



RODNEI ALVES MARQUES

**EXPOENTES DE HURST EM IMAGENS
INTERFEROMÉTRICAS DE UM CARCINOMA MAMÁRIO**

LAVRAS – MG

2019

RODNEI ALVES MARQUES

**EXPOENTES DE HURST EM IMAGENS INTERFEROMÉTRICAS DE UM
CARCINOMA MAMÁRIO**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

Prof(a). Dr(a). Thelma Sáfadi
Orientadora

**LAVRAS – MG
2019**

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Marques, Rodnei Alves.

Expoentes de Hurst em Imagens Interferométricas de um
Carcinoma Mamário / Rodnei Alves Marques. - 2019.

87 p. : il.

Orientador(a): Thelma Sáfyadi.

Tese (doutorado) - Universidade Federal de Lavras, 2019.
Bibliografia.

1. Análise multirresolução. 2. Transformada de wavelet. 3.
Expoente de Hurst. I. Sáfyadi, Thelma. II. Título.

RODNEI ALVES MARQUES

**EXPOENTES DE HURST EM IMAGENS INTERFEROMÉTRICAS DE UM
CARCINOMA MAMÁRIO
HURST EXPONENTS ON INTERFEROMETRIC IMAGES OF A BREAST
CARCINOMA**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

APROVADA em 08 de Novembro de 2019.

Prof. Dr. Renato Ribeiro de Lima	UFLA
Prof. Dr. Paulo Henrique Sales Guimarães	UFLA
Prof. Dra. Chang Chiann	USP
Prof. Dr. Roberto Alves Braga Junior	UFLA

Prof(a). Dr(a). Thelma Sáfadi
Orientadora

**LAVRAS – MG
2019**

AGRADECIMENTOS

A Deus, pela vida, saúde e determinação.

À Universidade Federal de Lavras (UFLA), em especial ao Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, pela oportunidade concedida para realização do doutorado.

Ao Instituto Federal de Minas Gerais (IFMG) campus Ouro Branco, pela concessão de licença remunerada para capacitação.

A professora Dra. Thelma Sáfy, orientadora, companheira e amiga em todas as etapas desta proposta, pela compreensão, paciência, sugestões e críticas, que muito contribuíram para o enriquecimento desta pesquisa.

Aos professores do Departamento de Estatística (DES): Dr. Devanil Jaques de Souza, Dr. Lucas Monteiro Chaves, Dr. Marcelo Silva de Oliveira, Dr. Renato Ribeiro de Lima, Dr. Paulo Henrique Sales Guimarães, Dr. Júlio Sílvio de Sousa Bueno Filho, Dr. Joel Augusto Muniz e Dr. Daniel Furtado Ferreira pelos conhecimentos compartilhados, auxílios nos trabalhos e pela harmoniosa convivência.

Ao meu pai José Braz, por acreditar em mim e por sempre me incentivar a novos desafios e nunca desistir. À minha mãe Euclídia de Jesus pelas orações e palavras de consolo.

À minha esposa, Camila Íris, que me deu tranquilidade, motivação e compreensão em todos os momentos desta etapa, sempre demonstrando seu amor e carinho, pois sem o seu apoio este projeto não seria possível, tampouco, participar do Programa de Doutorado em Estatística.

Aos amigos: Cristian Tiago Erazo Mendes, Rafael Agostinho Ferreira, Patrícia Mendes dos Santos e Vânia de Fátima Lemes de Miranda que, de alguma forma, contribuíram para a realização desta etapa.

Enfim, agradeço a todos que, pela presença, pelas palavras, pelo sorriso ou pela simples lembrança, deram-me coragem e determinação para alcançar meus objetivos.

Muito Obrigado!

*"Educação é uma descoberta progressiva de nossa própria ignorância".
(Voltaire)*

RESUMO

A presente pesquisa teve por objetivo desenvolver novas aplicações para análise de imagens baseadas na transformada discreta decimada e não decimada de *wavelet*. Foram estimados e verificado o comportamento, ao longo do tempo, dos expoentes de Hurst (direcional) com utilização a transformada de *wavelet* em 128 imagens interferométricas de um carcinoma mamário anaplásico canino, obtidas em intervalos regulares. O primeiro procedimento consiste em extrair partes das imagens que retratam tecidos cancerígenos e tecidos saudáveis e aplicar a metodologia de análise multirresolução em quatro níveis de resolução, utilizando a *wavelet* de Daubechies com oito momentos nulos. O segundo procedimento baseou-se em estimar os expoentes de Hurst nas 256 subimagens, com diferentes métodos utilizados para diminuir a correlação dos coeficientes *wavelet* e o efeito de possíveis *outliers* e, assim fazer uma análise temporal desses expoentes ao longo do tempo. Foi verificada a não sobreposição das distribuições nas séries em, pelo menos, uma direção quando comparadas subimagens do tecido cancerígeno e saudável, concluindo que a análise temporal dos expoentes de Hurst podem diferenciá-los. Com a aplicação dessa metodologia será possível obter uma melhora na qualidade da informação que pode ser retirada de imagens de tecidos cancerígenos, com o reconhecimento de padrões no expoente de Hurst, que podem ser usados, juntamente das análises clínicas, num processo de classificação ou detecção de tumores.

Palavras-chave: Análise multirresolução. Transformadas de wavelet. Expoente de Hurst.

ABSTRACT

This research aimed to develop new applications for image analysis based on the decimated and non-decimated discrete wavelet transform. We estimated and verified the behavior of the Hurst exponents (directional) using the wavelet transform in 128 interferometric images, obtained at regular intervals, of a canine anaplastic mammary carcinoma. The first procedure consists of extracting parts of the images that portray cancerous and healthy tissues and applying the multi-resolution analysis methodology at four levels of resolution using the Daubechies wavelet with eight null moments. The second procedure was based on estimating Hurst's exponents in the 256 sub-images, using different methods to decrease the correlation of wavelet coefficients and the effect of possible outliers, thus conducting a temporal analysis of these exponents over time. We verified the non-overlapping series distributions in at least one direction when comparing sub-images of cancerous and healthy tissue, concluding that the temporal analysis of Hurst exponents may differentiate them. By applying this methodology, it will be possible to improve the quality of the information obtained from images of cancerous tissues by recognizing patterns in the Hurst exponent that can be used along with a clinical analysis in the tumor classification or detection processes.

Keywords: Multiresolution analysis. Wavelet transform. Hurst exponent.

LISTA DE FIGURAS

Figura 3.1 – Análise Espectral com banco de filtros	20
Figura 3.2 – Transformada rápida de Fourier / Transformada de <i>wavelet</i>	21
Figura 3.3 – As <i>wavelets</i> de Haar	23
Figura 3.4 – <i>Wavelets</i> de Daubechies	24
Figura 3.5 – <i>Wavelet</i> mãe em diferentes escalas e localizações	26
Figura 3.6 – Representação gráfica de uma transformação multiescala com utilização da transformada decimada e a <i>wavelet</i> de Haar	29
Figura 3.7 – Representação gráfica de uma transformação multiescala com utilização da transformada não decimada	31
Figura 3.8 – Representação dos coeficientes <i>wavelet</i> bidimensionais	34
Figura 3.9 – Representação de um reservatório	35
Figura 4.1 – Aquisição das imagens	43
Figura 4.2 – Região Saudável e Região Cancerígena	44
Figura 4.3 – Procedimento de extração das subimagens a serem analisadas	45
Figura 5.1 – Decomposição de uma subimagem do tecido cancerígeno em (a) do tecido saudável em (b), no nível 1 e 4 de resolução com utilização DWT	46
Figura 5.2 – Gráficos dos expoentes de Hurst com utilização da DWT com a MQCW: (a) Gráficos das séries ao longo do tempo e (b) Densidades	48
Figura 5.3 – Gráficos dos expoentes de Hurst com utilização da DWT com Mid - energia: (a) Gráficos das séries ao longo do tempo e (b) Densidades	50
Figura 5.4 – Gráficos dos expoentes de Hurst com utilização da DWT com o método MEDL: (a) Gráficos das séries ao longo do tempo e (b) Densidades	52
Figura 5.5 – Gráficos dos expoentes de Hurst com utilização da DWT com o método MEDLA: (a) Gráficos das séries ao longo do tempo e (b) Densidades	54
Figura 5.6 – Gráficos dos expoentes de Hurst com utilização da DWT com estimador Trimean: (a) Gráficos das séries ao longo do tempo e (b) Densidades	56
Figura 5.7 – Decomposição de uma subimagem do tecido cancerígeno em (a) e do tecido saudável em (b), no nível 1 e 4 de resolução com utilização NDWT	57
Figura 5.8 – Gráficos dos expoentes de Hurst com utilização da NDWT com a MQCW: (a) Gráficos das séries ao longo do tempo e (b) Densidades	59

Figura 5.9 – Gráficos dos expoentes de Hurst com utilização da NDWT com Mid-energia: (a) Gráficos das séries ao longo do tempo e (b) Densidades	61
Figura 5.10 – Gráficos dos expoentes de Hurst com utilização da NDWT com o método MEDL: (a) Gráficos das séries ao longo do tempo e (b) Densidades.	63
Figura 5.11 – Gráficos dos expoentes de Hurst com utilização da NDWT com método MEDLA: (a) Gráficos das séries ao longo do tempo e (b) Densidades.	65
Figura 5.12 – Gráficos dos expoentes de Hurst com utilização da NDWT com estimador Trimean: (a) Gráficos das séries ao longo do tempo e (b) Densidades.	67
Figura 5.13 – Gráficos dos expoentes de Hurst com utilização da NDWT - MEDLA: (a) Densidade regiões saudáveis (b) Densidades regiões cancerígenas.	69
Figura 1 – Gráficos dos expoentes de Hurst das regiões sem carcinoma utilizando NDWT - MEDLA : (a) Gráficos das séries ao longo do tempo, (b) Dis- tribuições dos expoentes direcionais e (c) beanplots.	85
Figura 2 – Gráficos dos expoentes de Hurst das regiões com carcinoma utilizando NDWT - MEDLA : (a) Gráficos das séries ao longo do tempo, (b) Dis- tribuições dos expoentes direcionais e (c) beanplots.	87

LISTA DE TABELAS

Tabela 5.1 – Estatísticas descritivas dos expoentes de Hurst direcionais com utilização da DWT com a MQCW	47
Tabela 5.2 – Estatísticas descritivas dos expoentes de Hurst direcionais com utilização da DWT com Mid - energia	49
Tabela 5.3 – Estatísticas descritivas dos expoentes de Hurst direcionais com utilização da DWT com método MEDL	51
Tabela 5.4 – Estatísticas descritivas dos expoentes de Hurst direcionais com utilização da DWT com método MEDLA	53
Tabela 5.5 – Estatísticas descritivas dos expoentes de Hurst direcionais com utilização da DWT com estimador Trimean	55
Tabela 5.6 – Estatísticas descritivas dos expoentes de Hurst direcionais com utilização da NDWT com a MQCW	58
Tabela 5.7 – Estatísticas descritivas dos expoentes de Hurst direcionais com utilização da NDWT com Mid-energia	60
Tabela 5.8 – Estatísticas descritivas dos expoentes de Hurst direcionais com utilização da NDWT com método MEDL	62
Tabela 5.9 – Estatísticas descritivas dos expoentes de Hurst direcionais com utilização da NDWT com método MEDLA	64
Tabela 5.10 – Estatísticas descritivas dos expoentes de Hurst direcionais com utilização da NDWT com estimador Trimean	66
Tabela 1 – Estatísticas descritivas dos expoentes de Hurst direcionais (Regiões sem carcinoma - NDWT - MEDLA)	84
Tabela 2 – Estatísticas descritivas dos expoentes de Hurst direcionais (Regiões com carcinoma - NDWT - MEDLA)	86

LISTA DE SIGLAS

AMR Análise Multirresolução

BSL Biospeckle Laser

DWT Transformada Discreta de Wavelet Decimada

fBm Movimento Browniano Fracionário

FWT Transformada rápida de wavelet

GTME Estimador Trimean de Mid-energias

INCA Instituto Nacional do Câncer

MEDL Mediana do Logaritmo dos Coeficientes Wavelets ao quadrado

MEDLA Mediana do Logaritmo dos Coeficientes Wavelet médio ao quadrado

MQCW Média dos coeficientes wavelet ao quadrado

NDWT Transformada Discreta de Wavelet não Decimada

STFT Transformada de Fourier de Tempo Curto

SUMÁRIO

1	INTRODUÇÃO	11
2	REVISÃO BIBLIOGRÁFICA	14
3	REFERENCIAL TEÓRICO	19
3.1	Análise Fourier x análise wavelet	19
3.2	Funções wavelet	21
3.3	Transformada contínua de wavelet	24
3.4	Transformada discreta de wavelet	26
3.5	A análise multirresolução	27
3.6	Expoente de Hurst	34
3.6.1	Expoente de Hurst e análise wavelet	37
4	MATERIAL E MÉTODO	43
5	RESULTADOS E DISCUSSÕES	46
5.1	Análise temporal dos expoentes de Hurst com utilização da DWT	46
5.1.1	Estimação dos expoentes de Hurst com MQCW	47
5.1.2	Estimação dos expoentes de Hurst com Mid-energias	49
5.1.3	Estimação dos expoentes de Hurst com método MEDL	51
5.1.4	Estimação dos expoentes de Hurst com método MEDLA	53
5.1.5	Estimação dos expoentes de Hurst com estimador Trimean	55
5.2	Análise temporal dos expoentes de Hurst com utilização da NDWT	57
5.2.1	Estimação dos expoentes de Hurst com MQCW	58
5.2.2	Estimação dos expoentes de Hurst com Mid-energias	60
5.2.3	Estimação dos expoentes de Hurst com método MEDL	62
5.2.4	Estimação dos expoentes de Hurst com método MEDLA	64
5.2.5	Estimação dos expoentes de Hurst com estimador Trimean	66
6	CONCLUSÃO	70
	REFERÊNCIAS	71
	APENDICE A – Mid-energias	74
	APENDICE B – Teorema 3.1	77
	APENDICE C – Teorema 3.2	79
	APENDICE D – Teorema 3.3	82
	APENDICE E – Regiões similares - NDWT - MEDLA	84

1 INTRODUÇÃO

A utilização das transformadas de *wavelet* aumentou significativamente com uma grande variedade de aplicações nas diversas áreas do conhecimento, tais como: Acústica, Astronomia, Engenharia, Medicina, Física e Estatística. Especialmente na Estatística, dentre vários tópicos, o que mais se destaca é a filtragem de sinais, já que eles são corrompidos por ruídos durante sua transmissão.

Nos trabalhos estatísticos que envolvem *wavelets*, são abordados aspectos práticos e teóricos em filtragem de sinais biométricos, filtragem e compressão de sinais, detecção de anomalias em imagens e suavização de dados genômicos. A busca por representações esparsas de sinais trazem vantagens para aplicações envolvendo regularização, compreensão de dados e redução de ruídos.

A ideia por trás da estrutura de compreensão de dados, de acordo com Lima (2002), é a de remover a correlação presente e a partir de uma parte das informações, preencher a parte que esteja faltando. O requerimento básico é que se possa mudar rapidamente dos dados originais aos comprimidos e vice-versa. Em imagens existem correlações espaciais, que podem prever o valor de um pixel pelo pixel vizinho; correlações espectrais, capazes de prever a componente de frequência pelas frequências vizinhas e também correlação temporal, em que a maioria dos pixels de imagens simultâneas mudam pouco na direção do tempo.

Nessa perspectiva, os primeiros trabalhos que associaram *wavelet* com métodos estatísticos foram os de Donoho e Johnstone (1992), no qual a transformada de *wavelet* foi usada para transformar sinais densos em vetores esparsos com poucas entradas significativas.

A transformada de *wavelet* é uma técnica que permitiu um avanço significativo no processamento de imagens, que por sua vez, estimulou o estudo de bases ortonormais de *wavelet*, permitindo análise de imagens em várias escalas simultaneamente. Dessa forma, resolveu parte do problema de armazenamento e resolução.

O princípio da análise multirresolução proposto por Mallat (1989) diz que essa metodologia é essencial para a extração eficiente de informações a partir de uma observação. O objetivo dessa análise é oferecer uma infraestrutura para que as funções possam ser representadas em diferentes escalas ou resoluções para o caso das imagens. Essa metodologia permite decompor uma função $f(x, y)$, em uma combinação linear entre uma função escala e uma função *wavelet*, permitindo, assim, analisá-la em vários níveis de detalhes.

