiado pelo CNPq.

INSTITUTO DE FÍSICA

Serviço de Biblioteca e Informação

Tombo: 3898 ex. <sup>1</sup>

del. 18/08/03.

530.13

A485m

M

ex. 1

### FICHA CATALOGRÁFICA

Preparada pelo Serviço de Biblioteca e Informação  
do Instituto de Física da Universidade de São Paulo

Amaral, Selene da Rocha

Métodos Bayesianos e a Priori Multiescala em fMRI.  
São Paulo, SP - 2003

Dissertação (Mestrado) Universidade de São Paulo  
Instituto de Física. Departamento de Física Geral

Orientador: Prof. Dr. Nestor Caticha  
Área de Concentração: Física Geral

Unitermos: 1. Simulação;  
2. Inferência Bayesiana;  
3. fMRI.

USP/IF/SBI-051/2003



“ The actual science of logic is conversant at present only with things either certain, impossible, or entirely doubtful, none of which (fortunately) we have to reason on. Therefore the true logic for this world is the calculus of Probabilities, which takes account of the magnitude of the probability which is, or ought to be, in a reasonable man’s mind.”

**James Clerk Maxwell**

## Agradecimentos

Agradeço, de forma geral, a todas as pessoas que colaboraram para a realização deste trabalho e em particular:

- Ao Prof. Nestor Caticha pela orientação, pelas discussões e pelo incentivo sem o qual esta dissertação não seria possível;
- Ao Prof. Said Rabbani pelo apoio e interesse no desenvolvimento do trabalho;
- Ao Prof. Dráulio de Araújo pelo fornecimento dos dados reais, que engradeceram o trabalho e pela disposição em ajudar;
- Ao Roberto, Evaldo, Masayuki, Raphael e Henrique pela amizade e sobretudo pelo auxílio computacional;

Agradeço também ao CNPq pelo apoio financeiro.



## Resumo

Imagem funcional por ressonância magnética (fMRI) é uma técnica usada para estudar a localização de atividade no cérebro humano em funcionamento. Durante um experimento de fMRI, uma sequência de imagens por ressonância magnética é adquirida enquanto um indivíduo desempenha tarefas comportamentais específicas. Mudanças no sinal medido podem ser usados para identificar e caracterizar a atividade cerebral resultante da tarefa desempenhada.

Duas questões centrais na análise dos dados de fMRI adquiridos durante uma estimulação periódica são: i) como medir a resposta determinada experimentalmente em séries temporais de fMRI; e ii) como decidir se uma resposta aparente é significativa. Neste trabalho, nossos esforços são direcionados para tentar resolver a segunda questão. Assim, estudamos dois algoritmos inspirados em métodos bayesianos: o método de pixel independente e um novo método multigrid bayesiano. Os resultados obtidos em aplicações a dados artificiais do método de pixel independente, mostraram a importância da informação a priori no erro total de atribuição de atividade. A fim de direcionar o problema de atribuição a priori de um único pixel, propomos um esquema iterativo que origina-se do fato que pixels individuais tendem a estar agrupados. Ele é baseado em uma aplicação de multiescala de idéias bayesianas. Seu objetivo é construir uma distribuição de probabilidade a priori para atribuição de atividade a um dado pixel a partir das propriedades de atribuição de atividade a uma região de grande escala. Assim permite introduzir um prior local a partir das propriedades globais da imagem. A atribuição de atividade, o resultado principal do algoritmo, é proporcional à distribuição de probabilidade posterior no final da escala, aquele de um pixel. Estudamos dados artificiais e reais fornecidos de um simples experimento motor. Os resultados, ainda preliminares, são muito positivos. Também apontamos várias direções para pesquisa futura que podem possivelmente ajudar a fazer esta nova técnica útil em um nível prático.

## Abstract

Functional Magnetic Resonance Imaging (fMRI) is a technique used for studying the localization of activity in the working human brain. During a fMRI experiment, a sequence of magnetic resonance images is obtained while a subject performs specific behavioral tasks. Changes in the measured signal can be used to identify and characterize brain activity resultant of the task performance.

Two central questions in the analysis of fMRI data acquired during periodic stimulation are: i) how to measure the experimentally determined response in fMRI times series; and ii) how to decide whether an apparent response is significant. In this work we addressed the second question. Thus, we study two algorithms inspired in bayesian methods: the independent pixel method and a new multigrid bayesian method. The results obtained in applications to artificial data of the independent pixel algorithm showed the importance of the prior information in the overall error of attribution of activity. In order to address the problem of prior attribution of a single pixel we propose an iterative scheme that stems from the fact that individual pixels are not expected to be isolated. It is based on a multigrid application of bayesian ideas. Its objective is to build a prior probability distribution for assignment of activity to a given region from the properties of the attribution of activity to a larger scale region. Thus it permits introducing a local prior from the global properties of the image. The level of activity, the algorithm main result, is proportional to the posterior probability distribution in the finest scale, that of a pixel. We studied both artificial and real data acquired of a simple motor experiment. The results, while still preliminary are very positive. We also point out several directions for future research which may possibly help make this new technique useful on a practical level.



# Sumário

<b>1</b>	<b>Introdução</b>	<b>3</b>
<b>2</b>	<b>Princípios Fundamentais</b>	<b>6</b>
2.1	Ressonância Magnética Nuclear (NMR) . . . . .	6
2.1.1	Teoria Básica . . . . .	6
2.1.2	Relaxação e Detecção de Sinal . . . . .	12
2.1.3	Ecos de Spins . . . . .	16
2.2	Imagem por Ressonância Magnética (MRI) . . . . .	18
2.2.1	Gradientes de Campo Magnético e Espaço-k . . . . .	18
2.2.2	Técnicas de Imagem Convencional . . . . .	19
2.2.3	Técnica de EPI ( <i>Echo Planar imaging</i> ) . . . . .	21
2.2.4	Contraste de Imagem . . . . .	22
2.2.5	Instrumentação . . . . .	24
2.3	Imagem Funcional (fMRI) . . . . .	25
2.3.1	Método BOLD . . . . .	26
2.3.2	Arranjo Experimental em Bloco . . . . .	30
2.3.3	Análise Convencional de Dados . . . . .	33
2.4	Inferência Bayesiana - Análise de Dados . . . . .	35
2.4.1	Teoria de Probabilidade Bayesiana . . . . .	35
2.4.2	Estimativa de Parâmetro . . . . .	39
2.4.3	Teste de Hipótese . . . . .	40
<b>3</b>	<b>Métodos Preliminares</b>	<b>42</b>
3.1	Separação de Gaussianas . . . . .	42
3.1.1	Resultados de Simulações . . . . .	45
3.2	O Método de Pixel Independente . . . . .	48

<i>SUMÁRIO</i>	2
3.2.1 Resultados e Simulações . . . . .	51
<b>4 Método <i>Multigrid</i> Bayesiano</b>	<b>59</b>
4.1 Descrição do Método . . . . .	59
4.2 Resultados e Discussões de Simulações . . . . .	61
<b>5 Análise de Dados Reais</b>	<b>76</b>
5.1 Os dados de fMRI e a Inferência Inicial . . . . .	76
5.2 Método de Pixel Independente . . . . .	81
5.3 Método <i>Multigrid</i> Bayesiano . . . . .	84
<b>6 Conclusões e Perspectivas</b>	<b>89</b>



# Capítulo 1

## Introdução

O impacto de imagens médicas no campo da neurociência é considerável. Com o desenvolvimento de técnicas de imagens de tomografia computadorizada (CT) e de imagem por ressonância magnética foi possível ser mais específico em relação a posição do dano no cérebro de pacientes. A eletroencefalografia (EEG) abriu novas possibilidades para estudar funções cerebrais em indivíduos normais. No entanto foi a chegada das modalidades de imagem funcional de tomografia de emissão de pósitron (PET), tomografia computadorizada por emissão de um único fóton (SPECT), imagem funcional por ressonância magnética (fMRI), e magnetoencefalografia (MEG) que levaram a uma nova era no estudo das funções do cérebro [1, 2].

Particularmente, a Imagem por Ressonância Magnética (MRI), baseada no fenômeno de ressonância magnética nuclear, produz imagens do corpo humano com excelente contraste de tecido mole, permitindo os neurologistas distinguir entre a matéria branca e a cinzenta, e defeitos cerebrais tais como tumores. O desenvolvimento de agentes de contraste convenientes para estudos de MRI dinâmico, e melhoramentos na velocidade de produção de imagens, abriram a possibilidade de usar a técnica para estudos funcionais do cérebro. Em 1991, o primeiro experimento usando a MRI para estudar funções cerebrais foram feitas, adquirindo imagens do córtex visual enquanto o indivíduo era apresentado com um estímulo visual. Um agente de contraste foi usado neste primeiro estudo, mas não demorou até que o experimento fosse feito usando o sangue como um agente de contraste endógeno. A hemoglobina no sangue tem propriedades magnéticas diferentes dependendo se ela está oxigenada ou não; estas diferenças afetam o sinal registrado em imagens por RM. Adquirindo imagens de um indivíduo em repouso e durante a performance de uma tarefa específica, tornou-se possível adquirir imagem funcional do cérebro de modo com-

pletamente não invasivo.

Existem, no entanto, ainda áreas da técnica de MRI funcional que exigem refinamentos. O método de adquirir imagens rápidas por imagem eco planares, que é essencial para estudos dinâmicos rápidos, tem baixa qualidade. Os mecanismos intrínsecos da resposta de ativação observada não são bem entendidos, tanto que, nos últimos anos, há um interesse crescente em explorar a natureza fisiológica desta resposta e sua dependência na duração e tipo de estímulo neuronal.

Existe também, a questão envolvida na maneira que os dados são analisados. Por exemplo, como a detecção das áreas ativadas consiste em estabelecer quais as regiões do cérebro cujo fluxo sanguíneo se modificou de modo estatisticamente significativo durante a execução da tarefa, existe um problema sério que é a falta de consenso na comunidade de fMRI em como escolher os limites de significância para os testes de hipóteses aplicados, onde estes limites são geralmente valores arbitrários. Além disso, interpretar os resultados das classificações geralmente exige uma boa quantidade de informação *ad hoc* sem suporte dos dados, por exemplo, para sustentar que certos voxels (elementos unitários de uma imagem em 3D) aparentemente ativos estão verdadeiramente ativos enquanto que outros estão falsamente ativos.

O material apresentado neste trabalho trata da implementação de métodos bayesianos para a análise de dados de fMRI obtidos durante estudos de ativação do cérebro humano.

O segundo capítulo trata da teoria de imagem por ressonância magnética, incluindo as descrições quântica e clássica de ressonância magnética nuclear (NMR), de algumas técnicas que podem ser usadas para adquirir imagens de amostras biológicas, da origem do contraste em MRI e da técnica de fMRI. Esta última, é tratada em detalhes, descrevendo como a atividade cerebral afeta o contraste na imagem por RM e como as experiências são realizadas. Uma revisão de algumas técnicas convencionais usadas para analisar estatisticamente os dados de fMRI e uma introdução à inferência bayesiana é dada.

O terceiro capítulo que segue apresenta a avaliação feita da eficiência do teste de hipótese bayesiano para separar duas distribuições gaussianas independentes, que é a base do problema de localização e o desenvolvimento do método de pixel independente que detecta ativação de pixel a pixel, simplesmente aplicando o teste de hipótese utilizando uma distribuição de probabilidade a priori uniforme para todos os pixels. Os resultados de simulações com dados artificiais obtidos com diferentes atribuições de probabilidade para a atividade do pixel, antes de analisar os dados, são apresentados e discutidos.

O quarto capítulo apresenta os resultados centrais deste trabalho. Apresentamos e



estudamos um novo método estatístico eminentemente bayesiano para análise de dados de fMRI, o método *multigrid* bayesiano. Neste método, estimativas dos valores corretos das distribuições de probabilidade a priori são obtidas usando propriedades globais da imagem.

O quinto capítulo apresenta os resultados da análise de séries temporais de fMRI fornecidos de um simples experimento de estímulo motor em um voluntário normal, utilizando os dois métodos do Capítulo 3 e 4: o método de pixel independente e o método *multigrid* bayesiano. A descrição do protocolo de aquisição dos dados, a inferência dos parâmetros do modelo usado e a aplicação dos dois métodos bayesianos para detectar ativações em regiões do cérebro são tratadas.

E por fim, o sexto capítulo apresenta as conclusões e as perspectivas tiradas deste trabalho.

## Capítulo 2

# Princípios Fundamentais

Em 1946, o fenômeno de Ressonância Magnética Nuclear (NMR) foi observado independentemente por Purcell[3] e Bloch [4], trabalho pelo qual eles receberam o prêmio Nobel em 1952. Desenvolvimentos em NMR, incluindo ecos de spins e outras seqüências de pulsos de ondas de rádio (r.f.) conduziram ao uso difundido de espectroscopia por NMR, um método de análise da composição química de amostras. Em 1973 Paul Lauterbur usou gradientes de campo magnético para formar as primeiras imagens por ressonância magnética [5]. Desde então Imagem por Ressonância Magnética (MRI) tem sido utilizada em muitas aplicações biomédicas, químicas e de engenharia.

Na primeira parte deste capítulo descrevemos os princípios básicos de ressonância magnética nuclear e imagem por ressonância magnética. E ao final desta parte, mencionamos rapidamente sobre instrumentação em MRI, já que neste trabalho não lidamos com a sua implementação técnica. Na segunda parte, explicamos mais detalhadamente a técnica de Imagem Funcional por Ressonância Magnética (fMRI), como tais exames são feitos e como eles são geralmente analisados estatisticamente. Por fim, apresentamos uma introdução à inferência bayesiana, que é a base deste trabalho. Todos esses assuntos então, são apenas introduzidos neste capítulo.

## 2.1 Ressonância Magnética Nuclear (NMR)

### 2.1.1 Teoria Básica

Basicamente os núcleos atômicos são sistemas formados de muitas partículas, prótons e nêutrons acoplados por forças nucleares, possuindo um momento angular de spin  $J$ . As-



sociado a este momento angular existe um momento magnético nuclear  $\mu$ . Esses dois operadores vetoriais são paralelos e relacionados por:

$$\mu = \gamma \mathbf{J}, \quad (2.1)$$

sendo  $\gamma$  um parâmetro característico de cada espécie atômica, chamada de razão giromagnética. Afirmar que um núcleo atômico possui momento magnético é, de certa forma, como dizer que alguns núcleos se comportam como magnetos, isto é, eles são capazes de produzir um campo magnético nas suas vizinhanças.

É conveniente definir um operador de spin adimensional  $\mathbf{I}$ , através da seguinte relação:

$$\mathbf{J} = \hbar \mathbf{I}. \quad (2.2)$$

O autovalor do operador  $\mathbf{I}^2$  é dado por  $I(I+1)$ , em que  $I$  pode ser um número positivo inteiro ou semi-inteiro que representa o número quântico de spin nuclear. As projeções do momento angular  $\mathbf{I}$  em relação a um eixo de quantização são determinadas pelos possíveis valores que o número quântico magnético  $m_I$  pode ter:

$$m_I = I, I-1, \dots, -I, \quad (2.3)$$

existindo portanto  $2I+1$  valores para essas projeções. Um próton, por exemplo, tem um número quântico de spin  $I = \frac{1}{2}$ , havendo assim apenas duas orientações possíveis com relação a um campo magnético estático aplicado, com  $m_I = +\frac{1}{2}, -\frac{1}{2}$ . Os núcleos de  $^{14}\text{N}$ ,  $^{19}\text{F}$ ,  $^2\text{H}$ ,  $^{31}\text{P}$  têm número quântico de spin  $I = 1$ , com  $m_I = +1, 0, -1$ . Alguns núcleos, como  $^{12}\text{C}$ ,  $^{16}\text{O}$  têm  $I = 0$ , e esses átomos são insensíveis à presença de um campo magnético. Pela equação (2.2), o momento magnético nuclear  $\mu$  pode ser escrito como:

$$\mu = \gamma \hbar \mathbf{I}. \quad (2.4)$$

A aplicação de um campo magnético estático  $\mathbf{B}$  sobre um sistema de spins nucleares produz uma interação entre o sistema e o campo magnético. Essa interação é representada pelo operador hamiltoniano:

$$H = -\mu \cdot \mathbf{B}. \quad (2.5)$$

Considerando o campo magnético estático representado por  $B_0 \hat{k}$ , aplicado ao longo da direção  $z$ , podemos usar as equações (2.4) e (2.5) para reescrever o hamiltoniano  $H$  da

seguinte forma:

$$H = -\gamma\hbar B_0 I_z, \quad (2.6)$$

sendo  $I_z$  o componente z do operador de spin nuclear  $\mathbf{I}$ . Sabendo que qualquer componente do operador  $\mathbf{I}$  comuta com  $\mathbf{I}^2$  e, conseqüentemente, com o hamiltoniano Zeeman nuclear expresso pela Eq. (2.5), podemos especificar simultaneamente os auto-estados de  $I_z$ ,  $\mathbf{I}^2$  e  $H$ . Esses auto-estados serão designados pelos kets  $|Im_I\rangle$ , através das seguintes equações de autovalor:

$$H |Im_I\rangle = E_m |Im_I\rangle, \quad (2.7)$$

$$I_z |Im_I\rangle = m_I |Im_I\rangle, \quad (2.8)$$

$$\mathbf{I}^2 |Im_I\rangle = I(I+1) |Im_I\rangle. \quad (2.9)$$

Os autovalores  $E_m$  do hamiltoniano  $H$  são os níveis de energia Zeeman nuclear, dados pela relação:

$$E_m = -\gamma\hbar B_0 m_I. \quad (2.10)$$

Para um sistema de spin  $\frac{1}{2}$ , vimos que  $m_I = \pm\frac{1}{2}$ , existindo apenas dois níveis de energia:

$$E_{\pm} = \pm\frac{1}{2} \gamma\hbar B_0. \quad (2.11)$$

Aqui, observamos que a distância entre dois níveis adjacentes é igual a  $\gamma\hbar B_0$ . Cada nível de energia corresponde a uma diferente orientação do momento magnético com relação ao campo magnético  $\mathbf{B}$ . A presença desses níveis de energia pode ser detectada através de um espectro de absorção. O que é necessário é ter uma interação que possa causar transições entre níveis.

Para satisfazer a conservação de energia, a interação deve ser dependente do tempo. Essa interação é representada pelo hamiltoniano de perturbação sobre o hamiltoniano Zeeman. Para um sistema não perturbado, a diferença de energia  $\Delta E$  entre dois níveis adjacentes satisfaz à equação:

$$\Delta E = \hbar\omega = \gamma\hbar B_0, \quad (2.12)$$

o que nos permite escrever:

$$\omega = \gamma B_0 \quad \text{ou} \quad \nu_L = \frac{\gamma B_0}{2\pi}. \quad (2.13)$$



A equação (2.13) é denominada de *equação da ressonância*. Note que a condição de ressonância não depende da constante quântica de Planck  $\hbar$ . Assim sendo, esse assunto pode, igualmente, ser descrito classicamente. A frequência de ressonância  $\nu_L$  é chamada frequência de Larmor. A razão desse nome é que, classicamente, ela é igual à frequência de precessão do momento magnético em torno do campo magnético estático aplicado. É claro que quanto maior for a intensidade do campo estático  $B_0$  maior será a frequência de ressonância. De um modo geral, as frequências dos campos magnéticos oscilantes utilizadas nos espectrômetros de ressonância magnética nuclear, variam de alguns poucos MegaHertz ( $\approx 10MHz$ ) até centenas de MegaHertz ( $700MHz$ ), nos modernos espectrômetros de alta resolução.

Quando um sistema de prótons (spin  $\frac{1}{2}$ ) é submetido a um campo magnético estático, da forma descrita anteriormente, é esperado que a população de spins seja distribuída entre dois níveis de energia,  $E_+$  e  $E_-$ , obedecendo à estatística de Boltzmann. Com isso, na condição de equilíbrio térmico, o estado quântico que corresponde ao nível de energia mais baixo, (paralelos ao campo  $\mathbf{B}$ ), terá uma população de spins ligeiramente maior que aquela correspondente ao nível de energia mais alto  $|Im_I \rangle = |\frac{1}{2} - \frac{1}{2} \rangle$  (spins antiparalelos ao campo  $\mathbf{B}$ ). Designando-se por  $N_+$  e  $N_-$  as respectivas populações de spins nesses dois estados quânticos — e ainda considerando-se que a diferença de energia  $\Delta E$  seja menor  $k_B T$  — a razão entre as duas populações de spin, para os dois estados especificados, pode ser dada por:

$$\frac{N_-}{N_+} = e^{-\frac{\Delta E}{k_B T}} \approx 1 - \frac{\Delta E}{k_B T}. \quad (2.14)$$

Designando por  $n = N_- - N_+$  a diferença de população entre os dois níveis de energia, e por  $N = N_+ + N_-$  o número total de spins do sistema, temos:

$$\frac{n}{N} \approx \frac{\Delta E}{k_B T} = \frac{\gamma \hbar B_0}{2k_B T}, \quad (2.15)$$

mostrando que a diferença de população de spins entre os dois níveis de energia é diretamente proporcional à intensidade do campo magnético aplicado e inversamente proporcional à temperatura  $T$  do sistema.

A ressonância magnética nuclear é uma técnica que consiste em aplicar um campo magnético estático  $\mathbf{B}$  para fazer o desdobramento Zeeman entre os níveis de energia do sistema de spins, equação (2.6), e um campo magnético oscilante  $\mathbf{B}_1(t) = \mathbf{B}_p \cos \omega t$ , aplicado perpendicularmente ao campo estático, para induzir transições entre os níveis de energia do sistema quântico (absorção de energia do campo). Na presença do campo magnético

oscilante  $\mathbf{B}_1(t)$ , o hamiltoniano de perturbação é usualmente escrito como:

$$H^{(1)} = -\boldsymbol{\mu} \cdot \mathbf{B}_1(t) = -\gamma\hbar B_{px} I_x \cos \omega t, \quad (2.16)$$

sendo  $B_{px}$  a magnitude desse campo magnético oscilante, aplicado na direção do eixo x. O operador  $I_x$  tem elementos de matriz não-nulos entre os estados nucleares  $|Im\rangle$  e  $|Im'\rangle$ , de tal forma que transições entre os dois estados quânticos somente ocorrem quando o elemento de matriz  $\langle Im'|I_x|Im\rangle$  é diferente de zero. Tendo em vista que  $I_x = \frac{1}{2}(I_+ + I_-)$ , transições irão somente ocorrer quando a condição  $m' = m \pm 1$  for satisfeita, ou seja, as transições acontecem entre dois níveis de energia adjacentes.

Um átomo com momento magnético  $\boldsymbol{\mu}$  (nuclear ou eletrônico), quando submetido a um campo magnético  $\mathbf{B}$ , fica sujeito à ação de um torque  $\mathbf{N}$  igual à taxa de variação do seu momento angular  $\frac{d\mathbf{J}}{dt}$ . De acordo com a mecânica clássica, essa taxa de variação é geralmente expressa por:

$$\frac{d\mathbf{J}}{dt} = \boldsymbol{\mu} \times \mathbf{B}. \quad (2.17)$$

Com isso, usando a relação (2.1), podemos escrever:

$$\frac{d\boldsymbol{\mu}}{dt} = \boldsymbol{\mu} \times \gamma\mathbf{B}. \quad (2.18)$$

A soma vetorial de um conjunto de spins ou momentos magnéticos individuais de uma amostra, forma o que é chamado de magnetização  $\mathbf{M}$ . A magnetização de uma amostra é definida como a soma vetorial dos momentos magnéticos individuais por unidade de volume, isto é,

$$\mathbf{M} = \sum_{i=1}^N \boldsymbol{\mu}_i. \quad (2.19)$$

Assim, a taxa de variação do valor esperado da magnetização obedece à seguinte equação diferencial:

$$\frac{d\langle \mathbf{M} \rangle}{dt} = \langle \mathbf{M} \rangle \times (\gamma\mathbf{B}). \quad (2.20)$$

Vimos que para observar transições entre dois níveis de energia Zeeman de um sistema de spins é necessário aplicar um campo magnético oscilante perpendicular ao campo magnético estático. Desta forma, consideremos um campo magnético linearmente oscilante, com uma amplitude  $B_1$  e frequência angular  $\omega_z$ , girando em torno do eixo z. Se somente o componente circularmente polarizado de  $B_1$  girando em fase com o vetor magnetização



é considerado:

$$\mathbf{B}_h = B_1 \cos \omega_z t \hat{i} - B_1 \sin \omega_z t \hat{j}, \quad (2.21)$$

a equação (2.20) produz:

$$\frac{dM_x}{dt} = \gamma[M_y B_0 + M_z B_1 \sin \omega_z t]; \quad (2.22)$$

$$\frac{dM_y}{dt} = \gamma[M_z B_1 \cos \omega_z t - M_x B_0]; \quad (2.23)$$

$$\frac{dM_z}{dt} = \gamma[-M_x B_1 \sin \omega_z t - M_y B_1 \cos \omega_z t]. \quad (2.24)$$

Se uma condição inicial  $\mathbf{M}(0) = M_0 \hat{k}$  é definida, então as soluções para  $\mathbf{M}$  são:

$$M_x(t) = M_0 \sin \omega_1 t \sin \omega_z t; \quad (2.25)$$

$$M_y(t) = M_0 \sin \omega_1 t \cos \omega_z t; \quad (2.26)$$

$$M_z(t) = M_0 \cos \omega_1 t, \quad (2.27)$$

em que  $\omega_1 = \gamma B_1$ . Isto implica que aplicando um campo magnético oscilante de frequência  $\omega_z$ , a magnetização simultaneamente precessa sobre  $B_0$  em  $\omega_z$  e  $B_1$  em  $\omega_1$  como mostra a figura 2.1a.

Neste ponto é apropriado introduzir um novo sistema de referência para observar a evolução do vetor magnetização, o sistema rotativo, que rotaciona em torno do eixo  $z$  na frequência  $\omega_z$ . Se no sistema rotativo um sistema de eixos  $(x', y', z')$  é definido, então a equação (2.20) pode ser escrita como:

$$\frac{d\mathbf{M}}{dt} = \gamma \mathbf{M} \times \mathbf{B}_{ef} \quad (2.28)$$

em que

$$\mathbf{B}_{ef} = \left(B_0 - \frac{\omega}{\gamma}\right) \hat{k}' + B_1 \hat{i}', \quad \omega = \omega_z \quad (2.29)$$

e  $(\hat{i}', \hat{j}', \hat{k}')$  são vetores unitários na direção  $(x', y', z')$ . Esse campo magnético efetivo está representado na figura 2.1b.

Agora, aplicando sobre o sistema um pulso de r.f. no plano  $x'y'$  (direção do eixo  $x'$ ), com uma frequência angular  $\omega = \gamma B_0$  (condição de ressonância), temos que  $B_0 - \frac{\omega}{\gamma} = 0$  e o campo magnético efetivo fica reduzido a  $\mathbf{B}_{ef} = B_1 \hat{i}'$ . Com isso, o vetor magnetização irá precessionar num plano perpendicular ao eixo  $x'$ , ou seja, no plano  $y'z'$ , como ilustrado

na figura 2.1c.

Assim, se no instante inicial  $t = 0$ , a magnetização encontrar-se apontando no sentido positivo do eixo  $z$  (condição de equilíbrio térmico), ela poderá ser rotacionada em torno do eixo  $x'$  de qualquer ângulo  $\theta$  desejado. Esse ângulo de rotação é proporcional à duração do pulso de r.f. aplicado e é dado por:

$$\theta = \gamma B_1 \Delta t. \quad (2.30)$$

Com esse procedimento, a magnetização poderá ser colocada no plano  $xy$ , fazendo-se uma escolha adequada da duração  $\Delta t$  do pulso de r.f. de tal modo que o ângulo de rotação assuma o valor  $\theta = \frac{\pi}{2}$ . Fazendo-se  $\theta = \pi$  imediatamente após à aplicação do pulso, o vetor magnetização irá apontar numa direção antiparalela ao campo magnético estático. Valores típicos para a duração  $\Delta t$  são de alguns microsegundos.

### 2.1.2 Relaxação e Detecção de Sinal

Geralmente, para descrever o comportamento dos spins nucleares na presença de um campo magnético, devemos considerar as interações desses spins com a sua vizinhança. Aqui, vizinhança significa o movimento vibracional da rede cristalina, no caso de sólidos, ou o movimento browniano de átomos e moléculas, no caso de líquidos. Essas interações são responsáveis pelos mecanismos de relaxação do sistema de spins.

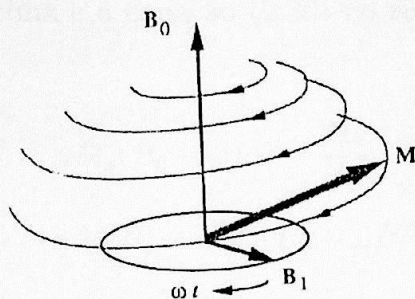
Associados a essas interações, dois parâmetros de grande importância, em especial na área de imagens, devem ser considerados: o *tempo de relaxação longitudinal*,  $T_1$ , também conhecido *tempo de relaxação spin-rede* e o *tempo de relaxação transversal*,  $T_2$ , ou *tempo de relaxação spin-spin*. Consideremos um sistema de spins em equilíbrio térmico com o seu meio, vetor magnetização alinhado com o campo magnético estático  $\mathbf{B}$ , aplicado ao longo do eixo  $z$ . Na condição de equilíbrio térmico, vimos que os valores médios  $M_x$  e  $M_y$  são nulos e  $M_z = M_0$ . No entanto, por um processo de perturbação do hamiltoniano do sistema, tal como descrito anteriormente, todos os três componentes do vetor magnetização podem ser considerados diferentes de zero num dado instante, e o processo de relaxação de spins pode ser estudado.

Na ausência de um campo de r.f., os componentes do vetor magnetização são descritos pelas seguintes equações diferenciais de Bloch:

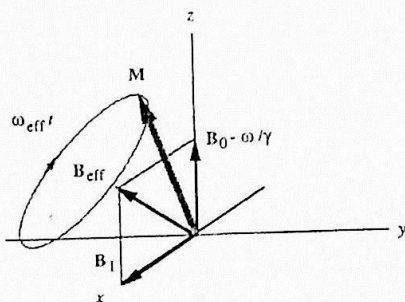
$$\frac{dM_z}{dt} = \frac{M_0 - M_z}{T_1}, \quad (2.31)$$



(a)



(b)



(c)

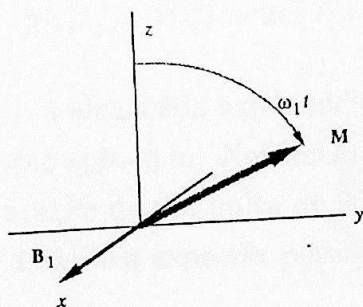


Figura 2.1: Precessão da magnetização (a) no sistema do laboratório, na presença de um campo longitudinal  $\mathbf{B}_0$ , e um campo transversal,  $\mathbf{B}_1$ , (b) no sistema rotativo na presença de um campo  $\mathbf{B}_{ef}$  e (c) no sistema rotativo quando  $B_0 = \omega/\gamma$  [6].

$$\frac{dM_x}{dt} = \frac{-M_x}{T_2}, \quad (2.32)$$

$$\frac{dM_y}{dt} = \frac{-M_y}{T_2}. \quad (2.33)$$

Combinando essas equações acima e a equação (2.20) no sistema rotativo as equações de Bloch passam a ser:

$$\frac{dM_x}{dt} = \gamma M_y (B_0 - \omega/\gamma) - \frac{M_x}{T_2}, \quad (2.34)$$

$$\frac{dM_y}{dt} = \gamma M_z B_1 - \gamma M_x (B_0 - \omega/\gamma) - \frac{M_y}{T_2}, \quad (2.35)$$

$$\frac{dM_z}{dt} = \gamma M_y B_1 - \frac{(M_z - M_0)}{T_1}. \quad (2.36)$$

Após um pulso de r.f. ser aplicado de tal modo que  $\theta = \gamma B_1 \Delta t = 90^\circ$ ,  $M_z = 0$ , a solução da equação (2.36) fica expressa por:

$$M_z(t) = M_0(1 - e^{-\frac{t}{T_1}}). \quad (2.37)$$

Essa equação nos mostra que o componente  $M_z$  da magnetização retorna ao seu estado de equilíbrio térmico da forma ilustrada na figura 2.2 do lado esquerdo. Se no instante inicial invertermos a magnetização, aplicando um pulso de r.f. com uma duração  $\Delta t$  de tal modo que  $\theta = \gamma B_1 \Delta t = 180^\circ$ ,  $M_z \neq 0$  a solução da equação (2.36) passa a ser

$$M_z(t) = M_0(1 - 2e^{-\frac{t}{T_1}}) \quad (2.38)$$

e a condição de equilíbrio térmico é alcançada exponencialmente.

Consideremos agora a relaxação spin-spin. Novamente, tomando-se o sistema de spins em equilíbrio térmico após a aplicação de um pulso de  $90^\circ$  ao longo do eixo  $x'$ , no sistema rotativo, a solução da equação (2.35) fica expressa por:

$$M_y(t) = M_0 e^{-\frac{t}{T_2}}. \quad (2.39)$$

Esse decréscimo exponencial do componente  $M_y$  com o tempo está ilustrado na figura 2.2.

Podemos notar pela equação (2.37), após a aplicação do pulso de  $90^\circ$ , transcorrido um tempo  $t = T_1$ , aproximadamente 63% da magnetização volta ao seu estado de equilíbrio



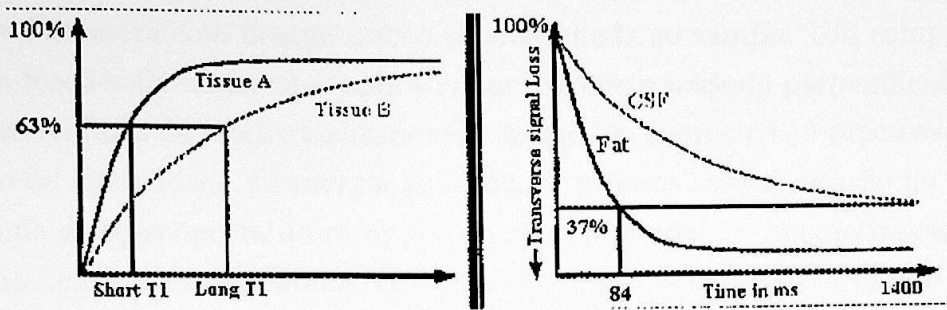


Figura 2.2: À esquerda: retorno da magnetização à condição de equilíbrio térmico, após a aplicação de um pulso de r.f. de  $90^\circ$ , para dois tipos de tecido. À direita: decaimento exponencial da magnetização transversal, após a aplicação de um pulso de r.f. de  $90^\circ$ , para a gordura e o CSF (fluido cerebrospinal) [7].

térmico. Já a equação (2.38), nos mostra que transcorrido um tempo  $t = T_2$ , apenas 37% da magnetização inicial  $M_z(0) = M_0$ , permanece apontando na direção do eixo  $y$  (figura 2.2 do lado direito).

Para detectar o sinal de NMR é necessário que uma bobina de r.f. seja colocada ao redor da amostra com seu eixo de simetria transversal ao campo magnético principal  $\vec{B}_0$ , na qual uma força eletromotriz (f.e.m.) oscilando na frequência de Larmor  $\omega_0$  é induzida com amplitude proporcional a  $M_x$ . O sinal da bobina é primeiro transformado para o sistema rotativo pela detecção sensível de fase. Isto envolve separadamente misturas da f.e.m. com dois sinais de referência, ambos oscilando na frequência de Larmor  $\omega_0$ , mas  $90^\circ$  fora de fase entre si. Assim, o sinal detectado na bobina tem a forma:

$$S(t) = S_0 \exp(-t/T_2) \cos(\Delta\omega t), \quad (2.40)$$

que após a detecção sensível de fase, tem componentes 'real' e 'imaginária':

$$S_R(t) = S_0 \exp(-t/T_2) \cos(\Delta\omega t) \quad (2.41)$$

$$S_I(t) = S_0 \exp(-t/T_2) \sin(\Delta\omega t) \quad (2.42)$$

em que  $\Delta\omega = \omega - \omega_0$ . Se  $\omega = \omega_0$  então o sinal passa ser a Eq.(2.39), do contrário, o sinal oscilará em uma frequência  $\Delta\omega$ . O sinal após a detecção sensível de fase - sinal produzido pela magnetização transversal - é conhecido como *Free Induction Decay* (FID) e por transformada de Fourier o mesmo sinal pode ser representado no domínio da frequência.

O fenômeno de NMR pode ser resumido classicamente. Um campo magnético estático,

$\vec{B}_0$ , polariza a amostra com magnetização total alinhada ao campo. Um campo magnético oscilando na frequência de Larmor aplicado por um curto período perpendicular a  $\vec{B}_0$  fará os prótons individuais da amostra absorverem energia do campo r.f., e precessarem em fase. Dependendo da intensidade da energia aplicada, os prótons movimentarão no plano  $xy$ , ou exatamente na direção oposta do campo magnético principal. A magnetização transversal induzirá uma f.e.m. em uma bobina receptora que eventualmente se tornará o sinal de RM e a transformada de Fourier deste sinal detectado produzirá sua saída no valor esperado de frequência.

### 2.1.3 Ecos de Spins

Uma das mais significantes contribuições para o desenvolvimento da ressonância magnética nuclear pulsada, principalmente na formação de imagens por ressonância magnética (discutida na próxima seção), foi dada por Hahn [8], com a descoberta dos ecos de spins.

A explicação que se segue é sobre o mecanismo de formação dos ecos de spins do ponto de vista fenomenológico. Suponhamos que no instante  $t = t_1 = 0$ , um pulso de r.f. de  $90^\circ$  seja aplicado ao longo do eixo  $x$  sobre o conjunto de spins (amostra) no sistema rotativo. O efeito desse pulso é fazer a magnetização girar de um ângulo de  $90^\circ$  em torno desse eixo. Transcorrido um tempo  $t_1$ , o vetor magnetização estará direcionada ao longo do eixo  $y$  (veja a figura 2.3). Nesse instante, devido à não homogeneidade do campo magnético estático aplicado, começa a haver uma defasagem da magnetização no plano  $xy$ . Essa defasagem pode ser vista como a formação de grupos de spins  $\delta M$  que precessionam com a mesma frequência, abrindo como uma espécie de leque nesse plano. Com isso, dentro de um intervalo de tempo muito curto, a magnetização total no plano  $xy$  será nula. Vamos supor que no instante  $t = \tau$ , um pulso de r.f. de  $180^\circ$  seja aplicado ao longo do eixo  $x$ . Assim, toda a magnetização que tenha retornado à condição de equilíbrio térmico, direção  $z$ , fica direcionada no sentido oposto ao campo magnético estático. Ao mesmo tempo, a magnetização remanescente no plano  $xy$  vão girar  $180^\circ$  em torno do eixo  $x$ . Conseqüentemente, após transcorrido um tempo  $\tau$ , a partir do instante em que o pulso de inversão foi aplicado, ou, equivalentemente, transcorrido um tempo  $2\tau$  após o instante inicial, a magnetização remanescente vai refocalizar-se, direcionando ao longo do eixo  $y$  negativo, formando assim um sinal de eco de spins.

A seqüência de pulsos ilustrada na figura 2.3 pode ser utilizada para medir o tempo de relaxação transversal  $T_2$ , que será discutida na próxima seção.



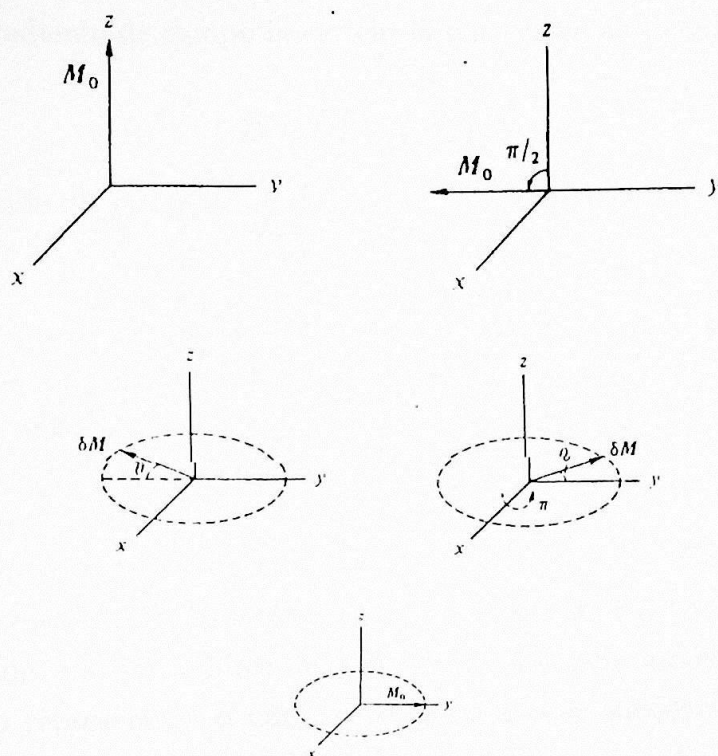


Figura 2.3: Processo de refocalização de spins, utilizando-se uma seqüência de pulsos  $90^\circ$ - $\tau$ - $180^\circ$  [9].

## 2.2 Imagem por Ressonância Magnética (MRI)

### 2.2.1 Gradientes de Campo Magnético e Espaço-k

Como foi mostrado na seção anterior, a equação fundamental de ressonância magnética é a equação de Larmor,  $\omega = \gamma B_0$ . Em um experimento de NMR, uma medida da frequência de precessão da magnetização fornece informação no campo experimentado por aquele grupo de spins. Por meio da manipulação da variação espacial do campo em um caminho conhecido, esta informação frequencial agora produz informação espacial.

Considere um gradiente de campo linear em  $\mathbf{B}$  que cresce ao longo do eixo  $x$ , tal que

$$\mathbf{B} = (B_0 + Gx) \mathbf{k} \quad (2.43)$$

onde  $G$  é a intensidade do gradiente. Isto resulta na equação de Larmor

$$\omega(x) = \gamma(B_0 + Gx) \quad (2.44)$$

ou na sua forma mais geral em três dimensões

$$\omega(\vec{r}) = \gamma(B_0 + \vec{G} \cdot \vec{r}) \quad (2.45)$$

Sob um gradiente de campo linear ao longo do eixo  $x$ , todo spin que se encontra em um valor particular de  $x$  precessará na mesma frequência. A forma do FID de tal amostra conterá componentes de cada um dos valores de  $x$  representados pela amostra, e o espectro de frequência portanto representará o número de spins que se encontram ao longo daquele plano

$$\rho(\omega) = \rho(x) = \int \rho(x, y) dy. \quad (2.46)$$

Este espectro simples portanto fornece a informação espacial do objeto sendo mapeado ao longo de uma dimensão. Para construir a imagem completa em 3D é necessário aplicar gradientes de campo variando no tempo.

Descrevendo o número de spins em uma localização particular  $\vec{r}$ , como a densidade de spins,  $\rho(\vec{r})$ , o sinal da amostra pode ser escrito

$$S(t) = \int \int \int \rho(\vec{r}) \exp[i\gamma \vec{G} \cdot \vec{r} t] d\vec{r} \quad (2.47)$$



O vetor de espaço recíproco é definido

$$\vec{k} = \gamma \vec{G} t \quad (2.48)$$

e a relação de Fourier entre a densidade de spin e o sinal torna-se

$$S(\vec{k}) = \iiint \rho(\vec{r}) \exp[i \vec{k} \cdot \vec{r}] d\vec{r} \quad (2.49)$$

$$\rho(\vec{r}) = \iiint S(\vec{k}) \exp[-i \vec{k} \cdot \vec{r}] d\vec{k} \quad (2.50)$$

Assim  $\vec{k}$  é a variável conjugada de  $\vec{r}$ . A resolução da imagem depende da extensão do espaço-k que é amostrado, e assim diferentes técnicas de imagem podem ser comparadas observando como cada uma delas revestem o espaço-k.

Existem técnicas de imagem que fazem uma amostragem de todas as três dimensões do espaço-k, mas muitas técnicas reduzem o problema a 2D aplicando a técnica de excitação seletiva.

### Excitação Seletiva de Fatias

A técnica que forma uma imagem de um único plano no objeto, por somente excitar os spins naquele plano, é conhecida como excitação seletiva. Para fazer isso um pulso de r.f. que somente afeta uma parte limitada do espectro de NMR é aplicado, na presença de um gradiente de campo linear ao longo da direção em que a fatia é selecionada (figura 2.4). Isso resulta somente na excitação daqueles spins cuja frequência de Larmor, que é imposta pelas suas posições, é a mesma em relação a frequência do pulso r.f. aplicada, ao passo que spins em outras fatias estão precessando muito lentamente ou muito rapidamente para absorver energia.

### 2.2.2 Técnicas de Imagem Convencional

As técnicas extensivamente usadas em MRI são todas baseadas em técnicas de Fourier, como a *Echo Planar imaging* (EPI) introduzida na próxima subseção. Entretanto, as primeiras imagens por RM utilizaram métodos de ponto e de linha e estas são descritas aqui, juntamente com a técnica de reconstrução por projeção.

A técnica de ressonância magnética de campo focalizado (*FONAR*) foi proposta por Damadian, e foi usada para produzir a primeira imagem do corpo inteiro em 1977 [10]. A base da técnica é criar um campo  $B_0$  focalizado que tem uma região central homogênea,

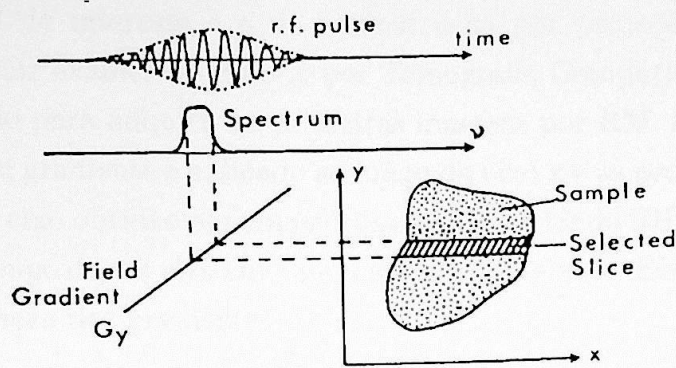


Figura 2.4: Excitação seletiva de uma fatia em uma amostra [7].

circundada por uma região altamente inhomogênea. Os sinais das regiões inhomogêneas terão um  $T_2^*$  (caracterizado pelo tempo  $T_2$  mais a contribuição da inhomogeneidade) muito curto e então poderão ser distinguidos do sinal da região homogênea. Explorando a região do campo homogêneo através do conjunto da amostra uma imagem pode ser formada. Este é procedimento que consome 10 minutos por fatia. As duas vantagens da técnica são sua simplicidade conceitual e a falta de uma exigência para o campo estático ser homogêneo sobre uma grande área. Pulsos de r.f. focalizados também podem ser usados para isolar uma pequena região, e tais métodos formam a base da espectroscopia localizada.

Como foi mostrada na subseção anterior, para mapear um objeto somente em uma dimensão, um único gradiente linear é suficiente para localizar posição diretamente do FID por transformada de Fourier. Assim, se a amostra tridimensional puder ser reduzida a um conjunto de componentes unidimensionais então a amostra inteira pode ser mapeada. Isto pode ser feito por excitação seletiva ou seleção de fatias, em duas dimensões. Primeiro, na presença de um gradiente ao longo do eixo z, um pulso seletivo é aplicado que saturam todos os spins externos ao plano de interesse. Então um gradiente é aplicado ao longo do eixo x, e esses spins não saturados são inclinados no plano transversal por um pulso seletivo de  $90^\circ$ . Imediatamente após o segundo pulso de r.f. a única região com alguma magnetização transversal coerente é a linha de intersecção dos dois planos selecionados. Um gradiente é agora aplicado ao longo do eixo y, e a evolução do FID registrada. Transformada de Fourier do FID fornece a densidade de prótons ao longo da linha. Repetindo a seleção de linha para todas as linhas no plano, uma imagem de todo o plano pode ser construída. Existem muitas variações da técnica de exame de linha, algumas das quais utilizam pulsos de  $180$  graus, mas elas são ineficientes se comparadas aos métodos de Fourier.



Um método final de interesse é o de reconstrução por projeção. Este é o método utilizado para construir exames de Raio-X por Tomografia Computadorizada (CT) [11], e foi o método utilizado para adquirir as primeiras imagens por RM. Seguindo a técnica de excitação seletiva, um gradiente é aplicado ao longo do eixo x e as projeções das densidades de spins sobre aquele eixo obtidas por transformada de Fourier do FID. Assim um gradiente linear é aplicado ao longo de um eixo em algum ângulo  $\theta$  ao eixo x. Este pode ser encontrado usando uma combinação dos gradientes x e y

$$G_y = G \sin \theta, \quad G_x = G \cos \theta. \quad (2.51)$$

O conjunto de projeções então pode ser colocado junto utilizando *back projection*. Esta distribui as densidades de spin medidas igualmente ao longo da linha normal ao eixo em que foram adquiridas (figura 2.5). Reconstruindo por projeções em todos os ângulos (0 a 180 graus) a imagem é formada. Existe entretanto um artefato de “borrão” por toda a imagem, onde uma tentativa foi feita para atribuir densidade de spin às áreas onde na verdade não existe nada, e artefatos de “estrelas” onde um número finito de projeções são usadas para definir estruturas de pontos. Isso pode ser corrigido utilizando uma técnica chamada *de back projection* filtrada, que convolvei cada um dos perfis com um filtro:

$$\frac{1}{(r - r_0)^2}. \quad (2.52)$$

Isso é feito no domínio de Fourier dividindo as projeções de Fourier pelo módulo do vetor  $\vec{k}$  (espaço recíproco), como definido no começo desta seção.

Reconstrução por projeção não é mais utilizada atualmente por razão do longo tempo de aquisição de dados e tem sido substituída por métodos que amostram o espaço-k mais uniformemente.

### 2.2.3 Técnica de EPI (*Echo Planar imaging*)

Enquanto que técnicas convencionais utilizadas em IRM coletam os dados para uma imagem de uma série de amostras discretas do sinal, EPI é uma técnica que forma uma imagem completa de um único FID.

A formação de imagens tomográficas exige codificação espacial em 3 dimensões. Em muitos casos, define-se o eixo-z como eixo de excitação seletiva de fatia, o eixo-x como o eixo de frequência e o eixo-y como a codificação de fase. A diferença entre EPI e imagem

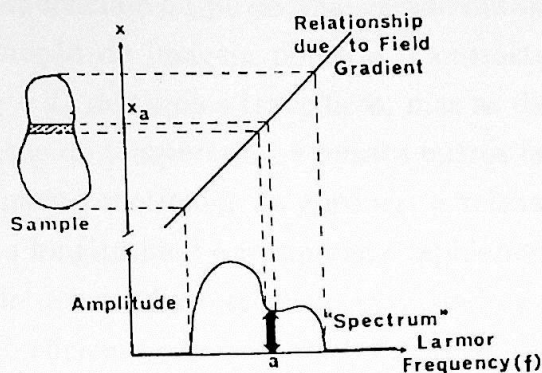


Figura 2.5: Um gradiente de campo na direção  $x$  produzindo fatias no plano  $yz$  que podem ser projetadas sobre o eixo  $x$  [7].

convencional ocorre na codificação espacial “no plano” remanescente.

As imagens ecoplanares podem ser adquiridas com seqüências de pulsos de ecos de spins, mencionadas na seção anterior e de ecos de gradiente. Para a técnica de eco de gradiente, o pulso de  $180^\circ$  formando eco de spins é omitido nestas seqüências e o sinal é refocado somente pelos gradientes. Desligando o gradiente ‘readout’ rapidamente, todo o espaço- $k$  pode ser amostrado antes que a relaxação spin-spin (governada pelo tempo  $T_2$ ) atenua a magnetização transversal. A codificação é utilizada para amostrar  $k_y$ .

A técnica de EPI é geralmente difícil de implementar, exigindo um hardware especializado, entretanto ela tem a vantagem de ser uma técnica de imagem muito rápida, capaz de capturar movimento de órgãos como o coração, e dinamicamente ativação em imagens cerebrais. Na Seção 2.3 discutiremos um pouco mais sobre EPI.

## 2.2.4 Contraste de Imagem

A base de contraste é a densidade de spins dos prótons por todo o objeto. Se não existe nenhum spin presente em uma região não é possível conseguir um sinal de RM. A densidade de spins dos prótons depende da quantidade de água e de gordura existentes em tecidos no corpo humano. Como o osso possui baixa densidade de spins, MRI torna-se neste caso uma escolha menos conveniente para a formação de imagem do esqueleto que radiografias e CT por Raios-X. Já que existe uma pequena diferença em densidade de spins entre muitos outros tecidos no corpo, outros mecanismos de contraste deve ser empregado. Estes são geralmente baseados na variação em valores de  $T_1$  e de  $T_2$  para diferentes tecidos. Estes



valores significativamente influenciam o tipo de sinal gerado durante a MRI e assim contribui enormemente para a formação de imagem por RM. Contraste do tecido é afetado não somente por valores de  $T_1$  e  $T_2$  de tecidos específicos, mas as diferenças na intensidade do campo magnético, mudanças de temperatura e muitos outros fatores.

Devido ao lento movimento molecular da gordura, a relaxação spin-rede ocorre rapidamente e a magnetização longitudinal é recuperada rapidamente. O vetor magnetização total retorna ao seu estado de equilíbrio térmico conduzindo a um tempo  $T_1$  curto para a gordura. A água não é tão eficiente quanto a gordura em recuperação em  $T_1$  devido à alta mobilidade das moléculas de água. O núcleo da água não cede sua energia à rede (tecido ao redor) tão rapidamente quanto a gordura, e portanto leva mais tempo para a magnetização longitudinal retornar à condição de equilíbrio térmico resultando em um tempo  $T_1$  longo.

O decaimento  $T_2$  é dependente da interação do núcleo e a troca de energia com núcleos vizinhos. A gordura tem uma troca de energia muito eficiente e portanto tem um  $T_2$  relativamente longo.

Os valores do tempo  $T_1$  e  $T_2$  em vários tecidos no corpo humano são explorados por dois parâmetros, TR e TE, geralmente medidos em milissegundos. TR significa tempo de repetição, ou o tempo transcorrido entre sucessivos pulsos de r.f.; TE significa tempo de atraso de eco, ou intervalo de tempo entre pulsos de r.f. e a medida do primeiro eco.

Nas seqüências de pulsos descritas a seguir, é considerado que o método de imagem utilizado é EPI, no entanto métodos idênticos ou similares podem ser usados com outras técnicas de imagem por RM.

### Contraste em $T_1$

Uma imagem da densidade de prótons pode ser ponderada aplicando um pulso de r.f. que satura a magnetização longitudinal antes de formar a imagem. Spins que retornam rapidamente à condição de equilíbrio térmico terão maior  $M_z$  disponível antes de formar imagem que aqueles que retornam lentamente ao equilíbrio. Este efeito fica claro se são formadas rapidamente imagens da mesma fatia, ou conjunto de fatias, pois o pulso de excitação da fatia previamente adquirida afeta a magnetização disponível para a fatia atual. Mais comumente no entanto, imagens ponderadas em  $T_1$  são formadas por pulsos de IR (Inversão Recuperação). Como foi visto na Subseção 2.1.2, um pulso de  $180^\circ$  inverterá a magnetização longitudinal, ao mesmo tempo não produzirá nenhuma magnetização transversal. O retorno ao equilíbrio de  $M_z$  é governado pela Eq.(2.38). A magnetização é permitida retornar ao equilíbrio térmico por um tempo TI (Tempo de Inversão), depois que a imagem

é formada usando um pulso de  $90^\circ$ , e geralmente gradientes de imagem. A quantidade de sinal disponível dependerá da razão de retorno de  $M_z$ . Se a amostra tem spins com vários tempos diferentes de relaxação, é possível escolher TI tal que o sinal de spins com uma curva de retorno seja completamente nulo, ao mesmo tempo que fornece um bom contraste entre spins com outras curvas de retorno.

### Contraste em $T_2^*$

O problema em medir  $T_2$  com uma seqüência que suporta em um echo de gradiente (como EPI faz) é que existe o efeito de inomogeneidades do campo que elimina a magnetização transversal na mesma maneira que a relaxação spin-spin o faz. Se os spins em um único voxel não experimentam exatamente o mesmo campo, então a coerência de suas magnetizações será reduzida. O efeito combinado de relaxação spin-spin e um campo inomogêneo em magnetização transversal é caracterizado por outro tempo  $T_2^*$ , e o decréscimo do sinal é governado pela equação similar à equação 2.39

$$S = S_0 \exp(-t/T_2^*). \quad (2.53)$$

De fato, imagens ponderadas em  $T_2^*$  são desejáveis para aplicações de IRM funcional, como é explicado na próxima seção. Para mudar a ponderação em  $T_2^*$  de uma imagem, é necessário somente mudar TE. Quanto mais longo TE, maior a ponderação em  $T_2^*$ .

### Contraste em $T_2$

Como foi mencionado na Subseção 2.1.3, é possível obter o valor real de  $T_2$  refocando o efeito de campo inomogêneo sobre a magnetização transversal usando um eco de spins. Este tem o efeito de refocar o sinal após transcorrido um tempo  $2\tau$ . O contraste na imagem é dependente de  $T_2$ , utilizando um longo TE.

## 2.2.5 Instrumentação

Um equipamento de MRI é constituído de quatro componentes: magneto, bobinas de gradiente, transmissor e receptor de r.f., e o computador. Todo o controle do equipamento é feito por um computador. Um diagrama simplificado do sistema é mostrado na figura 2.6.

Desta forma, os pulsos de r.f. e de gradiente são especificados e depois são amplificados e enviados para as bobinas. O sinal de NMR, uma vez que tenha tido a fase sensivelmente



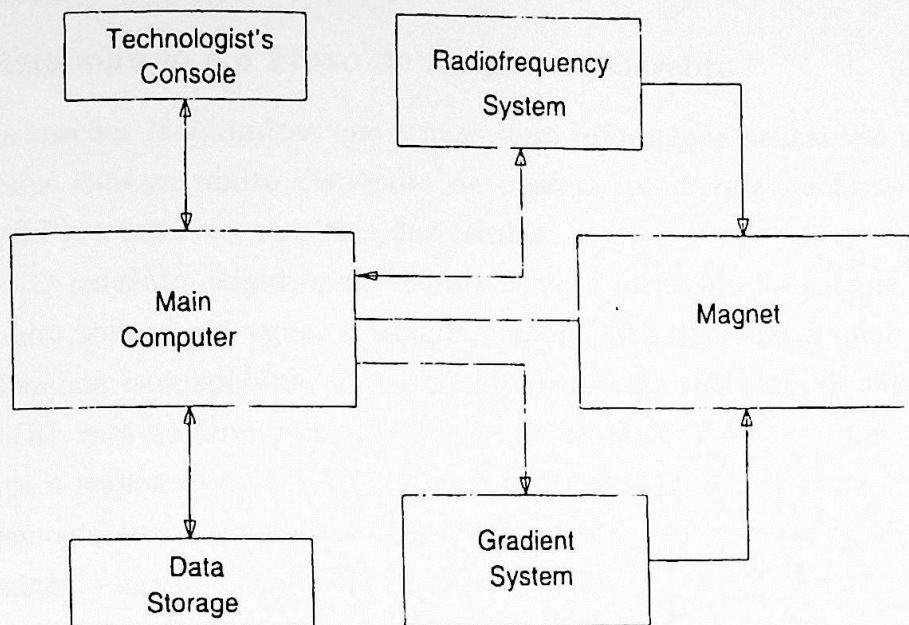


Figura 2.6: Diagrama esquemático de um sistema de MRI [13].

detectada, é alterado para um sinal digital por um conversor analógico-digital. O sinal digital é então enviado para um processador de imagem por transformada de Fourier e a imagem, formada por pixels, se em 2D, ou voxels, se em 3D, é mostrada no monitor.

Os dados brutos, que são os sinais antes da transformada de Fourier, são armazenados para que seja feita a aplicação de correções dos dados em pós processamento. Para permitir o uso rápido de transformada de Fourier, matrizes de tamanhos  $2^n$  são geralmente usadas.

## 2.3 Imagem Funcional (fMRI)

Diferentes técnicas permitem observar a atividade neuronal no homem. A magnetoencefalografia (MEG) e a eletroencefalografia (EEG) procuram analisar os sinais elétricos ligados à atividade dos neurônios (variações dos potenciais de ação-PA, correntes pós-sinápticas), que fornecem preciosas informações sobre a cronometria das diferentes áreas envolvidas. Para localizar as regiões implicadas, a técnica de imagem funcional por ressonância magnética (fMRI) baseia-se na observação das conseqüências locais dessas atividades elétricas sobre a circulação sanguínea e o consumo energético do cérebro.