Esta pesquisa teve por objetivo utilizar a análise multirresolução, que forma um sistema ortonormal de suporte compacto, com aplicação das transformadas decimada e não decimada de *wavelet*, para analisar o comportamento dos expoentes de Hurst, ao longo do tempo, em uma sequência de imagens interferométricas de um carcinoma mamário anaplástico canino retiradas em intervalos regulares e obtidas por meio do método Biospeckle laser. A decomposição foi feita nas três direções (horizontal, vertical e diagonal). Além disso, verificar se a análise temporal dos expoentes de Hurst, também, é capaz de diferenciar imagens de tecido cancerígeno do tecido saudável. E por fim, discutir o efeito do uso de diferentes métodos utilizados para estimar os expoentes de Hurst, na série temporal.

Para atingir esse objetivo, foi extraído da sequência de 128 imagens (640 x 486 pixels), subimagens (64 x 64 pixels) da região cancerígena e da região saudável, totalizando 256 subimagens. Processar as subimagens utilizando a análise multirresolução proposta por Mallat (1989) com a função *wavelet* de Daubechies com oito momentos nulos (D8) e quatro níveis de resolução que, de acordo com Wang e Shiyong (2000), são as *wavelets* que melhor caracterizam os detalhes por meio da *wavelet* mãe.

O instituto nacional do câncer americano estima que uma, em cada oito mulheres, em algum momento, será diagnosticada com câncer de mama durante sua vida, (Altekruse et al., 2010). Atualmente, a mamografia é a principal estratégia para detecção precoce de tumores. Entretanto, a interpretação radiológica das imagens é uma tarefa difícil, uma vez que a aparência mamográfica do tecido saudável é altamente variável e complexa e os sinais de câncer precoce são frequentemente pequenos e sutis.

De acordo com pesquisadores do Instituto Nacional do câncer brasileiro, os diagnósticos falsos estão relacionados, muitas vezes, à má qualidade do exame, (INCA, 2013). Para realizar uma mamografia, o técnico posiciona a paciente no equipamento e radiografa a mama. A imagem é revelada e analisada por um médico que emite um laudo com diagnóstico. Uma falha em qualquer etapa, pode ocasionar erro. Estima-se que entre dez e trinta por cento dos cânceres poderiam ter sido detectados mais precocemente e muitas dessas falhas de diagnósticos podem ser atribuídas, principalmente, a fatores humanos, (Martin et al., 1979).

Nesse sentido, essa pesquisa justifica-se, principalmente, por contribuir com a melhora da qualidade da informação que pode ser retirada das imagens de células cancerígenas, com o reconhecimento de padrões nos expoentes de Hurst, permitindo, assim, estabelecer junto de

análises clínicas, conclusões mais confiáveis que podem ser usadas como, por exemplo, num processo de classificação ou detecção de tumores precocemente.

Para facilitar a introdução do assunto, será apresentada uma breve revisão bibliográfica cronológica sobre *wavelets*, desde a primeira citação feita por Haar (1910), passando pelas pesquisas realizadas por Mallat (1989) e Daubechies (1989), que impulsionaram a comunidade científica com processamento digital de imagens, até os dias atuais, com seu uso no processamento de imagens radiográficas.

No referencial teórico é apresentada uma breve introdução sobre as transformadas de Fourier e *wavelet*, para responder às seguintes questões: O que é transformada de Fourier? O que são *wavelets*? O que é transformada de *wavelet*? O que é análise de multirresolução (AMR)? O que é expoente de Hurst? E por fim, será retratado o cálculo do expoente de Hurst com a implementação de diferentes métodos para sua estimação, propostos para diminuir a correlação e o efeito de possíveis *outliers*.

2 REVISÃO BIBLIOGRÁFICA

O primeiro trabalho sobre wavelet data de 1910 na tese de doutorado de Alfred Haar, na qual ele desenvolve análises escalonadas. Haar apresentou em sua tese um sistema ortonormal gerado por meio de rotações e translações de uma dada função. Esse sistema oferece a vantagem de ser composto de funções de suporte compacto (funções não nulas em um intervalo limitado e fechado da reta), capazes de localizar sinais no domínio da frequência e no domínio do tempo.

As *wavelets* de Haar, embora sejam de suporte compacto, não são continuamente diferenciáveis. Com intuito de resolver essa dificuldade, entre as décadas de 20 e 30, Schauder e Luzin apresentaram uma base de funções contínuas. A base de Schauder apresenta vantagens sobre a de Haar, pois permite obter uma série para a derivada em termos da base de Haar. Anos mais tarde, o foco das pesquisas passou a abranger métodos de obtenção das bases, a partir de funções originais.

No início da década de 80, Alex Grossmann e Jean P. Morlet introduziram o conceito de *wavelets*, enquanto analisavam sinais geofísicos destinados à exploração de petróleo e gás. Eles constataram que, com repentinas variações do sinal, a análise de Fourier não era eficiente.

Em 1985, Stéphane Mallat impulsionou as pesquisas na área por meio de seu trabalho em processamento digital de imagens, particularmente com os filtros espelhados em quadratura e com algoritmos piramidais.

O método de Mallat permite construir a maioria das *wavelets* ortogonais. A função de escala foi introduzida em 1989 e seu princípio fundamental é analisar o sinal por meio da combinação de uma função escala (passa-baixa) e *wavelets* (passa-alta). Essa ideia é essencial na codificação em sub-bandas e na análise multirresolução.

Dando continuidade aos trabalhos de Mallat, Meyer (1990) construiu as primeiras *wavelets* não triviais, continuamente diferenciáveis. A distinção entre as *wavelets* de Haar e de Meyer é a diferenciabilidade contínua. Porém, as *wavelets* de Meyer não são de suporte compacto.

O trabalho de Haar foi completado em 1988 por Ingrid Daubechies em seu artigo *Orthonormal Bases of compactly supported wavelets*, no qual foi apresentado um conjunto de bases ortonormais de *wavelets* suaves com suporte compacto. Sua pesquisa foi um divisor de águas nessa área, uma vez que ela estabeleceu uma base sólida para o desenvolvimento de métodos *wavelet*, apresentando à comunidade científica o grande potencial de decompor sinais (imagens e séries temporais) em suas componentes elementares ao longo de diferentes escalas.

Os trabalhos de Mallat e Daubechies estão entre as contribuições mais relevantes e citadas na literatura científica. Enquanto Daubechies construiu uma base ortonormal de funções *wavelet*, Mallat identificou um importante conceito de análise multirresolução (AMR), permitindo a obtenção de algoritmos rápidos para as transformadas discretas de *wavelet* por meio de seus filtros.

A partir desses trabalhos, as *wavelets* têm sido utilizadas nas mais diferentes áreas da Física, tais como caracterização de sinais acústicos, radar e sonar, análise de tons musicais, óptica e eletromagnetismo. Também são utilizados nos conteúdos matemáticos, como fractais, modelagem de sistemas lineares, modelagem geométrica, solução de equações diferenciais ordinárias e parciais (OLIVEIRA, 2007b).

Outra importante linha de pesquisa que tem estimulado métodos *wavelet* é a busca por representações esparsas de sinais. Elas melhoram aplicações envolvendo redução de ruídos e compressão de dados. O trabalho pioneiro que envolve *wavelets* com métodos estatísticos é o de Donoho e Johnstone (1992), no qual as transformadas de *wavelet* são exploradas para transformar sinais densos em vetores esparsos com poucas entradas significativas.

Donoho e Johnstone (1992) propõem um procedimento para se obter, a partir de séries originais de tempo, as séries alisadas (*denoised*), que são possíveis pontos de partida para aplicação de métodos de previsão. Arino e Vidakovic (1995) propuseram um procedimento alternativo aos modelos temporais, no qual se baseia a utilização de *wavelets* para fracionar a série original em subs-séries de alta frequência e baixa frequência, para posteriormente obter mais detectabilidade do comportamento cíclico da série original.

Nos conteúdos estatísticos, as *wavelets* têm sido utilizadas para estimação de densidade, detecção de pontos de mudança e de bordo, problemas inversos, análise de séries temporais, regressão não paramétrica e especialmente redução ou anulação de ruídos em sinais.

Ainda sobre os conceitos estatísticos, as *wavelets* também foram usadas por Shannon (1948) para construção do sinal, a partir do processo de amostragem. O teorema proposto por Shannon estabelece, sob certas condições, que amostras de um sinal podem conter todas as informações a ele associado. Portanto, o sinal pode ser reconstruído a partir de amostras. Mas esse estudo se aplica somente para sinais que não possuem componentes espectrais para frequência acima de uma dada frequência.

Nos últimos anos, tem ganhado muita atenção na comunidade científica, os trabalhos que utilizam procedimentos de filtragem de sinais por meio de encolhimento dos coeficientes

wavelet. Nesses trabalhos, são abordados principalmente aspectos práticos e teóricos sobre transformada de *wavelet* em filtragem de sinais biométricos; filtragem e compressão de imagens; detecção de anomalias em imagens e suavização de dados genômicos.

No sentido de detecção de anomalias em imagens, Braga et al. (2012) utilizaram a técnica biospeckle laser para analisar imagens no domínio da frequência utilizando a *wavelet* de Morlet. Eles observaram que tecidos cancerosos apresentavam altas frequências, enquanto tecidos saudáveis revelavam baixas frequências. Também foram analisadas áreas de transição, onde eram manifestadas frequências baixas e intermediárias.

Nicolis et al. (2011) também utilizaram a transformada de *wavelet* 2D em mamografias digitalizadas para classificar tumores como maligno ou benigno, com base nos expoentes de Hurst. Segundo esses autores, o tecido mamário benigno tende a ser mais irregular e com menores expoentes de Hurst. Em contrapartida, tecidos mamários malignos tendem a ser mais regulares e com expoentes de Hurst maiores. Ainda nesse trabalho, os autores propõem como alternativa para reduzir os efeitos de *outliers*, uma mudança no cálculo do expoentes de Hurst. Nesse trabalho os expoentes de Hurst se mostraram discriminatórios e levaram a um classificador que toma decisões corretas em mais de 85% dos casos. Esse resultado também coincide com o trabalho realizado por Ramírez-Cobo e Vidakovic (2013), em que os mesmos dados foram usados para a mesma classificação, porém, com uma abordagem baseada em multifractalidade.

Em outro trabalho, Ramírez-Cobo et al. (2011) propuseram a transformada de *wavelet* 2D de escala mista. Nesse trabalho, os espectros baseados em *wavelets* 2D foram aplicados para explorar e classificar os dados de radares de chuva com base em seu mecanismo de geração de tempestades. Ao contrário da tradicional *wavelet* 2D, em que a extensão de matrizes retangulares complica significativamente o algoritmo, as transformadas de *wavelet* 2D de escala mista são mais simples. De acordo com os autores, como a transformada de *wavelet* proposta é aplicada primeiro nas linhas e depois nas colunas, pode-se lidar não apenas com imagens retangulares, mas também com diferentes bases. Dessa forma, o espectro *wavelet* de escala mista é capaz de avaliar a irregularidade espacial mesmo quando a escala sobre as duas coordenadas diferem.

Jeon et al. (2014) também utilizaram a transformada de *wavelet* 2D de escala mista no contexto da avaliação de regularidades em imagens mamográficas. Segundo esses autores, essa transformação é implementada por operações matriciais e garante baixo custo computacional e

uma transformada inversa direta. O procedimento baseou-se na avaliação da regularidade dos tecidos de “fundo” das imagens, em vez de características da mamografia, como microcalcificações e massas tumorais. A partir do modelo de regressão logística verificou-se que estimativas robustas dos expoentes de Hurst possuem poder suficiente para diferenciar casos benignos de malignos.

Ainda com o objetivo de melhorar os diagnósticos em imagens de células cancerígenas, Ramírez-Cobo e Vidakovic (2013) apresentaram como nova ferramenta, o espectro multifractal baseado em *wavelet* 2D, para análises de assinaturas multi-fractais em imagens de mamografias digitais. Nesse trabalho, ao invés de se concentrar em microcalcificações, como é prevalente nas pesquisas, também foi explorado os “fundos” da mamografia e a estrutura fina do tecido. Esses autores descobriram que existem, em termos de assinaturas multifractais, notáveis diferenças existentes no fundo da mamografia de tecidos saudáveis versus tecidos com câncer. Então, como os antecedentes da mamografia não são normalmente usados na prática diagnóstica, essa abordagem adiciona um novo classificador para ser combinado com outras informações clínicas.

A decomposição em coeficientes *wavelet* é uma ferramenta de fundamental importância para o processamento de imagens e tem cooperado para criação de novas técnicas. Outra ferramenta que auxilia na análise de imagens é o expoentes de Hurst, o qual é uma medida de autossimilaridade que capta informações da imagem que não são visíveis a olho humano.

Para a estimação do expoentes de Hurst, Abry et al. (2000) utilizaram o logaritmo da média dos coeficientes *wavelet* ao quadrado ($\log_2 \overline{d_j^2}$). No entanto, esse cálculo é afetado pela não normalidade dos coeficientes *wavelet* e também pela correlação existente entre eles. Para corrigir esse viés, os pesquisadores propuseram diversas metodologias.

Soltani et al. (2004) definiram “Mid-energias” para reduzir o viés provocado pela correlação. De acordo com essa abordagem, cada nível de multirresolução é dividido em duas partes iguais e os coeficientes correspondentes de cada parte são pareados, elevados ao quadrado e calculados a sua média. Isso produziu um efeito de quase decorrelação, superando consideravelmente os estimadores que usavam $\log_2 \overline{d_j^2}$.

Ainda com objetivo de gerar estimadores mais robustos para o expoente de Hurst, Kang e Vidakovic (2017) propuseram os métodos MEDL e MEDLA, os quais diminuíram a correlação entre os coeficientes *wavelet*. O primeiro método é baseado na relação entre a mediana do logaritmo do quadrado dos coeficientes *wavelet*. O segundo, na relação entre a mediana do logaritmo da média entre dois coeficientes *wavelet* elevados ao quadrado sorteados aleatoria-

mente. Segundo os autores, as medianas são mais robustas a potenciais *outliers* que possam ocorrer, quando a transformação logarítmica de um coeficiente *wavelet* ao quadrado é tomada e a magnitude do coeficiente é próxima de zero.

Também nesse sentido, Feng e Vidakovic (2017) utilizaram o estimador *Trimean* no cálculo do expoente de Hurst para reduzir os efeitos de *outliers*. O estimador *Trimean* foi definido como a média ponderada entre a mediana e os quantis simétricos à mediana da distribuição. A metodologia foi aplicada a um conjunto real de dados de comportamento de resposta pupilar, para extrair características significativas de idosos com e sem deficiência visual. Os pesquisadores concluíram, no que diz respeito ao conjunto de dados, que a magnitude do diâmetro da pupila depende mais da luz ambiente do que da patologia ocular inerente. Já em relação ao estimador *Trimean*, o método reduziu a correlação dos coeficientes *wavelet*, aumentando a precisão dos expoentes de Hurst.

3 REFERENCIAL TEÓRICO

Nesta seção, será apresentada uma introdução sobre as transformadas de Fourier e as transformada de *wavelet*, com suas vantagens e desvantagens. Também será apresentada a análise multirresolução proposta por Mallat (1989) para sinais unidimensionais e bidimensionais. Por fim, apresenta-se uma explicação sobre o cálculo do expoente de Hurst, com o uso de métodos alternativos para estimar a variância dos coeficientes *wavelet*, para tentar diminuir a correlação e o efeito de possíveis *outliers* que possam surgir com a transformação logarítmica.

3.1 Análise Fourier x análise wavelet

A análise de Fourier estabelece que um sinal pode ser representado por uma soma infinita de termos em senos e cossenos, chamada expansão de Fourier. Uma das vantagens da transformada de Fourier é o fato de poder determinar todas as frequências presentes no sinal.

No entanto, não existe uma relação temporal, ou seja, apesar de determinar o conteúdo de frequências presentes no sinal, não há nenhuma informação sobre em qual intervalo de tempo elas ocorrem. A transformada de Fourier não fornece uma análise temporal, apenas frequencial.

Algumas soluções foram encontradas com objetivo de ter conjuntamente uma análise frequencial e temporal. Uma delas foi a Transformada de Fourier de Tempo Curto (STFT - *Short Time Fourier Transform*), também conhecida como transformada de Gabor. A intenção da STFT é introduzir um novo parâmetro de frequência local para observar, por meio de uma curta janela, o sinal supostamente estacionário (OLIVEIRA, 2007b).