### 2.3.1 Método BOLD

#### Metabolismo e o Fluxo de Sangue no Cérebro

As reações bioquímicas que transmitem informação neural via neurotransmissores e PA, todas causam muito dispêndio de energia, produzida na forma de adenosina trifosfato (ATP), a reserva energética das células.

O oxigênio exigido pelo metabolismo é fornecido no sangue. Como o oxigênio não é muito solúvel em água, o sangue contém uma proteína à qual o oxigênio pode se ligar, chamada hemoglobina. A parte importante da molécula de hemoglobina, para efeito de fMRI, está no átomo de ferro, limitado em uma estrutura orgânica, e é este mesmo átomo que é responsável pela coloração sangüínea. Quando uma molécula de oxigênio se liga à hemoglobina, é conhecida como oxi-hemoglobina (Hb-O<sub>2</sub>), e quando não existe nenhuma oxigênio ligado é chamada deoxi-hemoglobina (deoxi-Hb).

Para continuar com a alta energia que o cérebro exige, a entrega de oxigênio e o fluxo de sangue para este órgão são muito grandes. Embora o peso do cérebro seja somente 2% do peso do corpo, sua taxa de consumo de oxigênio é 20% da taxa do corpo, e de fluxo de sangue 15%. O fluxo de sangue para a massa cinzenta, que é uma área rica em sinapses, é da ordem de 10 vezes aquele para a massa branca por unidade de volume. A regulação do fluxo de sangue regional é pouco entendida, mas é conhecido que a atividade neural localizada resulta em um rápido crescimento seletivo no fluxo de sangue naquela área.

#### Contraste BOLD em Imagens de RM

Como o fluxo de sangue está estritamente relacionado à atividade neural, medida do fluxo sangüíneo cerebral regional (rCBF), ele é útil para estudar funções cerebrais. Existe um mecanismo de contraste, muito sensível, que depende do nível de oxigenação sangüínea, conhecido como contraste dependente do nível de oxigênio no sangue (BOLD). Os mecanismos fundamentais do contraste BOLD estão ainda indeterminados, entretanto existem hipóteses para explicar as mudanças de sinal observado.

A deoxi-hemoglobina é uma molécula paramagnética ao passo que a oxi-hemoglobina é diamagnética. A presença da molécula de deoxi-hemoglobina em um vaso sangüíneo causa uma diferença de susceptibilidade entre o vaso e o tecido vizinho. Tais diferenças de susceptibilidade causam defasagem do sinal de RM [14], levando a uma redução no valor de  $T_2^*$ . Em um experimento de imagem ponderada em  $T_2^*$ , a presença de deoxi-hemoglobina nos vasos sangüíneos provoca um escurecimento da imagem naqueles voxels



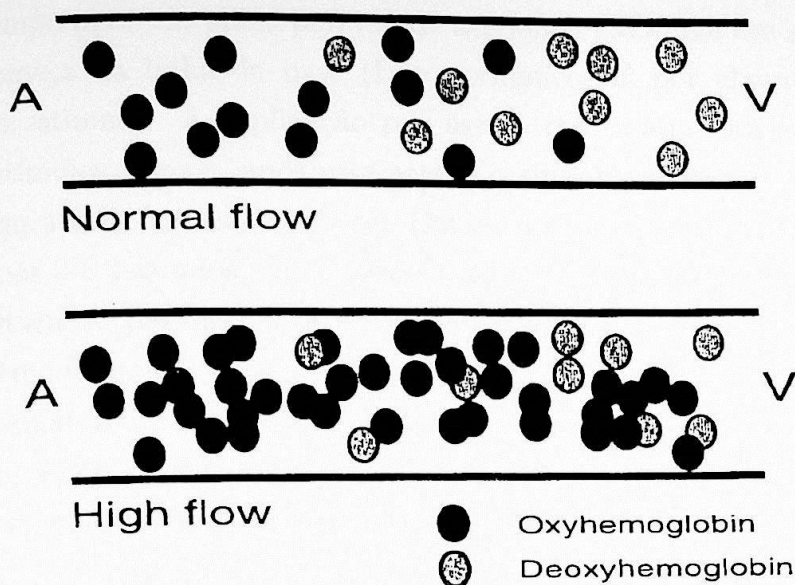


Figura 2.7: Uma ativação cerebral provoca localmente o aumento do fluxo sanguíneo sem aumento do consumo de oxigênio. Conseqüentemente, o oxigênio contido no sangue venoso (V) aumenta, resultando em um aumento na intensidade do sinal de RM [20].

contendo vasos [15, 16]. Visto que oxi-hemoglobina é diamagnética e não produz a mesma defasagem, mudanças em oxigenação do sangue podem ser observadas como mudanças de sinal em imagens ponderadas em  $T_2^*$  [17, 18, 19].

Seria esperado que em atividade neural, como consumo de oxigênio é maior, que a concentração de deoxi-hemoglobina no sangue também aumentasse, e o sinal de RM decrescesse. No entanto o que é observado é um crescimento em sinal, implicando em uma redução de deoxi-hemoglobina. Isto porque em atividade neural, o crescimento em fluxo de sangue do cérebro excede o crescimento em consumo de oxigênio, havendo um aumento de oxi-hemoglobina (figura 2.7). Assim o efeito volume em atividade neural é uma redução em deoxi-hemoglobina paramagnética, e um crescimento em sinal.

Resultados de estudos de PET e espectroscopia próximo do infravermelho (NIRS) tem ajudado a entender estes mecanismos. Tais estudos tem mostrado um aumento da razão da oxi-hemoglobina pela deoxi-hemoglobina em ativação. Um aumento na quantidade total de hemoglobina é também observado, refletindo no aumento em volume de sangue em ativação [21].

A evolução temporal das mudanças no sinal BOLD, representando as mudanças que a técnica detecta, é conhecida como *função de resposta hemodinâmica* (HRF) ao estímulo aplicado. A figura (2.8) ilustra um modelo de resposta hemodinâmica como função do

tempo para um único período de atividade (30 s) em um pixel da imagem. O sinal medido começa na linha de base (I) e permanece lá por alguns segundos após a apresentação do estímulo. A explicação por esse atraso ainda está sob investigação. Em exames de estímulos visuais, após esse atraso o sinal decresce por um curto intervalo de tempo, na figura essa fase não aparece. Então, o sinal começa a crescer (II), atingindo um máximo após 3-8 segundos. Se a apresentação do estímulo continuar, o sinal estabiliza-se (III). A altura do patamar está associado, como um dos fatores, com a intensidade da resposta hemodinâmica, ou seja, com o grau de atividade neural. Quando a estimulação termina, o sinal ainda permanece elevado por um tempo. Então o sinal começa a decrescer (IV) lentamente, geralmente seguido por uma depressão abaixo da linha de base. Enfim, o sinal retorna a linha de base (V).

A forma da resposta hemodinâmica é hipotética, pois suas características ainda estão incompletas. Mas, nos últimos anos, o interesse em explorar a natureza fisiológica do sinal BOLD e sua dependência na duração e no tipo de estímulo neural vem crescendo enormemente.

### Mapeamento Funcional

Para estudar funções cerebrais via fMRI é necessário adquirir repetidamente imagens do cérebro, enquanto é apresentado um estímulo ao indivíduo ou ele é solicitado a realizar alguma tarefa. O sucesso do exame depende de três aspectos: da seqüência de pulsos aplicada, do arranjo experimental e da forma como os dados são analisados.

As principais fontes de ruído em MRI são o ruído térmico no receptor e o ruído do "corpo" que descreve flutuações aleatórias de spins que induzem f.e.m na bobina receptora. Esse ruído da imagem é caracterizado pela razão do sinal BOLD pelo ruído (SNR) e a razão do contraste pelo ruído (CNR). A SNR pode distinguir para diferentes tecidos na imagem e a CNR é calculada entre dois tecidos particulares.

Mudanças no sinal devido à ativação dependem da intensidade do campo magnético estático aplicado. Isto porque diferenças de susceptibilidade produzem um efeito de defasagem muito grande do sinal em campos muito altos. Mas em campos típicos da ordem de  $1.5T$ , as alterações no sinal são muito sutis significando em torno de 3 – 4% do sinal que reflete a linha de base [23]. Quando a intensidade do campo cresce a CNR cresce, ou seja, a intensidade do contraste BOLD cresce mais rapidamente que o ruído do sistema, de modo que campos muito intensos são desejáveis [24].

O aspecto mais importante da seqüência de pulsos é que se deve adquirir imagens



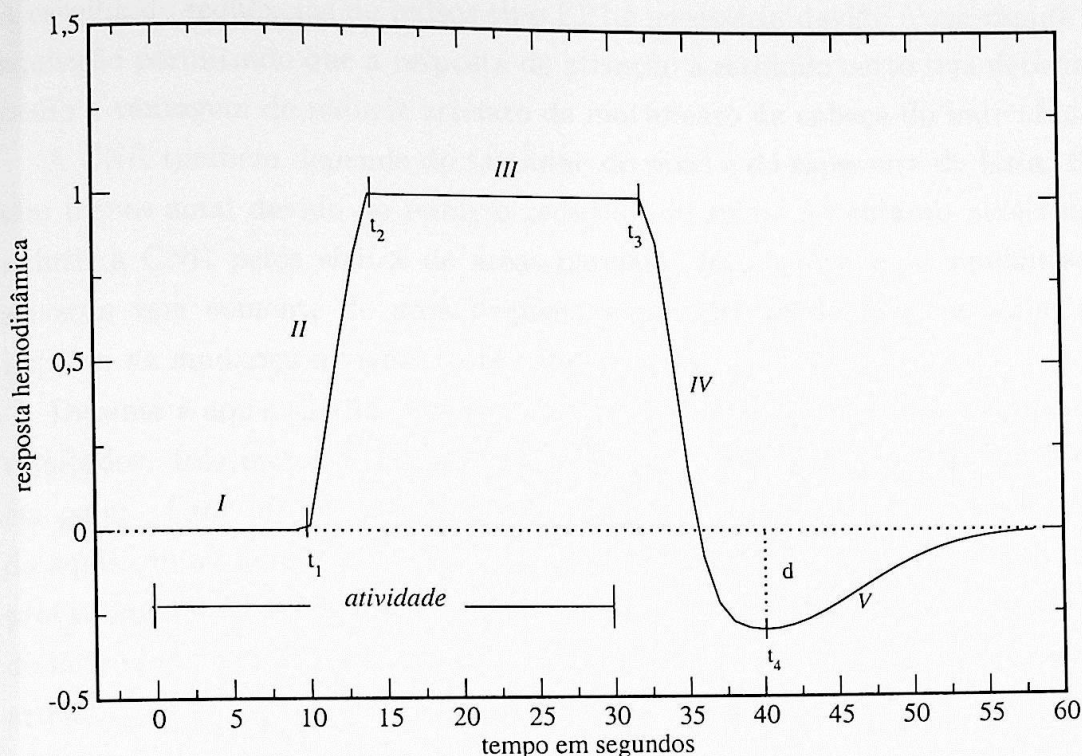


Figura 2.8: Modelo da resposta hemodinâmica. A parametrização usada para modelar a função de resposta hemodinâmica (HRF) foi tirada de [22], mas com os parâmetros fixos e sugeridos como segue:

$$h(t) = \begin{cases} 0.0 & , t < t_1 & (I) \\ (1/0.6) \exp\left(-\left(\frac{t-t_2}{T_2}\right)^2\right) - 0.6 & , t \in [t_1, t_2[ & (II) \\ 1.0 & t \in [t_2, t_3[ & (III) \\ (1+d) \exp\left(-\left(\frac{t-t_3}{3}\right)^2\right) - d & t \in [t_3, t_4[ & (IV) \\ -(d) \exp\left(-\left(\frac{t-t_4}{9}\right)^2\right) & t \in [t_4, (T_{at} + T_{rep})[ & (V) \end{cases}$$

onde os pontos temporais estão relacionados aos parâmetros do modelo, da seguinte forma:  $t_1 = T_1 = 10$  s;  $t_2 = T_1 + T_2 = 15$  s;  $t_3 = T_{at} + T_3 = T_{at} + 2$ ;  $t_4 = T_{at} + T_3 + T_4 = T_{at} + 10$ . Sendo que  $T_1$  representa o parâmetro de atraso anterior (início do estímulo);  $T_2$  é o parâmetro de subida do sinal hemodinâmico (durante o estímulo);  $T_3$  é o parâmetro de atraso posterior (após o término da ativação);  $T_4$  é o parâmetro de descida do sinal (retornando a base de linha) e  $d$  é a amplitude da depressão do sinal, onde utilizamos  $d = 0.3$ . O tempo de ativação  $T_{at}$  e o tempo de repouso  $T_{rep}$  são parâmetros fornecidos de acordo com o protocolo de aquisição dos dados, de forma que nesta figura, utilizamos  $T_{at} = T_{rep} = 30$  segundos.

ponderadas em  $T_2^*$ . Isto significa que um echo de gradiente é geralmente o mais utilizado. A escolha de seqüências de pulsos tipo EPI é necessária devido à sua rápida velocidade de aquisição permitindo que a resposta de ativação a estímulo curto seja detectada e também tendo a vantagem de reduzir artefato de movimento da cabeça do indivíduo.

A CNR também depende do tamanho do pixel e da espessura da fatia. Pixels menores têm menos sinal devido ao número reduzido de spins, no entanto pixels maiores podem reduzir a CNR pelos efeitos de áreas parciais. Isso ocorre se as mudanças no sinal sob ativação vem somente de uma pequena região dentro do pixel, e assim produz menos impacto na mudança no sinal total naquele pixel.

Durante a aquisição das imagens existem vários efeitos fisiológicos que podem afetar os resultados. Isso inclui os batimentos cardíacos, a respiração e o movimento do indivíduo em geral. Este último problema pode ser atacado de duas maneiras, ou no momento da aquisição ou no pós-processamento das imagens. Estratégias de pós-processamento são provavelmente a melhor maneira para tentar solucionar estes problemas [25]. O movimento do indivíduo também pode reduzir o contraste da SNR em imagens funcionais, e introduzir artefato nos mapas de ativação mesmo se o movimento não está correlacionado com o estímulo.

Outra fonte de artefato em imagens funcionais é o sinal que vem da drenagem de vasos sanguíneos. Como as imagens de echo de gradiente são sensíveis a vasos de diâmetros de micrometros a milímetros, pode ser difícil distinguir entre sinais do tecido e aqueles das veias, que poderiam estar uma distância grande do local de ativação [26]. Uma maneira de reduzir o sinal de grandes vasos seria usar uma seqüência de echo de spins. Este é sensível a efeitos somente de  $T_2$  e elimina os efeitos de defasagem dos grandes vasos [27, 28]. O uso de uma seqüência de echo de spins resulta em uma redução no sinal BOLD autêntico do tecido, de modo que muitas vezes é melhor adquirir um conjunto separado de imagens que são sensíveis a grandes vasos sanguíneos, e usar isto para decidir se rejeita o sinal.

Muito poderia ainda ser dito sobre artefatos e otimização de imagens cerebrais em MRI, mas não é nosso objetivo neste trabalho.

### 2.3.2 Arranjo Experimental em Bloco

Atualmente a fMRI é utilizada principalmente como uma ferramenta de pesquisa, mas promissoras aplicações clínicas incluem um planejamento para o tratamento cirúrgico, como ferramenta de mapeamento pré-cirúrgico [29].

A proposta de muitos exames de fMRI é localizar regiões do cérebro que estão envolvidas



na execução de uma determinada tarefa e estudar os padrões espaço-temporais de ativação. Com o objetivo de conseguir atribuir mudanças no sinal a mudanças na função cerebral, estas mudanças necessitam ser muito diferentes da flutuação gradativa no sinal resultante de instabilidade instrumental.

Neste trabalho consideramos apenas o arranjo de apresentação de estímulo mais comum, conhecido como arranjo experimental em bloco. Nele, blocos de várias imagens são adquiridos sucessivamente, alternando-se períodos de estímulo (A) com períodos de repouso (R). Uma proposta é fazer a duração destes períodos ser longa o suficiente para acomodar a resposta hemodinâmica, e assim um valor em torno de 30 segundos é geralmente escolhido. As áreas no cérebro que são ativadas pela tarefa A e não pela tarefa R podem então ser identificadas olhando para as regiões onde a intensidade da imagem segue o arranjo do estímulo. Este arranjo pode ser considerado como uma função de referência e assim a análise de fMRI consiste em determinar a resposta (em termos do contraste BOLD) resultante desta função.

Os primeiros estudos concentraram no sistema visual [23]. Aqui as tarefas eram A: olhar para uma fonte luminosa, e R: olhar para a escuridão, servindo como um período de descanso para a tarefa visual sendo estudada. Estes estudos de estímulo visual identificaram corretamente o córtex visual primário, já estudados por PET e por registros neurais em animais, e portanto serviram para verificar a fMRI.

Geralmente as duas tarefas são arranjadas de tal maneira que elas se diferenciem significativamente somente naquela função que está sendo investigada. Por exemplo, se a proposta do exame é identificar regiões ativadas por funções motoras, as condições A e R devem consistirem de atividades exclusivamente associadas à motricidade, evitando a contaminação por outros processos cerebrais.

Uma dificuldade com o arranjo de tais estudos é que outras funções cerebrais involuntárias, tais como atenção, intenção de fazer ou agir, e emoções, tais como frustração com a tarefa, podem interferir com o sinal de interesse. O nível de envolvimento de cada uma destas funções nas duas tarefas pode diferenciar entre indivíduos ou mesmo entre ocasiões em uma maneira imprevisível. Um outro problema é que a mudança na oxigenação do sangue devido à função de interesse pode ser mínima se o indivíduo acha a tarefa fácil. Por fim, pode ser completamente difícil controlar ou mesmo descobrir até que ponto o indivíduo estava realmente desempenhando a tarefa ou estava distraído com outros pensamentos; pedindo para o indivíduo pressionar botões como parte de suas tarefas ou entrevistando o indivíduo após a aquisição dos dados são maneiras de tentar monitorar o desempenho.

### Modelo de Séries Temporais

O sinal BOLD de um pixel  $p$  qualquer da imagem em um instante  $t$  pode ser descrito simplesmente como

$$s(p, t) = s_{rep}(p, t) + s_{ativ}(p, t) + \Delta(p, t), \quad (2.54)$$

em que  $s_{rep}$  representa a intensidade do sinal durante o período de repouso, ou a linha de base,  $s_{ativ}$  representa a intensidade do sinal durante o período de estímulo, e  $\Delta$  inclui as flutuações devidas à combinação dos vários artefatos possíveis.

Neste trabalho, utilizamos um modelo simples para representar esse sinal. Para cada pixel  $p(1, \dots, P)$ , a série temporal  $(y_{pt}, t = 1, \dots, T)$  dos sinais BOLD, sendo  $T$  é o número de pontos no tempo, é suposto obedecer a seguinte relação linear:

$$y_{pt} = b_p + h_t A_p a_p + \epsilon_{pt} \quad \epsilon_{pt} \sim N(0, \sigma^2) \quad (2.55)$$

$$p = 1, \dots, P, t = 1, \dots, T.$$

Aqui, a variável  $b_p$  corresponde à componente  $s_{rep}$ , que é tipicamente bem determinada dos dados; os arranjos experimentais que incorporam períodos de repouso no início e no fim do experimento fazem estimativas de  $b_p$  particularmente precisas. A componente  $s_{ativ}$  foi caracterizada pela variável  $h_t$  que representa a função de referência para o estímulo, pelo parâmetro  $A_p$  que é a amplitude desta função de referência e por  $a_p$  que é um número binário, ou seja, quando  $a_p = 0$  representa inatividade e quando  $a_p = 1$  representa atividade. E por fim, um ruído gaussiano independente,  $\epsilon_{pt}$ , é suposto modelar a componente  $\Delta$ .

Vimos que o arranjo experimental aplicado pode ser considerado como uma função de referência. Para arranjos em bloco, é comum representar essa função de referência como simplesmente uma função “box-car” que é 1 durante a condição A e  $-1$  durante a condição R, como veremos na próxima subseção. Mas, na realidade, como a função de referência leva em conta a forma da resposta hemodinâmica como descrita anteriormente, ela está longe de ser um “box-car” ideal. Portanto, a variável  $h_t$  na equação 2.55, representa o modelo da HRF como ilustrado na figura (2.8).

A série temporal em um pixel pode apresentar deriva (*drift*). O sinal BOLD medido pode assumir valores linearmente maiores ao longo do tempo. Este efeito provavelmente tem uma variedade de causas, mas estas ainda não estão completamente identificadas. De acordo com Genovese [30] esse efeito de deriva pode exibir estruturas espaciais interessantes. No modelo que utilizamos, equação 2.55, consideramos que esse efeito nas séries temporais



tenha sido corrigido em um pré-processamento.

### 2.3.3 Análise Convencional de Dados

Existem três etapas para a análise dos dados brutos de qualquer exame de fMRI. Primeiramente, existem os passos de pré-processamento, que podem ser aplicados aos dados para melhorar a detecção de áreas ativadas na imagem. Estes incluem os ajustes das imagens, para corrigir o movimento da cabeça do indivíduo durante o exame, e a aplicação de filtros aos dados para melhorar a SNR, além de corrigir o efeito de deriva. Depois, a análise estatística, que detecta os pixels na imagem que mostram uma resposta ao estímulo, é feita. E finalmente, as imagens de ativação são apresentadas com significância superior a um certo limiar.

Porém, nesta subseção nos concentramos apenas na análise estatística dos dados, comentando alguns métodos existentes para tentar detectar a ativação.

Uma variedade de métodos são empregados para analisar estatisticamente dados de fMRI, com o objetivo de identificar quais as regiões do cérebro cujo fluxo sanguíneo se modificou de modo significativo em resposta a uma determinada tarefa.

Um dos métodos mais simples para obter resultados de um arranjo experimental em bloco é realizar uma subtração. Isso é feito tirando a média de todas as imagens adquiridas durante os períodos de ativação, e subtraindo a média de todas as imagens em repouso. A desvantagem de tal técnica é que ela é extremamente sensível ao movimento da cabeça do indivíduo, resultando em grandes quantidades de artefatos na imagem.

O método de subtração não produz um peso estatístico aos resultados finais, de modo que em vez da subtração direta é mais comum usar um teste-t de *Student*. A estatística deste teste é calculada de pixel a pixel, para uma série temporal  $y$ , usando a fórmula

$$t = \frac{\bar{y}_a - \bar{y}_r}{\sigma_{\bar{y}_a - \bar{y}_r}} \quad (2.56)$$

onde

$$\sigma_{\bar{y}_a - \bar{y}_r} = \sqrt{\frac{\sum(y_a - \bar{y}_a)^2 + \sum(y_r - \bar{y}_r)^2}{n_a + n_r - 2} \left( \frac{1}{n_a} + \frac{1}{n_r} \right)}, \quad (2.57)$$

é a variância total. O índice 'a' refere-se a  $n_a$  imagens adquiridas durante o período de ativação, e 'r' refere-se a  $n_r$  imagens durante o período de repouso.

## Técnicas de Correlação

Um método estatístico muito comum, principalmente para arranjos experimentais em bloco, é o coeficiente de correlação da série temporal do pixel com uma função de referência definida para aproximar a forma da resposta hemodinâmica [31]. Este é menos sensível a outras mudanças fisiológicas durante o exame, e ao movimento. Para a série temporal  $y$ , e a função de referência  $r$ , o coeficiente de correlação de um pixel é calculado como:

$$cc = \frac{\sum(y - \bar{y})(r - \bar{r})}{\sqrt{\sum(y - \bar{y})^2 \sum(r - \bar{r})^2}}, \quad (2.58)$$

e assume um valor 1 para correlação perfeita, um valor 0 para nenhuma correlação, e um valor -1 indica uma anti-correlação perfeita entre os dois sinais.

A escolha de uma forma da função de referência é vital para o sucesso desta técnica em detectar ativações. É comum aproximar a forma da resposta hemodinâmica com uma função "box-car", mas como já mencionamos, esta aproximação não é realística. Outras escolhas são possíveis, mas ainda simples, como função seno, ou mesmo usar a resposta de pixels de outras regiões da imagem.

Uma aproximação mais geral e um das mais usadas para a resposta hemodinâmica, é a convolução da densidade de Poisson com a HRF, como sugerida por Friston [32], que leva em conta o atraso e o crescimento suave do fluxo sanguíneo. Como em geral cada fatia do volume da imagem não é adquirida ao mesmo tempo, a intensidade relativa da ativação no instante em que cada fatia foi adquirida é modelada, convoluindo a função de referência com a função de Poisson. Então a partir desta série, os coeficientes de correlação podem ser calculados de fatia a fatia, construindo a HRF a partir dos pontos apropriados na série temporal modelada.

Além da função de Poisson, também são utilizadas função tipo Gama e Gaussiana.

Lange e Zeger [33] usa um modelo de regressão para a série temporal do pixel no domínio espectral em que a forma da HRF é permitida variar. A vantagem deste método é que ajustando as séries temporais neste domínio, o impacto de autocorrelação e *drift* nos resultados é reduzido. Para mais discussões sobre este método, observe [34, 35].

Todos estes métodos convencionais comentados acima, são aplicados mais efetivamente para arranjos experimentais em bloco e o modelo da resposta hemodinâmica é completamente simples e não leva em conta várias características importantes do sistema. Além disso, a escolha de limites de significância usados para apresentar as regiões ativadas na imagem é geralmente arbitrária, parecendo não ter nenhum fundamento para esta quan-



tificação formal.

## 2.4 Inferência Bayesiana - Análise de Dados

Nesta seção uma apresentação geral é feita sobre probabilidade e inferência bayesiana, pois nos próximos capítulos estas teorias são aplicadas ao problema de separação de gaussianas e ao problema central deste trabalho: detecção de regiões do cérebro ativadas durante determinada tarefa imposta ao sujeito.

### 2.4.1 Teoria de Probabilidade Bayesiana

Para um melhor entendimento sobre este assunto, é importante dizer que 'aleatoriedade' na teoria bayesiana representa nossa incapacidade de prever algo que, por sua vez, reflete nossa falta de conhecimento sobre o sistema de interesse. Ou seja, uma probabilidade é atribuída de acordo com a informação que temos sobre este sistema.

O teorema de Bayes fornece a regra para atualizar a opinião sobre uma hipótese  $H$  dado os dados  $D$ , e a informação *background* relevante (contexto)  $I$ :

$$p(H|D, I) = \frac{p(H|I) \times p(D|H, I)}{p(D|I)}. \quad (2.59)$$

Aqui, a quantidade  $p(H|I)$  é chamada de probabilidade a priori, ou simplesmente priori, que é a informação que temos sobre a verdade da hipótese antes da análise dos dados corrente. Isto é modificado pela verossimilhança, ou  $p(D|H, I)$ , e produz a probabilidade posterior,  $p(H|D, I)$ , representando nosso estado de conhecimento sobre a verdade da hipótese levando em conta a efeito dos dados no contexto  $I$ . O último termo,  $p(D|I)$ , não depende explicitamente da hipótese  $H$ , e pode ser considerado como uma constante de normalização. A informação  $I$  é a conjunção de todas as outras proposições relevantes para determinar  $p(H|I)$  e  $p(D|I)$ .

Note que todas estas probabilidades são condicionais - elas especificam o grau de nossa opinião em alguma (s) proposição segundo a suposição que algumas outras proposições são verdadeiras.

O teorema de Bayes é uma simples consequência da regra do produto da teoria de probabilidade. A regra do produto fornece a probabilidade da conjunção lógica 'and' de

duas proposições A e B, escrito como A,B – isto é:

$$p(A, B|I) = p(A|B, I) \times p(B|I) = p(B|A, I) \times p(A|I). \quad (2.60)$$

A equação 2.59 é obtida rearranjando os termos na igualdade acima. Usando a regra do produto diretamente, nós podemos estender o teorema de Bayes para múltiplas atualizações seqüenciais:

$$\begin{aligned} p(H|D_1, D_2, D_3, I) &= \frac{p(H|I) \times p(D_1, D_2, D_3|H, I)}{p(D_1, D_2, D_3|I)} = \\ &= \frac{p(H|I) \times p(D_1|H, I) \times p(D_2|D_1, H, I) \times p(D_3|D_2, D_1, H, I)}{p(D_1|I) \times p(D_2|D_1, I) \times p(D_3|D_2, D_1, I)}. \end{aligned} \quad (2.61)$$

Note a dificuldade aqui. Quando cada nova parte dos dados é fatorada, seu efeito é condicional em todos os dados previamente considerados. Esta dificuldade é geralmente superada quando supomos que os dados são independentes. Isto significa que, dado  $I$ , conhecendo que  $D_2$  é verdadeiro nada nos diz sobre  $D_1$ , e vice-versa. Logo,  $D_2$  não contém nenhuma informação relacionada a  $D_1$  que já não esteja presente em  $I$ : matematicamente, nós podemos escrever isto como

$$p(D_2|D_1, I) = p(D_2|I) \quad e \quad p(D_1|D_2, I) = p(D_1|I). \quad (2.62)$$

Assim, a regra do produto se reduz a:

$$p(D_1, D_2|I) = p(D_1|I) \times p(D_2|I). \quad (2.63)$$

E, assim, quando os múltiplos dados  $D_i$  são independentes, a versão de atualização múltipla do teorema de Bayes se reduz a:

$$p(H|D_1, D_2, D_3 \dots) = \frac{p(H|I) \times p(D_1|H, I) \times p(D_2|H, I) \times p(D_3|H, I) \dots}{p(D_1|I) \times p(D_2|I) \times p(D_3|I) \dots}. \quad (2.64)$$

Esta equação simplifica extremamente o problema de aplicar os dados em uma hipótese. Mas ela deve ser usada com cuidado, pois a independência condicional não é sempre válida.

A segunda regra fundamental da teoria da probabilidade é a regra da soma. A regra da soma fornece a probabilidade da disjunção 'ou' de duas proposições A e B, escrito como A+B:

$$p(A + B|I) = p(A|I) + p(B|I) - p(A, B|I). \quad (2.65)$$



Esta é a regra geral da soma e ela é uma das mais úteis em aplicações.

As regras da soma e do produto são as regras quantitativas necessárias para que a teoria da probabilidade seja consistente com a lógica aristotélica. Então quando nossas hipóteses são verdadeiras ou falsas, o cálculo de probabilidade se reduz ao cálculo de predicado. Jaynes [36] dá um excelente desenvolvimento mostrando como estas regras podem ser obtidas das condições fundamentais de uma opinião sensata.

Um conjunto de  $n$  hipóteses é dito ser mutuamente exclusivos com respeito a  $I$  se

$$p(H_i, H_j|I) = 0 \quad \text{para } i \neq j. \quad (2.66)$$

Isto é válido quando  $I$  é tal que dois dos  $H_j$  não podem ser simultaneamente verdadeiros. É equivalente a dizer:

$$p(H_i|H_j, I) = p(H_j|H_i, I) = 0 \quad \text{para } i \neq j. \quad (2.67)$$

Assim a regra da soma se reduz a:

$$p(H_1 + H_2|I) = p(H_1|I) + p(H_2|I). \quad (2.68)$$

Um conjunto de hipóteses é dito ser completo com respeito a  $I$  se:

$$p(H_1 + H_2 + \dots + H_n|I) = 1. \quad (2.69)$$

Isto é válido quando  $I$  é tal que pelo menos um dos  $H_j$  deve ser verdadeiro. Alternativamente, podemos indicar isto como  $(H_1 + H_2 + \dots + H_n) = V$ .

Quando um conjunto de  $H_j$  é mutuamente exclusivo e completo dado a informação  $I$ , nós podemos usar a regra da soma 2.68 para eliminar uma proposição somando sobre todas as possibilidades. Esta eliminação de proposição pela soma é chamada de marginalização:

$$\begin{aligned} \sum_{j=1}^n p(D, H_j|I) &= p(D, H_1|I) + p(D, H_2|I) + \dots + p(D, H_n|I) \\ &= p(H_1|D, I)p(D|I) + p(H_2|D, I)p(D|I) + \dots + p(H_n|D, I)p(D|I) \\ &= [p(H_1|D, I) + p(H_2|D, I) + \dots + p(H_n|D, I)]p(D|I) \\ &= p(H_1 + H_2 + \dots + H_n|D, I)p(D|I) \quad \text{exclusividade} \\ &= p(V|D, I)p(D|I) \quad \text{completude} \\ &= p(D|I) \end{aligned} \quad (2.70)$$

que fornece o denominador para o teorema de Bayes, equações 2.59.

Generalizando a equação acima, com  $n \rightarrow \infty$ , a função de distribuição de probabilidade  $p(D, H_j|I)$ , torna-se tecnicamente uma função de densidade de probabilidade (pdf). Com o objetivo de uniformizar nossa notação entre o caso contínuo e caso discreto, usaremos ' $p$ ' para denotar probabilidades. Assim para o caso contínuo, a equação acima torna-se:

$$\int_{-\infty}^{+\infty} p(D, H|I)dH = p(D|I). \quad (2.71)$$

e a equação 2.69 torna-se:

$$\int_{-\infty}^{+\infty} p(H|D, I)dH = 1. \quad (2.72)$$

A marginalização é também uma técnica muito poderosa em análise de dados pois nos permite lidar com os efeitos de parâmetros de incômodo. Estes são parâmetros que claramente afetam a análise, mas que não são de nenhum interesse final no cálculo corrente. Marginalizando sobre o conjunto de valores possíveis de parâmetros mutualmente exclusivo e completo, nós podemos considerar seus efeitos ao eliminá-los do resultado final.

As regras da soma e do produto, a independência condicional, e a marginalização fornecem as ferramentas básicas para atualizar a opinião bayesiana. Mas com o objetivo de aplicar estas ferramentas nós primeiro devemos ter a verossimilhança  $p(D|H, I)$  dos dados de acordo com cada hipótese e as priori  $p(H|I)$  da hipótese independente dos dados.

As probabilidades a priori têm sido ignoradas por estatísticos. Isto é evidentemente devido ao fato que, até recentemente, existia pouca concordância em como especificar consistentemente tais probabilidades, e da natureza subjetiva das probabilidades a priori. Trabalhos recentes sobre a testabilidade de informações a priori, como por exemplo princípios de máxima entropia, fizeram muito para corrigir esta situação. Informações mínimas a priori para muitos dos tipos de parâmetro padrão estão agora em acordo, mas a priori informativa é ainda uma área de pesquisa em atividade. Na prática, para qualquer problema real, existe conhecimento suficiente sobre o sistema de interesse para especificar pouca probabilidade a priori normalizável. Se a probabilidade a posteriori é ainda fortemente dependente da priori considerada, então mais dados devem ser obtidos, ou as probabilidades a priori devem ser melhoradas baseadas na consideração cuidadosa do sistema de interesse.

O teorema de Bayes é somente uma parte da inferência bayesiana como desenvolvida por Laplace, Jeffreys, Jaynes, e sucessores. Outros conceitos de probabilidade, tais como, as regras do produto e da soma, o conceito de independência condicional e a técnica de



marginalização são também necessárias. Nenhum destes conceitos podem ser usados até que se possa extrair probabilidades a priori de uma proposição do problema; portanto o uso de probabilidades a priori é uma das características diferenciadas da inferência bayesiana.

### 2.4.2 Estimativa de Parâmetro

A estimativa de vários parâmetros é um problema comum em análise de dados; alguns destes parâmetros são de interesse para o problema em questão, mas outros não.

Para fazer a melhor estimativa de um conjunto de  $r$  parâmetros  $\{\omega_j\}$ , isto é, os valores que foram efetuados quando os dados foram retirados, precisamos calcular o pdf posterior:  $p(\{\omega_j\}|D, I)$ . Então, do teorema de Bayes (equação 2.59), temos:

$$p(\{\omega_j\}|D, I) \propto p(\{\omega_j\}|I) \times p(D|\{\omega_j\}, I). \quad (2.73)$$

Note que nós omitimos o denominador  $p(D|I)$ , pois ele não envolve  $\{\omega\}$  explicitamente, e substituímos o sinal de igualdade por um de proporcionalidade. Se for preciso, nós podemos avaliar a constante de normalização ausente pela equação 2.69:

$$\int_{-\infty}^{+\infty} p(\{\omega_j\}|D, I) d\omega_1 \dots d\omega_r = 1. \quad (2.74)$$

É conveniente trabalhar com o logaritmo do pdf posterior, já que ele é uma função monotônica e portanto o seu máximo ocorre no mesmo lugar que aquele de  $p$ . Então a equação 2.73 passa a ser:

$$L = \ln[p(\{\omega_j\}|D, I)] = \ln[p(\{\omega_j\}|I) \times p(D|\{\omega_j\}, I)]. \quad (2.75)$$

Uma das maneiras de inferir os valores dos parâmetros, dado os dados e a informação *background* relevante, é encontrar o máximo de  $L$ . Isto porque, a probabilidade (densidade) associada com qualquer conjunto particular de valores para os parâmetros é uma medida de quanto nós acreditamos que eles realmente se encontram na vizinhança daquele limite, assim a exigência de encontrar o máximo de  $L$  para os valores ótimos  $\{\omega_{j0}\}$  pode ser expressa pelas equações simultâneas:

$$\left. \frac{\partial L}{\partial \omega_j} \right|_{\{\omega_{j0}\}} = 0, \quad (2.76)$$

em que,  $j = 1, 2, \dots, r$ . Para sermos rigorosos, nós devemos também checar o sinal da segunda derivada parcial para assegurar que  $\{\omega_{j0}\}$  representa um máximo ao invés de um mínimo ( ou um ponto de inflexão):

$$\left. \frac{\partial^2 L}{\partial \omega_j^2} \right|_{\{\omega_{j0}\}} < 0. \quad (2.77)$$

Podemos também ter o caso em que estamos interessadas apenas em um dos parâmetros, por exemplo  $\omega_1$ , por algum motivo intrínseco no problema em questão, de forma que todos os outros parâmetros são simplesmente um *nuisance*, então nós precisamos apenas calcular a probabilidade posterior do nosso parâmetro de interesse,  $\omega_1$ :  $p(\omega_1|D, I)$ . Como vimos na subseção anterior, isto pode ser obtido através da técnica de marginalização (2.71):

$$p(\omega_1|D, I) = \int_{-\infty}^{+\infty} p(\{\omega_j\}|D, I) d\omega_2 \dots d\omega_r. \quad (2.78)$$

Um problema que surge quando queremos encontrar o máximo do pdf posterior, é que nem sempre esta otimização pode ser feita analiticamente, mas muitas vezes utilizando o computador. Para o caso unidimensional (um parâmetro), uma maneira de resolver este problema é fazer o gráfico da probabilidade posterior, em que o eixo-x representa os valores possíveis do parâmetro e o eixo-y a avaliação da probabilidade posterior para cada um deles. Esta aproximação direta também pode ser usada para o problema de dois parâmetros, mas não serve para o caso a partir de três parâmetros, sendo exigido o uso de métodos numéricos mais eficientes. Detalhes sobre os procedimentos de otimização podem ser vistos em Sivia [37], que inclui exemplos para o problema de dois parâmetros em uma linguagem bayesiana.

### 2.4.3 Teste de Hipótese

Os princípios exigidos para testar uma hipótese  $H_1$  são os mesmos daqueles usados para estimar parâmetros, ou seja, tudo é obtido pela aplicação das regras da soma e do produto da teoria de probabilidade.

O teste de hipótese bayesiano é muito simples. Suponhamos que queremos testar  $H_1$  que pode ser alguma suposição sobre o mecanismo físico que gera os dados. Mas, fundamentalmente, causalidade física não é um ingrediente essencial do problema: o que é essencial é que exista somente algum tipo de conexão lógica entre a hipótese e os dados. Assim, para quantificar quanto nós acreditamos que esta proposição é verdadeira, baseada



nos dados  $D$  e em toda a informação *background* relevante  $I$ , nós necessitamos calcular a probabilidade posterior  $p(H_1|D, I)$ , através do teorema de Bayes, como mostramos na equação 2.59. Agora, se queremos somente o mérito relativo de  $H_1$  comparado à outra hipótese  $H_2$ , então necessitamos dividir a probabilidade posterior de uma pela outra:

$$O(H_1|D, I) \equiv \frac{p(H_1|D, I)}{p(H_2|D, I)}. \quad (2.79)$$

Aqui, a razão da probabilidade que  $H_1$  é verdadeira pela probabilidade que  $H_2$  é verdadeira leva um nome técnico. Nós a chamamos de '*odds*' sobre a proposição  $H_1$  simbolizada pelo termo do lado direito da equação.

Para estimar os '*odds*' na equação 2.79, devemos substituir o numerador pela equação 2.59 e aplicar o teorema de Bayes ao denominador; isto fornece:

$$\frac{p(H_1|D, I)}{p(H_2|D, I)} = \frac{p(H_1|I) \times p(D|H_1, I)}{p(H_2|I) \times p(D|H_2, I)}, \quad (2.80)$$

porque o termo  $p(D|I)$  de cima cancela com o de baixo. Se esta razão-*odds* for maior que um, então nós escolhemos a hipótese  $H_1$ ; se ela for menor que um, escolhemos  $H_2$ ; e se ela for da ordem de um, então os dados correntes são insuficientes para fazer um julgamento informativo.

Como foi dito anteriormente, para atribuir as probabilidades envolvendo as medidas experimentais,  $p(D|H_1, I)$  e  $p(D|H_2, I)$  na equação acima, deve existir uma conexão lógica entre os dados e as hipóteses  $H_1$  e  $H_2$ : quanto menor esta conexão, menor a probabilidade correspondente.

## Capítulo 3

# Métodos Preliminares

A localização de regiões ativadas no cérebro é vista em fMRI como um problema de classificação, onde cada pixel é classificado como ativo ou inativo com respeito à tarefa de interesse. Sendo assim, neste capítulo aplicamos o teste de hipótese bayesiano, introduzido no capítulo anterior, para resolver este problema. Mas, antes disto, na primeira seção, avaliamos a eficiência deste teste, lidando com um problema de separação de duas distribuições gaussianas independentes, que é a base do problema de localização. E na segunda seção, desenvolvemos o método de pixel independente que detecta ativação de pixel a pixel, simplesmente aplicando o teste de hipótese a dados simulados de fMRI.

O resultado principal apresentado neste capítulo é que o desempenho (medido com a probabilidade de atribuição correta de atividade) depende fortemente do valor do a priori usado. Isto sugere o tema central deste trabalho. No próximo capítulo os esforços serão direcionados para obter estimativas dos valores corretos dos a priori usando propriedades globais da imagem.

### 3.1 Separação de Gaussianas

Iniciamos o estudo propondo um problema simples de classificação: decidir em qual distribuição gaussiana pertence uma dada variável aleatória. Para tal, aplicamos o método bayesiano simples de teoria da decisão, o teste de hipótese, como descrito no capítulo anterior.

Consideremos um conjunto de dados  $\{r_i, i = 1, \dots, n\}$ , que representa  $n$  variáveis aleatórias geradas com probabilidade  $p$  de estarem distribuídos de acordo com uma gaussiana de média  $m_1$  e largura  $\sigma_1$ ,  $G1(p, m_1, \sigma_1)$ , e com probabilidade  $(1 - p)$  de estarem em



uma outra gaussiana de média  $m_2$  e largura  $\sigma_2$ ,  $G2(1 - p, m_2, \sigma_2)$ . Se queremos decidir em qual destas distribuições gaussianas, independentes,  $G1$  ou  $G2$  pertence uma dada variável aleatória,  $r_{n+1}$ , podemos aplicar o teste de hipótese, calculando a razão das probabilidades posterior, como na equação (2.80), que fica:

$$\frac{p(H_1|r_{n+1}, \{\omega_j\})}{p(H_2|r_{n+1}, \{\omega_j\})} = \frac{p(r_{n+1}|H_1, \{\omega_j\}) \times p(H_1|\{\omega_j\})}{p(r_{n+1}|H_2, \{\omega_j\}) \times p(H_2|\{\omega_j\})}, \quad (3.1)$$

em que  $H_1$  é a hipótese de  $r_{n+1}$  pertencer à distribuição gaussiana  $G1(p, m_1, \sigma_1)$ ;  $H_2$  é a hipótese de  $r_{n+1}$  pertencer a distribuição  $G2(1 - p, m_2, \sigma_2)$ ;  $\{\omega_j, j = 1, \dots, 5\} = \{p, m_1, m_2, \sigma_1, \sigma_2\}$  é o conjunto de parâmetros desconhecidos a serem estimados.

As densidades de probabilidades diretas,  $p(r_{n+1}|H_1, \{\omega_j\})$  e  $p(r_{n+1}|H_2, \{\omega_j\})$  da equação (3.1), são as probabilidades do  $n$ -ésimo primeiro dado ter um valor  $r_{n+1}$ , condicionada à hipótese  $H_1$  ou  $H_2$  ser verdadeira e os parâmetros  $\{\omega_j\}$  serem conhecidos. E elas são dadas, respectivamente, por

$$p(r_{n+1}|H_1, \{\omega_j\}) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(r_{n+1}-m_1)^2}{2\sigma_1^2}} \quad (3.2)$$

e

$$p(r_{n+1}|H_2, \{\omega_j\}) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(r_{n+1}-m_2)^2}{2\sigma_2^2}}. \quad (3.3)$$

Podemos considerar que antes de analisar os dados, que se os valores dos parâmetros  $\{\omega_j\}$  são conhecidos, a variável  $r_{n+1}$  tem probabilidade  $p$  de ocorrer em  $G1$ , ou seja,  $p(H_1|\{\omega_j\}) = p$  e conseqüentemente, tem probabilidade  $(1 - p)$  de ocorrer em  $G2$ , ou seja,  $p(H_2|\{\omega_j\}) = 1 - p$ .

Portanto, substituindo esses pdfs a priori e as equações (3.2) e (3.3) na equação (3.1), nós podemos escrever a razão das probabilidades posterior como

$$\frac{p(H_1|r_{n+1}, \{\omega_j\})}{p(H_2|r_{n+1}, \{\omega_j\})} = \frac{p\sigma_2 e^{-\frac{(r_{n+1}-m_1)^2}{2\sigma_1^2}}}{(1-p)\sigma_1 e^{-\frac{(r_{n+1}-m_2)^2}{2\sigma_2^2}}}. \quad (3.4)$$

Como sabemos, se esta razão for menor que um,  $r_{n+1}$  é classificado como uma variável aleatória de acordo com  $G1$ , senão nós a classificamos como estando em  $G2$ .

### Determinação dos Parâmetros do Modelo

Como podemos perceber, só é possível fazer toda a análise acima se conhecermos  $\{\omega_j\}$ , senão, precisamos estimá-los. Para fazer isto, necessitamos calcular a densidade de probabilidade posterior de  $\{\omega_j\}$  que permite obter a informação mais completa possível dos dados sobre os parâmetros:

$$p(\{\omega_j\}|\{r_i\}, I) \propto p(\{r_i\}|\{\omega_j\}, I) \times p(\{\omega_j\}|I), \quad (3.5)$$

Note que omitimos o denominador  $p(\{r_i\}|I)$ , já que podemos considerá-lo como uma constante de normalização. A densidade de probabilidade a priori,  $p(\{\omega_j\}|I)$ , o último termo do lado direito desta equação, representa o que nós sabemos sobre os parâmetros dados somente a informação,  $I$ , que estamos lidando com um processo gaussiano. Para expressar total ignorância a priori sobre esses parâmetros, nós utilizamos um pdf uniforme simples:

$$p(\{\omega_j\}|I) = \begin{cases} \text{constante, para } \sigma_1 \text{ e } \sigma_2 > 0, \\ 0, & \text{outro caso} \end{cases} \quad (3.6)$$

Se admitirmos que a probabilidade do  $i$ -ésimo dado de obter um valor  $r_i$  é simplesmente dada pela soma de duas gaussianas (quando os valores dos parâmetros são conhecidos), como segue:

$$p(r_i|\{\omega_j\}, I) = \frac{p}{\sqrt{2\pi\sigma_1}} e^{-\frac{(r_i-m_1)^2}{2\sigma_1}} + \frac{1-p}{\sqrt{2\pi\sigma_2}} e^{-\frac{(r_i-m_2)^2}{2\sigma_2}}, \quad (3.7)$$

e que todos os dados  $\{r_i, i = 1, \dots, n\}$  são independentes, então a função *likelihood* é o produto das probabilidades de obter os dados individuais, dada como:

$$p(\{r_i\}|\{\omega_j\}, I) = \prod_{i=1}^n p(r_i|\{\omega_j\}, I) = \prod_{i=1}^n \left[ \frac{p}{\sqrt{2\pi\sigma_1}} e^{-\frac{(r_i-m_1)^2}{2\sigma_1}} + \frac{1-p}{\sqrt{2\pi\sigma_2}} e^{-\frac{(r_i-m_2)^2}{2\sigma_2}} \right]. \quad (3.8)$$

De acordo com a equação (3.5), se nós multiplicarmos a equação acima pelo priori da equação (3.6), nós obteremos o seguinte para o logaritmo da densidade de probabilidade posterior  $L$ :

$$L = \ln[p(\{\omega_j\}|\{r_i\}, I)] = \text{constante} + \sum_{i=1}^n \ln \left[ \frac{p}{\sqrt{2\pi\sigma_1}} e^{-\frac{(r_i-m_1)^2}{2\sigma_1}} + \frac{1-p}{\sqrt{2\pi\sigma_2}} e^{-\frac{(r_i-m_2)^2}{2\sigma_2}} \right], \quad (3.9)$$

onde, é claro, o pdf posterior é zero para os parâmetros  $(\sigma_1, \sigma_2)$  negativos.

Como foi visto na Seção 2.4.2, para obtermos as melhores estimativas dos parâmetros,



$\{\omega_{j0}\} = \{p_0, m_{10}, m_{20}, \sigma_{10}, \sigma_{20}\}$ , uma maneira é calcular as derivadas parciais do logaritmo deste pdf posterior. Por conseguinte,

$$\left. \frac{\partial L}{\partial \omega_j} \right|_{\{\omega_{j0}\}} = \frac{\partial}{\partial \omega_j} \left[ \sum_{i=1}^n \ln \left( \frac{p}{\sqrt{2\pi}\sigma_1} e^{-\frac{(r_i - m_1)^2}{2\sigma_1^2}} + \frac{1-p}{\sqrt{2\pi}\sigma_2} e^{-\frac{(r_i - m_2)^2}{2\sigma_2^2}} \right) \right] \Bigg|_{\{\omega_{j0}\}} = 0, \quad (3.10)$$

em que  $j = 1, \dots, 5$ . Para resolver simultaneamente estas cinco equações não-lineares, nós utilizamos o método gradiente, também conhecido como *Steepest Descent*. Este é definido pelo algoritmo iterativo

$$\omega_j^{k+1} = \omega_j^k + \lambda \frac{\partial L}{\partial \omega_j^k},$$

em que  $\lambda$  é um escalar fixo em um valor muito pequeno, de modo a não saltar o máximo da função, no nosso caso,  $L$ , e ir para o infinito. Pesquisamos a partir de um ponto inicial  $\omega_j^k$  ao longo da direção do gradiente positivo até encontrarmos um máximo nesta linha, que é  $\omega_j^{k+1}$ . Fazemos isto para todos os parâmetros  $\{\omega_j\} = \{p, m_1, m_2, \sigma_1, \sigma_2\}$ .

### 3.1.1 Resultados de Simulações

Para avaliar o limite do teste de hipótese, mencionado nesta seção, realizamos simples simulações que consistiram em gerar ( $n = 10000$ ) variáveis aleatórias gaussianas distribuídas independentemente em  $G1(p, m_1, \sigma_1)$  e em  $G2(1-p, m_2, \sigma_2)$ . Para isso, usamos o método de Box-Muller que gera variáveis gaussianas ( $r_1$  e  $r_2$ ) a partir de variáveis aleatórias uniformes entre 0 e 1 ( $\eta, x_1$  e  $x_2$ ). Dadas estas variáveis, obtemos  $r_1$  e  $r_2$  a partir da condição de transformação:

$$\begin{aligned} & \text{se } \eta < p \\ r_1 &= \sigma_1 \sqrt{-2 \ln(x_1)} \cos(2\pi x_2) + m_1 \\ & \text{senão se } \eta > p \\ r_2 &= \sigma_2 \sqrt{-2 \ln(x_1)} \cos(2\pi x_2) + m_2. \end{aligned}$$

Esta condição foi repetida até que a  $n$ -ésima variável aleatória gaussiana,  $r_n$ , fosse gerada. Dessa forma, geramos duas situações diferentes de gaussianas: a primeira com  $G1(0.4, 2, 1)$  e  $G2(0.6, 6, 2)$  e a outra com  $G1(0.3, 3, 1)$  e  $G2(0.7, 4, 2)$ , como ilustradas na figura (3.1).

De posse dos dados,  $\{r_i, i = 1, \dots, 10000\}$ , tirados de cada uma dessas situações, estimamos os seus parâmetros  $\{\omega_j\} = \{p, m_1, m_2, \sigma_1, \sigma_2\}$  resolvendo numericamente as cinco equações simultâneas, expressas na equação (3.10), pelo método gradiente, como descrito

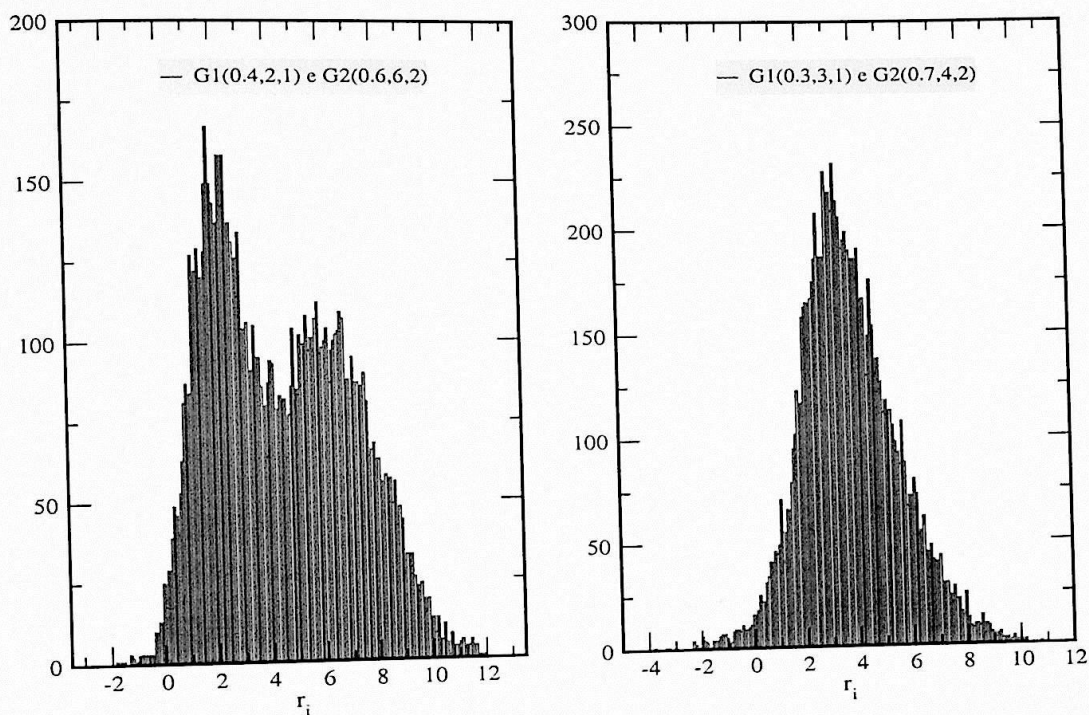


Figura 3.1: Duas situações geradas de duas distribuições gaussianas. À esquerda, uma situação em que duas distribuições estão bem separadas e à direita, uma outra situação em que não é possível distingui-las aparentemente.



Tabela 3.1: Valores estimados dos parâmetros  $\{\omega_j\} = \{p, m_1, m_2, \sigma_1, \sigma_2\}$  para cada situação gerada de gaussianas.

	primeira situação	segunda situação
$\{\omega_j\}$ reais	$\{0.4, 2, 6, 1, 2\}$	$\{0.3, 3, 4, 1, 2\}$
$\{\omega_{j0}\}$ estimados	$\{0.403, 1.994, 6.055, 1.006, 1.985\}$	$\{0.282, 3.014, 3.959, 0.935, 1.971\}$

Tabela 3.2: Erro cometido pelo teste de hipótese bayesiano em separar duas distribuições gaussianas para as duas situações geradas de gaussianas.

	primeira situação	segunda situação
erro cometido	0.0849	0.31

previamente. Na tabela (3.1) são mostrados os valores dos parâmetros usados nas duas situações e suas respectivas estimativas. Podemos notar que esses resultados refletem quão bem o modelo de duas gaussianas que utilizamos, equação (3.8), reproduz os dados.

Nosso próximo passo foi gerar mais ( $n = 10000$ ) novas variáveis gaussianas,  $\{r_t, t = 1, \dots, n\}$  da mesma forma que geramos o conjunto  $\{r_i\}$ . Então, utilizando os valores dos parâmetros estimados para cada situação gerada, aplicamos o teste de hipótese, equação (3.4), para cada variável gaussiana  $r_t$ . Feito isto, calculamos o número de variáveis classificadas erradas pelo método. Na tabela (3.2) são mostrados o erro total cometido usando o teste de hipótese para as duas situações.

Esses resultados mostram que, quando as duas gaussianas estão nitidamente separadas, como na primeira situação, o teste de hipótese bayesiano consegue distingui-las com erro em torno de 8% e quando as distribuições se aproximam, como na segunda situação em que uma gaussiana está praticamente dentro de outra, a distinção entre as duas classes (ou distribuições) torna-se mais difícil, mas, mesmo assim em torno de 70% das variáveis são classificadas corretamente pelo teste. Desta forma, podemos dizer que este método é razoavelmente eficiente para decidir entre duas classes. Mas isso pode ser simplesmente explicado pelo fato que em um teste de hipótese bayesiano: a probabilidade de errar é sempre menor que a probabilidade de acertar. Ou seja, para o caso que estamos lidando

$$p(H_1 = -1 | \{r_t\} \in 1) < p(H_1 = -1 | \{r_t\} \in -1).$$

Estudar o problema de separação de gaussianas foi importante, pois serviu como uma preparação para analisar dados mais complexos, como os dados de fMRI, onde se tem que decidir se uma dada região do cérebro está em repouso ( $H_1=-1$ ) ou em atividade neuronal ( $H_2=1$ ), durante a execução de alguma tarefa imposta ao sujeito, que é o assunto da próxima seção.

## 3.2 O Método de Pixel Independente

Devido à eficiência do teste de hipótese para o problema estudado na seção anterior, resolvemos aplicá-lo também para o problema de detecção de ativações em regiões do cérebro. Assim, surgiu o método de pixel independente que consiste em testar duas hipóteses para cada pixel da imagem:  $H_1$  para a hipótese que o pixel está ativo com respeito à tarefa de interesse e  $H_0$  para hipótese que o mesmo pixel está inativo. Novamente, para cada pixel  $p$  precisamos calcular a razão das probabilidades posterior:

$$\frac{p(H_1|d_p, I)}{p(H_0|d_p, I)} = \frac{p(d_p|H_1, I)}{p(d_p|H_0, I)} \times \frac{p(H_1|I)}{p(H_0|I)}. \quad (3.11)$$

Aqui, o último termo do lado direito representa nossa preferência, independente do pixel, entre as duas hipóteses antes da análise dos dados. O conjunto de dados discretos  $d_p = \{d_{p1}, \dots, d_{pT}\}$  representa a série temporal dos sinais medidos por RM corrompidos com ruído:

$$d_{pt} = y_{pt} = f_{pt} + \epsilon_{pt}, \quad (3.12)$$

$$p = 1, \dots, P, t = 1, \dots, T,$$

sendo que,  $f_{pt}$  representa o modelo do sinal de RM. Considerando o modelo de séries temporais da equação (2.55), descrito no capítulo anterior:

$$y_{pt} = b_p + A_p h_t a_p + \epsilon_{pt}, \quad \epsilon_{pt} \sim N(0, \sigma^2), \quad (3.13)$$

podemos dizer que:

$$f_{pt} = b_p + A_p h_t a_p. \quad (3.14)$$

Desta forma, podemos definir  $H_0$  como sendo a hipótese que  $a_p = 0$

$$H_0 \equiv f_p = b_p \quad (3.15)$$



e  $H_1$  a hipótese que  $a_p = 1$

$$H_1 \equiv f_p = b_p + A_p h, \quad (3.16)$$

em que,  $f_p = \{f_{p1}, \dots, f_{pT}\}$  é o conjunto dos sinais modelo para cada pixel e  $h = \{h_1, \dots, h_T\}$  o modelo da resposta hemodinâmica, fixo para todos os pixels da imagem, como mostrado na figura (2.8).