Tudo se passa como se o sinal fosse “fatiado” em vários trechos e em cada trecho é analisada a decomposição espectral, resultando em um espectro local. Observa-se, então, uma sequência de “fotos” do espectro evoluindo no tempo, mostrada na expressão (3.1).

$$\dots F(w, t_{-2}), F(w, t_{-1}), F(w, t_0), F(w, t_1), F(w, t_2) \dots \quad (3.1)$$

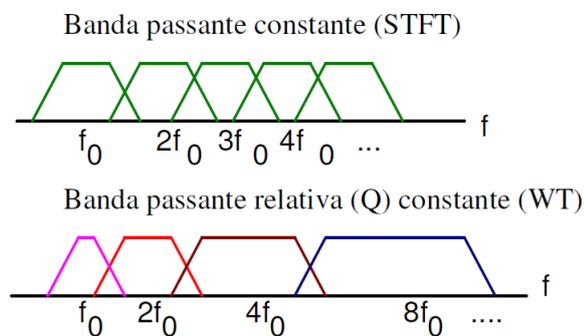
O detalhe é que uma vez fixada a janela para STFT, a resolução no tempo (t) e na frequência (f) permanece contante em todo plano (txf). Um exemplo em que a análise de Fourier apresenta dificuldade ocorre no estudo de sinais sísmicos. A maior parte do tempo o sismógrafo registra um sinal estacionário (sinal base) e na ocorrência de terremotos um sinal curto, rápido e de frequência elevada aparece, não sendo eficaz a análise pela transformada rápida de Fourier.

Relacionada à essa teoria, a formulação do Princípio da Incerteza de Heisenberg, desenvolvida por Gabor (1946), estabelece que é impossível saber a frequência exata e o tempo exato de ocorrência dessa frequência, num sinal. Assim, uma maneira alternativa de se analisar esse tipo de sinal, chamada Análise de Multirresolução (AMR), é observá-lo em frequências diferentes com resoluções diferentes.

Assim, impõe-se uma alta resolução no tempo e baixa resolução na frequência, para frequências mais altas; uma resolução frequencial alta e resolução temporal baixa para frequências mais baixas. Isso se deve ao fato de sinais com componentes em alta frequência apresentarem rápidas alterações no domínio temporal, enquanto que sinais com componentes de baixa frequência possuem alterações mais lentas no domínio temporal.

Como alternativa para solucionar o problema no plano tempo-frequência, surgiu a transformada de *wavelet*, que é feita similarmente a STFT no que tange à multiplicação do sinal por uma função, que agora passa a ser uma função *wavelet* e não mais uma janela. A transformada é calculada separadamente por segmentos diferentes, como mostra a Figura (3.1)

Figura 3.1 – Análise Espectral com banco de filtros



Fonte: Oliveira (2007a)

As transformadas de *wavelet* servem como instrumentos de decomposição ou quebra de sinais, para assim, possibilitar a análise dos dados em diferentes domínios de frequências com a resolução de cada componente ligada a sua escala. Ademais, na análise *wavelet*, podem-se usar funções em domínio finito, tornando-as convenientes na aproximação de dados descontínuos.

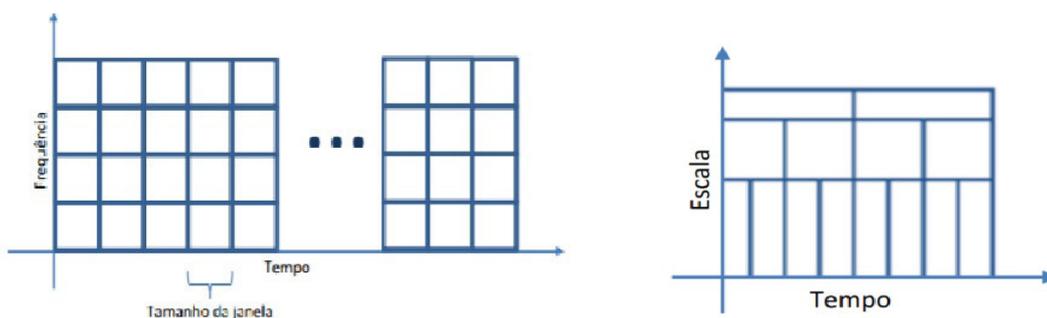
Se as funções de interesse forem imagens, curvas ou superfícies as *wavelets* oferecem uma técnica sofisticada para representar os níveis de detalhes presentes, e o princípio embaixador dessa construção é o uso de dilatações e translações. Um exemplo interessante é visto na mudança de escalas em mapas. Em escalas menores, veem-se detalhes, mas se perde no

comportamento global. Numa escala maior, tem-se uma visão global, mas com menor precisão. Isso faz das transformadas de *wavelet* uma excelente ferramenta na análise de dados.

A análise de Fourier é uma das origens da análise de *wavelet*. Portanto, as duas análises possuem similaridade, mas também diferenças importantes. A primeira semelhança é a reversibilidade, ou seja, elas permitem avançar e voltar em sinais transformados. Outra semelhança corresponde às funções base, que estão no domínio da frequência.

A principal diferença é que a análise de *wavelet* tem a capacidade de decompor os dados em escala de tempo e frequência, o que permite decompor um sinal em várias escalas. Outra diferença é a variação das janelas. Por exemplo: para isolar as discontinuidades de um sinal, seriam possíveis funções de bases curtas, mas, para se obter análise de frequência detalhada, seriam mais interessantes funções de base longas. Assim, as transformações *wavelet* permitem fazer as duas coisas ao mesmo tempo. Na Figura 3.2, tem-se a representação dessas diferenças.

Figura 3.2 – Transformada rápida de Fourier / Transformada de *wavelet*



Fonte: Nason (2008)

3.2 Funções wavelet

Wavelets são funções matemáticas (ψ) com variantes contínuas e discretas, capazes de separar dados em suas diferentes componentes frequenciais e extraem cada componente com uma resolução adequada à sua escala. Elas têm vantagens em relação à análise de Fourier, que, por sua vez, analisa o sinal como um todo, desconsiderando sinais com discontinuidades e variações bruscas. De acordo com Percival e Constantine (2000), para ser considerada uma *wavelet*, a função deve atender as seguintes características:

1. A área total sob a curva da função é zero:

$$\int_{-\infty}^{\infty} \psi(t) dt = 0 ; \quad (3.2)$$

2. A energia da função é finita, ou seja:

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt < L \quad L \in \mathbb{N} . \quad (3.3)$$

Definição:

Uma *wavelet* é uma função $\psi(x) \in L^1(\mathbb{Z}) \cap L^2(\mathbb{Z})$, tal que a família de funções:

$$\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k), \quad \text{com } j \text{ e } k \in \mathbb{Z}, \quad (3.4)$$

seja uma base ortonormal para $L^2(\mathbb{Z})$ (OLIVEIRA, 2007b).

Existem inúmeras *wavelets* interessantes. Em alguns casos, a apresentação envolve a função de escala, relacionada com a *wavelet*, por vezes referida como *wavelet*-pai. A seguir, serão apresentadas, como exemplo, a *wavelet* de Haar (1910) e também as *wavelets* de Daubechies (1989), que é o conjunto de *wavelets* mais usado.

- *Wavelet* de Haar

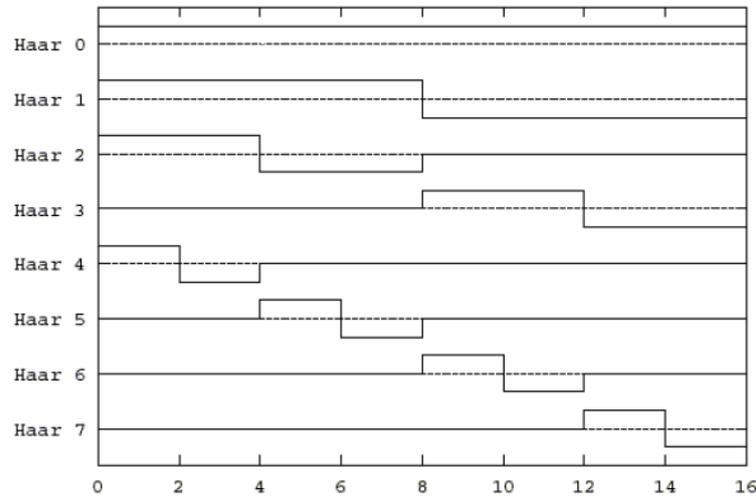
A primeira referência sobre *wavelets* apareceu na tese de doutorado de Alfred Haar (1910), na qual se discute a análise escalonada. As *wavelets* de Haar, embora de suporte compacto, não são diferenciáveis, como pode ser observado pela sua expressão:

$$\psi(x) = \begin{cases} 1, & \text{se } 0 \leq x < \frac{1}{2} \\ -1, & \text{se } \frac{1}{2} \leq x < 1 \\ 0, & \text{c.c.} \end{cases} \quad (3.5)$$

A *wavelet* de Haar pode ser utilizada em imagens com mudanças bruscas de contrastes, pois nesses casos, bases constantes por partes podem ser mais adequadas. De acordo com Lima (2002), elas permitem recuperar os detalhes que seriam perdidos com a diminuição de resolução quando se passa de uma escala para a seguinte, ou seja, elas medem o incremento de informação entre duas escalas consecutivas.

A *wavelet* de Haar é a única com suporte compacto para o qual se tem uma forma analítica fechada. As *wavelets* de Daubechies possuem suporte compacto e são diferenciáveis. Entretanto, não se conhece uma forma analítica e elas são calculadas numericamente. A Figura 3.3 mostra uma decomposição da *wavelet* de Haar em oito escalas e localizações.

Figura 3.3 – As *wavelets* de Haar



Fonte: Oliveira (2007a)

- Wavelets de Daubechies

Um dos maiores desafios da teoria *wavelet* foi a construção de uma família de *wavelets* ortogonais e de suporte compacto, desafio esse superado por Ingrid Daubechies (1989) que construiu um dos mais usados conjuntos de *wavelets*. Definida como:

$$\psi(x) = \phi(2x) - \phi(2x - 1), \quad (3.6)$$

sendo:

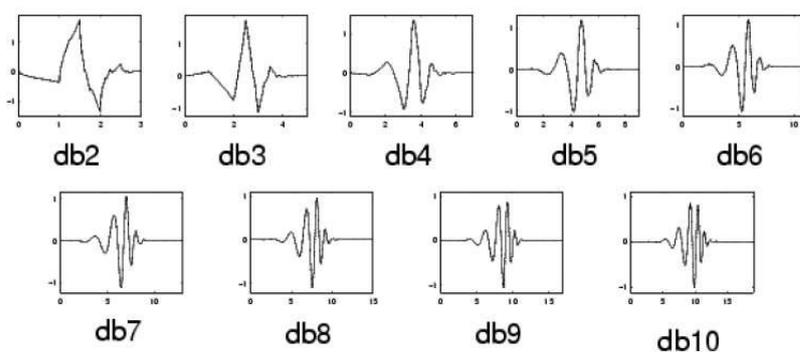
$$\phi(x) = \begin{cases} 1, & \text{se } 0 \leq x < 1 \\ 0, & \text{c.c.} \end{cases} \quad (3.7)$$

Assim são construídas as *wavelets* de Ingrid Daubechies sendo, $N\phi$ *wavelet* que possuem todos os momentos até a ordem $N - 1$ nulos, ou seja,

$$\int_{-\infty}^{\infty} x^l N\phi(x) dx = 0, \quad l = 0, 1, \dots, N - 1. \quad (3.8)$$

A *wavelet* de Haar pode ser considerada um caso particular das *wavelets* de Daubechies, quando $N = 1$. As *wavelets* de Daubechies também são muito aplicadas ao problema de processamento de imagens digitais, principalmente a *wavelet* 2ϕ , conhecida como Db4 (*wavelet* de Daubechies com quatro momentos nulos). Segundo Wang (2000), as *wavelets* de Daubechies são as que melhor caracterizam os detalhes através das *wavelets* mãe, que ficam nas sub-bandas de baixas resoluções (os coeficientes *wavelet*), enquanto nas *wavelets* pai ficam caracterizadas a tendência do sinal. Na Figura 3.4, são apresentados exemplos de algumas *wavelets* de Daubechies.

Figura 3.4 – *Wavelets* de Daubechies



Fonte: Oliveira (2007a)

3.3 Transformada contínua de wavelet

Como dito anteriormente, a transformada de *wavelet* constitui uma ferramenta moderna de análise e processamento de imagens, codificação em sub-bandas, análise multirresolução para visão artificial em computadores e modelagem de sistemas variantes no espaço-tempo. A transformada de *wavelet* é feita similarmente à transformada de Fourier, no que diz respeito à multiplicação do sinal por uma função, que nesse caso será uma função *wavelet* e não mais senoides, como na transformada de Fourier.

Outra forma de expressar a característica de regularidade é dizer que a transformação *wavelet* é um operador local no domínio do tempo, ou seja, ela decompõe um sinal em uma função no domínio do tempo e no domínio da frequência.

A transformada contínua de *wavelet* (CWT) pode ser escrita similarmente à transformada de Fourier (FRANCIS, KIM, 2013):

$$W(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(x) \Psi\left(\frac{x-b}{a}\right) dx, \quad (3.9)$$

em que a e b são parâmetros reais de escala e translação, respectivamente. A função $\Psi(x)$, conhecida como *wavelet* mãe (*wavelet* principal), modifica-se no decorrer da transformada.

O parâmetro de translação (b) indica que a função $\Psi(x)$ foi transladada no eixo x a uma distância equivalente a b . Esse termo corresponde à informação do tempo no domínio transformado ou também, relacionado a localização de uma “janela”, cujo conteúdo é analisado. Já o parâmetro (a) causa uma mudança de escala, aumentando se $a > 1$ e diminuindo se $a < 1$, ou seja, é o processo de compressão e dilatação do sinal. Assim, altas escalas correspondem a uma visão global do sistema, enquanto baixas escalas correspondem a uma visão mais detalhada.

A transformada de *wavelet* também pode ser definida como o produto interno da função $f(x)$ com a função base $\psi_{a,b}(x)$ multiplicada por uma constante, isto é:

$$W(a,b) = \langle f(x), \psi_{a,b}(x) \rangle = \int_{-\infty}^{\infty} f(x) \psi_{a,b}(x) dx, \quad (3.10)$$

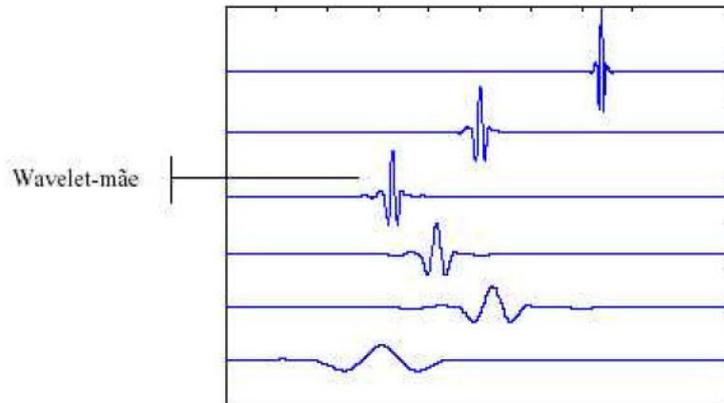
em que,

$$\psi_{a,b}(x) = \frac{1}{\sqrt{|a|}} \Psi\left(\frac{x-b}{a}\right). \quad (3.11)$$

A transformada de *wavelet* não tem apenas uma função base como a transformada de Fourier. Pelo contrário, pode ter infinitas funções bases. Existe um grande número de funções que podem ser consideradas *wavelet* mãe.

O termo $\frac{1}{\sqrt{|a|}}$ é um fator de normalização da energia do sinal e $\psi_{a,b}(x)$ é uma transformada definida por um mapeamento afim unitário. Essas *wavelets* são versões transladadas (b) e dilatadas/comprimidas (a) de uma mesma onda protótipo, chamada *wavelet* mãe. Na Figura 3.5 é mostrada uma *wavelet* mãe em diferentes escalas e localizações.

Figura 3.5 – Wavelet mãe em diferentes escalas e localizações



Fonte: Oliveira (2007a)

A invariância por mudança de escala é uma propriedade importante dessa transformação. Por exemplo, uma mudança de escala no sinal de uma variável $f(t)$ resulta em um escalonamento da *wavelet*, mesmo sendo de duas variáveis, ou seja, se $f(t)$ tem a transformada de *wavelet* $CWT(a, b)$ para certa função *wavelet* ψ , então $f(\alpha t)$, com $\alpha \in \mathbb{R}$ e $\alpha \neq 0$ tem a transformada de *wavelet* dada por $|\alpha|^{-\frac{1}{2}}CWT(\alpha a, \alpha b)$.

3.4 Transformada discreta de wavelet

Na prática, a transformada contínua de *wavelet* não é muito funcional, pois demandaria realizações de infinitas translações e escalonamentos, necessitando de muito tempo e recursos computacionais. De acordo com Oliveira (2007a), as formas discretas são mais fáceis de serem implementadas e também mais atraentes do ponto de vista computacional.