Uma vez definidas as hipóteses, precisamos atribuir as probabilidades envolvendo as medidas experimentais,  $p(d_p|H_1, I)$  e  $p(d_p|H_0, I)$  da equação (3.11). Para isso, pegamos a diferença entre os dados e o modelo de cada hipótese. Se conhecemos os modelos verdadeiros, então essa diferença é apenas o ruído. Como o modelo da equação (3.13) supõe que o ruído é descrito por uma distribuição de probabilidade conjunta gaussiana independente, isto é, para cada pixel  $p$  a probabilidade a priori de um conjunto de valores do ruído  $(\epsilon_{p1}, \dots, \epsilon_{pT})$ , admitindo que a variância  $\sigma^2$  é conhecida, é dada por

$$p(\epsilon_{p1}, \dots, \epsilon_{pT} | \sigma, I) \propto \prod_{t=1}^T \left[ \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\epsilon_{pt}^2}{2\sigma^2}\right) \right], \quad (3.17)$$

então as probabilidades diretas que nós devemos obter os dados  $d_p = (d_{p1}, \dots, d_{pT})$ , dado os parâmetros, são proporcionais às funções *likelihood*:

$$p(d_p|H_1, A_p, b_p, \sigma, I') \propto L_1(A_p, b_p, \sigma) = \prod_{t=1}^T \sigma^{-1} \exp\left\{-\frac{1}{2\sigma^2}[d_{pt} - f_{pt}]^2\right\},$$

$$L_1(A_p, b_p, \sigma) = (\sigma)^{-T} \times \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=1}^T [d_{pt} - (b_p + A_p h_t)]^2\right\}, \quad (3.18)$$

$$p(d_p|H_0, A_p, b_p, \sigma, I') \propto L_0(A_p, b_p, \sigma) = (\sigma)^{-T} \times \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=1}^T [d_{pt} - b_p]^2\right\}, \quad (3.19)$$

sendo  $I = I' + (A_p, b_p, \sigma)$ . Assim, a equação 3.11 pode ser escrita como:

$$\frac{p(H_1|d_p, A_p, b_p, \sigma, I')}{p(H_0|d_p, A_p, b_p, \sigma, I')} = \frac{L_1(A_p, b_p, \sigma)}{L_0(A_p, b_p, \sigma)} \times \frac{p(H_1|A_p, b_p, \sigma, I')}{p(H_0|A_p, b_p, \sigma, I')}. \quad (3.20)$$

Substituindo as equações (3.18) e (3.19) na equação acima, e escrevendo em termos do

logaritmo, a razão das probabilidades posterior fica expressa como:

$$\ln \left[ \frac{p(H_1|d_p, A_p, b_p, \sigma, I)}{p(H_0|d_p, A_p, b_p, \sigma, I)} \right] = -\frac{l_1}{2\sigma^2} + \frac{l_0}{2\sigma^2} + \ln \left[ \frac{p(H_1|A_p, b_p, \sigma, I)}{p(H_0|A_p, b_p, \sigma, I)} \right], \quad (3.21)$$

em que,  $l_1 = \sum_{t=1}^T [d_{pt} - (b_p + A_p h_t)]^2$  e  $l_0 = \sum_{t=1}^T [d_{pt} - b_p]^2$ . Logo, podemos classificar o pixel da seguinte forma: se

$$\frac{l_0 - l_1}{2\sigma^2} > \ln \left[ \frac{p(H_0|A_p, b_p, \sigma, I)}{p(H_1|A_p, b_p, \sigma, I)} \right], \quad (3.22)$$

o pixel é classificado como ativo com respeito a uma determinada tarefa, caso contrário ele é considerado como inativo. Note que o termo do lado direito reflete nossa preferência relativa a priori para as hipóteses alternativas, dado os parâmetros do modelo. Para ser imparcial, nós devemos atribuir uma probabilidade igual a um para todos os pixels. Já o termo do lado esquerdo, é a diferença entre o log-likelihood de  $H_0$  com o de  $H_1$ . Ela é uma medida de quão bem o modelo de cada uma das hipóteses é evidenciada nos dados.

A análise acima é válida sempre que os valores dos parâmetros do modelo são conhecidos  $\{b_p, A_p\}$ . O parâmetro  $\{b_p\}$  é a linha de base e, como vimos, é bem determinada dos dados, uma aproximação é tirar a média do sinal sobre o tempo em cada pixel  $p$ , isto é:

$$b_p = \frac{1}{T} \sum_{t=1}^T d_{pt}. \quad (3.23)$$

A amplitude da resposta hemodinâmica,  $A_p$ , pode ser inferida dos dados, dado que a hipótese  $H_1$  é verdadeira e os parâmetros  $\{b, \sigma^2\}$  conhecidos, usando o teorema de Bayes:

$$p(A_p|H_1, d_p, b_p, \sigma, I) = \frac{L_1(A_p, b_p, \sigma) \times p(A_p|H_1, b_p, \sigma, I)}{p(d_p|H_1, b_p, \sigma, I)}, \quad (3.24)$$

sendo  $p(A_p|H_1, b_p, \sigma, I)$  a informação que temos antes da análise dos dados sobre a amplitude  $A_p$ , dados os parâmetros e a hipótese  $H_1$  e  $L_1(b_p, A_p, \sigma)$  é a mesma função likelihood calculada na equação (3.18). O termo do denominador pode ser omitido, já que não depende explicitamente de  $A_p$  e pode ser considerado como uma constante de normalização.

Como foi mencionado na subseção (2.4.2), para encontrar o máximo do pdf posterior da equação (3.24) para um pixel, precisamos simplesmente calcular este pdf para vários valores possíveis de  $A_p$ . Para isto, é numericamente melhor calcular primeiro o logaritmo dessa equação para todos os pontos da rede; e então subtrair o número maior de cada um



e “exponenciá-lo”. Se a normalização for necessária, a pdf pode ser multiplicado por uma constante de modo que a área sobre a curva seja igual a um.

Quando a variância  $\sigma^2$  não é conhecida de antemão, ela muitas vezes é considerada como uma parâmetro de incômodo e pode ser eliminada utilizando a regra da marginalização, equação (2.71). O  $\sigma$  é restrito à valores positivos e além disso é um parâmetro de escala. Para expressar completa ignorância sobre o valor de uma escala, nós usamos o prior de Jeffreys  $1/\sigma$  [38]. Multiplicando este prior pelas funções *likelihood* das equações (3.18) e (3.19) e integrando sobre todos os valores positivos de  $\sigma$ , obtem-se as seguintes funções *likelihood*:

$$p(d_p|H_1, A_p, b_p, I) \propto L_1(A_p, b_p) = \int_0^\infty (\sigma)^{-1} L_1(A_p, b_p, \sigma) d\sigma \quad (3.25)$$

$$p(d_p|H_0, A_p, b_p, I) \propto L_0(A_p, b_p) = \int_0^\infty (\sigma)^{-1} L_0(A_p, b_p, \sigma) d\sigma. \quad (3.26)$$

### 3.2.1 Resultados e Simulações

Para entender o problema de localização de regiões ativadas no cérebro e avaliar o quanto o método de pixel independente é eficiente para resolver este problema, realizamos várias simulações de dados de fMRI. Para que estas correspondessem a dados reais, simulamos uma experiência típica em fMRI utilizando o seguinte protocolo de aquisição de imagens: foram geradas 60 imagens ( $T = 60$ ) consistindo de  $(m \times m)$  pixels, com um tempo de repetição (TR), intervalo entre as imagens, de 3 segundos; usamos um arranjo experimental em bloco, com três períodos de estímulo e três de repouso, cada período de 10 imagens.

Iniciamos o estudo construindo uma rede de  $(50 \times 50)$  pixels, com 100 destes ativos, como ilustrada na figura (3.2). Para cada pixel  $p$  ( $p = 1, \dots, 2500$ ), geramos uma série temporal com a linha de base  $b_p = 100$  e a amplitude  $A_p = 3$  (quando  $a_p = 1$ ), de acordo com a equação (3.13). Assim, na figura (3.3) é mostrada a série temporal gerada com  $\sigma = 6.0$  para um pixel em atividade e para um em repouso.

O fato dessa rede consistir de pixels que obedecem a uma mesma série temporal, nos possibilita fazer uma análise da interferência do ruído para detectar ativação utilizando o método de pixel independente. Com este objetivo, geramos séries temporais com vários valores da variância ( $\sigma = 1, 2, \dots, 11$ ) e calculamos o número de erros cometido pelo método para cada valor de  $\sigma$  gerado. Para analisar essas séries temporais, procedemos da seguinte forma: admitindo que os parâmetros do modelo, ( $A_p = 3, b_p = 100$ ) e a variância são conhecidos, aplicamos a condição da razão das probabilidades posterior, equação (3.22), para cada pixel da imagem, usando a probabilidade a priori certa para  $H_1$ , isto

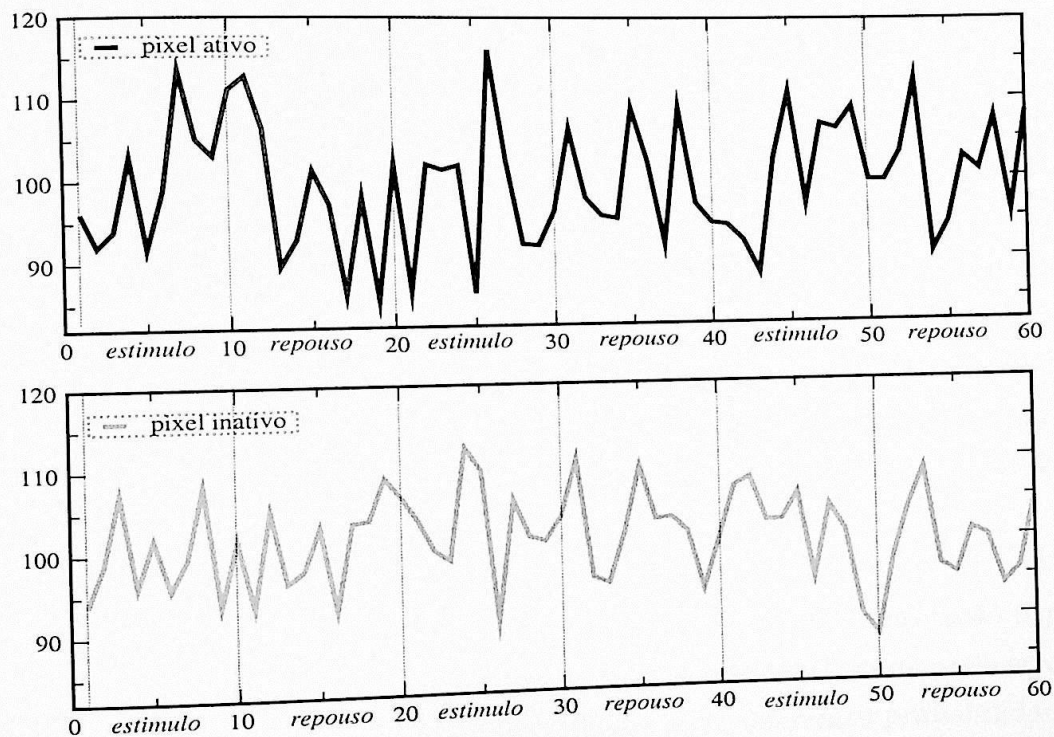
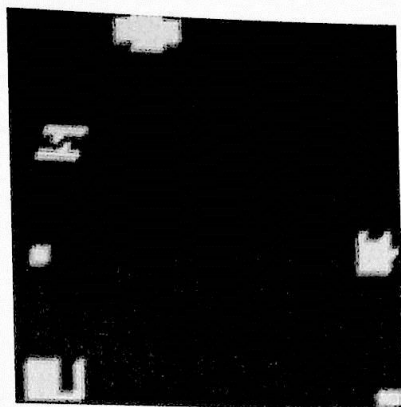


Figura 3.3: Duas séries temporais simuladas com  $\sigma = 6$  e  $b_p = 100$  para um pixel em atividade e um em repouso. As linhas verticais mostram a separação entre os períodos, e os períodos são indexados no eixo horizontal ao longo do índice de imagem (1 – 60). O eixo vertical refere-se aos valores do sinal de RM gerado. Devido ao valor alto usado para a variância do sinal, que não é possível notar diferença entre a série temporal gerada com  $A_p = 3$  para um pixel ativo (painel de cima), e a série temporal gerada para um pixel em repouso (o painel de baixo).



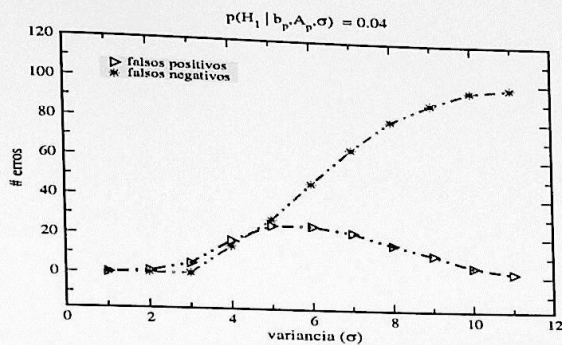


Figura 3.4: Resultado da análise da interferência do ruído em detectar ativação pelo método de pixel independente, sabendo que  $b_p = 100$ ,  $A_p = 3$  e atribuindo probabilidade a priori certa para a atividade do pixel,  $p(H_1|A_p, b_p, \sigma, I) = 0.04$ .

é,  $p(H_1|A_p, b_p, \sigma, I) = 0.04$ ; calculamos quantos deles são falsos positivos (pixels inativos classificados como ativos) e quantos são falsos negativos (pixels ativos classificados como inativos) de acordo com o método de pixel independente. Na figura (3.4) são mostrados os resultados desta análise.

Esses resultados mostram que quando a variância é menor que a amplitude  $A_p$  do modelo de resposta hemodinâmica, o método de pixel independente detecta todos os pixels ativos e praticamente todos os inativos, mas a medida que  $A_p$  vai sendo superada pela variância, o número de falsos positivos e negativos crescem. A partir de  $\sigma = 7$ , os falsos positivos vão diminuindo de forma que praticamente todos os pixels inativos e nenhum ativo são detectados em  $\sigma = 11$ . Isto acontece porque os dados são corrompidos por um ruído muito maior que o sinal, de forma que o método de pixel independente classifica todos os pixels da imagem como inativos.

Note que toda essa análise foi feita utilizando a probabilidade a priori certa para  $H_1$ . Assim, junto com a análise acima, testamos a importância da informação a priori para classificar os pixels em ativos ou inativos, detectando o erro total (número de erros de falsos positivos e falsos negativos) cometido pelo método, para diferentes probabilidades a priori para  $H_1$  usadas. Assim, os quatro primeiros gráficos da figura (3.5) mostram o número de erros de falsos positivos e de falsos negativos em relação ao  $\sigma$  gerado, usando diferentes probabilidades a priori para a atividade do pixel,  $p(H_1|A_p, b_p, \sigma, I) = \{0.01, 0.04, 0.1, 0.5\}$ , e o gráfico (3.5e) mostra o número de erros total cometido de acordo com as probabilidades a priori atribuídas a hipótese  $H_1$ , usando séries temporais geradas com  $\sigma = 6$ . É importante ressaltar que, devemos evitar atribuir probabilidade a priori 0 e 1, pois estes não são modificados pelos dados por meio da função *likelihood*,  $L_1(A_p, b_p, \sigma)$ .

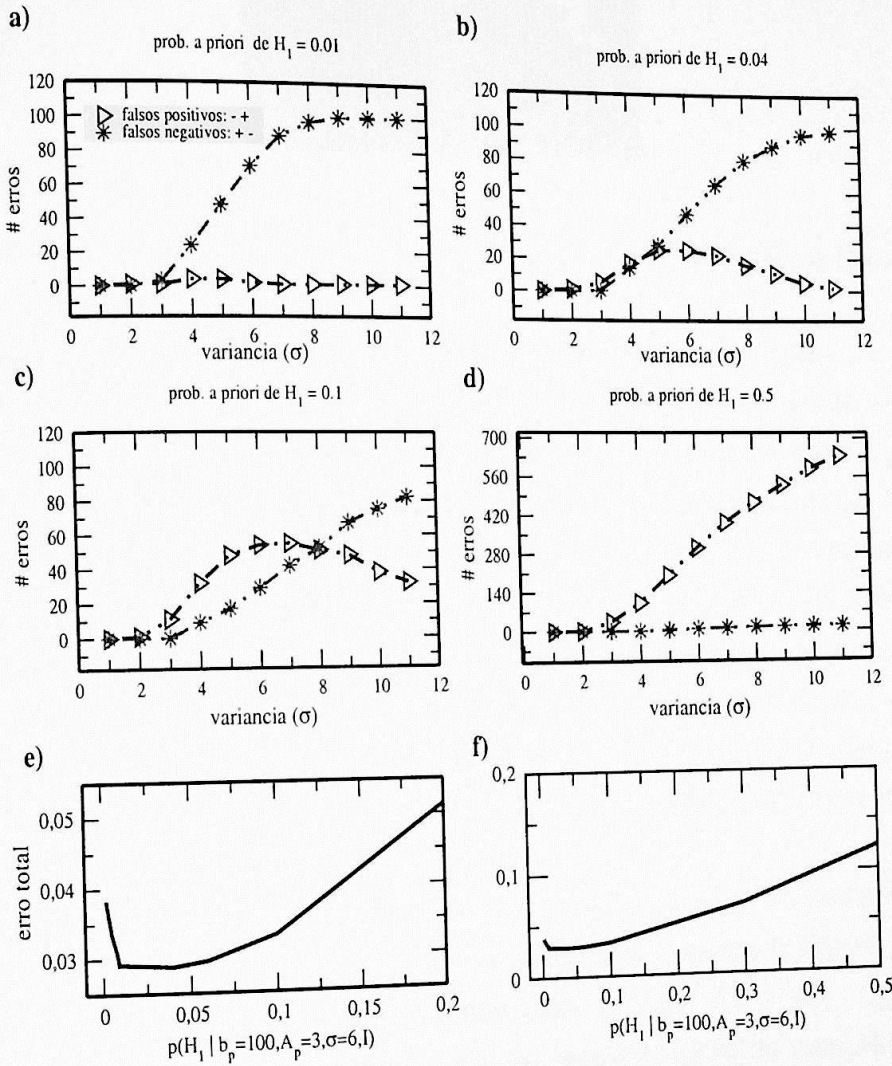


Figura 3.5: Resultados obtido pelo método de pixel independente na rede de  $(50 \times 50)$  pixels, figura (3.2), com  $b_p = 100$ ,  $A_p = 3$ , para vários valores gerados da variância do ruído e atribuindo diferentes probabilidades a priori para a atividade do pixel,  $p(H_1 | A_p, b_p, \sigma, I)$ .



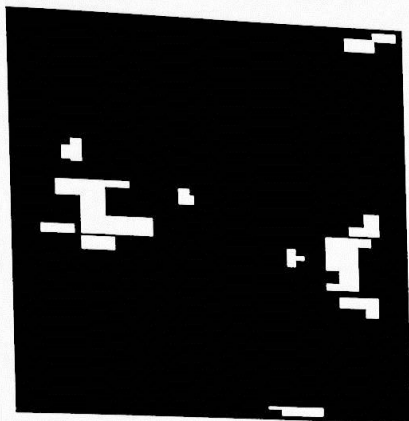


Figura 3.6: Imagem de  $(128 \times 128)$  pixels gerada com 5% de pixels ativos (pontos brancos).

Esses resultados mostram que de acordo com a informação que temos a priori sobre o estado do pixel, o resultado da detecção muda. Isto fica claro, se visualizarmos o gráfico (3.5e), onde existe um erro mínimo, em torno de 4%, quando usamos uma a priori próxima da correta e esse erro vai a 12%, quando levamos em conta apenas a função *likelihood*, com  $p(H_1|A_p, b_p, \sigma, I) = 0.5$ , gráfico (3.5f). Note que neste caso, ou seja, quando não temos preferência a priori para nenhuma das hipóteses  $H_0$  ou  $H_1$ , o número de falsos negativos é muito menor em relação a priori verdadeira, mas em compensação, os falsos positivos são muito maiores, gráfico (3.5d). E à medida que diminuimos a probabilidade a priori de  $H_1$ , o número de falsos positivos também diminui, devido à crescente preferência para  $H_0$  e os falsos negativos aumentam, até que todos os pixels são classificados como inativos quando  $p(H_1|A_p, b_p, \sigma, I) = 0.01$ , gráfico (3.5a).

No próximo passo foi repetir toda essa análise para uma rede de tamanho maior, consistindo de  $(2^7 \times 2^7)$  pixels com 5% da rede ativada, como mostra a figura (3.6).

As séries temporais são as mesmas geradas para a rede  $(50 \times 50)$ , figura (3.3), com os valores de  $b_p = 100$  e  $A_p = 3$ . Assim, geramos séries temporais com diferentes valores da variância do ruído, ( $\sigma = 2, 3, 4, \dots, 12$ ), e calculamos o número de erros cometido pelo método de pixel independente assumindo uma probabilidade diferente a priori para a atividade do pixel,  $p(H_1|A_p, b_p, \sigma, I)$ , a cada valor de  $\sigma$  gerado; Na figura (3.7) são mostrados os resultados obtidos para esta rede. Estes resultados apenas reforçam os obtidos para a rede menor, mostrando que nosso estado de conhecimento após a análise dos dados é, no entanto, fortemente dependente do que conhecemos ou admitimos antes dos resultados. Portanto, as probabilidades posterior são diferentes.

Na figura (3.8) mostramos o melhor resultado obtido pelo método de pixel independente

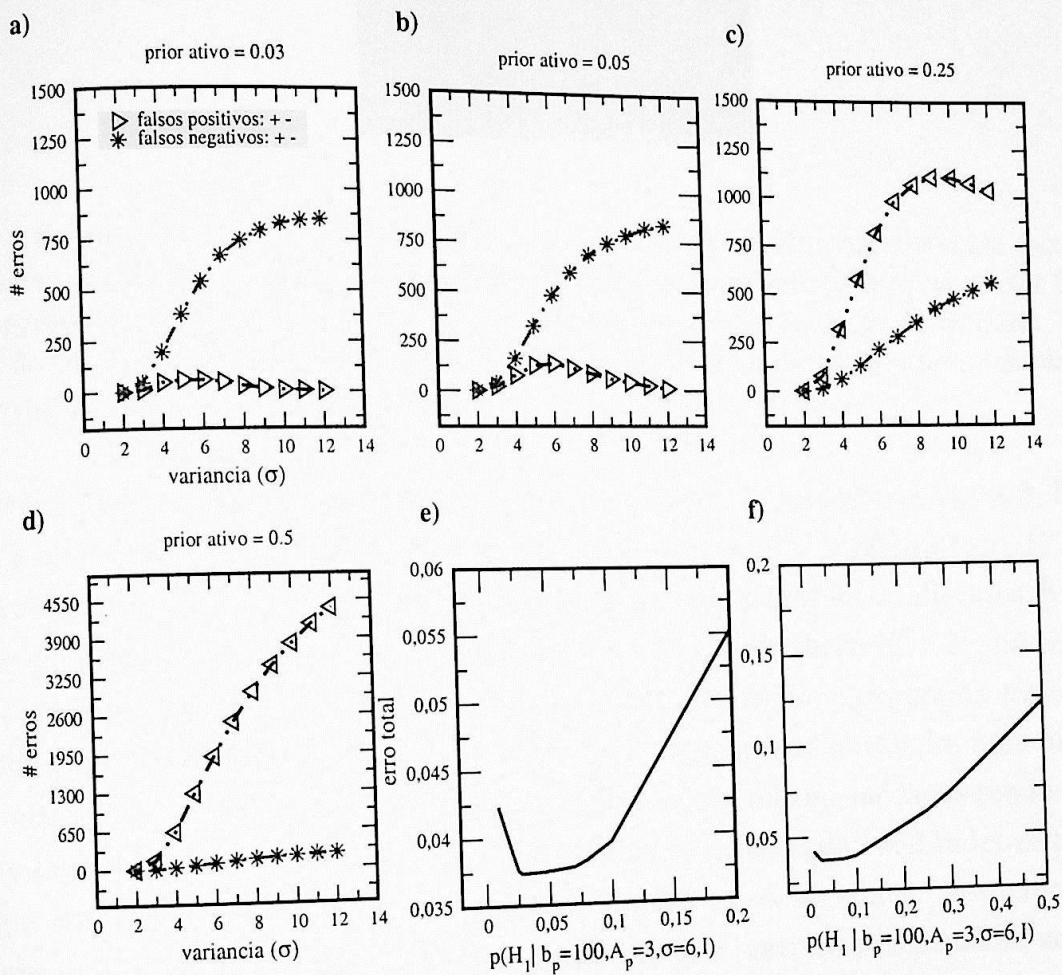


Figura 3.7: Resultados obtido pelo método de pixel independente em classificar pixels como ativos ou inativos na rede de  $(2^7 \times 2^7)$  da figura (3.6), para vários valores gerados da variância do ruído,  $\sigma$ , e assumindo diferentes probabilidades a priori para a atividade do pixel,  $p(H_1 | A_p, b_p, \sigma, I)$ .



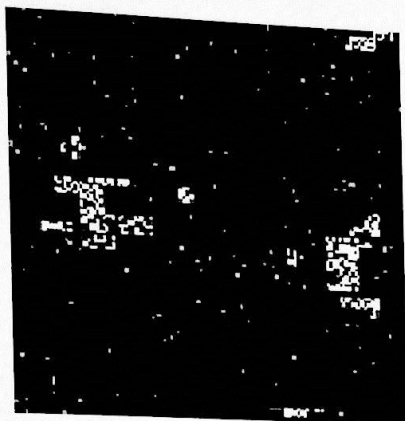


Figura 3.8: Imagem obtida pelo método de pixel independente para detectar pixels ativos na imagem da figura (3.6), utilizando os valores dos parâmetros que as séries temporais foram geradas,  $A_p = 3$ ,  $b_p = 100$ ,  $\sigma = 6$  e atribuindo  $p(H_1|A_p, b_p, \sigma, I) = 0.05$ . Os vários pontos de cor cinza claro que aparecem no fundo da imagem correspondem aos pixels falsos positivos e os pontos cinza escuros são os pixels falsos negativos.

para detectar pixels ativos na rede de  $(2^7 \times 2^7)$ , usando os parâmetros  $A_p = 3$ ,  $b_p = 100$  e  $\sigma = 6$ , e atribuindo a probabilidade a priori de  $H_1$  verdadeira,  $p(H_1|A_p, b_p, \sigma, I) = 0.05$ .

Para estimar os parâmetros  $(b_p, \sigma)$ , que até agora admitimos ser conhecidos de antemão, fizemos um histograma dos dados gerados com  $\sigma = 6$  para a rede de  $(2^7 \times 2^7)$  pixels da figura (3.6), e o ajustamos por uma distribuição gaussiana, utilizando o programa *xmgrace*, como ilustrado na figura (3.9). Como podemos ver, a curva de ajuste obtida foi uma distribuição gaussiana de média 100 e  $\sigma = 6$ . É interessante notar, que mesmo os dados sendo ajustados, e muito bem, por uma única distribuição gaussiana, o método de pixel independente ainda consegue detectar pixels ativos, ainda que com limitações, como vimos. Neste momento fica claro que, o problema de classificar pixels como ativos e inativos se assemelha ao problema estudado na seção anterior, de separação de gaussianas.

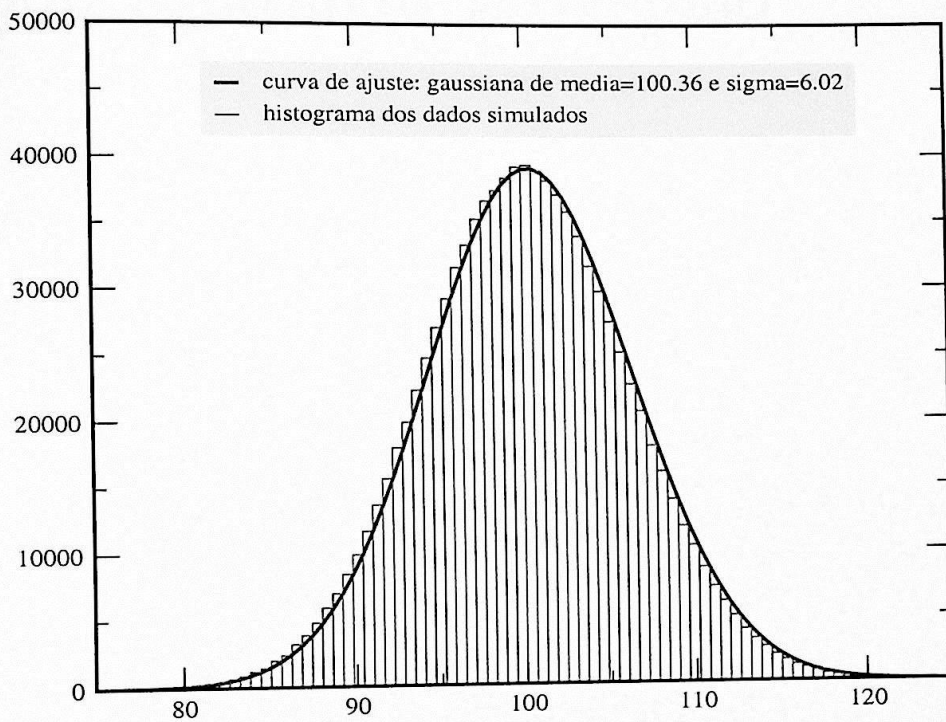


Figura 3.9: Histograma dos dados gerados com os parâmetros,  $b_p = 100$ ,  $A_p = 3$ ,  $\sigma = 6$  para a rede de  $(2^7 \times 2^7)$  pixels e sua curva de ajuste obtida pelo programa *xmgrace*.



## Capítulo 4

# Método *Multigrid* Bayesiano

Devido aos resultados obtidos na Seção (3.2), que mostraram a importância da informação a priori para classificar os pixels da imagem como ativos ou inativos, e ao fato que ativação real tende a estar mais agrupada que ativação artificial causada por ruído, desenvolvemos, neste capítulo, idéias de um método inspirado em técnicas de multiescalas. Seu objetivo é construir uma distribuição de probabilidade a priori para a atribuição de atividade a um dado pixel a partir das propriedades globais da imagem. Uma vez feito isso, o nível de atividade é proporcional à distribuição de probabilidade posterior. Assim, na primeira seção deste capítulo, descrevemos o método acima mencionado e na segunda seção, finalizamos com resultados obtidos por esse método a dados simulados de fMRI.