A discretização ocorre apenas no domínio dos parâmetros (escala e translação), não na variável independente do sinal (tempo ou espaço). As transformadas discreta de *wavelet* foram introduzidas para superar os obstáculos das *wavelets* contínuas. Elas são transladadas e escalonadas em intervalos discretos, com uma pequena modificação na *wavelet* contínua, fazendo $a = N^j$ e $b = kMa$, $j, k \in \mathbb{N}$, ou seja,

$$\psi_{j,k}(x) = \frac{1}{\sqrt{|N^j|}} \Psi \left(\frac{x - kMN^j}{N^j} \right), \quad (3.12)$$

em que j e k são inteiros, $N > 1$ é um parâmetro de dilatação fixo e M é o fator de translação o qual depende do fator de dilatação.

A *wavelet* mãe resultante é uma matriz de valores $H_{j,k}$ relacionada à *wavelet* contínua $\psi(x)$ pela expressão:

$$h_{j,k} = N^{-\frac{j}{2}} \psi\left(\frac{x}{N^j} - kM\right) . \quad (3.13)$$

Quanto mais próximo de 1 o valor de N , mais a versão discreta se aproxima da versão contínua. Normalmente, são utilizados $N = 2$ e $M = 1$, como na equação (3.4), para facilitar os cálculos computacionais e obter um número de coeficientes $h_{j,k}$ não muito grande e, ainda, conservando toda informação original de $f(x)$.

De forma similar, a transformada de *wavelet* discreta é uma matriz de coeficientes $W_{j,k}$ dados por:

$$w_{j,k} = \sum_{k=1}^m \sum_{j=1}^n f(k) h_{j,k}^* , \quad (3.14)$$

em que $h_{j,k}^*$ é o conjugado complexo e $f(k)$ é a sequência de amostras obtidas de $f(x)$.

Na prática, como é muito difícil obter uma expressão para a transformada contínua, emprega-se a versão discreta, calculada por computador. De acordo com Meyer (1990), as *wavelets* não constituem mais um capítulo da ciência mas, sim, uma ferramenta matemática fundamental tal como integração por partes.

Um dos principais motivos da popularidade da transformada de *wavelet* é a existência de um algoritmo que permite calcular os coeficientes, com esforço de ordem $O(n)$, em que n é o tamanho dos dados originais. Esse algoritmo é conhecido como transformada rápida de *wavelet* - FWT.

3.5 A análise multirresolução

As primeiras bases ortonormais de *wavelets* aparentemente eram muito complicadas. O cenário foi alterado, a partir da análise de multirresolução (ARM), formulada por Mallat e Meyer (1989). Mallat, ao trabalhar com *wavelet* de Meyer pela primeira vez, estava analisando imagens em várias escalas simultaneamente. De acordo com Lima (2002), isso estimulou o estudo de bases ortonormais de *wavelets* como ferramenta para descrever matematicamente o incremento na informação necessário para ir de uma informação grosseira para uma aproximação com maior resolução.

O princípio da representação multiescala proposta por Mallat é um ingrediente essencial na extração eficiente de informações a partir de uma observação. Além disso, é um dos métodos para construção de *wavelets* ortogonais.

O principal objetivo da AMR é oferecer uma infraestrutura para que as funções em $L^2(\mathbb{N})$ possam ser representadas em diferentes escalas (ou níveis de resolução). Assim, uma AMR pode ser definida por uma família de subespaços encaixados.

Definição:

Uma análise de multirresolução em $L^2(\mathbb{R})$ consiste numa sequência de subespaços fechados $V_m \subset L^2(\mathbb{R})$, satisfazendo as seguintes relações, (OLIVEIRA, 2007b):

1. $V_m \subset V_{m+1}$;
2. $f(t) \in V_m \subset L^2(\mathbb{R}) \Leftrightarrow f(2t) \in V_{m+1}$;
3. $\text{clos}^1 \cup V_m = L^2(\mathbb{R})$;
4. $\cap V_m = 0$;
5. $\exists \phi(t) \in V_0$ tal que $\phi(t - n)$ é uma base ortonormal para V_0 .

A análise multirresolução proposta por Mallat (1989) faz uma combinação de dois filtros passa-baixo e passa-alto, o filtro passa-baixo é uma função escala, denotada por $\phi(t)$ e o filtro passa-alto uma *wavelet* $\psi(t)$. Essa é uma ideia essencial na codificação em sub-bandas e na análise multirresolucional. Com isso, as duas equações centrais são:

$$\psi_{j,k}(t) = 2^{\frac{j}{2}} \psi(2^j t - k) \quad e \quad \phi_{j,k}(t) = 2^{\frac{j}{2}} \phi(2^j t - k) . \quad (3.15)$$

Assim como na análise de Fourier, o objetivo é escrever uma função $f(t)$ como combinação linear de outras funções. Então $f(t)$ pode ser representada por funções *wavelet*, como:

$$f(t) = \sum_k c_{j_0,k} \phi_{j_0,k}(t) + \sum_{j=j_0}^{\infty} \sum_k d_{j,k} \psi_{j,k}(t) , \quad (3.16)$$

em que

$$c_{j_0,k} = \int f(t) \phi_{j_0,k}(t) dt \quad e \quad d_{j,k} = \int f(t) \psi_{j,k}(t) dt , \quad (3.17)$$

são chamados coeficientes *wavelet*, sendo $c_{j_0,k}$ o coeficiente de escala (passa-baixa) e $d_{j,k}$ os coeficientes de detalhe (passa-alta).

¹ $\text{Clos} \cup$ - denota a união dos conjuntos A_m , incluindo todos os pontos limites em $L^2(\mathbb{R})$

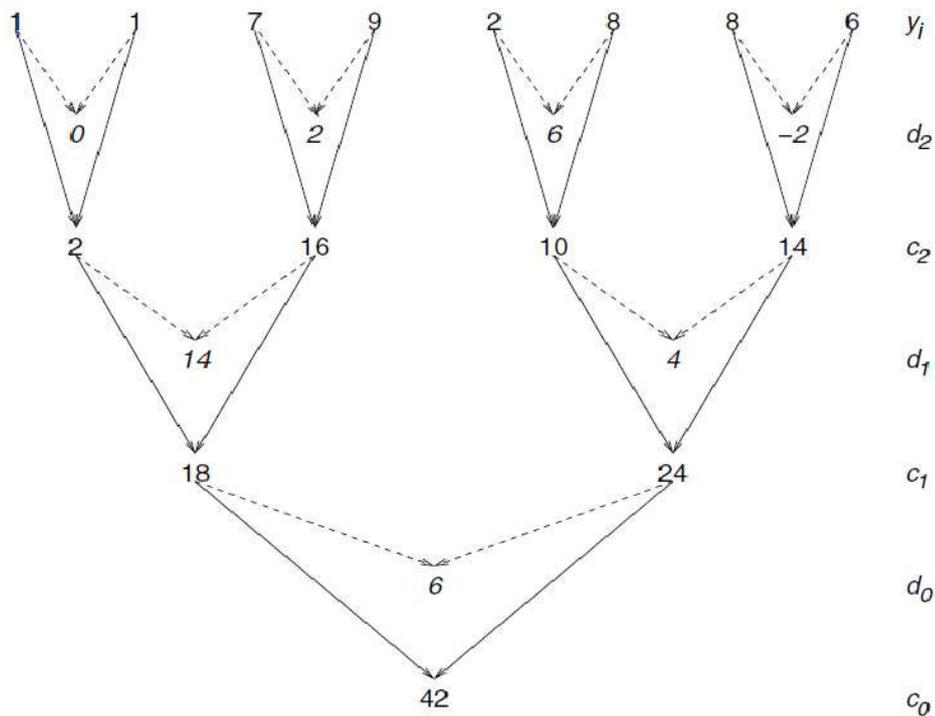
De acordo com a equação (3.15), a transformada discreta de *wavelet* é decimada (DWT), pois seleciona os valores filtrados pares, de modo que em cada escala o número de coeficientes é a metade do número de coeficientes da escala anterior.

Para melhor compreensão, seja uma sequência unidimensional dada por:

$$y = (y_1, y_2, y_3, \dots, y_n) = (1, 1, 7, 9, 2, 8, 8, 6) .$$

Como são oito elementos, temos, então, $j = 3$, pois, $2^3 = 8$. Para obter os coeficientes $d_{2,k}$, basta fazer as subtrações $d_{2,1} = 1 - 1$, $d_{2,2} = 9 - 7$, $d_{2,3} = 8 - 2$ e $d_{2,4} = 6 - 8$, para obter os coeficientes $c_{2,k}$, basta fazer as somas $c_{2,1} = 1 + 1$, $c_{2,2} = 7 + 9$, $c_{2,3} = 2 + 8$ e $c_{2,4} = 8 + 6$. Esse processo, chamado algoritmo piramidal, é repetido sucessivamente até obter os coeficientes c_0 e d_0 . A Figura 3.6 ilustra esses cálculos com utilização da *wavelet* de Haar de uma forma gráfica.

Figura 3.6 – Representação gráfica de uma transformação multiescala com utilização da transformada decimada e a *wavelet* de Haar



Fonte: Nason (2008)

Para aumentar o número de coeficientes em cada nível de resolução, é proposto a transformada discreta de *wavelet* não decimada (NDWT). A ideia é conservar ambas decimações, pares e ímpares, em cada nível e manter esse procedimento ao longo de um algoritmo piramidal. De acordo com Nason (2008), o maior problema é que esse procedimento acaba com a estrutura ortogonal e torna-se redundante.

Segundo Kang e Vidakovic (2017), a utilização da NDWT provoca um aumento na correlação dos coeficientes *wavelet* e um viés no cálculo do expoente de Hurst (que será discutido posteriormente), quando comparada a DWT. No entanto, a variância do estimador é menor, o tamanho do sinal ou imagem de entrada pode ser arbitrário, dentre outras vantagens.

Outra diferença entre DWT e NDWT, de acordo com Morettin (2014), é que a DWT é dependente da escolha da origem dos dados. Por exemplo, uma translação dos dados resulta num conjunto distinto de coeficientes *wavelet*, o que não acontece com a NDWT.

Então, para conservar toda informação mantém-se os coeficientes *wavelet* obtidos pela DWT sobre o conjunto original de dados, depois, deve-se transladá-los e transformá-los. Nesse caso, os elementos do vetor original serão deslocados uma posição para esquerda. Para formalizar, considere dois operadores D_0 e D_1 , sobre uma sequência x_l da seguinte maneira:

$$(D_0x)_l = x_{2l} \text{ e } (D_1x)_l = x_{2l+1} . \quad (3.18)$$

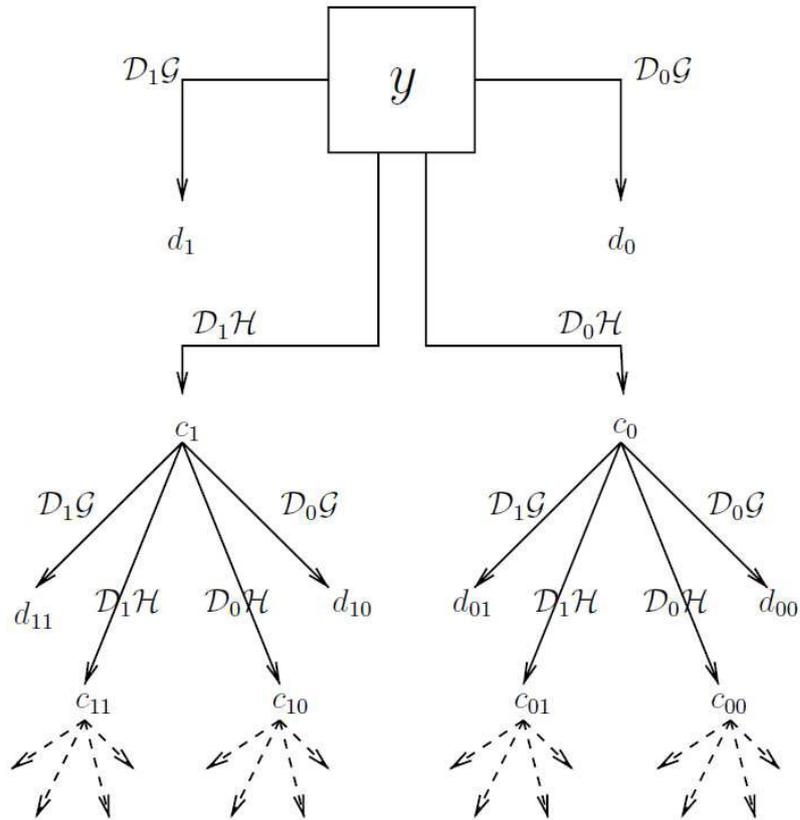
A DWT usa o operador D_0 para obter os coeficientes *wavelet* $d_{j,k}$ e $c_{j,k}$, em NDWT vamos utilizar os dois operadores. Portanto, o algoritmo para obter NDWT procede da seguinte forma (MORETTIN, 2014):

1. Com os dados $Y = (y_0, y_1, y_2, \dots, y_{n-1})$, aplica-se D_0Hy e D_1Hy para obter d_0 e d_1 , cada uma dessas frequências contendo $\frac{n}{2}$ coeficientes, resultando n coeficientes na escala mais fina;
2. Aplica-se aos dados D_0Ly e D_1Ly para obter os coeficientes c_0 e c_1 , resultando n coeficientes suaves;
3. Para a próxima escala (2^{J-j}), aplica-se D_0H e D_1H a D_0Ly e D_1Ly , cada um com $\frac{n}{4}$ coeficientes, totalizando n coeficientes;
4. Para obter os coeficientes suaves $c_{i,j}$, aplica-se D_0L e D_1L aos coeficientes c_0 e c_1 , obtendo n coeficientes;

5. Continuando dessa maneira, na escala 2^{J-j} , terá 2^j conjuntos de coeficientes, cada um com $n2^{-j}$ elementos, $j = 0, 1, 2, \dots, j-1$, logo o número total será $n2^{-j} \cdot 2^j = n$.

A Figura 3.7 mostra um diagrama que ilustra a operação feita pela transformada não decimada de *wavelet* (NDWT) em um sinal de entrada unidimensional y .

Figura 3.7 – Representação gráfica de uma transformação multiescala com utilização da transformada não decimada



Fonte: Nason (2008)

Até o momento foram descritas as análises multirresolução unidimensional, mas ela também pode ser estendida para dimensões maiores. Uma imagem pode ser representada por uma função matricial $f(x,y)$, para o qual as entradas (x,y) representam intensidades de cores dos pixels no ponto (x,y) , ou seja, para cada elemento (x,y) na imagem (pixel) temos um valor $f(x,y)$ correspondente à intensidade luminosa naquele ponto. Por exemplo, na equação (3.19) o número 157 corresponde à intensidade luminosa no ponto (1,3)

$$f(x,y) = \begin{bmatrix} (0,1) & (0,2) & (0,3) \\ (1,1) & (1,2) & (1,3) \\ (2,1) & (2,2) & (2,3) \end{bmatrix} = \begin{bmatrix} 124 & 133 & 54 \\ 48 & 203 & 157 \\ 1 & 89 & 214 \end{bmatrix}. \quad (3.19)$$

A segmentação de imagens é uma das tarefas mais críticas presentes na análise de imagens e consiste na partição do conjunto de dados de modo a produzir estruturas com mais informações. A utilização dessas informações constitui uma abordagem adequada para representar regiões de uma imagem com características percebidas pelos humanos.

Para descrever regiões texturizadas, utilizam-se métodos de extração de características, os quais visam obter uma representação simplificada, porém mais significativa para cada região. É como se o conhecimento de pequenos trechos equiespaçados da imagem fossem suficientes para realizar uma intercalação “perfeita” de vazios entre os trechos. Essa análise de imagens é uma ferramenta fundamental nos dias atuais. Apesar dos desempenhos de CPU’S, discos e canais de transmissão terem evoluído bastante, ainda surgem alguns problemas de armazenamento e resolução, já que um vídeo de sinal tem aproximadamente 30 imagens por segundo.

A grande parte das informações utilizadas são altamente correlacionadas. Portanto, parece ser possível compressão sem perda de informação e também que se possa mudar rapidamente dos dados originais aos comprimidos e vice-versa.

A principal ideia por trás da compressão dos dados é a de remoção da correlação presente entre eles, ou seja, os dados sendo correlacionados, a partir de uma parte dos dados podemos preencher a parte que esteja faltando. Existem as correlações espaciais onde se pode prever o valor de um pixel numa imagem olhando para o pixel vizinho; as correlações espectrais onde se pode prever uma componente de frequência olhando para frequências vizinhas; e as correlações temporais onde a maioria dos pixels de duas imagens mudam muito pouco na direção do tempo (LIMA, 2002).