### 4.1 Descrição do Método

Enquanto que o método de pixel independente simplesmente aplica o teste de hipótese utilizando uma distribuição de probabilidade a priori uniforme para todos os pixels, o método *multigrid* bayesiano gera estimativas da distribuição de probabilidade a priori para as hipóteses  $H_1$  e  $H_0$ , tais como definidas nas respectivas equações (3.16) e (3.15), utilizando uma escala hierárquica. A descrição do método é a seguinte: suponham uma região da imagem de tamanho  $(2^n \times 2^n)$ , construímos uma escala hierárquica crescente de  $(n + 1)$  níveis, sendo que cada nível, a matrix de pixels é dividida em partes iguais de tamanho  $(2^q \times 2^q)$ , sendo que cada valor de  $q(0, 1, \dots, n)$  representa o nível da escala. Cada elemento dessa matrix, formado por uma região  $R = \left(\frac{2^n}{2^q} \times \frac{2^n}{2^q}\right)$  pixels, é representado pela média das séries temporais sobre aquela região. Então, a nova série temporal da região  $R$  é calculada

da seguinte forma:

$$Y_{ijt}^q = \frac{1}{R} \sum_{p=1}^R d_{pt} = \frac{1}{R} \sum_{p=1}^R (f_{pt} + \epsilon_{pt}), \quad (4.1)$$

$$t = 1, \dots, T, \quad i, j = 1, 2, \dots, 2^q,$$

em que  $(d_{pt}, t = 1, \dots, T)$  é a série temporal de um pixel  $p$  e  $f_{pt}$  é o modelo do sinal dado pela equação (3.14). Para cada nível, a distribuição de probabilidade posterior é calculada para cada região da matriz e usada como a distribuição a priori do nível seguinte, sendo que os extremos da escala,  $(q = 0)$  e  $(q = n)$  possuem, respectivamente, uma distribuição de probabilidade posterior da ativação global da imagem e uma distribuição posterior que descreve a ativação de cada pixel.

Assim, para cada valor de  $q$ , calculamos a probabilidade posterior de cada hipótese para cada elemento de matriz,  $Y_{ij}^q(Y_{ij1}^q, \dots, Y_{ijT}^q)$ :

$$p_{post}^q(H_1 | Y_{ij}^q, b_p, A_p, \sigma, I) = \frac{p^q(Y_{ij}^q | H_1, b_p, A_p, \sigma, I) \times p_{prior}^q(H_1 | b_p, A_p, \sigma, I)}{p^q(Y_{ij}^q | b_p, A_p, \sigma, I)} \quad (4.2)$$

$$p_{post}^q(H_0 | Y_{ij}^q, A_p, b_p, \sigma, I) = \frac{p^q(Y_{ij}^q | H_0, b_p, A_p, \sigma, I) \times p_{prior}^q(H_0 | b_p, A_p, \sigma, I)}{p^q(Y_{ij}^q | b_p, A_p, \sigma, I)}, \quad (4.3)$$

sendo a constante de normalização:

$$p^q(Y_{ij}^q | A_p, b_p, \sigma, I) = p^q(Y_{ij}^q | H_1, b_p, A_p, \sigma, I) \times p_{prior}^q(H_1 | b_p, A_p, \sigma, I) + p^q(Y_{ij}^q | H_0, b_p, A_p, \sigma, I) \times p_{prior}^q(H_0 | b_p, A_p, \sigma, I). \quad (4.4)$$

As probabilidades diretas,  $p^q(Y_{ij}^q | H_1, b_p, A_p, \sigma, I)$  e  $p^q(Y_{ij}^q | H_0, b_p, A_p, \sigma, I)$ , são proporcionais às funções *likelihood* e similares às equações (3.18) e (3.19):

$$L_1^q(b_p, A_p, \sigma) = (\sigma)^{-T} \times \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=1}^T \left[Y_{ijt}^q - \left(\frac{1}{R} \sum_{p=1}^R b_p + \frac{1}{R} \sum_{p=1}^R A_p h_t\right)\right]^2\right\} \quad (4.5)$$

$$L_0^q(b_p, A_p, \sigma) = (\sigma)^{-T} \times \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=1}^T \left[Y_{ijt}^q - \frac{1}{R} \sum_{p=1}^R b_p\right]^2\right\}. \quad (4.6)$$

Atualizamos as distribuições de probabilidades a priori no próximo nível a partir das



probabilidades posterior no nível anterior:

$$p_{prior}^{q+1}(H_1|b_p, A_p, \sigma, I) = p_{post}^{q+1}(H_1|Y_{ij}^q, b_p, A_p, \sigma, I) \quad (4.7)$$

$$p_{prior}^{q+1}(H_0|b_p, A_p, \sigma, I) = p_{post}^{q+1}(H_0|Y_{ij}^q, b_p, A_p, \sigma, I). \quad (4.8)$$

No nível zero, essas probabilidades a priori são atribuídas de acordo com a informação a priori que temos a respeito da atividade global da imagem.

Note que à medida que a escala cresce, as distribuições a priori tornam-se cada vez menos uniformes sobre as regiões da imagem. Isto porque, as informações fornecidas pelos dados em um nível da escala são sempre levadas em consideração, como informações a priori, no nível seguinte.

Ao final da escala, em ( $q = n$ ), obtemos o nível de atividade de cada pixel de acordo com sua probabilidade posterior.

## 4.2 Resultados e Discussões de Simulações

Esta seção tem como objetivo tentar otimizar os resultados obtidos na subseção (3.2.1), pelo método de pixel independente, para a rede de ( $2^7 \times 2^7$ ) pixels da figura (3.6), gerada com séries temporais de amplitude  $A_p = 3$  e linha de base  $b_p = 100$ . Assim, aplicamos a esses dados simulados o método *multigrid* bayesiano, descrito na seção anterior.

Iniciamos o estudo analisando os dados para diferentes valores gerados da variância do ruído e com diferentes informações a priori sobre a atividade global da imagem,  $p_{prior}^0(H_1|b_p, A_p, \sigma, I)$ , como segue: geramos séries temporais com diferentes variâncias,  $\sigma = \{1, 1.5, 2, 3, \dots, 12\}$ , admitindo os parâmetros  $(A_p, b_p, \sigma)$  conhecidos, a cada valor de  $\sigma$  gerado, atribuímos uma probabilidade a priori diferente para  $H_1$  no nível zero da escala,  $p_{prior}^0(H_1|b_p, A_p, \sigma, I)$ , e calculamos as probabilidades posterior para  $H_1$  e  $H_0$ , de acordo com as respectivas equações (4.2) e (4.3), para todos os níveis da escala, ( $q = 0, \dots, 7$ ); no final da escala ( $q = 7$ ), em que obtemos a probabilidade posterior para  $H_1$  para cada pixel,  $p_{post}^7(H_1|Y_{ij}^7, b_p, A_p, \sigma, I)$ , calculamos os erros de falsos positivos e de negativos cometidos pelo método, utilizando um limite de significância para a classificação, ou seja, um pixel só é classificado como ativo, se sua probabilidade posterior de  $H_1$  for maior que um certo limiar, caso contrário, ele é classificado como inativo.

Assim, na figura (4.1) mostramos os resultados dessa análise, em que usamos um limiar de 0.90. Note nos três primeiros gráficos que, para séries temporais geradas com valores

de  $\sigma$  menores ou iguais que a amplitude  $A_p = 3$ , o método *multigrid* bayesiano consegue detectar todos os pixels inativos, independente da nossa opinião sobre a atividade global da imagem, ao passo que os pixels ativos são melhores detectados à medida que a amplitude vai sendo superada pela variância, de modo que para todos os valores de  $\sigma$  menores que 8, o melhor resultado é obtido quando atribuímos uma distribuição de probabilidade a priori muito alta para a atividade global da imagem, ou seja,  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.95$ , figura (4.1a); entre  $\sigma = 8$  e  $\sigma = 9$ , um priori que reflete imparcialidade sobre as duas hipóteses,  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.5$ , melhora bastante a detecção de pixels inativos, compensando o aumento do número de falsos negativos, figura (4.1b); a partir de  $\sigma = 10$ , quando usamos um priori bem mais baixo para  $H_1$ , ou seja,  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.3$ , figura (4.1c), a quantidade de falsos positivos é menor ainda, compensando novamente o aumento de falsos negativos, de modo que mesmo para um ruído muito grande com  $\sigma = 12$ , o método ainda consegue detectar 65% dos pixels ativos na imagem.

Para uma variância alta,  $\sigma = 8$ , a figura (4.1d) deixa claro que o melhor resultado é obtido quando usamos  $p_{prior}^0(H_1|b_p, A_p, \sigma, I)$  em torno de 0.8; e para  $\sigma = 6$ , figura (4.1e), quanto mais alto o priori usado, mais eficiente é o método *multigrid* bayesiano. Desta forma, constatamos que as propriedades globais da imagem são realmente importantes para que se obtenha bons resultados por este método e que ele é capaz de detectar pixels ativos em imagens corrompidas por muito ruído.

É interessante visualizar a seqüência de imagens obtida pelo método com  $p_{prior}^0(H_1|b_p = 100, A_p = 3, \sigma = 6, I) = 0.95$ , figura (4.2). Cada imagem foi obtida em um nível  $q$  da escala hierárquica, que ilustra o nível de atividade, proporcional à probabilidade posterior para  $H_1$ , sobre cada região  $R$  da matriz ( $2^q \times 2^q$ ). Essa seqüência mostra claramente que mesmo usando uma probabilidade a priori alta para atividade global da imagem, esta vai sendo modificada pelos dados a cada nível da escala, de tal forma que no final desta, em ( $q = 7$ ), obtemos uma probabilidade a priori que reflete todas as informações relevantes tiradas nos níveis anteriores, que é então usada para obter uma imagem que apresente para cada pixel uma probabilidade posterior para  $H_1$ .

Para melhor visualização, na figura (4.3) mostramos esse resultado junto com a rede gerada da figura (3.6).

Para verificar o resultado da figura (4.1e), outra seqüência de imagens obtida com um priori diferente, ( $p_{prior}^0(H_1|b_p = 100, A_p = 3, \sigma = 6, I) = 0.5$ ), é mostrada na figura (4.4). Note que como a imagem obtida em ( $q = 0$ ) tem um nível de atividade global muito baixo, a probabilidade posterior para  $H_1$  no nível seguinte é praticamente nula sobre



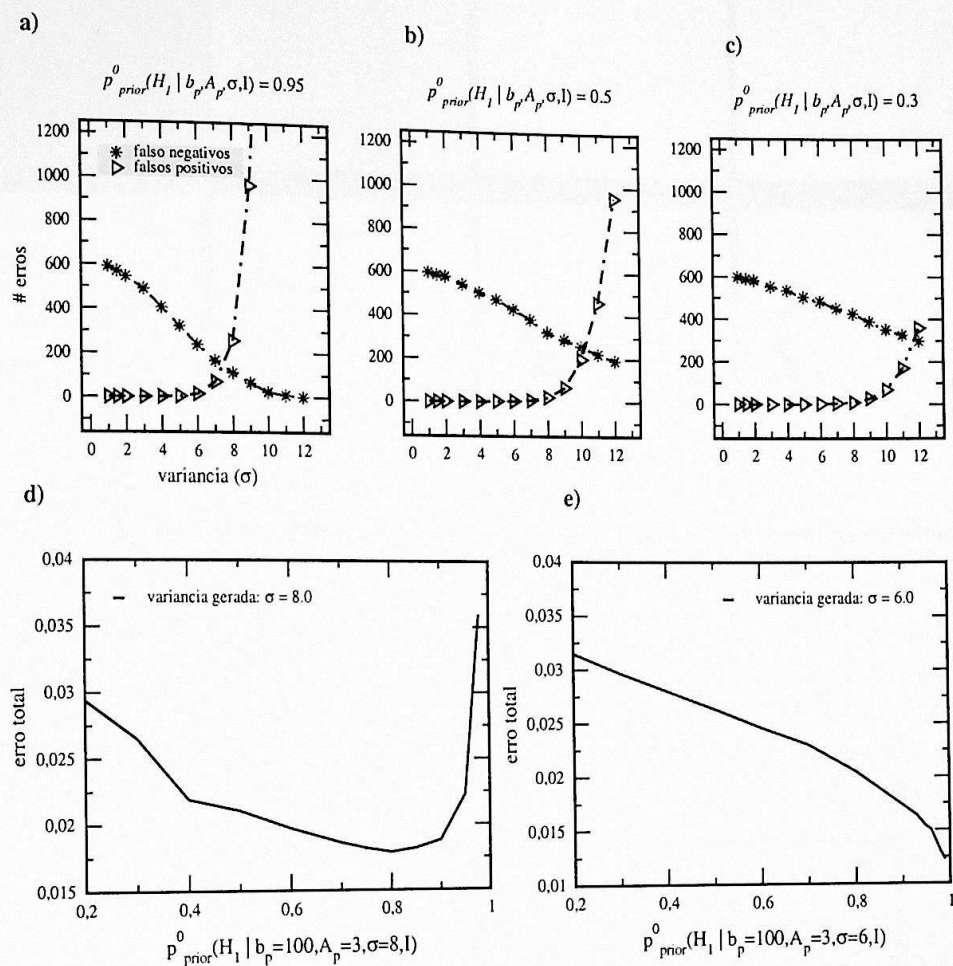


Figura 4.1: Resultados obtidos para a rede de  $(2^7 \times 2^7)$  pixels da figura (3.6), gerada com séries temporais de amplitude  $A_p = 3$ , linha de base  $b_p = 100$  e diferentes valores da variância do ruído,  $\sigma = \{1, 1.5, 2, 3, \dots, 12\}$ , utilizando o método *multigrid* bayesiano para classificar os pixels como ativos ou inativos, com um limiar de 0.90 e atribuindo (a) uma probabilidade a priori muito alta para  $H_1$ ,  $p_{prior}^0(H_1 | b_p, A_p, \sigma, I) = 0.95$ , (b) um priori que reflete imparcialidade,  $p_{prior}^0(H_1 | b_p, A_p, \sigma, I) = 0.5$ , e (c) um priori baixo para  $H_1$ ,  $p_{prior}^0(H_1 | b_p, A_p, \sigma, I) = 0.3$ . Em (d) e (e) mostramos o priori ideal para  $H_1$  com um ruído gerado, respectivamente, com  $\sigma = 8$  e  $\sigma = 6$ .



Figura 4.2: Sequência de imagens obtida na escala hierárquica do método *multigrid* bayesiano, utilizando  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.95$  e assumindo os parâmetros conhecidos,  $b_p = 100$ ,  $A_p = 3$ ,  $\sigma = 6$ . Cada imagem, obtida em um nível  $q$  da escala, ilustra o nível de atividade, proporcional a probabilidade posterior para  $H_1$ , sobre cada região  $R$  da matriz ( $2^q \times 2^q$ ). O nível de atividade é representado nas imagens em escala de cinza.

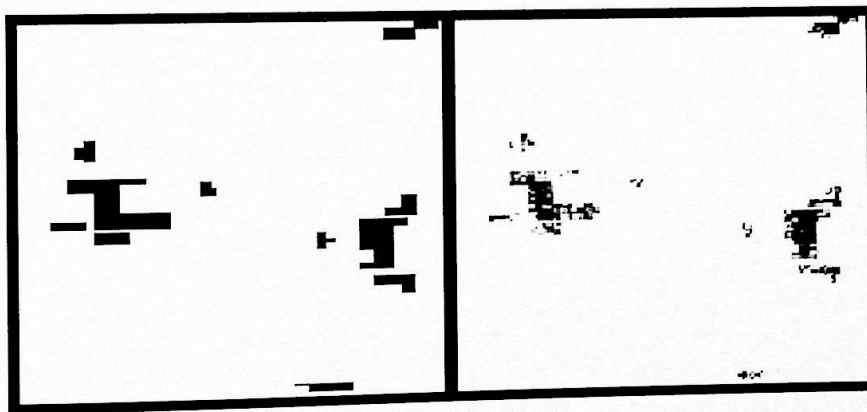


Figura 4.3: Duas imagens de  $(2^7 \times 2^7)$  pixels. À esquerda ilustra a imagem original, gerada com 5% de pixels ativos (pontos pretos), e à direita ilustra a imagem obtida da detecção de pixels ativos da imagem à esquerda, resultante do último nível da escala hierárquica crescente do método *multigrid* bayesiano, usando  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.95$  e admitindo  $b_p = 100$ ,  $A_p = 3$ ,  $\sigma = 6$  conhecidos. O nível de atividade de cada pixel da imagem, à direita, é representado em escala de cinza.





Figura 4.4: Seqüência de imagens obtida na escala hierárquica do método *multigrid* bayesiano, utilizando  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.5$  e considerando os parâmetros conhecidos,  $b_p = 100$ ,  $A_p = 3$ ,  $\sigma = 6$ . Cada imagem ilustra, em escala de cinza, o nível de atividade de acordo a probabilidade posterior para  $H_1$  sobre cada região  $R$  da matriz  $(2^q \times 2^q)$ .

cada região  $R$  da matriz  $(2 \times 2)$ , e permanece assim até em  $q = 3$ . Somente a partir de  $q = 4$ , quando uma região aparece com um nível de atividade muito baixo, as regiões ativadas da imagem vão sendo detectadas. Com isso, a imagem final resultante em  $q = 7$  apresenta pixels ativos com níveis de atividade mais baixos e com mais falsos negativos que a imagem final da figura (4.2), obtida com um priori igual a 0.95. Logo, confirmamos que para séries temporais geradas com uma variância que é o dobro da amplitude,  $\sigma = 6$ , o melhor resultado é obtido com um a priori que reflete nossa preferência quase total para a atividade global da imagem.

Até agora, admitimos que a variância, ( $\sigma = 6$ ), é conhecida, e como na subseção (3.2.1), ela já foi estimada pelos dados considerados aqui, como mostramos na figura(3.9). Resolvemos analisar o desempenho do método *multigrid* quando consideramos valores da variância. Com este objetivo, analisamos as séries temporais com  $p(H_1|b_p, A_p, \sigma, I) = 0.95$  e diferentes valores da variância, ( $\sigma = 2, 3, \dots, 12$ ), calculando no último nível da escala, em  $q = 7$ , os erros de falsos positivos e de negativos cometidos pelo método, com um limiar

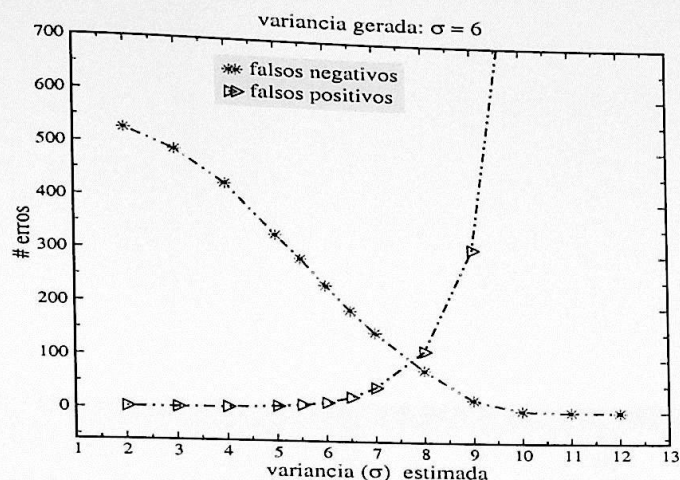


Figura 4.5: Resultado obtido da análise de séries temporais geradas com  $b_p = 100$ ,  $A_p = 3$ ,  $\sigma = 6$ , através do método *multigrid* bayesiano, utilizando  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.95$ , um limiar de 0.90 e considerando diferentes valores de  $\sigma$ .

de 0.90. Os resultados são mostrados na figura (4.5).

Note que essa figura é semelhante à figura (4.1e) obtida com o  $\sigma$  correto, de forma que se compararmos a duas figuras, podemos observar que quando subestimamos o valor da variância, os falsos negativos e positivos não são muito diferente de quando usamos a variância correta. Mas quando  $\sigma$  é superestimado, os erros de falsos positivos e negativos diminuem muito, sendo que em ( $\sigma = 8$ ), são detectados mais de 90% dos pixels ativos e de inativos na imagem e em ( $\sigma = 12$ ), são detectados todos os pixels ativos e cerca de 50% dos pixels inativos na imagem. Logo, podemos dizer que quando superestimamos o ruído próximo ao seu valor correto, o método *multigrid* é mais eficiente em detectar ativação na imagem.

Vale a pena observar a seqüência de imagens obtida na escala hierárquica crescente, para  $\sigma$  superestimado em 8, com  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.95$ , figura (4.6). Aqui, podemos perceber claramente as regiões formadas em cada nível da escala, apresentando níveis de atividade de acordo com suas probabilidades posterior de  $H_1$ . No primeiro nível da escala, a região global da imagem apresenta praticamente o mesmo nível de atividade imposta pelo a priori usado. Se fixarmos o olhar para a imagem final, obtida em  $q = 7$ , percebemos que são detectados mais pixels ativos com  $\sigma = 8$  do que com o  $\sigma$  correto da figura (4.2), compensando os falsos positivos que aparecem com mais nitidez, mas com baixos níveis de atividade.

Analizamos a evolução da distribuição posterior da hipótese  $H_1$  na escala hierárquica, de





Figura 4.6: Seqüência de imagens obtida na escala hierárquica do método *multigrid* bayesiano, com  $b_p = 100$ ,  $A_p = 3$ ,  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.95$  e superestimando a variância em 8. Cada imagem ilustra, em escala de cinza, o nível de atividade de acordo a probabilidade posterior para  $H_1$  sobre cada região  $R$  da matriz  $(2^q \times 2^q)$ .

acordo com o valor de  $\sigma$  considerado, para alguns pixels classificados corretamente e alguns falsos positivos e negativos, obtidos na figura (4.5) com um priori ( $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.95$ ). Desta forma, fizemos um gráfico de  $p_{post}^q(H_1|Y_{ij}^q, b_p, A_p, \sigma, I)$ , com respeito ao nível  $q$  da escala, para cada valor de  $\sigma$  assumido, ( $\sigma = 5, 6, 7, 12$ ), como mostra a figura (4.7). Repare que o priori usado no nível  $q = 0$ , foi colocado em todos os gráficos com o objetivo de mostrar a mudança ocorrida na probabilidade posterior nesse nível, e que apenas na figura (4.7d), não mostramos os pixels inativos classificados corretamente.

Quando conhecemos a variância,  $\sigma = 6$ , figura (4.7b), observamos que logo no primeiro nível da escala, a probabilidade posterior de  $H_1$  diminui para um pouco mais de 0.6, e quando subestimamos  $\sigma = 5$ , figura (4.7a), essa probabilidade posterior diminui mais ainda, indo para um pouco mais de 0.4, mas se  $\sigma$  é superestimado, figura (4.7c), a probabilidade posterior diminui muito pouco e quase não muda quando  $\sigma = 12$ , figura (4.7d). A medida que o nível da escala cresce, figura (4.7a), (4.7b), (4.7c), a probabilidade posterior vai diminuindo e indo praticamente a zero em  $q = 4$ , para quase todos os pixels considerados, sendo que neste último nível, na figura (4.7c), cada um desses pixels tem uma probabilidade posterior diferente. Ao passo que quando  $\sigma = 12$ , figura (4.7d), a probabilidade é alta até em  $q = 2$ , e a partir do nível seguinte, os pixels começam a se diferenciarem pelas suas probabilidades posterior atribuídas a  $H_1$ .

Note que em todos os gráficos, no final da escala, aparecem pixels falsos positivos com baixos níveis de atividade, isto ocorre devido ao limiar alto usado para calcular os erros cometidos pelo método.

## Análise do Parâmetro $A_p$ do Modelo

Vamos analisar agora a amplitude  $A_p$  da resposta hemodinâmica. Para isso, considerando os parâmetros,  $b_p = 100, \sigma = 6$ , conhecidos, analisamos as séries temporais através do método *multigrid*, atribuindo diferentes probabilidades a priori para  $H_1$  no nível zero, ( $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = \{0.3, 0.5, 0.95\}$ ), para diferentes valores considerados da amplitude, ( $A_p = \{1.2, \dots, 5\}$ ). Com isso, no último nível da escala, em  $q = 7$ , novamente, calculamos os falsos positivos e negativos usando um limiar de 0.90. Estes resultados são mostrados na figura (4.8).

Como podemos observar na figura (4.8a), quando usamos  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.3$ , o método tem melhor desempenho com a amplitude subestimada em 2, com mais de 98% dos pixels classificados corretamente e a medida que aumentamos o valor de  $A_p$ , os falsos positivos diminuem e os negativos aumentam, de modo que quando superestimamos  $A_p = 5$ ,



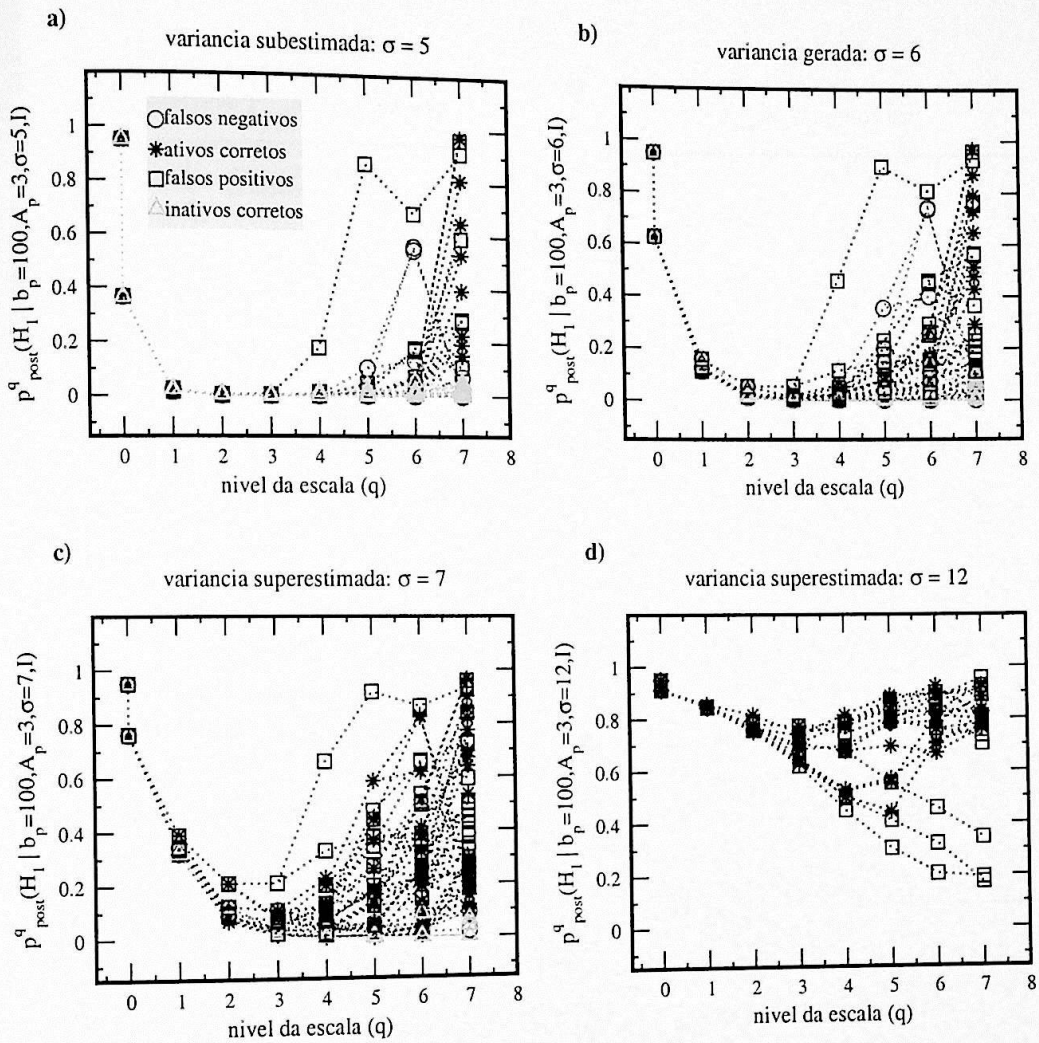


Figura 4.7: Evolução da distribuição posterior,  $p^q_{post}(H_1|Y_{ij}^q, b_p, A_p, \sigma, I)$ , na escala hierárquica do método *multigrid* bayesiano, de alguns pixels corretamente e falsamente classificados, de acordo com os resultados da figura (4.5), considerando (a)  $\sigma = 5$ , (b) variância correta,  $\sigma = 6$ , (c)  $\sigma = 7$  e (d)  $\sigma = 12$ .

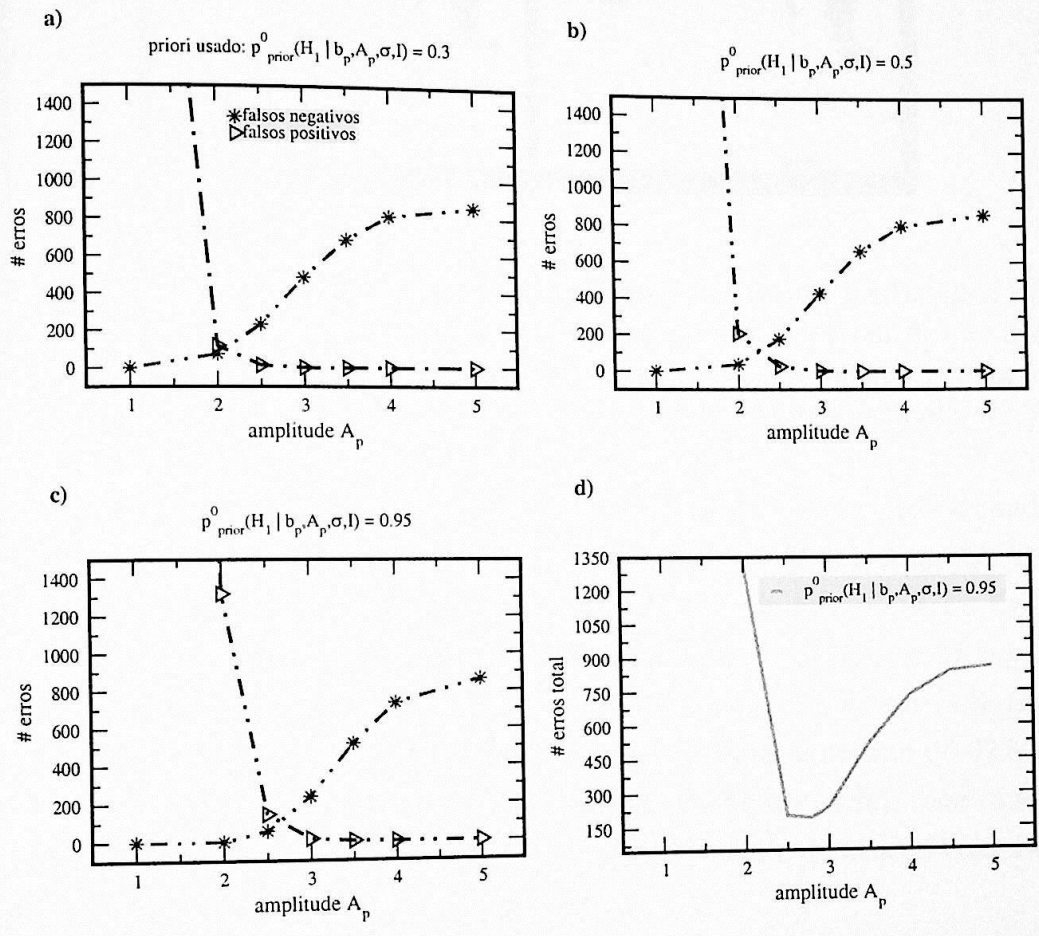


Figura 4.8: Resultado obtido da análise de séries temporais geradas com  $b_p = 100$ ,  $A_p = 3$ ,  $\sigma = 6$ , através do método *multigrid* bayesiano, atribuindo diferentes distribuições a priori para a atividade global da imagem,  $p_{prior}^0(H_1 | b_p, A_p, \sigma, I)$ , para diferentes valores da amplitude  $A_p$ , utilizando um limiar de 0.90.



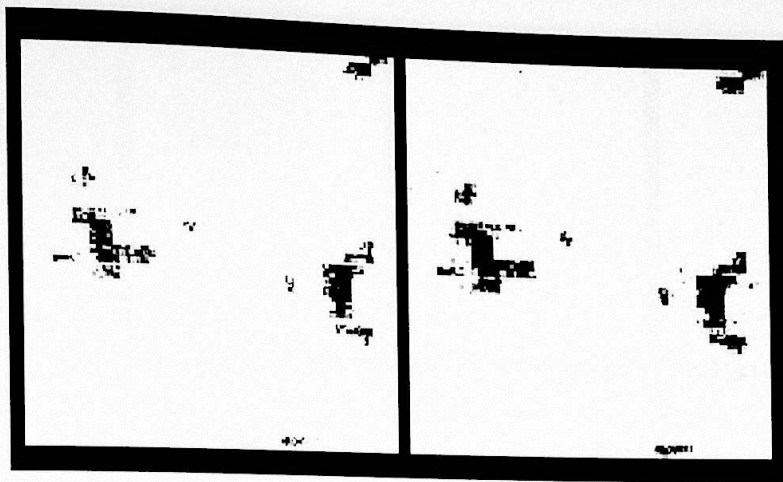


Figura 4.9: Duas imagens resultantes do último nível da escala hierárquica do método *multigrid* bayesiano, para séries temporais geradas com  $b_p = 100$ ,  $A_p = 3$ ,  $\sigma = 6$  e atribuindo  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.95$ . À esquerda, a imagem obtida com a amplitude correta,  $A_p = 3$  e à direita, a imagem obtida com a amplitude subestimada em 2.5 .

o método *multigrid* detecta todos os pixels inativos e nenhum pixel ativo. Quando usamos um a priori mais alto para a atividade global, figura (4.8b) e (4.8c), o melhor resultado é obtido com  $A_p = 2.5$ , amplitude mais próxima da correta, de modo que na figura (4.8c), quando atribuímos um a priori muito alta para a atividade global, ( $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.95$ ), o resultado obtido com  $A_p = 2$ , não é tão bom quanto na figura (4.8a), devido ao aumento de falsos positivos, onde o método *multigrid* detecta em torno de 92% dos pixels inativos. Isto fica claro quando resumimos o resultado da figura (4.8c), mostrando os erros totais cometidos pelo método com  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.95$ , de acordo com a amplitude considerada, figura (4.8d).

Assim, podemos afirmar que quando subestimamos a amplitude da resposta hemodinâmica, próxima à correta, o método *multigrid* bayesiano depende fortemente das propriedades globais da imagem. Na figura (4.9) mostramos a imagem final obtida com  $A_p = 2.5$  e  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.95$ , junto com a imagem final obtida com  $A_p$  correto da figura (4.3). Estes resultados mostram que a imagem obtida com a amplitude subestimada apresenta muito mais pixel ativos que quando usamos a amplitude correta, todavia os falsos positivos são mais acentuados naquela.

Para reforçar esses resultados, comparamos a seqüência de imagens obtida com a amplitude subestimada em 2 e  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.5$ , figura (4.10), com outra seqüência de imagens obtida com o mesmo valor da amplitude, mas agora atribuindo uma probabilidade



Figura 4.10: Sequência de imagens obtida na escala hierárquica do método *multigrid* bayesiano, com  $b_p = 100$ ,  $\sigma = 6$ ,  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.5$  e subestimando a amplitude em 2. Cada imagem ilustra, em escala de cinza, o nível de atividade de acordo a probabilidade posterior para  $H_1$  sobre cada região  $R$  da matriz  $(2^q \times 2^q)$ .

a priori muito alta para a atividade global da imagem, isto é,  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.95$ , figura (4.11). Podemos perceber claramente, que mesmo usando uma amplitude errada, o resultado da detecção pode ser otimizado por meio da informação que temos a priori sobre a atividade global da imagem.

#### Determinação do Parâmetro $A_p$

Nosso próximo passo foi escolher alguns pixels na imagem e estimar sua amplitude  $A_p$  dos dados. Como vimos na Seção (3.2), se conhecemos os valores de  $b_p$  e  $\sigma$ , a amplitude pode ser estimada encontrando o máximo da sua probabilidade posterior, pela equação (3.24). Para isto, calculamos o logaritmo dessa equação para vários valores da amplitude, entre 0 e 6, utilizando uma probabilidade a priori que reflete total ignorância sobre o parâmetro  $A_p$ , isto significa que atribuímos uma probabilidade a priori uniforme, como na equação (3.6). E então confeccionamos a curva da probabilidade posterior de  $A_p$  não normalizada, mas escalada, da forma descrita na Seção (3.2), para que seu maior valor seja um. Na figura



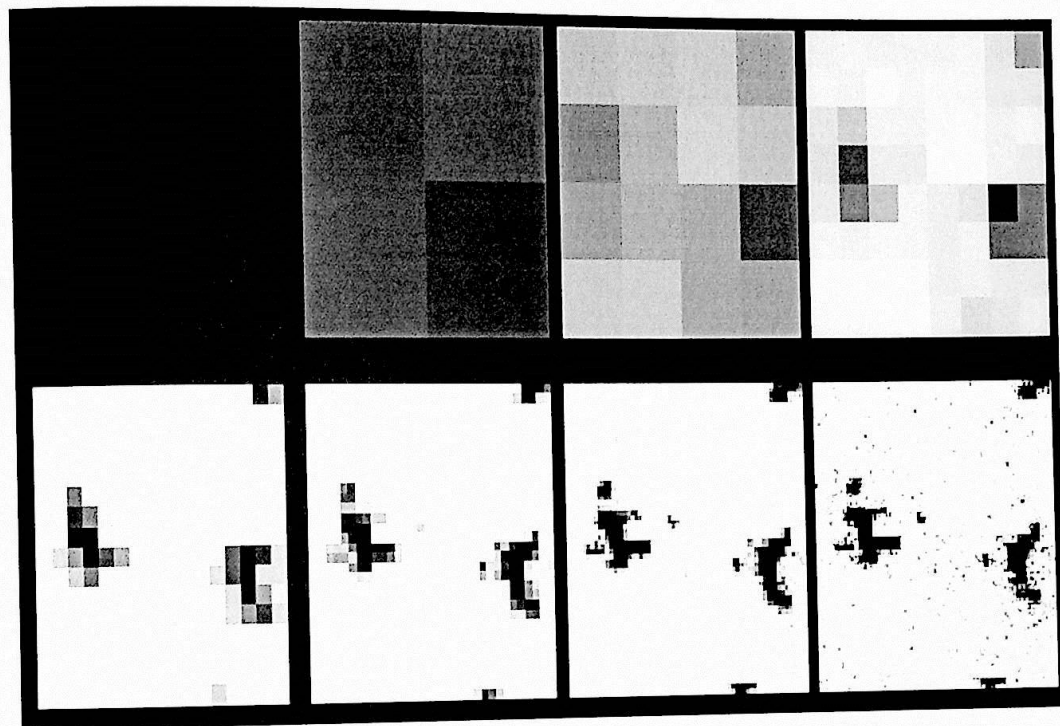


Figura 4.11: Seqüência de imagens obtida na escala hierárquica do método *multigrid* bayesiano, com  $b_p = 100$ ,  $\sigma = 6$ ,  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.95$  e subestimando a amplitude em 2. Cada imagem ilustra, em escala de cinza, o nível de atividade de acordo a probabilidade posterior para  $H_1$  sobre cada região  $R$  da matriz  $(2^q \times 2^q)$ .

(4.12) mostramos os resultados da estimativa da amplitude para seis pixels da imagem.

O gráfico (4.12a), mostra duas curvas do pdf posterior não normalizado correspondentes a dois pixel classificados corretamente pelo método *multigrid*, sendo que para o pixel inativo, o máximo do pdf posterior se encontra em  $A_p = 1$  e para o pixel ativo, encontramos um máximo em  $A_p = 2.5$ . No gráfico (4.12b), o pdf posterior resultante para um pixel inativo classificado como ativo, tem seu maior valor em  $A_p = 2$ , e para um pixel falso negativo o máximo do pdf posterior se encontra em  $A_p = 1.5$ . O gráfico (4.12c), o máximo do pdf posterior se encontra em  $A_p = 2.5$ , tanto para o pixel ativo como para o pixel inativo classificados corretamente. Por meio destes resultados, podemos perceber que não é possível tirar conclusões sobre o estado do pixel, simplesmente estimando sua amplitude.



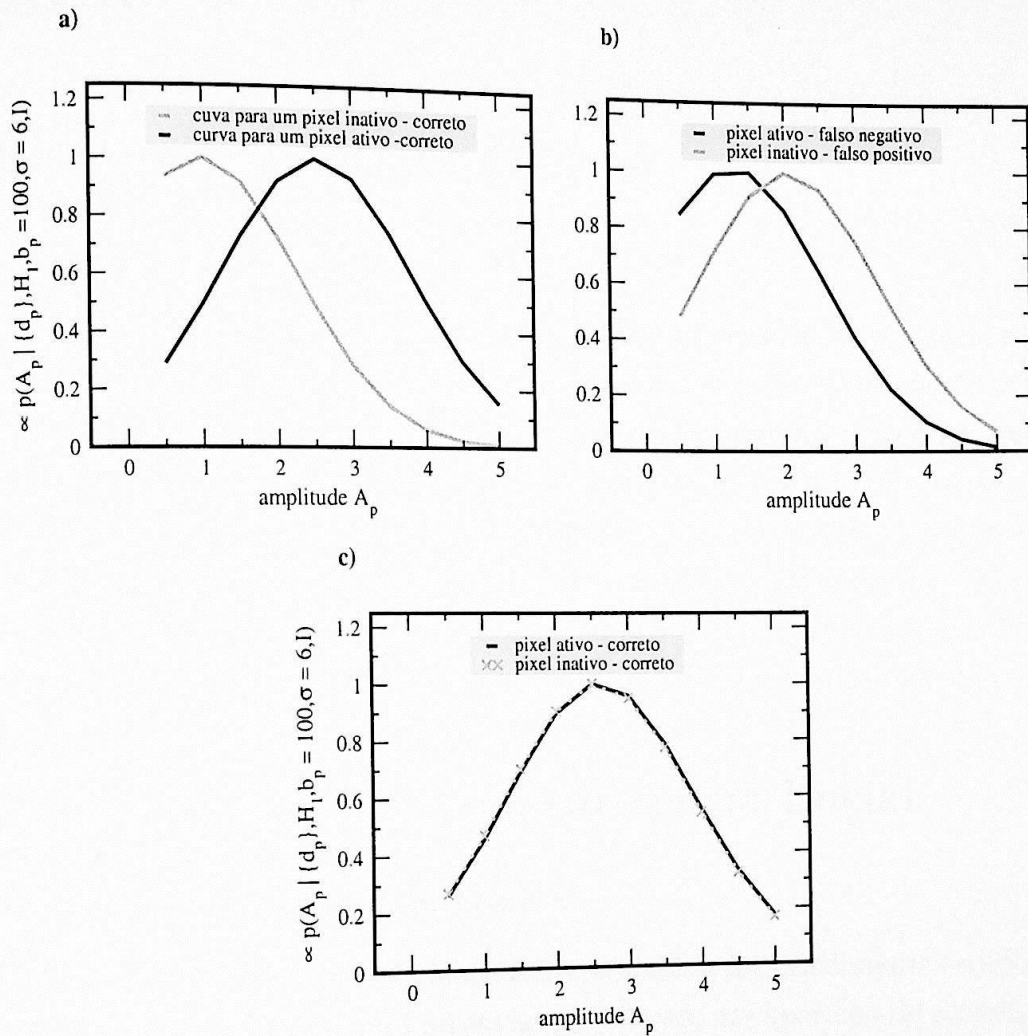


Figura 4.12: Resultados da estimativa do parâmetro  $A_p$ , para alguns pixels ativos e inativos da rede ( $2^7 \times 2^7$ ), utilizando o máximo do pdf posterior não normalizado.

## Capítulo 5

# Análise de Dados Reais

Este capítulo tem como objetivo testar dois métodos bayesianos, descritos anteriormente, com dados reais de fMRI: o método de pixel independente e o método *multigrid* bayesiano. Assim, analisamos dados fornecidos de um simples experimento motor em que o indivíduo movimentava voluntariamente os dedos das duas mãos. Na primeira seção deste capítulo, resumimos os métodos experimentais usados na aquisição desses dados e inferimos os parâmetros do modelo de séries temporais da equação (3.13). Nas duas últimas seções mostramos, respectivamente, os resultados obtidos pelo método de pixel independente e pelo método *multigrid* bayesiano, utilizando os parâmetros inferidos na seção anterior.

### 5.1 Os dados de fMRI e a Inferência Inicial

#### Protocolo de Aquisição dos Dados

O conjunto de dados reais que utilizamos neste trabalho, foi gentilmente cedido pelo Prof. Dráulio de Araújo [39]. Estes dados foram adquiridos em um aparelho de Ressonância Magnética Siemens, modelo Somatom Vision, de 1.5 Tesla, com bobina de cabeça de quadratura e polarização circular comercialmente disponível (em atividade regular no HCFMRP - Serviço de Radiodiagnóstico). Uma seqüência de 66 imagens Eco Planares (EPI), ponderadas em  $T_2^*$ , consistindo de  $128 \times 128$  pixels foi registrada usando os seguintes parâmetros: TR/TE - 3840/40 ms; ângulo -  $90^\circ$ ; FOV (*Field of view*) - 220 mm; largura da fatia - 3 mm. O arranjo experimental consistiu de seis períodos de repouso, e cinco de estímulo, cada época formada de 6 imagens (25 segundos). Foram descartadas as 2 primeiras imagens, devido à artefatos de saturação de  $T_1$ . Uma correção de movimento foi aplicada aos dados



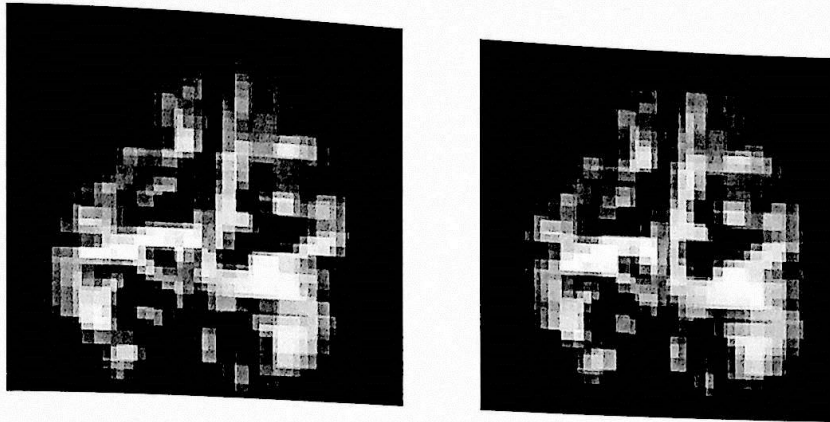


Figura 5.1: A imagem do lado esquerdo foi adquirida durante o repouso e a do lado direito durante a tarefa motora.

antes da análise estatística. Este é um passo comum de pré-processamento (observe, por exemplo, [30]).

Duas imagens, uma adquirida em repouso e a outra em atividade são mostradas na figura (5.1). “A olho nu” não conseguimos detectar diferença entre elas.

## Inferência Inicial

*Modelo da resposta hemodinâmica:* Nos capítulos anteriores, analisamos séries temporais utilizando o modelo de resposta hemodinâmica descrito na figura (2.8), com os parâmetros  $T_{at} = 30s$  e  $T_{rep} = 30s$ . Mas aqui, de acordo com o protocolo de aquisição, a resposta hemodinâmica foi modelada com  $T_{at} = 25s$  e  $T_{rep} = 25s$ .

*Inferência:* De posse dos dados, precisamos estimar a variância do ruído  $\sigma$  e a linha de base  $b_p$ , que como vimos, são os parâmetros do modelo de séries temporais da equação (3.13). Para isto, calculamos a linha de base  $b_p$  de acordo com a equação (3.23), que é a média do sinal sobre o tempo em cada pixel  $p$ , e subtraímos a mesma em cada dado  $d_{pt}$ , de forma que obtemos um conjunto de dados  $d'_p = \{d'_{p1}, \dots, d'_{pT}\}$  de média zero, isto é,

$$d'_{pt} = d_{pt} - b_p, \quad p = 1, 2, \dots, P, t = 1, 2, \dots, T. \quad (5.1)$$

Assim, da mesma forma que na subseção (3.2.1), fizemos o histograma dos dados  $D' = \{d'_1, \dots, d'_P\}$  e ajustamos o mesmo por uma distribuição gaussiana, utilizando o programa *xmgrace*, como ilustra a figura (5.2). Este resultado mostra que a curva de ajuste obtida foi uma distribuição gaussiana de média aproximadamente zero e variância 11, que são,

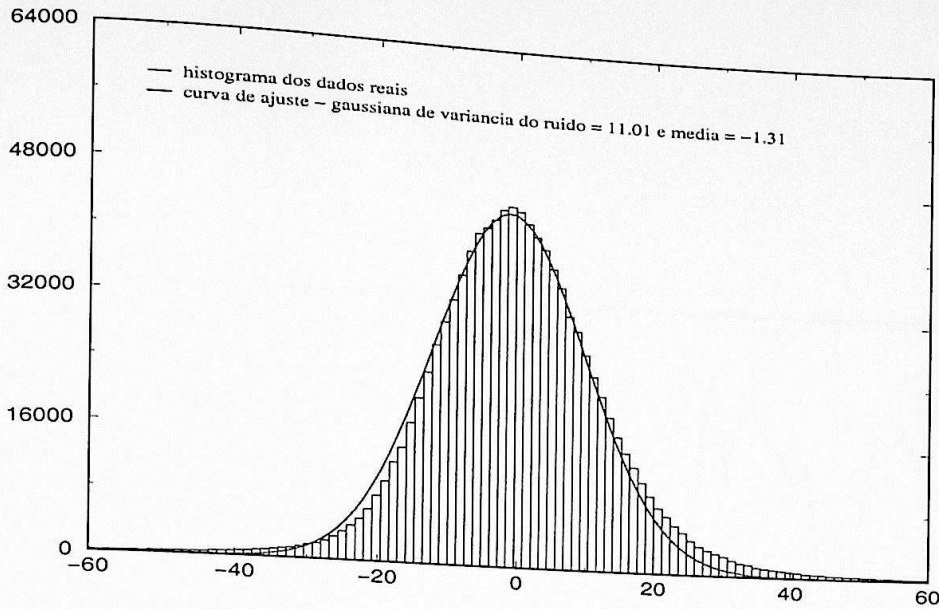


Figura 5.2: Histograma dos dados reais normalizados,  $D' = \{d'_1, \dots, d'_P\}$ , e sua curva de ajuste obtida com  $\sigma \sim 11$  e  $b_p \sim 0$ .

respectivamente, as estimativas da linha de base e da variância do ruído. Note que esse dados reais são semelhantes aos dados simulados no Capítulo 3, que também são regidos por uma distribuição gaussiana.

A figura (5.3) mostra as séries temporais normalizadas,  $(d'_{pt}, t = 1, \dots, 64)$ , tiradas de dois pixels com linhas verticais dividindo as condições. Mudanças de sinal que coincidem com o arranjo experimental são evidentes na primeira série temporal mas não na segunda.

A partir dos valores estimados de  $b_p$  e  $\sigma$ , estimamos a amplitude  $A_p$  dos dados  $d'_p$ , encontrando o máximo da sua probabilidade posterior, para todos os pixels da imagem, pela equação (3.24):

$$p(A_p|H_1, d'_p, b_p, \sigma, I) = \frac{L_1(b_p, A_p, \sigma) \times p(A_p|H_1, b_p, \sigma, I)}{N}, \quad (5.2)$$

sendo

$$L_1(A_p, b_p, \sigma) = (\sigma)^{-T} \times \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=1}^T [d'_{pt} - A_p h_t]^2\right\};$$

$H_1$  é a hipótese que o pixel apresenta atividade durante a execução da tarefa motora ( $a = 1$ );  $N$  é a constante de normalização; e  $p(A_p|H_1, b_p, \sigma, I)$  é a probabilidade a priori para  $H_1$ . Para isto, como na Seção (4.2), calculamos o logaritmo dessa equação para vários



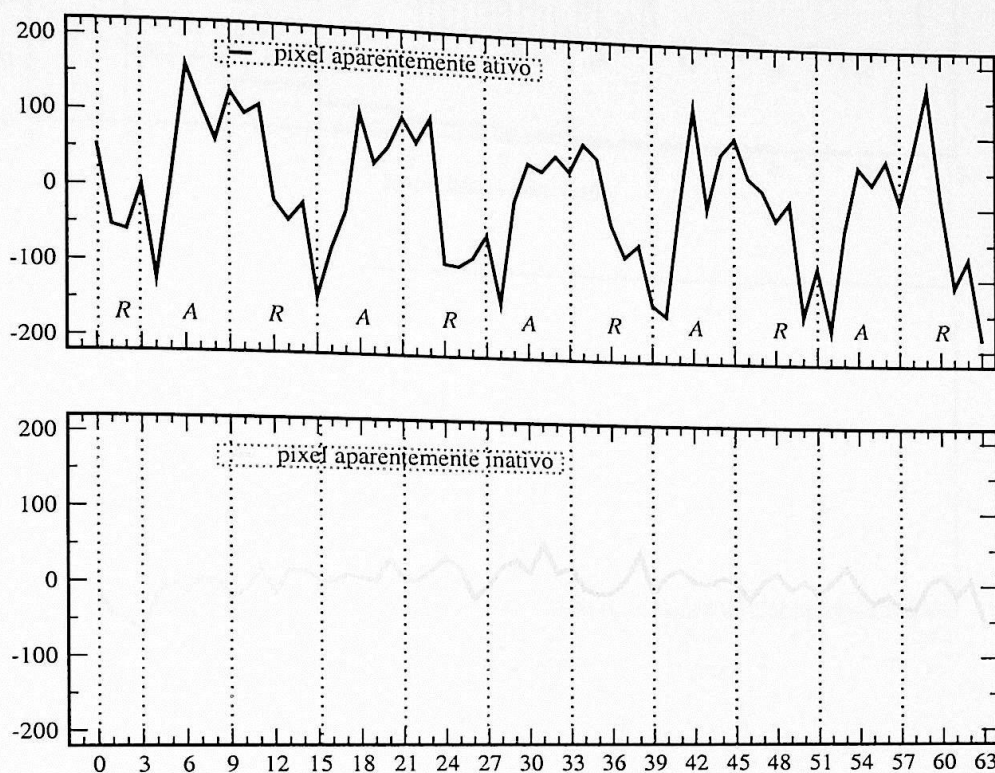


Figura 5.3: Duas séries temporais normalizadas de um experimento motor, *finger tapping*. As linhas verticais mostram a separação entre os períodos (R- repouso e A- ativação), e os períodos são indexados no eixo horizontal ao longo do índice de imagem (0 – 63). O eixo vertical refere-se aos valores do sinal de RM medido. A série temporal no painel de cima mostra um pixel aparentemente ativo; note a correspondência entre a tarefa motora e o padrão de mudança de sinal; ao contrário do painel de baixo, que mostra pouco evidência de atividade.

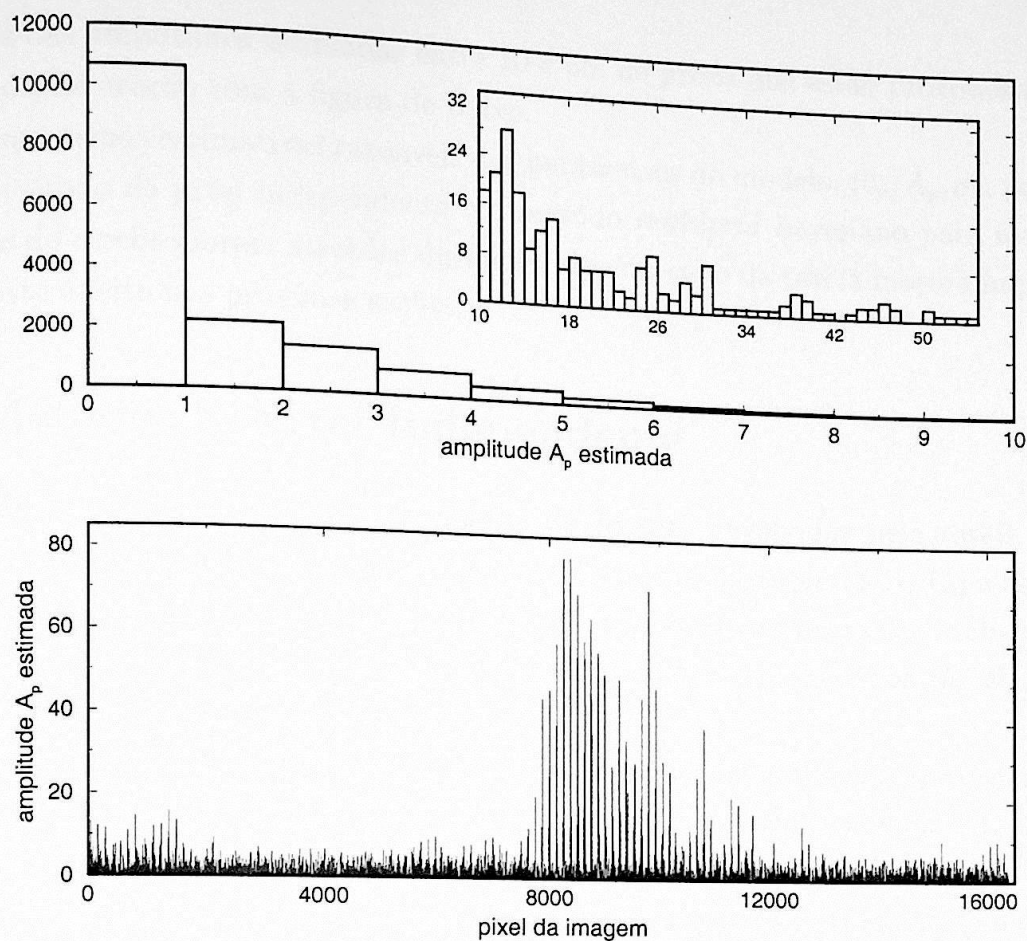


Figura 5.4: resultado da estimativa do parâmetro  $A_p$  para todos os pixels da imagem, utilizando o máximo do pdf posterior não normalizado. Abaixo, mostramos todos os pixels com suas respectivas amplitude estimadas, é notável a presença de um foco de ativação entre o pixel 8000 ao 11000, aproximadamente. Acima, ilustramos o histograma tirado do gráfico abaixo.

valores da amplitude, usando uma probabilidade a priori que reflete total ignorância sobre o parâmetro  $A_p$ , isto significa que atribuímos uma probabilidade a priori uniforme, como expresso na equação (3.6). Com isso, pegamos o valor da amplitude correspondente ao máximo da probabilidade posterior de  $A_p$  não normalizada, para cada pixel da imagem e colocamos em um histograma. O resultado disto é mostrado na figura (5.4).

O gráfico de baixo da figura (5.4) mostra como a amplitude da resposta hemodinâmica está distribuída na imagem, de acordo com a nossa estimativa. Com este gráfico, fizemos um histograma que foi separado em dois, mostrado no gráfico de cima: o histograma maior mostra amplitudes estimadas entre 0 e 10, em que 95% dos pixels tem uma amplitude pequena em relação ao valor da variância do ruído, entre 0 e 4; o gráfico interno mostra o



histograma das amplitudes estimadas entre 10 e 55, de pixels que estão próximos à região de atividade, de acordo com a figura de baixo.

Agora que temos estimativas razoáveis dos parâmetros do modelo,  $(b_p, A_p, \sigma)$ , podemos, utilizar o método de pixel independente e o método *multigrid* bayesiano para identificar regiões do cérebro foram ativadas durante o desempenho da tarefa motora imposta ao sujeito, e isto é feito nas próximas seções.

## 5.2 Método de Pixel Independente

Nesta seção, para tentar localizar as regiões do cérebro envolvidas pela tarefa motora primária, usamos o método de pixel independente, descrito na Seção (3.2). Com esse objetivo, aplicamos em cada pixel  $p$  ( $p = 1, \dots, 16384$ ), a condição da razão das probabilidades posterior, equação (3.22), dado os dados,  $(d'_{pt}, t = 1, \dots, 64)$ , os parâmetros,  $(b_p, A_p, \sigma)$ , e as probabilidades a priori para as hipóteses  $H_1$  e  $H_0$ .

Primeiramente, analisamos os dados reais assumindo uma amplitude igual para todos os pixels da imagem. Para isto, utilizamos  $A_p = 20$ , devido ao resultado da figura (5.4), que mostra que com este valor de amplitude podemos encontrar alguma atividade na imagem. Desta forma, na figura (5.5) apresentamos dois mapas estatísticos obtidos pelo método de pixel independente utilizando o valor estimado da variância,  $\sigma = 11$ , onde em cada mapa foi atribuído uma probabilidade diferente a priori para a hipótese do pixel estar ativo com respeito a tarefa motora,  $p(H_1|A_p, b_p, \sigma, I) = \{0.01, 0.5\}$ . Note que quando somos imparciais,  $p(H_1|A_p, b_p, \sigma, I) = 0.5$ , ou seja, quando não temos preferência a priori para nenhuma das hipóteses  $H_0$  ou  $H_1$ , o número de pixels classificados como ativos cresce, em relação a um a priori muito mais baixo usado para a atividade do pixel,  $p(H_1|A_p, b_p, \sigma, I) = 0.01$ . Ou seja, o método detecta mais pixels fora da imagem quando diminuimos nossa preferência sobre  $H_1$ .