Numa imagem típica, o que se percebe são grandes regiões onde os valores dos pixels são muito próximos ou até mesmo iguais. Então, os coeficientes *wavelet* a elas associadas ou são nulos ou mínimos. Já em regiões de fronteira, próximas aos contornos onde os valores dos pixels variam muito, é que realmente teremos uma variação local significativa nos valores dos pixels. Portanto, coeficientes *wavelet* significativos.

Um algoritmo para a transformada de *wavelet* bi-dimensional explora a propriedade de separabilidade decorrente da sua construção pelo produto tensorial. Logo, tanto o algoritmo de decomposição como o de reconstrução, podem ser implementados em duas dimensões, com aplicação dos algoritmos unidimensionais aplicados nas linhas e nas colunas da imagem.

Portanto, em uma transformada bidimensional observam-se três diferentes tipos de funções *wavelet* mãe, (MORETTIN, 2014):

$$\begin{aligned}\Phi_l(x,y) &= \phi(x) \phi(y) , \\ \Psi^v(x,y) &= \psi(x) \phi(y) , \\ \Psi^h(x,y) &= \phi(x) \psi(y) , \\ \Psi^d(x,y) &= \psi(x) \psi(y) ,\end{aligned}\tag{3.20}$$

em que Ψ^v , Ψ^h e Ψ^d são *wavelets* verticais, horizontais e diagonais, respectivamente, e a função escala Φ_l , também chamada de *wavelet* pai, representa a parte suave da imagem.

Logo, uma função $f(x,y) \in \mathbb{R}^2$ pode ser escrita na forma:

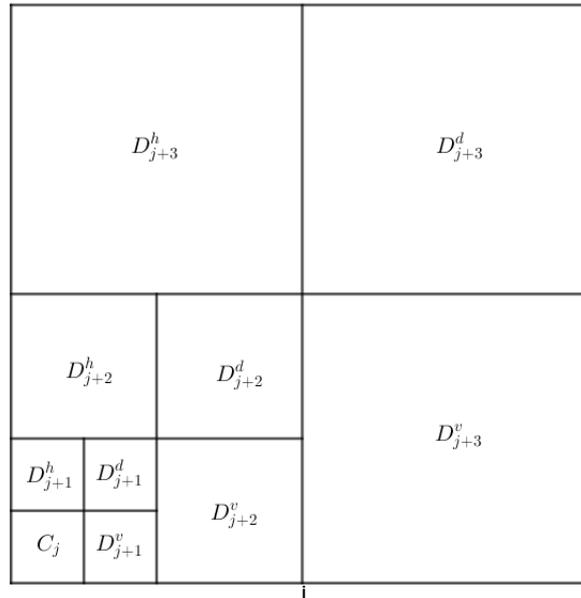
$$f(x,y) = \sum_k c_{j_0,k} \Phi_{j_0,k}(x,y) + \sum_{j=j_0}^{\infty} \sum_k \sum_{\mu=h,v,d} d_{j,k}^{\mu} \Psi_{j,k}^{\mu}(x,y) ,\tag{3.21}$$

com os coeficientes *wavelet* dados por:

$$c_{j_0,k} = \int_{\mathbb{R}^2} f(x,y) \Phi_{j_0,k}(x,y) dx dy ,\tag{3.22}$$

$$d_{j,k}^{\mu} = \int_{\mathbb{R}^2} f(x,y) \Psi_{j,k}^{\mu}(x,y) dx dy .\tag{3.23}$$

Os coeficientes suaves e de detalhe obtidos em cada nível do algoritmo serão matrizes, que podem ser representadas conforme apresentado na Figura 3.8, na qual são representados três níveis de resolução obtidos da AMR bidimensional decimada.

Figura 3.8 – Representação dos coeficientes *wavelet* bidimensionais

Fonte: Morettin (2014)

Pode-se observar na Figura 3.8 que o quadrado superior esquerdo (D_{j+3}^h) representa os detalhes horizontais; o superior direito (D_{j+3}^d), os detalhes diagonais; e o inferior direito (D_{j+3}^v), os detalhes verticais. O menor quadrado inferior esquerdo (C_j) representa a parte suave.

Na decomposição *wavelet* padrão, as operações ocorrem apenas em uma dimensão, ou seja, começa-se por se aplicar a transformada apenas nas linhas e depois nas colunas da imagem.

3.6 Expoente de Hurst

Este importante expoente, de acordo com Piscarev (2017), foi introduzido por Harold Edwin Hurst, hidrólogo britânico que fez parte do projeto de uma barragem no rio Nilo, no Egito. Para a construção, era necessário estimar o fluxo de água e a necessidade de escoamento. Inicialmente, acreditava-se que o fluxo de água era uma variável aleatória, portanto um processo estocástico.

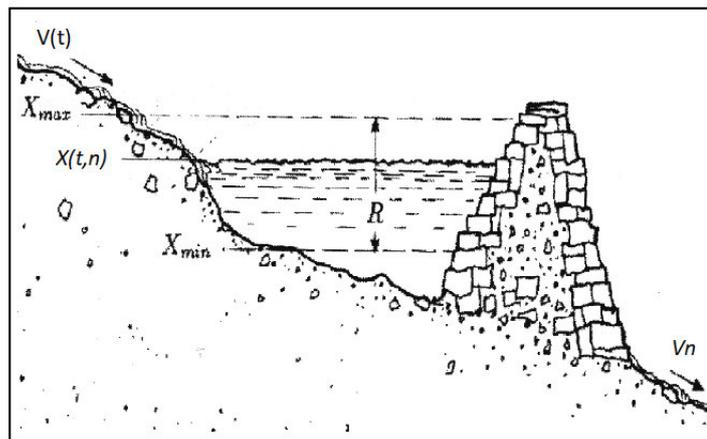
No entanto, Hurst estudou os registros sobre as inundações do Nilo de nove séculos e encontrou um padrão. O comportamento de anos com grande inundação seguido de anos com grande inundação, era denominado comportamento persistente. Se um ano de pouca inundação era seguido de um ano de pouca inundação era denominado antipersistente.

Esse foi o ponto de partida do estudo. Portanto, valores com volumes de água maiores do que a média eram substituídos por valores ainda maiores. Depois disso, o processo alterava sua direção, e valores num nível abaixo da média eram substituídos por outros ainda menores. Havia ciclos com uma duração não periódica.

De acordo com Hurst, era esperado que o alcance (*range*) dessas oscilações mudariam de acordo com o tamanho do intervalo de tempo usado para medida. Caso as séries temporais seguissem o movimento Browniano fracionário (fBm), descrito por Kolmogorov e formalizado por Mandelbrot e Van Ness (1968), o alcance (R) mudaria com a raiz quadrada do tempo (\sqrt{T}). Para padronizar a medida durante todo o tempo, ele dividiu o alcance pelo desvio padrão S , também chamada de análise de reescalonamento ou análise R/S .

A questão é: como determinar um modelo ideal baseado nos registros de descargas, ou seja, um reservatório que nunca deve transbordar ou esvaziar. A Figura 3.9 é uma representação do reservatório, sendo $v(t)$ o volume de água em um dado ano t , V_n a quantidade média de água liberada do reservatório, num período de n anos, para manter uma quantidade de água armazenada.

Figura 3.9 – Representação de um reservatório



Fonte: FEDER (1989)

Nas condições ideais, o volume é dado por:

$$V_n = \frac{1}{n} \sum_{t=1}^n V(t) . \quad (3.24)$$

Este será o volume médio descarregado do reservatório por ano. Seja a variável $X(t, n)$ a variável soma acumulada, até o ano t , da diferença entre a descarga no reservatório e o volume

médio, isto é:

$$X(t, n) = \sum_{i=1}^t [V(i) - \bar{V}_n]. \quad (3.25)$$

O volume de armazenamento necessário para manter a descarga média no período, denotado por R , é a diferença entre o valor máximo e o mínimo das descargas acumuladas. Portanto, R pode ser denotado pela expressão:

$$R(n) = \text{Max}_{(1 < t < n)} X(t, n) - \text{Min}_{(1 < t < n)} X(t, n), \quad (3.26)$$

sendo:

$R(n)$ = variação entre o conteúdo máximo e mínimo;

$\text{Max}X(t, n)$ = nível máximo;

$\text{Min}X(t, n)$ = nível mínimo;

n = período de tempos em anos;

t = determinado ano do período.

Hurst estudou fenômenos naturais, tais como vazão de rios e anéis de árvores, usando a razão adimensional R/S . R é a medida entre os níveis máximo e mínimo e S , o desvio padrão. Essa razão permite comparar o reescalonamento de séries temporais dos mais variados tipos de fenômenos naturais. Que podem ser descritas por uma lei de potência, de acordo com a seguinte relação empírica, dada por (HURST, 1965):

$$\frac{R}{S} = kN^H, \quad (3.27)$$

sendo:

$\frac{R}{S}$ = reescalonamento da série de dados;

k = uma constante;

N = número de observações

H = expoente de Hurst.

Aplicando o logaritmo em ambos os lados na equação (3.27), obtém-se:

$$\log \left(\frac{R}{S} \right) = H \log N + \log k \quad (3.28)$$

A estimativa do expoente de Hurst H será a taxa de variação resultante da regressão linear entre $\log \left(\frac{R}{S} \right)$ e $\log N$.

3.6.1 Expoente de Hurst e análise wavelet

Do ponto de vista estatístico, a análise dos expoentes de Hurst pode ser interpretada como um resumo informativo dos dados e isso pode ser feito não aplicando as técnicas diretamente aos dados, mas nos coeficientes *wavelet*. Em uma decomposição *wavelet* o expoente de Hurst conecta os diferentes níveis de decomposição, de modo mais específico, conecta o comportamento dos coeficientes *wavelet* e o expoente de singularidade (RAMÍREZ-COBO e VIDA KOVIC, 2013).

O expoente de Hurst, de acordo com Nicolis et al. (2011), é uma medida padrão de escala regular que pode ser conectada a medidas de memória longa, dimensão e fractalidade em sinais e imagens. Essa avaliação pode ser feita tanto no domínio dos pixels como também no domínio transformado, geralmente no domínio escala/frequência como Fourier e *wavelets*.

Quando as diferenças funcionais individuais intrínsecas não podem ser descritas por estatísticas básicas ou por tendências, o expoente de Hurst pode ser uma boa opção. Medições de alta frequência e imagens obtidas de várias fontes possuem um grau de autos-similaridade e escala regular inerente. Então, quando os dados parecem um ruído, o expoente de Hurst pode ser o único recurso informativo que possa resumi-los.

A noção de autos-similaridade foi estendida a processos aleatórios. Especificamente, de acordo com Kang e Vidakovic (2017), um processo estocástico $\{X(t), t \in \mathbb{R}^d\}$ é autos-similar com expoente escala (ou expoente de Hurst) H , se para algum $\lambda \in \mathbb{R}$:

$$X(\lambda t) \stackrel{d}{=} \lambda^H X(t), \quad (3.29)$$

onde $\stackrel{d}{=}$ é entendido como igualdade em distribuição.

Se os processos possuem uma estrutura estocástica, o expoente de Hurst torna-se um parâmetro em um modelo estatístico bem definido e pode ser estimado. O movimento Browniano fracionário (fBm) é importante e bem compreendido para dados de escala. Sua importância segue do fato de ser um processo único de Gauss com incrementos estacionários, isto é, auto-similar, de acordo com a equação (3.29).

De acordo com Kang e Vidakovic (2017), o fBm tem um (pseudo)-espectro da forma $S(w) \propto |w|^{-(2H+1)}$ e, conseqüentemente, as log-magnitudes dos coeficientes detalhes em diferentes resoluções em uma decomposição *wavelet* possuem uma relação linear. Então, para

estimar o expoente de Hurst numa decomposição *wavelet*, são necessários somente os coeficientes *wavelet*.

Logo, quando um fBm com expoente de Hurst é mapeado para o domínio *wavelet* por uma DWT ou NDWT, os coeficientes *wavelet* satisfazem as seguintes propriedades:

- (i) d_j , coeficiente *wavelet* do nível j , segue uma distribuição normal com média zero e variância $\sigma_0^2 2^{-j(2H+1)}$, onde σ_0^2 é a variância do coeficiente de detalhe no nível zero;
- (ii) Uma sequência de coeficientes *wavelet* no nível j é estacionária;
- (iii) A covariância entre dois coeficientes de qualquer nível de detalhe diminui exponencialmente à medida que a distância entre eles aumenta.

De acordo com a propriedade (i) uma relação entre os coeficientes *wavelet* e o expoente de Hurst é dada por:

$$\log_2 \mathbb{E}\{d_j^2\} = -j(2H + 1) + 2 \log_2 \sigma_0. \quad (3.30)$$

Uma estimativa para $\mathbb{E}\{d_j^2\}$ é a média dos coeficientes *wavelet* ao quadrado (MQCW).

$$\overline{d_j^2} = \frac{1}{n_j} \sum_{i=1}^{n_j} d_{j,k}^2, \quad (3.31)$$

sendo, n_j o número de coeficientes *wavelet* no nível j . Então, o expoente de Hurst pode ser estimado como:

$$\hat{H} = -\frac{\hat{\beta}}{2} - \frac{d}{2}, \quad (3.32)$$

em que $\hat{\beta}$ é a taxa de variação da regressão linear entre os pares $(j, \log_2 \overline{d_j^2})$ com d relacionado à dimensão dos dados.

Abry et al. (2000) calculou a variância dos coeficientes *wavelet* $\mathbb{E}\{d_j^2\}$ assumindo que os coeficientes *wavelet* são normais independentes e identicamente distribuídos no nível j . Mas esse método é afetado pela não normalidade de d_j e também pela correlação entre coeficientes *wavelet*, que resulta em viés de ponderação.

Para reduzir esse viés, Soltani et al. (2004) definiram Mid-energia, como:

$$D_{j,k} = \frac{d_{j,k}^2 + d_{j,k+n_j/2}^2}{2}, \quad k = 1, 2, \dots, n_j/2. \quad (3.33)$$

De acordo com essa abordagem, cada nível de resolução é dividido em duas partes iguais e os coeficientes correspondentes são pareados, elevados ao quadrado e os novos coeficientes calculados de acordo com a média. Isso produz efeito de quase decorrelação e os autores concluem, como demonstrado no Apêndice A, que $D'_{j,k} = -\log_2 D_{j,k}$ seguem uma distribuição Gumbel, ou seja:

$$D'_{j,k} = -\log_2 D_{j,k} \sim G\left((2H+1)j - \log_2 C + \frac{\gamma}{\log 2}, \frac{\pi^2}{6\log^2 2}\right). \quad (3.34)$$

Portanto para estimar o expoente de Hurst é usado a seguinte relação:

$$E(D'_{j,k}) = \overline{-\log_2 D_{j,k}} = (2H+1)j - \log_2 C + \frac{\gamma}{\log 2}. \quad (3.35)$$

Sendo $\hat{\beta}$ a inclinação da reta resultado da regressão linear entre os pares $(j, \overline{-\log_2 D_{j,k}})$, logo o expoente de Hurst pode ser estimado como:

$$\hat{H} = -\frac{\hat{\beta}}{2} - \frac{d}{2}, \quad (3.36)$$

com d relacionado a dimensão dos dados.

Kang e Vidakovic (2017) propuseram o método MEDL, que aplica a mediana do logaritmo dos coeficientes *wavelet* elevados ao quadrado. Nesse método, o expoente de Hurst (\hat{H}) pode ser estimado de acordo com o teorema 3.1, cuja demonstração encontra-se no Apêndice B.

Teorema 3.1

Seja y_j^* a mediana do logaritmo de d_j^2 , em que d_j é um coeficiente *wavelet* no nível j de uma transformação *wavelet*, em um fBm com expoente de Hurst H . Então, a mediana da população é dada por:

$$y_j^* = -\ln 2(2H+1)j + k, \quad (3.37)$$

em que k é uma constante independente de j . O expoente de Hurst pode ser estimado como:

$$\hat{H} = -\frac{\hat{\beta}}{2\ln 2} - \frac{d}{2}, \quad (3.38)$$

e $\hat{\beta}$ é a inclinação da reta resultado da regressão linear entre os pares (j, \hat{y}_j^*) e d relacionado a dimensão dos dados.

Os autores também propuseram o método MEDLA, que aplica a mediana do logaritmo da média dos pares de coeficientes *wavelet* elevados ao quadrado, para reduzir a correlação entre os coeficientes. O método MEDLA é semelhante ao proposto por Soltani et al. (2004), exceto que, em vez de fixar as combinações de pares, são amostrados aleatoriamente m pares, cujos membros estão a uma distância de pelo menos $q_j = 2^{J-j}$, que diminuam de acordo com cada nível de resolução. Assim, o expoente de Hurst pode ser estimado de acordo com o teorema 3.2, cuja demonstração encontra-se no Apêndice C.