A partir desses resultados, nosso próximo passo foi comparar o mapa estatístico obtido na figura (5.5) com  $p(H_1|A_p, b_p, \sigma, I) = 0.01$  com outros dois mapas, cada um obtido com um valor de amplitude maior que o primeiro. Assim, na figura (5.6) apresentamos os três mapas estatísticos obtidos com  $\sigma = 11$ ,  $p(H_1|A_p, b_p, \sigma, I) = 0.01$ , onde cada um deles foi obtido assumindo uma amplitude diferente,  $A_p = \{20, 30, 50\}$ . Podemos notar que quando assumimos uma amplitude que é um pouco menos que o triplo da variância do ruído,  $A_p = 30$ , o método de pixel independente classifica mais pixels inativos e menos como ativos em relação a figura obtida  $A_p = 20$ , e esta diferença aumenta quando  $A_p = 50$ , de

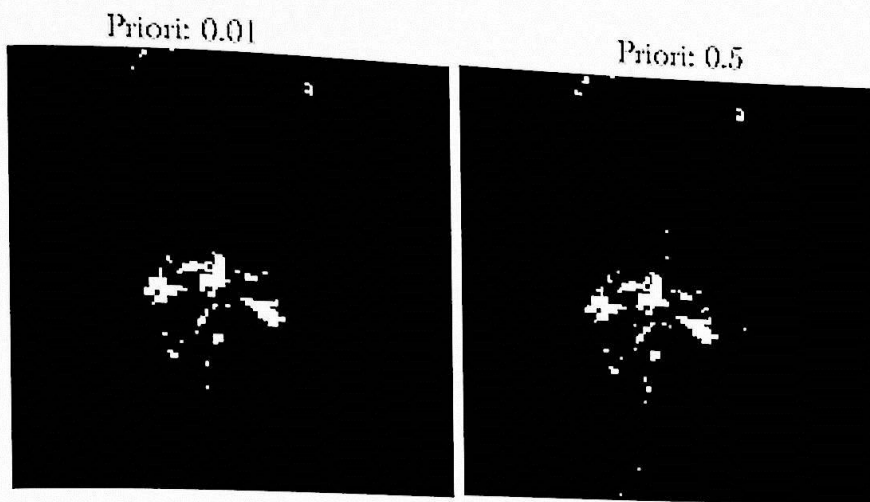


Figura 5.5: Mapa estatístico obtido pelo método de pixel independente, utilizado o valor estimado da variância,  $\sigma = 11$ , amplitude uniforme,  $A_p = 20$  e atribuindo, à esquerda, uma probabilidade a priori muito baixo para a atividade do pixel na imagem,  $p(H_1|A_p, b_p, \sigma, I) = 0.01$ , e, à direita, uma probabilidade a priori imparcial sobre o estado do pixel,  $P(H_1|A_p, b_p, \sigma, I) = 0.5$ .

modo que nesta última corremos o risco de aumentar o número de falsos negativos, assim, apenas os pixels ativos que apresentem respostas mais robustas durante a tarefa motora são detectados com aquela amplitude.

Consideremos agora o valor estimado da amplitude em cada pixel da imagem, de acordo com o resultado obtido na figura (5.4). Desta forma, analisamos os dados utilizando, novamente,  $p(H_1|A_p, b_p, \sigma, I) = 0.01$ ,  $\sigma = 11$  e o parâmetro  $A_p$  estimado em cada pixel. Na figura (5.7) mostramos o mapa estatístico resultante dessa análise. Note que este resultado é parecido com o obtido com  $A_p = 20$ , figura (5.5), mas o número de pixels classificados como ativos é maior.

Para ajudar na localização das regiões ativadas no cérebro, fizemos a superposição do mapa estatístico obtido com  $A_p = 50$ , figura (5.6), a uma imagem. Este resultado é apresentado na figura (5.8). Atribuímos uma coloração vermelha na imagem, em pixels onde o método detectou atividade.

Note que com este valor de amplitude, o método de pixel independente detecta ativações em dois pontos isolados e em três regiões do cérebro, em que estas, provavelmente, são responsáveis pela tarefa motora.



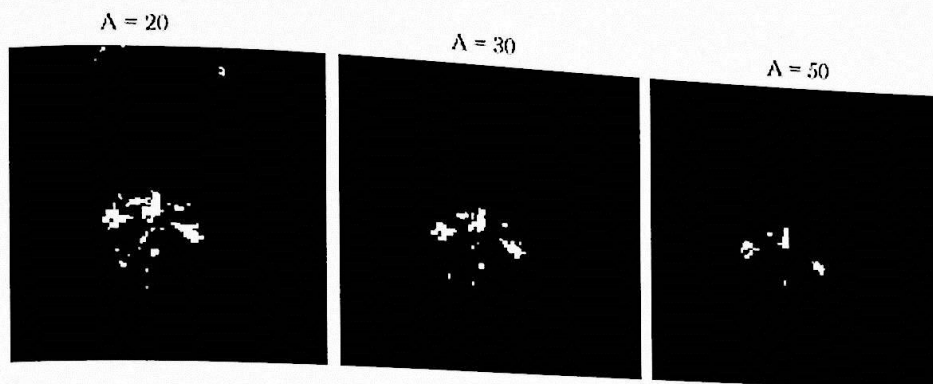


Figura 5.6: Três mapas estatísticos obtidos pelo método de pixel independente, utilizado o valor estimado da variância,  $\sigma = 11$ , uma probabilidade a priori muito baixo para a atividade do pixel na imagem,  $p(H_1|A_p, b_p, \sigma, I) = 0.01$ , e uma amplitude igual para todos os pixels, da esquerda para direita,  $A_p = 20, 30$  e  $50$ , respectivamente.

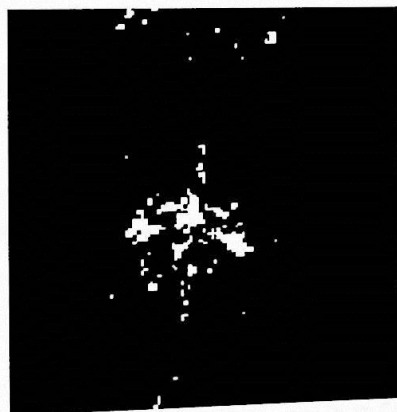


Figura 5.7: Mapa estatístico obtido pelo método de pixel independente, utilizado o valor estimado da variância,  $\sigma = 11$ , uma probabilidade a priori muito baixo para a atividade do pixel na imagem,  $p(H_1|A_p, b_p, \sigma, I) = 0.01$ , e a amplitude estimada em cada pixel.



Figura 5.8: A imagem ilustra em vermelho regiões onde detectamos prováveis ativações durante a tarefa motora, pelo método de pixel independente, utilizando  $\sigma = 11$ ,  $A_p = 50$  e atribuindo um a priori muito baixo para a atividade do pixel na imagem,  $p(H_1|A_p, b_p, \sigma, I) = 0.01$ .

### 5.3 Método *Multigrid* Bayesiano

Com a mesma finalidade da seção anterior, analisamos os dados reais,  $(d_{pt}^j, t = 1, \dots, T)$ , da Seção (5.1), pelo método *multigrid* bayesiano, capítulo 4, usando os valores estimados dos parâmetros  $(A_p, \sigma)$  e atribuindo uma probabilidade a priori para a atividade global da imagem,  $p_{prior}^0(H_1|b_p, A_p, \sigma, I)$ .

Novamente, começamos o estudo analisando os dados utilizando  $A_p = 20$ ,  $\sigma = 11$ , e atribuindo uma a priori muito alta para a atividade global,  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.95$  e uma a priori mais baixa,  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.5$ . Na figura (5.9) mostramos os dois mapas estatísticos obtidos no último nível da escala hierárquica ( $q = 7$ ). Note que nos dois mapas aparecem apenas três regiões com probabilidades muito altas de estarem ativadas e mostrando que para este valor de amplitude, as propriedades globais da imagem são praticamente irrelevantes para detectar ativação.

Desta forma, resolvemos analisar os dados usando uma amplitude diferente,  $A_p = 10$ , e atribuindo diferentes probabilidades a priori para a atividade global,  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = \{0.95, 0.5, 0.05\}$ , figura (5.10). Cada mapa estatístico mostrado na figura foi obtido usando uma priori diferente. Comparando estas figuras, observamos que o mapa estatístico obtido com  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.05$ , é mais claro, com menos pontos espalhados que os outros mapas obtidos com um priori maior.



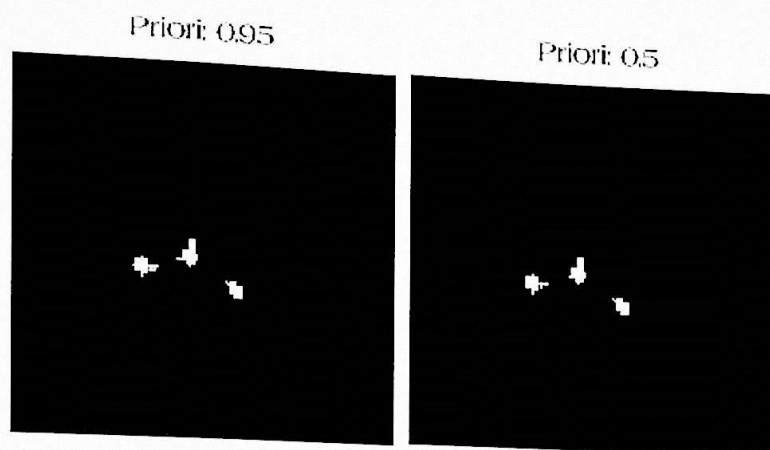


Figura 5.9: Dois mapas estatísticos obtidos pelo método *multigrid* bayesiano no último nível da escala hierárquica,  $q = 7$ , utilizando  $\sigma = 11$ ,  $A_p = 20$  e atribuindo, à esquerda, uma probabilidade a priori muito alta para atividade global da imagem,  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.95$ , e, à direita, uma probabilidade a priori imparcial sobre o estado global da imagem,  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.5$ .

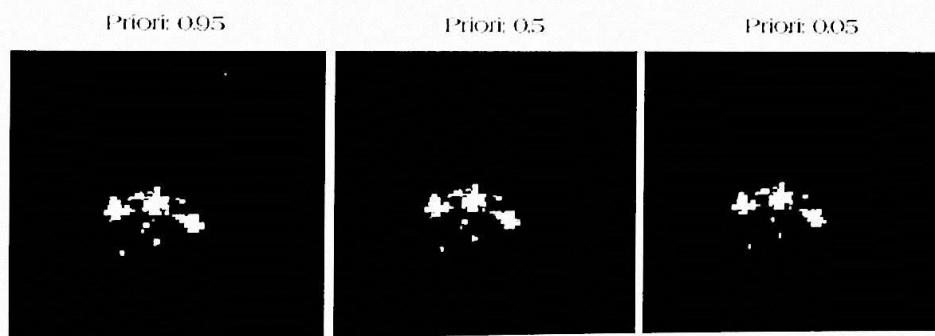


Figura 5.10: Três mapas estatísticos obtidos pelo método *multigrid* bayesiano no último nível da escala hierárquica,  $q = 7$ , utilizando  $\sigma = 11$ ,  $A_p = 10$  e atribuindo, da esquerda para a direita, uma probabilidade a priori muito alta para a atividade global da imagem,  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.95$ , um probabilidade a priori imparcial sobre o estado global da imagem,  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.5$ , e um priori muito baixo para atividade global da imagem,  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.05$ .

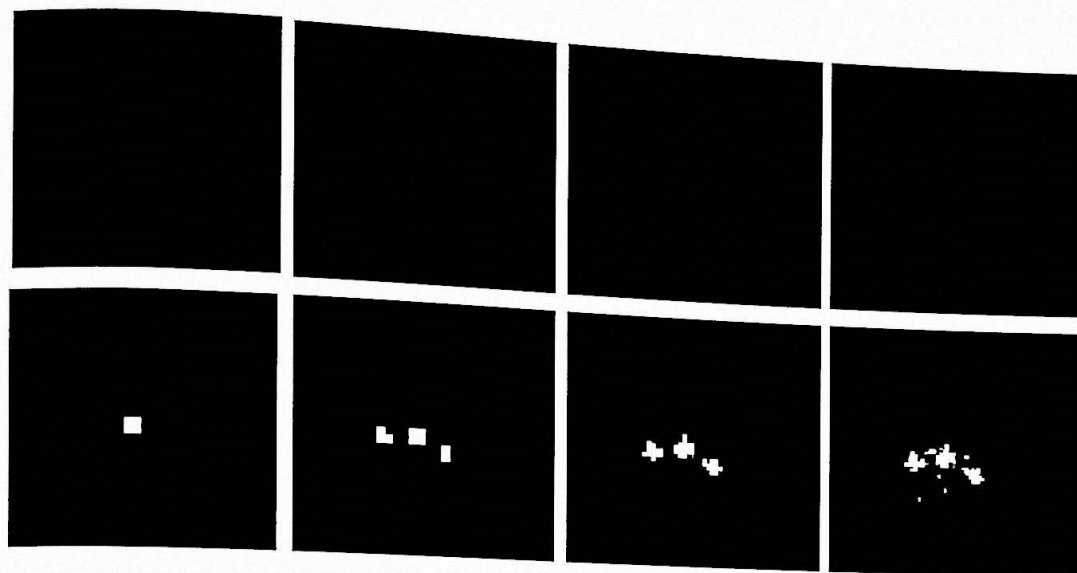


Figura 5.11: Sequência de imagens obtida na escala hierárquica do método *multigrid* bayesiano, utilizando  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.05$ ,  $\sigma = 11$  e amplitude uniforme,  $A_p = 10$ . Cada imagem ilustra, em escala de cinza, o nível de atividade de acordo a probabilidade posterior para  $H_1$  sobre cada região  $R$  da matriz ( $2^q \times 2^q$ ).

Com isso, na figura (5.11) mostramos a seqüência de imagens obtida na escala hierárquica, usando  $A_p = 10$ ,  $\sigma = 11$  e  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.05$ . Note que a imagem obtida em  $q = 0$  tem um nível de atividade global praticamente nulo e permanece assim até em  $q = 3$ . Somente em  $q = 4$ , uma única região aparece com um nível de atividade muito alto, e partir do nível seguinte, três regiões são classificadas como ativadas com nível de atividade aumentados à medida que a escala cresce. Assim, o mapa estatístico resultante no nível  $q = 7$  apresenta estruturas e alguns pontos isolados com níveis de atividade muito altos.

No próximo passo foi atribuir  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.95$  para analisar os dados usando diferentes amplitudes,  $A_p = \{12, 18, 25, 30\}$  e  $\sigma = 11$ . O mapa estatístico resultante de cada amplitude é mostrado na figura (5.12). Note que o resultado obtido com  $A_p = 12$  é similar ao da figura (5.11) com  $A_p = 10$ , aparecendo as mesmas estruturas e uma quantidade menor de pontos isolados. A medida que aumentamos a amplitude, o tamanho destas estruturas diminui, de modo que quando usamos  $A_p = 30$ , estas são muito menores ocorrendo um provável aumento de falsos negativos na imagem.

Se compararmos todos os resultados obtidos acima, observamos que o método *multigrid* detecta as mesmas estruturas da imagem, com probabilidades muito altas de estarem



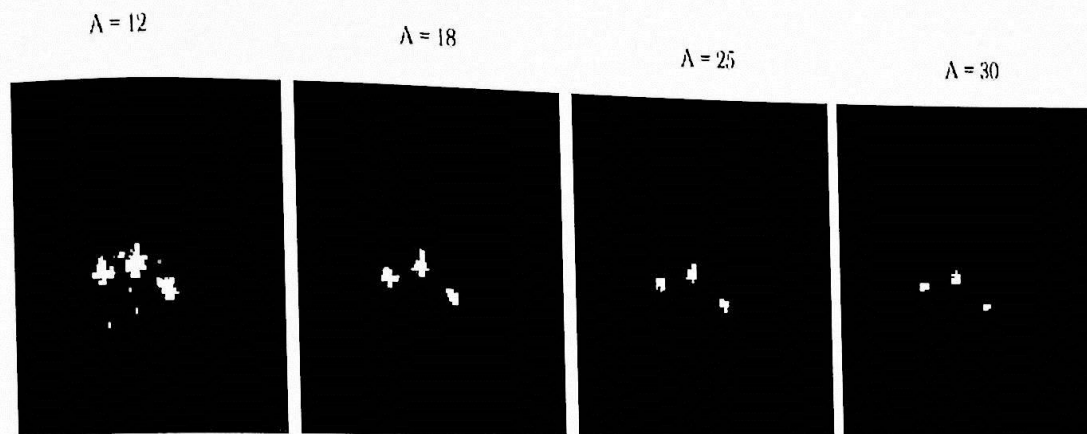


Figura 5.12: Mapas estatísticos obtidos no último nível da escala hierárquica ( $q = 7$ ) do método *multigrid* bayesiano, utilizando  $p_{prior}^0(H_1|b_p, A_p, \sigma, I) = 0.95$ ,  $\sigma = 11$  e uma amplitude igual para todos os pixels.

ativadas.

Portanto, para encontrar a localização dessas regiões ativadas no cérebro, escolhemos três mapas estatísticos obtidos, respectivamente, com  $A_p = 18, 20$  e  $25$ , para serem superpostos em uma imagem. Assim, atribuímos uma coloração avermelhada em cada pixel da imagem de acordo com sua probabilidade posterior de estar ativo. Os resultados são mostrados na figura (5.13). Estas estruturas mostradas em vermelho, correspondem, provavelmente, às áreas motoras do córtex cerebral humano.

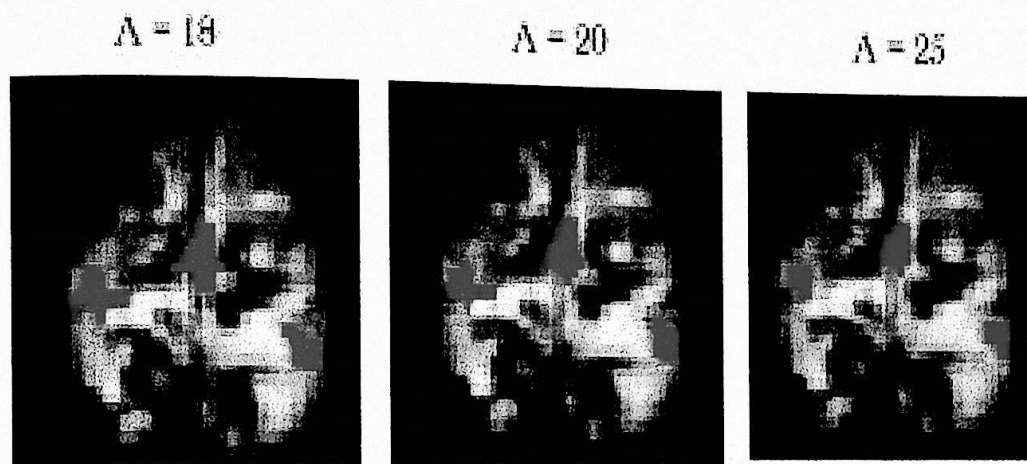


Figura 5.13: Imagens de ativação originárias de um experimento motor, analisadas pelo método *multigrid* bayesiano, usando um a priori muito alto para a atividade global,  $P_{prior}^0(H_1|A_p, b_p, \sigma, I) = 0.95$ , a variância estimada,  $\sigma = 11$ , e uma amplitude uniforme na imagem,  $A_p = \{18, 20, 25\}$ . O nível de atividade de cada pixel, proporcional à probabilidade posterior é representado em tons de vermelho.



## Capítulo 6

# Conclusões e Perspectivas

Neste trabalho, com o objetivo de identificar regiões ativadas no cérebro durante a execução de alguma tarefa imposta ao sujeito, via imagem funcional por ressonância magnética (fMRI), utilizando um arranjo experimental em bloco, apresentamos e implementamos dois métodos bayesianos: o de pixel independente e o método *multigrid* bayesiano.

Em simulações com dados artificiais, os resultados obtidos pelo método de pixel independente, que detecta ativação de pixel a pixel simplesmente aplicando o teste de hipótese bayesiano, mostraram a importância da informação a priori, antes da análise dos dados, para a classificação dos pixels na imagem como ativos ou inativos. A partir disso e usando o fato que ativação real tende a estar mais agrupada que ativação artificial causada por ruído, propusemos o método *multigrid* bayesiano. Seu objetivo é obter estimativas dos valores corretos das distribuições a priori usando propriedades globais da imagem, e assim obtivemos resultados muito melhores. Constatamos que este método é mais sensível em detectar ativação, com erros de falsos positivos muito baixos e que ele é capaz de detectar pixels ativos em imagens corrompidas por muito ruído.

Em dados reais fornecidos de um experimento simples de estímulo motor, em que a resposta do sinal BOLD medido é mais robusta, os resultados obtidos pelo método de pixel independente, utilizando diferentes amplitudes da resposta hemodinâmica e atribuindo um a priori muito baixo para a atividade do pixel, apresentaram algumas estruturas parecidas, mas com muitos pontos isolados na imagem, exceto quando usamos uma amplitude muito alta,  $A_p = 50$ , em que aquelas estruturas apresentaram-se mais claras na imagem. Já com o método *multigrid* bayesiano, praticamente todos os resultados obtidos apresentaram estruturas parecidas e mais definidas, que parecem corresponder às regiões corticais de projeção motora do cérebro.

Ainda resta o seguinte problema, para implementar o método *multigrid* no estágio atual é necessário estimar certos parâmetros, tais como  $A_p$ ,  $\sigma$ ,  $b_p$ . Estes valores não são necessários no que tange à resposta final, que visa atribuir níveis de atividade neural. Do ponto de vista de inferência Bayesiana não passam de parâmetros de incômodo. Portanto para eliminá-los devemos integrar sobre todos os seus possíveis valores, com seu peso adequado que é a sua distribuição a priori. Este procedimento pode ser muito custoso do ponto de vista numérico e portanto pode haver necessidade de implementações de aproximações que o tornem factível.

Sendo assim, podemos esperar que o método *multigrid* bayesiano se transforme em uma eficiente ferramenta. Os resultados são preliminares e incentivam a continuar nesta linha. O nosso objetivo principal é desenvolver esta ferramenta até um nível em que possa ser usada dentro de um ambiente de trabalho hospitalar. Há uma certa distância entre o estágio atual e a ferramenta, e percorrer esta distância implica em resolver certos problemas tanto teóricos quanto práticos.

Há necessidade de quantificar o sucesso deste método e compará-lo a outros já desenvolvidos. Também é necessário a implementação da técnica em três dimensões, para análise de séries temporais em voxels da imagem, para um ganho maior de informações espaciais. Estudar e incluir a especificidade da resposta hemodinâmica, tanto sua forma funcional quanto a determinação dos parâmetros envolvidos, uma vez que é sabido que a forma temporal desta varia através do cérebro e é muito sensível à vasculatura local.



## Referências Bibliográficas

- [1] Posner, M., Raichler, M. (1994) Images of Mind. New York. *Scientific American Library*.
- [2] Mazoyer, B. (1987) La tomographie par émission de positons. *Bulletin de la Société Française de Physique* **64**,3-9.
- [3] Purcell, E.M., Torrey, H.C. and Pound, R.V.(1946) Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Phys. Rev.* **69**,37-38.
- [4] Bloch, F., Hansen, W.W. and Packard, M. (1946) Nuclear Induction. *Phys. Rev.* **69**,127.
- [5] Lauterbur, P.C. (1973) Image Formation by Induced Interactions: Examples Employing Nuclear Magnetic Resonance. *Nature* **242**,190-191.
- [6] Callaghan, P. T. (1991) Principles of Nuclear Magnetic Resonance Microscopy, Clarendon Press, Oxford.
- [7] King, Margareth M. Basic Principles of MRI.
- [8] Hans, E. L. (1950) Spins Echoes. *Phys. Rev.* **80**,580.
- [9] Slichter, C.P. (1989) Principles of Magnetic Resonance.
- [10] Damadian, R., Goldsmith, M. and Minkoff, L. (1977) FONAR Image of the Live Human Body. *Physiol. Chem.* **9**,97-100.
- [11] Hounsfield, G.N. (1973) Computerized Transverse Axial Scanning (Tomography). *Br. J. Radiol.* **46**,10116-1022.
- [12] Gowland, P. and Mansfield, P. (1993) accurate Measurement of  $T_1$  in vivo in Less Than 3 Seconds Using Echo-Planar Imaging. *Magn. Reson. Med.* **30**,351-354.

- [13] Lazar, Nicole A., Eddy, William F., Genovese, C. R. and Welling, J. (2001) Statistical Issues in fMRI for Brain Imaging. *International Statistical Review* **69**,105-127.
- [14] Thulborn, K. R. Waterton, J. C., Mathews, P. M. and Radda, G. K. (1982) Oxygen Dependence of the Transverse Relaxation Time of Water Protons in Whole Blood at High Field. *Biochim. Biophys. Acta.* **714**,265-270.
- [15] Ogawa, S., Lee, T. M., Nayak., A. S. and Glynn, P. (1990) Oxygenation-Sensitive Contrast in Magnetic Resonance Image of Rodent Brain at High Magnetic Fields. *Magn. Reson. Med.* **14**,68-78.
- [16] Ogawa, S. and Lee, T. M. (1990) Magnetic Resonance Imaging of Blood Vessels at High Fields: In Vivo and in Vitro Measurements and Image Simulation. *Magn. Reson. Med.* **16**,9-18.
- [17] Ogawa, S., Lee, T. M., Kay, A. R. and Tank, D. W. (1990) Brain Magnetic Resonance Imaging with Contrast dependent on Blood Oxygenation. *Proc. Natl. Acad. Sci. USA* **87**,9868-9872.
- [18] Turner, R., Le Bihan, D., Moonen, C. T. W., Despres, D. and Frank, J. (1991) Echo-Planar Time Course MRI of Cat Brain Oxygenation Changes. *Magn. Reson. Med.* **22**,159-166.
- [19] Jezzard, P., Heinemann, F., Taylor, J., Despres, D., Wen, H., Balaban, R. S., and Turner, R. (1994) Comparison of EPI Gradient-Echo Contrast Changes in Cat Brain Caused by Respiratory Challenges with Direct Simultaneous Evaluation of Cerebral Oxygenation via a Cranial Window. *NMR in Biomed.* **7**,35-44.
- [20] Cohen, M. S. and Bookheimer, S. Y. (1994) Localization of Brain Function using Magnetic Resonance Imaging. *TINS* **17**,268-276.
- [21] Villringer, A., Planck, J., Hock, C., Schleinkofer, L. and Dirmagl, U. (1993) Near Infrared Spectroscopy (NIRS): A New Tools to Study Hemodynamic Changes During Activation of Brain Function in Human Adults. *Neurosci. Lett.* **154**,101-104.
- [22] Gossl, C., Auer, D. P., and Fahrmeir, L. (2001) Bayesian Modeling of the Hemodynamic Response Function in BOLD fMRI. *NeuroImage* **14**,140-148.



- [23] Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., Kennedy, D. N., Hoppel, B. E., Cohen, M. S., Turner, R., Cheng, H. M., Brady, T. J. and Rosen, B. R. (1992) Dynamic Magnetic Resonance Imaging of Human Brain Activity During Primary Sensory Stimulation. *Proc. Natl. Acad. Sci. USA* **89**,5675-5679.
- [24] Turner, R., Jessard, P., Wen, H., Kwong, K. K., Le Bihan, D., Zeffiro, T. and Balaban, R. S. (1993) Functional Mapping of the Human Visual Cortex at 4 and 1.5 Tesla Using Deoxygenation Contrast EPI. *Magn. Reson. Med.* **29**,277-279.
- [25] Le, T. H and Hu, X. (1996) Retrospective Estimation and Correction of Physiological Artifacts in fMRI by Direct Extraction of Physiological Activity from MR Data. *Magn. Reson. Med.* **35**,290-298.
- [26] Gao, J. H., Miller, I., Lai, S., Xiong, J. and Fox, P. T. (1996) Quantitative Assessment of Blood Inflow Effects in Functional MRI Signals. *Magn. Reson. Med.* **36**,314-319.
- [27] Fisel, C. R., Ackerman, J. L., Buxton, R. B., Garrido, L., Belliveau, J. W., Rosen, B. R. and Brady, T. J. (1991) MR Contrast Due to Microscopically Heterogeneous Magnetic Susceptibility: Numerical Simulations and Applications to Cerebral Physiology. *Magn. reson. Med.* **17**,336-347.
- [28] Hykin, J., Bowtell, R., Mansfield, P., Glover, P., Coxon, R., Worthington, B. and Blumhardt, L. (1994) Functional Brain Imaging using EPI at 3T. *Magma* **2**,347-349.
- [29] Gandle, A.J., Hill, D.L.G., Stundholme, C. et al. (1994) Combined and 3D rendered multimodal data for planning skull base surgery: a prospective evaluation. *Neurosurgery* **35**,463-471.
- [30] Genovese, C. R. (2000). A bayesian time-course model for functional magnetic resonance imaging data (with discussion). *Journal of the American Statistical Association* **95**,691-719.
- [31] Bandettini, P. A., Jesmanowicz, A., Wong, E. C. and Hyde, J. S. (1993) Processing Strategies for Time-Course Data Sets in Functional MRI of the Human Brain. *Magn. Reson. Med.* **30**,161-173.
- [32] Friston, K. J., Jezzard, P. and Turner, R. (1994) Analysis of Functional MRI Time-Series. *Human Brain Mapping* **1**,153-171.

- [33] Lange, N. and Zeger, S. (1997) Non-linear Fourier time series analysis for human brain mapping by functional magnetic resonance imaging. *Journal of the Royal Statistical Society C Applied Statistics* **46**,1-29.
- [34] Eddy, W. F. (1997) Comment on Lange and Zeger. *Journal of the Royal Statistical Society C* **46**,19-20.
- [35] Genovese, C. R. (1997) Comment on Lange and Zeger. *Journal of the Royal Statistical Society C* **46**,23-24.
- [36] Jaynes, E. T. Probability Theory: The Logic of Science Incompleto
- [37] Sivia, D. S. (1996) Data Analysis - A Bayesian Tutorial. Clarendon Press-Oxford.
- [38] Jeffreys, E. T. (1983) Theory of Probability. Oxford University Press, London.
- [39] Prof. Dr. Dráulio Barros de Araújo – Laboratório de Biomagnetismo -Departamento de Física e Matemática-FFCLRP-USP.