Teorema 3.2

Sejam d_{j,k_1} e d_{j,k_2} dois coeficientes wavelet no nível j , em posições k_1 e k_2 respectivamente, de uma NDWT de um fBm com expoente de Hurst H . Suponha que $|k_1 - k_2| > q_j$, em que q_j é a distância mínima de separação que depende do nível j e da base *wavelet* selecionada. Seja y_j^* a mediana do logaritmo de $[\frac{d_{j,k_1}^2 + d_{j,k_2}^2}{2}]$. Então a mediana da população é dada por:

$$y_j^* = -\ln 2(2H + 1)j + k \quad , \quad (3.39)$$

em que k é uma constante independente de j . Assim, o expoente de Hurst pode ser estimado como:

$$\hat{H} = -\frac{\hat{\beta}}{2\ln 2} - \frac{d}{2} \quad , \quad (3.40)$$

e $\hat{\beta}$ é a inclinação da reta resultado da regressão linear entre os pares (j, y_j^*) e d relacionado à dimensão dos dados.

Embora a mediana não seja influenciada por *outliers*, ela pode se comportar de maneira inesperada como resultado de seu comportamento não suave. O fato de a mediana não ser universalmente o melhor estimador de resistência a *outliers*, forneceu uma motivação para examinar alternativas que são intermediárias no comportamento entre média suave, mas atípica, e a mediana muito insensível à inconsistência, mas não suave (FENG e VIDAKOVIC, 2017).

Nesse sentido, Feng e Vidakovic (2017) propuseram uma nova metodologia chamada estimador *Trimean* de Mid-energia (GTME), na qual a média dos coeficientes *wavelet* é substituída pela média ponderada da mediana da distribuição e dos dois quantis simétricos a mediana.

Dessa forma, a generalização do estimador *Trimean* é definido como a média ponderada da mediana $Y_{\frac{1}{2}}$ e os dois quantis Y_p e Y_{1-p} com $p \in (0, \frac{1}{2})$:

$$\hat{\mu} = \frac{\alpha}{2}Y_p + (1 - \alpha)Y_{\frac{1}{2}} + \frac{\alpha}{2}Y_{1-p}. \quad (3.41)$$

Os pesos para os dois quantis, Y_p e Y_{1-p} , são os mesmos e $\alpha \in [0, 1]$. Um caso especial, chamado estimador Tukey's *Trimean*, que obteve melhores resultados, considera $\alpha = \frac{1}{2}$ e $p = \frac{1}{4}$, (Feng e Vidakovic, 2017). Para calcular esse estimador, primeiro os dados são classificados em ordem crescente. Em seguida, é tomado o valor no qual um quarto da sequência esteja abaixo dele (primeiro quantil), o valor em que metade esteja abaixo dele (mediana) e o valor em que três quartos da sequência estejam abaixo dele. Por fim, calculamos a média ponderada desses valores. Denotando o estimador Tukey's *Trimean* por $\hat{\mu}_T$, então:

$$\hat{\mu}_T = \frac{1}{4}Y_{\frac{1}{4}} + \frac{1}{2}Y_{\frac{1}{2}} + \frac{1}{4}Y_{\frac{3}{4}}. \quad (3.42)$$

Nesse método proposto em 3.42, os autores geram $N/2$ Mid-energias

$$D_{j,k} = \frac{d_{j,k}^2 + d_{j,k+N/2}^2}{2}, \quad k = 1, 2, 3, \dots, N/2, \quad (3.43)$$

e em cada nível são selecionados $D_{j,k}$ coeficientes a uma distância M para formar M grupos. Assumem que $\frac{N/2}{M}$ coeficientes, dentro de cada subgrupo são independentes. Desse modo os subgrupos formados são da seguinte forma:

$$\begin{aligned} \text{Grupo 1} &: \{D_{j,1}, D_{j,1+M}, D_{j,1+2M}, \dots, D_{j,N/2-M+1}\} \\ \text{Grupo 2} &: \{D_{j,2}, D_{j,2+M}, D_{j,2+2M}, \dots, D_{j,N/2-M+2}\} \\ &\vdots \\ \text{Grupo } M &: \{D_{j,M}, D_{j,2M}, D_{j,3M}, \dots, D_{j,N/2}\}. \end{aligned} \quad (3.44)$$

A aplicação desse método para estimar o expoentes de Hurst (\hat{H}) pode ser descrito pelo teorema 3.3, cuja demonstração se encontra no Apêndice D.

Teorema 3.3

Seja $\hat{\mu}_{j,i}$ o estimador *Trimean* geral baseado em

$$\{D_{j,i}, D_{j,i+M}, D_{j,i+2M}, \dots, D_{j,N/2-M+i}\} = D(i, j), \quad (3.45)$$

em que $D(j, i)$ para $1 \leq i \leq M$ e $1 \leq j \leq J$ é o i -ésimo subgrupo de Mid-energias no nível j em uma NDWT de um fBm de tamanho N com expoente de Hurst H . Então a distribuição assintótica de $\hat{\mu}_{j,i}$ é normal, ou seja:

$$\hat{\mu}_{j,i} \sim N(c(\alpha, p)\lambda_j, \frac{2M}{N}f(\alpha, p)\lambda_j^2), \quad (3.46)$$

em que

$$c(\alpha, p) = \frac{\alpha}{2} \ln \left(\frac{1}{p(1-p)} \right) + (1-\alpha) \ln 2,$$

$$f(\alpha, p) = \frac{\alpha(1-2p)(\alpha-4p)}{4p(1-p)} + 1 \quad e \quad \lambda_j = \sigma^2 \cdot 2^{-(2H+1)j}$$

σ é o desvio padrão dos coeficientes *wavelet* no nível zero. O expoente de Hurst pode ser estimado:

$$\hat{H} = \frac{1}{M} \sum_{i=1}^M \hat{H}_i = \frac{1}{M} \sum_{i=1}^M \left(-\frac{\hat{\beta}_i}{2} - \frac{d}{2} \right) = -\frac{\bar{\beta}}{2} - \frac{d}{2}. \quad (3.47)$$

em que $\bar{\beta} = \frac{1}{M} \sum_{i=1}^M \hat{\beta}_i$ é a média das inclinações da regressão linear dos pares $(j, \log_2(\hat{\mu}_{i,j}))$ nos grupos $i = 1, 2, 3, \dots, M$ e d está relacionado a dimensão dos dados.

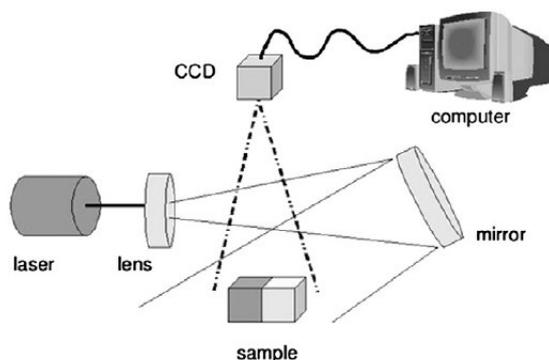
Essas metodologias também podem ser estendidas para outras dimensões. No caso bi-dimensional, o expoente de Hurst direcional pode ser calculado com uma pequena mudança relacionada à dimensão, fazendo $d = 2$ nas direções horizontal, vertical e diagonal. j é a escala (níveis de resolução) e β_*^i a taxa de variação, de acordo com o método adotado para estimar os expoentes de Hurst, nas direções horizontal, vertical e diagonal.

4 MATERIAL E MÉTODO

As imagens interferométricas do carcinoma mamário anaplástico canino foram obtidas por meio de um método óptico, que usa o padrão de interferência formado quando um material biológico é iluminado por um laser, chamada técnica Biospeckle laser (BSL). Um feixe de laser HeNe com 10 mw, 632 nm, foi expandido para iluminar o tumor dirigido por um espelho, uma câmera e um computador são responsáveis pela montagem das imagens.

Na seção de iluminação, foram armazenadas 128 imagens em níveis de cinza (8 bits), relativas ao speckle dinâmico, em intervalos de 0,08 segundos. A Figura 4.1 mostra uma representação da técnica Biospeckle.

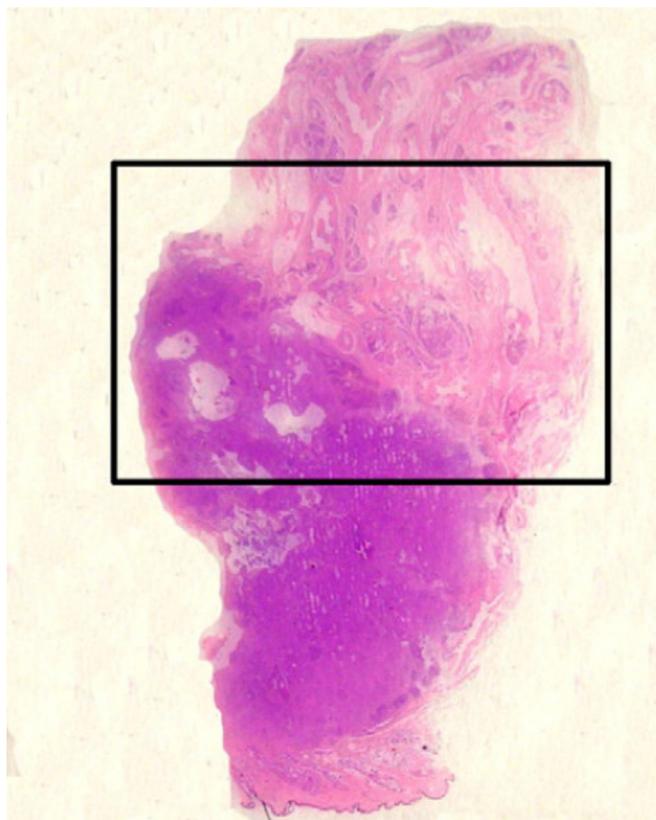
Figura 4.1 – Aquisição das imagens



Fonte: Braga et al. (2012)

Foram extraídas das 128 imagens interferométricas (640 x 486 pixels) do carcinoma, subimagens (64 x 64 pixels) com região composta somente por tecido cancerígeno e subimagens composta somente por tecido saudável, totalizando 256 subimagens. As localizações dos tecidos cancerígenos e saudáveis foram baseadas na pesquisa feita por Braga et al. (2012), com análise espectral nas imagens interferométricas, em que foram observadas altas frequências em tecidos cancerígenos e baixas frequências em tecidos saudáveis. Na Figura 4.2, são mostradas as regiões com tecido saudável em rosa claro e tecido cancerígeno em rosa escuro, de acordo com Braga et al. (2012).

Figura 4.2 – Região Saudável e Região Cancerígena



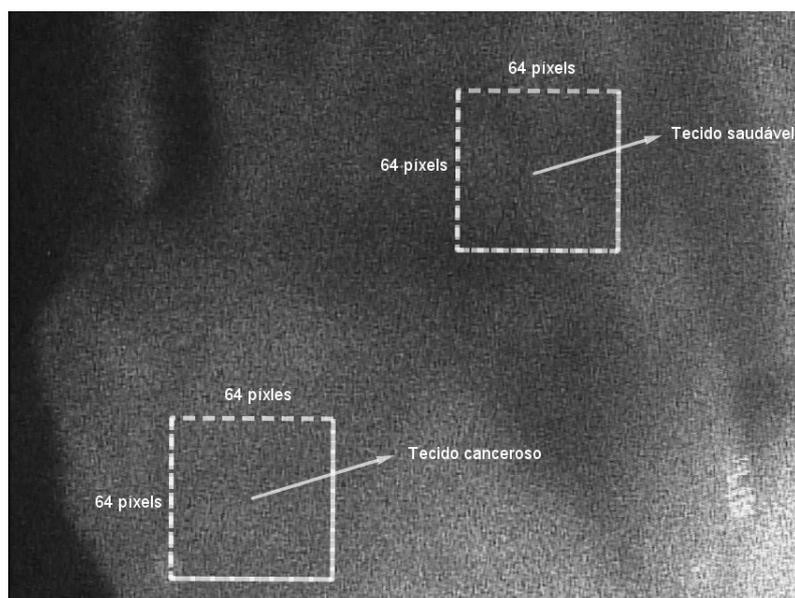
Fonte: Braga et al. (2012)

O procedimento adotado para processamento das subimagens interferométricas de células cancerígenas e saudáveis foi a análise multirresolução de *wavelet*, proposta por Mallat (1989), pois se mostra muito eficiente na extração de detalhes em imagens.

A análise multirresolução se justifica pela representação da imagem com poucos coeficientes e com bom armazenamento de informações das bordas e arestas. A decomposição foi feita com utilização da transformada discreta de *wavelet* decimada DWT e não decimada NDTW, com a implementação da *wavelet* de Daubechies com oito momentos nulos (Daub8) e com quatro níveis de resolução.

A decomposição foi feita via transformada de *wavelet* 2D nas três direções (v, h e d) e verificado o comportamento dos expoentes de Hurst, ao longo do tempo, na sequência das subimagens e nas três direções mencionadas. Por fim, foram discutidos os efeitos da média dos coeficientes *wavelet* ao quadrado, Mid-energias, os métodos MEDL e MEDLA e o estimador *Trimean*, usados para estimar os expoentes de Hurst, na série ao longo do tempo. Na Figura 4.3, são apresentadas as localizações das subimagens a serem analisadas.

Figura 4.3 – Procedimento de extração das subimagens a serem analisadas



Fonte: Adaptação Braga et al. (2012)

O processamento das imagens e as análises das séries foram feitas com utilização do *software R* (Core Team, 2018). Para executar a decomposição *wavelet* das imagens, foram utilizados os pacotes de ferramentas *bmp* (Jefferis, 2017), *Waveslim* (Whitcher, 2015), *Wavethresh* (Nason, 2016) e para as análises das séries, os pacotes *beanplot* (Kampstra, 2008) e *fBasics* (Wuertz, Setz e Chalabi, 2017).

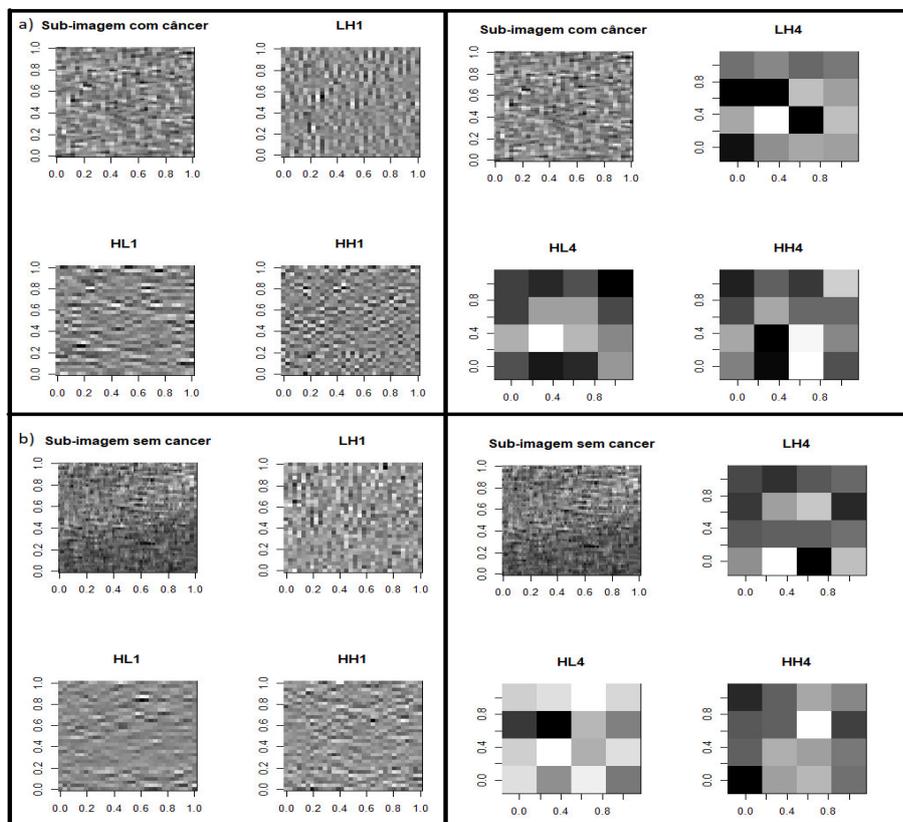
5 RESULTADOS E DISCUSSÕES

A análise dos resultados está dividida em duas subseções, a primeira com utilização da transformada decimada de *wavelet* e a segunda com a transformada não decimada de *wavelet*.

5.1 Análise temporal dos expoentes de Hurst com utilização da DWT

Nesta seção, serão mostrados os resultados com a utilização da média dos coeficientes *wavelet* ao quadrado (MQCW), os métodos MEDL e MEDLA, Mid-energias e o estimador *Trimean*, para a estimação dos expoentes de Hurst. As subimagens foram decompostas com implementação das transformadas decimada de *wavelet* (DWT), em quatro níveis de resolução. Na Figura 5.1, apresenta-se uma subimagem do tecido cancerígeno em (a) e do tecido saudável em (b) decomposta com utilização da transformada decimada de *wavelet* no nível 1 e no nível 4 de resolução. Percebe-se que, com o aumento no nível de resolução melhora-se no nível de detalhes.

Figura 5.1 – Decomposição de uma subimagem do tecido cancerígeno em (a) do tecido saudável em (b), no nível 1 e 4 de resolução com utilização DWT



Para cada método utilizado, para estimar o expoente de Hurst, são analisados os gráficos das séries temporais nas três direções, as estatísticas descritivas e os gráficos das distribuições empíricas em subimagens do tecido cancerígeno e saudável.

5.1.1 Estimação dos expoentes de Hurst com MQCW

Na Tabela 5.1, são apresentadas as estatísticas descritivas das séries de expoentes de Hurst, ao longo do tempo, em subimagens da região cancerígena e saudável, nas três direções (horizontal, vertical e diagonal), com utilização para sua estimativa, a transformada decimada de wavelet e a média dos coeficientes *wavelet* ao quadrado (MQCW). Pode-se destacar que existe uma variação dos expoentes de Hurst, ao longo do tempo, em subimagens interferométricas obtidas por meio da técnica Biospeckle, ou seja, mesmo se tratando de subimagens da mesma região, observa-se que, se tiradas em tempos distintos, podem gerar diferentes expoentes de Hurst.

Uma variabilidade maior em torno do valor médio na direção diagonal, comparada às outras direções, foi observada nas duas regiões. Por consequência, um maior coeficiente de variação. Outra observação a ser destacada é sobre as médias dos expoentes de Hurst em tecidos saudáveis, que apresentaram valores maiores nas direções horizontal e vertical, comparado às médias dos expoentes do tecido cancerígeno. Por fim, em relação às medidas de assimetria e curtose, mostram distribuições com características assimétricas e leptocúrtica dos expoentes de Hurst, ao longo do tempo.

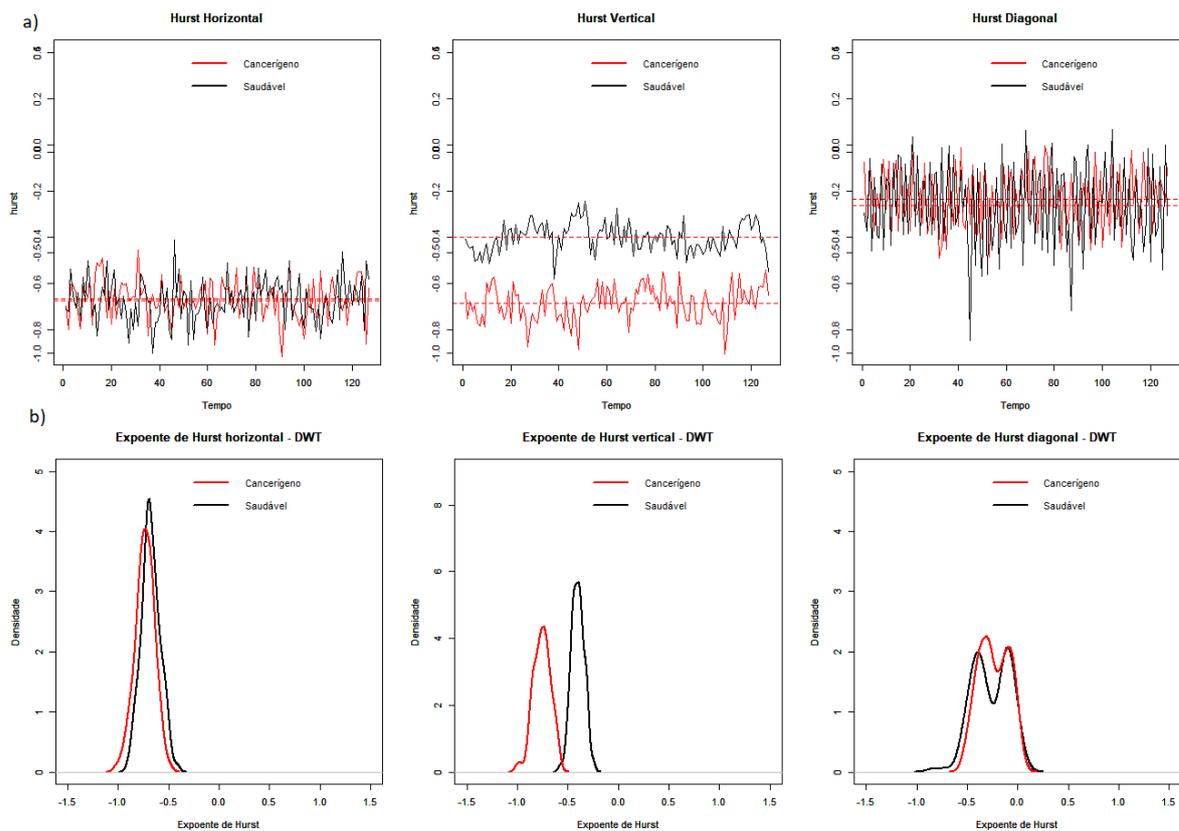
Tabela 5.1 – Estatísticas descritivas dos expoentes de Hurst direcionais com utilização da DWT com a MQCW

Expoentes (Tumor)	Média	Mediana	Variância	CV	Assimetria	Curtose
Horizontal	-0,7326	-0,7327	0,0092	0,1313	-0,2381	0,0350
Vertical	-0,7551	-0,7522	0,0072	0,1127	-0,2429	-0,1335
Diagonal	-0,2329	-0,2485	0,0217	0,6332	-0,0149	-1,1741
Expoentes (Saudável)	Média	Mediana	Variância	CV	Assimetria	Curtose
Horizontal	-0,6748	-0,6828	0,0086	0,1374	0,1427	-0,2082
Vertical	-0,3996	-0,4007	0,0039	0,1571	-0,0138	-0,2504
Diagonal	-0,2620	-0,2890	0,0335	0,6990	-0,2738	-0,6124

Nas Figura 5.a e 5.b, são apresentados os gráficos das séries e das distribuições empíricas dos expoentes de Hurst, ao longo do tempo, nas três direções. Foram feitas sobreposições das distribuições dos expoentes de Hurst com tecidos cancerígenos (em vermelho) e saudáveis (em

preto). Percebe-se uma interseção nas curvas, principalmente nas direções horizontal e diagonal, o que evidencia o poder do expoente de Hurst de diferenciar subimagens interferométricas de tecido cancerígeno do saudável, detectada pelo expoente vertical. Resultados semelhantes foram obtidos por Nicolis et al. (2011), realizadas com imagens radiográficas de tumores em diferentes mamas.

Figura 5.2 – Gráficos dos expoentes de Hurst com utilização da DWT com a MQCW: (a) Gráficos das séries ao longo do tempo e (b) Densidades



Na próxima seção, serão apresentados os resultados utilizando a transformada decimada de *wavelet* e com implementação da Mid-energia para estimação dos expoentes de Hurst. De acordo com essa abordagem, cada nível de resolução é dividido em duas partes iguais e os coeficientes *wavelet* correspondentes são pareados, elevados ao quadrado e os novos coeficientes calculados em média.

5.1.2 Estimação dos expoentes de Hurst com Mid-energias

São apresentadas na Tabela 5.2 as estatísticas descritivas das séries de expoentes de Hurst, ao longo do tempo, nas três direções com utilização da transformada decimada de *wavelet* e implementação da Mid-energia para sua estimação, nas subimagens do tecido cancerígeno e saudável. Esse método foi proposto por Soltani et al. (2004) para reduzir o viés de ponderação no expoente de Hurst que pode ocorrer pela não normalidade dos coeficientes *wavelet* e pela correlação existente entre eles.

Destaca-se, assim como ocorrido na aplicação da MQCW, a variação dos expoentes de Hurst em torno de um valor médio, ao longo do tempo. Além disso, observa-se também, de acordo com a Tabela 5.2, um aumento dos valores médios e das variâncias das séries e quando se utiliza Mid-energia nas duas regiões, conseqüentemente um aumento do coeficiente de variação.

Vale salientar também que, comparando os valores médios dos expoentes de Hurst na tabela 5.2, subimagens do tecido saudável produzem expoentes maiores do que subimagens do tecido cancerígeno, fato que também ocorreu com o uso da média dos coeficientes *wavelet* ao quadrado. Ademais, as medidas de assimetria e curtose revelam que persiste o comportamento assimétrico e leptocúrtico das distribuições nas duas regiões.

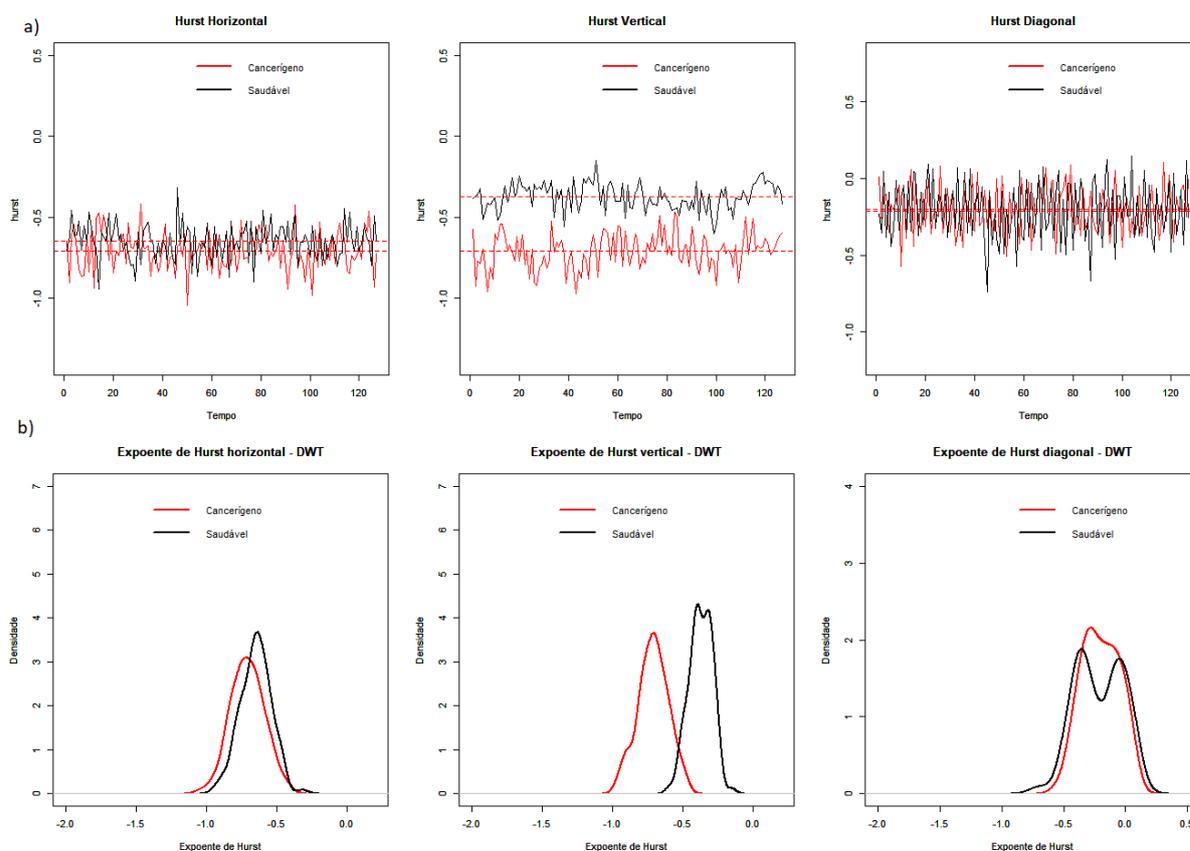
Tabela 5.2 – Estatísticas descritivas dos expoentes de Hurst direcionais com utilização da DWT com Mid-energia

Expoentes (Tumor)	Média	Mediana	Variância	CV	Assimetria	Curtose
Horizontal	-0,7080	-0,7091	0,0142	0,1688	-0,0112	-0,1813
Vertical	-0,7109	-0,7096	0,0115	0,1512	-0,1239	-0,3475
Diagonal	-0,1999	-0,2007	0,0233	0,7647	-0,0240	-1,8797
Expoentes (Saudável)	Média	Mediana	Variância	CV	Assimetria	Curtose
Horizontal	-0,6493	-0,6486	0,0118	0,1678	-0,0802	0,0209
Vertical	-0,3724	-0,3727	0,0066	0,2189	-0,2105	-0,2941
Diagonal	-0,2119	-0,2334	0,0368	0,9061	-0,1418	-0,8863

São apresentados na Figura 5.3b os gráficos das distribuições empíricas e das séries dos expoentes de Hurst, ao longo do tempo, nas direções horizontal, vertical e diagonal. Percebe-se que, como obtido por acordo Jeon, Nicolis e Vidakovic (2014), estimativas robustas do expoente de Hurst podem diferenciar subimagens de tecido cancerígenos e saudável, já que as distribuições não se sobrepõem em todas as direções.

Informação que pode ser observada com a análise da Figura 5.3a, que mostram médias individuais das séries próximas nas direções horizontal e diagonal e médias afastadas na direção vertical, mostrando que a diferenciação entre tecido cancerígeno e saudável foi detectada pelo expoente de Hurst vertical. Além disso, nota-se também uma mistura de distribuições na direção diagonal.

Figura 5.3 – Gráficos dos expoentes de Hurst com utilização da DWT com Mid - energia: (a) Gráficos das séries ao longo do tempo e (b) Densidades



Na próxima seção, serão apresentados os resultados com utilização da transformada decimada de *wavelet* e implementação do método MEDL, que aplica a mediana do logaritmo dos coeficientes *wavelet* elevados ao quadrado, para estimar o expoente de Hurst.

5.1.3 Estimação dos expoentes de Hurst com método MEDL

São apresentadas na Tabela 5.3, as estatísticas descritivas das séries de expoentes de Hurst direcionais, ao longo do tempo, com a utilização da transformada decimada de *wavelet* em subimagens do tecido cancerígeno e saudável com implementação do método MEDL, para sua estimação. Esse método foi proposto por Kang e Vidakovic (2017) para tentar reduzir a correlação existente entre os coeficientes *wavelet*.

Pode-se destacar uma variação dos expoentes Hurst, ao longo do tempo, na três direções e nas duas regiões (cancerígena e saudável). Comparada com as estatísticas descritivas com utilização da MQCW e Mid-energia, Tabelas 5.1 e 5.2, tanto para subimagens do tecido cancerígeno e saudável, percebe-se um aumento nos valores médios e uma diminuição da variância das séries de expoentes de Hurst, por consequência uma diminuição do coeficiente de variação em todas as direções. Isso pode ter ocorrido pelo fato de a mediana não ser influenciada por possíveis *outliers* que possam ocorrer com a transformação logarítmica nos coeficientes *wavelet* (FENG e VIDAKOVIC, 2017).

Verifica-se também que, como ocorrido nos métodos já aplicados, tecidos saudáveis tendem a apresentar expoentes de Hurst maiores que tecidos cancerígenos nas direções horizontal e vertical. Em relação às medidas de assimetria e curtose nota-se, na Tabela 5.3, com aplicação do método MEDL, que as distribuições continuam assimétricas e leptocúrticas, como verificado com o uso da MQCW e de Mid-energia para a estimação dos expoentes de Hurst.

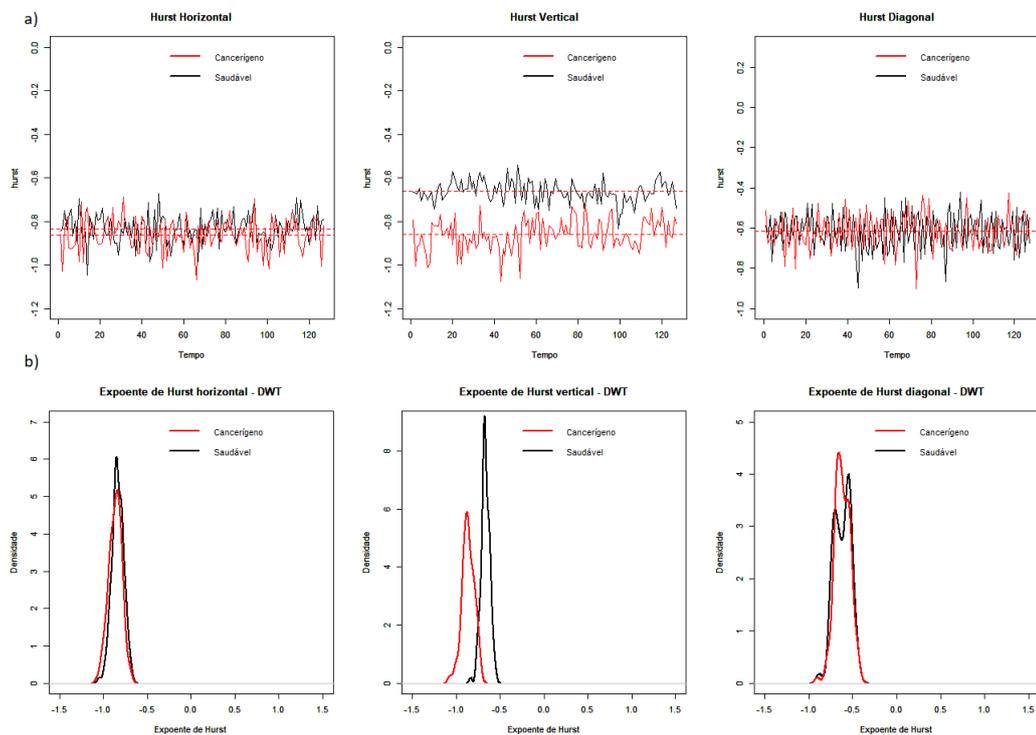
Tabela 5.3 – Estatísticas descritivas dos expoentes de Hurst direcionais com utilização da DWT com método MEDL

Expoentes (Tumor)	Média	Mediana	Variância	CV	Assimetria	Curtose
Horizontal	-0,8621	-0,8592	0,0054	0,0855	-0,2200	-0,2200
Vertical	-0,8607	-0,8642	0,0048	0,0808	-0,2997	0,1664
Diagonal	-0,6164	-0,6197	0,0072	0,1381	-0,1414	0,0517
Expoentes (Saudável)	Média	Mediana	Variância	CV	Assimetria	Curtose
Horizontal	-0,8329	-0,8374	0,0045	0,0806	-0,1731	-0,0427
Vertical	-0,6635	-0,6671	0,0023	0,0734	-0,2084	0,4652
Diagonal	-0,6173	-0,6152	0,0089	0,1529	-0,2416	-0,5847

Na Figura 5.4b, são apresentados os gráficos das distribuições empíricas dos expoentes de Hurst, ao longo do tempo, nas três direções, com aplicação do método MEDL para estimar os expoentes de Hurst. Como ocorrido quando utilizou-se a MQCW e Mid-energia, houve uma interseção nas curvas nas direções horizontal e diagonal, evidenciando um comportamento semelhante em subimagens do tecido saudável e cancerígeno nessas direções. Na direção vertical, apesar de uma diferença menor entre os valores médios, percebe-se uma diferenciação maior entre tecido saudável e cancerígeno com uso do método MEDL, consequência de uma menor variância dos expoentes de Hurst, comparada aos métodos anteriores.

Isso também pode ser evidenciado com a análise dos gráficos das séries (Figura 5.4a), que por sua vez, mostram valores médios individuais dos expoentes próximos nas direções horizontal e diagonal e afastados na direção vertical. Por fim, nota-se, como nos casos anteriores, uma mistura de distribuições na direção diagonal.

Figura 5.4 – Gráficos dos expoentes de Hurst com utilização da DWT com o método MEDL: (a) Gráficos das séries ao longo do tempo e (b) Densidades



Na próxima seção, serão apresentados os resultados utilizando a transformada decimada de *wavelet* com implantação do método MEDLA para estimação dos expoentes de Hurst.

5.1.4 Estimação dos expoentes de Hurst com método MEDLA

O método MEDLA foi proposto por Kang e Vidakovic (2017) para reduzir a correlação existente entre os coeficientes *wavelet* em cada nível de resolução. O método aplica a mediana do logaritmo da média de pares de coeficientes *wavelet* ao quadrado, selecionados aleatoriamente em cada nível de resolução. Os m pares selecionados devem estar a uma distância mínima de $q_j = 2^{J-j}$, que diminuam de acordo com cada nível j de resolução.

Na Tabela 5.4, são apresentadas as estatísticas descritivas das séries de expoentes de Hurst direcionais, ao longo do tempo, em subimagens do tecido cancerígeno e saudável, com a utilização da transformada decimada de *wavelet* e o método MEDLA para sua estimação. Como esse método envolve uma aleatorização, os gráficos apresentados na figura 5.5 e a Tabela 5.4 são resultados de uma simulação, em que, para cada subimagem foram gerados 100 expoentes de Hurst direcionais e tomados em média. Destaca-se, assim como ocorrido nos casos anteriores, existência de variação nas três direções, em subimagens do tecido cancerígeno e saudável.

Comparado ao método MEDL, que também utiliza a mediana, nota-se além de valores médios próximos, também uma diminuição da variância e por consequência menores coeficientes de variação, nas três direções. Também em comparação com uso da MQCW e Mid-energia, houve diminuição na variância e por isso, se apresenta até o momento como o método com menor variação das séries de expoentes de Hurst. Ademais, as medidas de assimetria e curtose (Tabela 5.7), reforçam, como nos casos anteriores, características assimétricas e leptocúrticas das distribuições empíricas em todas as direções.

Tabela 5.4 – Estatísticas descritivas dos expoentes de Hurst direcionais com utilização da DWT com método MEDLA

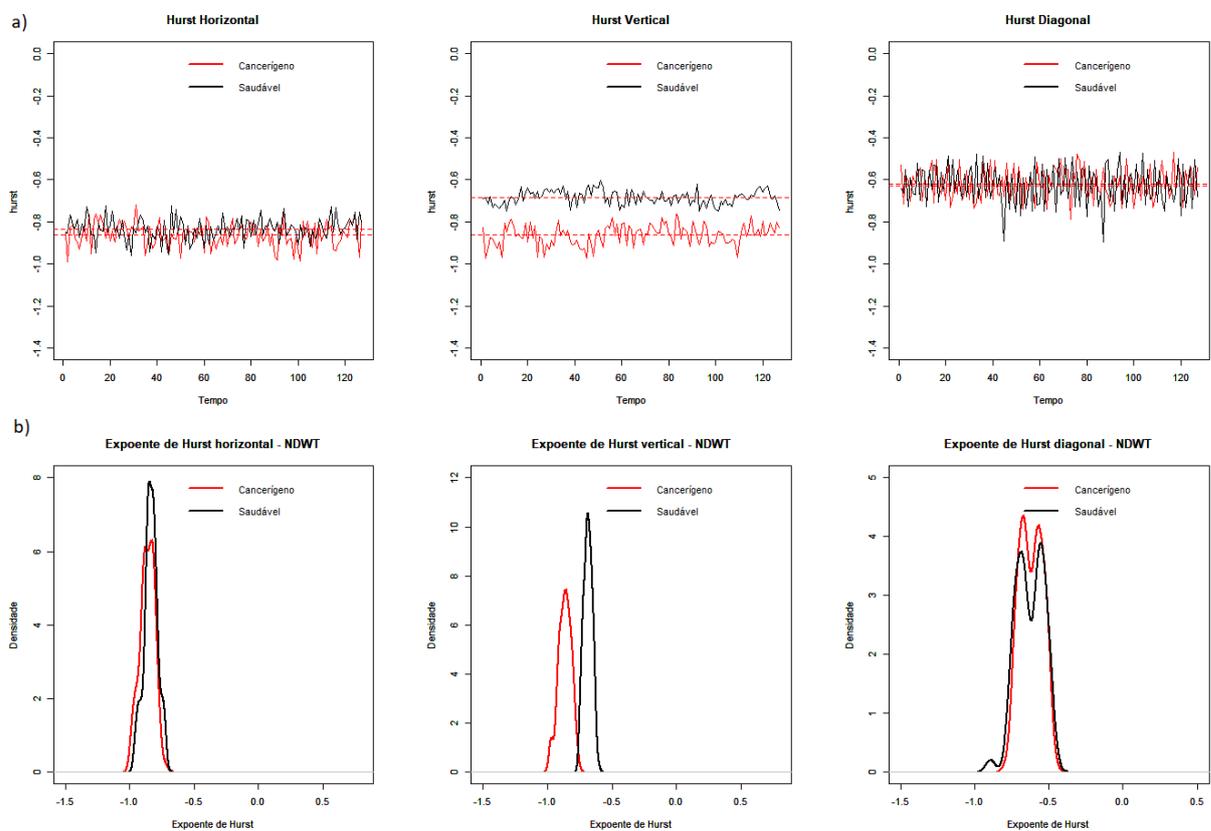
Expoentes (Tumor)	Média	Mediana	Variância	CV	Assimetria	Curtose
Horizontal	-0,8609	-0,8567	0,0030	0,0646	-0,2052	-0,4386
Vertical	-0,8641	-0,8622	0,0024	0,0567	-0,2436	-0,4458
Diagonal	-0,6173	-0,6246	0,0055	0,1206	0,0111	-1,0977
Expoentes (Saudável)	Média	Mediana	Variância	CV	Assimetria	Curtose
Horizontal	-0,8344	-0,8303	0,0028	0,0636	-0,2010	-0,1141
Vertical	-0,6835	-0,6813	0,0011	0,0488	-0,0385	-0,7571
Diagonal	-0,6257	-0,6339	0,0083	0,1463	-0,3180	-0,4982

Os gráficos das séries e das distribuições empíricas dos expoentes de Hurst direcionais, com aplicação do método MEDLA, são apresentados na Figura 5.5. Como ocorrido nos casos

anteriores, nota-se sobreposição das curvas nas direções horizontal e diagonal e um afastamento na direção vertical. Isso evidencia que, a análise temporal dos expoentes de Hurst em imagens interferométricas são capazes de distinguir tecido cancerígeno do tecido saudável.

Essa informação também pode ser verificada com a análise das séries (Figura 5.5a), que por sua vez, mostram médias individuais próximas nas direções horizontal e diagonal e médias afastadas na direção vertical.

Figura 5.5 – Gráficos dos expoentes de Hurst com utilização da DWT com o método MEDLA: (a) Gráficos das séries ao longo do tempo e (b) Densidades



Na próxima seção, serão apresentados os resultados utilizando a transformada decimada de *wavelet* com implementação do estimador *Trimean* para o cálculo dos expoentes de Hurst. Nesse estimador, a média dos coeficientes *wavelet* ao quadrado é substituída pela média ponderada da mediana da distribuição e os dois quantis simétricos a mediana.

5.1.5 Estimação dos expoentes de Hurst com estimador Trimean

Para produzir os resultados, com utilização do estimador *Trimean*, foram gerados 4 subgrupos de coeficientes *wavelet* em cada nível de resolução e calculado a média *Trimean* em cada subgrupo para estimação dos expoentes de Hurst. Na Tabela 5.5, são apresentadas as estatísticas descritivas das série de expoentes de Hurst direcionais, ao longo do tempo, com utilização da transformada decimada de *wavelet* em subimagens do tecido cancerígeno e saudável, com o uso do estimador *Trimean*. Esse método foi proposto por Feng e Vidakovic (2017) com o objetivo de reduzir a correlação entre os coeficientes *wavelet* e o efeito de possíveis *outliers* que possam surgir com a transformação logarítmica.

Como ocorrido nos métodos anteriores, nota-se uma variação dos expoentes de Hurst, ao longo do tempo. Mas comparada com as estatísticas descritivas utilizando a MQCW e os métodos MEDL e MEDLA para estimar os expoentes de Hurst, percebe-se, com implementação do estimador *Trimean*, um aumento da variância e das médias das séries. No entanto, variâncias menores que Mid-energias.

As medidas de assimetria e curtose, apresentadas na Tabela 5.5, mostram que as distribuições empíricas mantêm-se assimétricas e leptocúrticas.

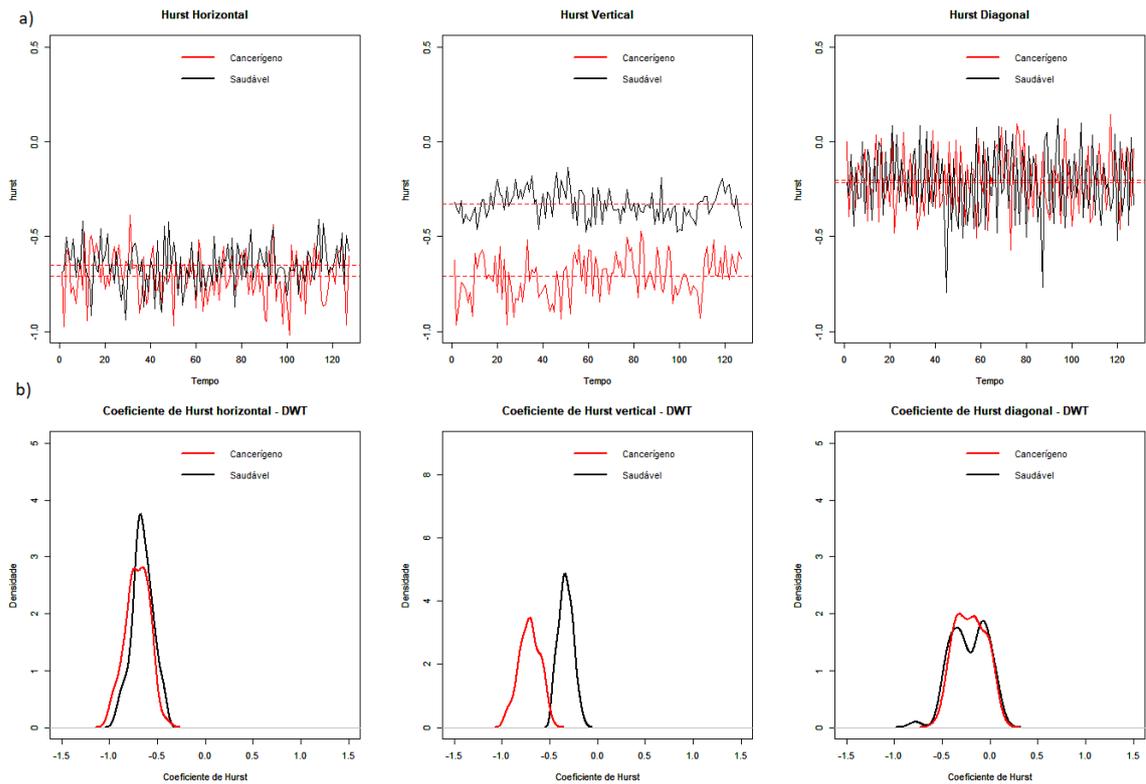
Tabela 5.5 – Estatísticas descritivas dos expoentes de Hurst direcionais com utilização da DWT com estimador Trimean

Expoentes (Tumor)	Média	Mediana	Variância	CV	Assimetria	Curtose
Horizontal	-0,6605	-0,6555	0,0098	0,1504	-0,2708	-0,0148
Vertical	-0,6673	-0,6594	0,0074	0,1294	-0,3133	-0,1071
Diagonal	-0,1539	-0,1692	0,0228	0,9817	-0,0052	-0,1968
Expoentes (Saudável)	Média	Mediana	Variância	CV	Assimetria	Curtose
Horizontal	-0,5967	-0,6037	0,0094	0,1628	0,2205	-0,0098
Vertical	-0,3057	-0,3096	0,0038	0,2025	0,0072	-0,4568
Diagonal	-0,1765	-0,1910	0,0348	1,0572	-0,2561	-0,7515

Na Figura 5.6b, são apresentados os gráficos das distribuições empíricas dos expoentes de Hurst, ao longo do tempo, nas três direções, com aplicação do estimador *Trimean* para estimá-lo. Como verificado nos métodos anteriores, houve sobreposição das distribuições nas direções horizontais e diagonais. A diferenciação entre tecido saudável e cancerígeno foi detectada pelos expoentes de Hurst verticais, já que não se observa sobreposição expressiva.

Essas informações também podem ser observadas com a análise das respectivas séries (Figura 5.6a) que mostram médias individuais próximas nas direções horizontal e diagonal e afastadas na direção vertical. Por fim, observa-se a presença de *outliers* e uma mistura de distribuições na direção diagonal.

Figura 5.6 – Gráficos dos expoentes de Hurst com utilização da DWT com estimador Trimean: (a) Gráficos das séries ao longo do tempo e (b) Densidades

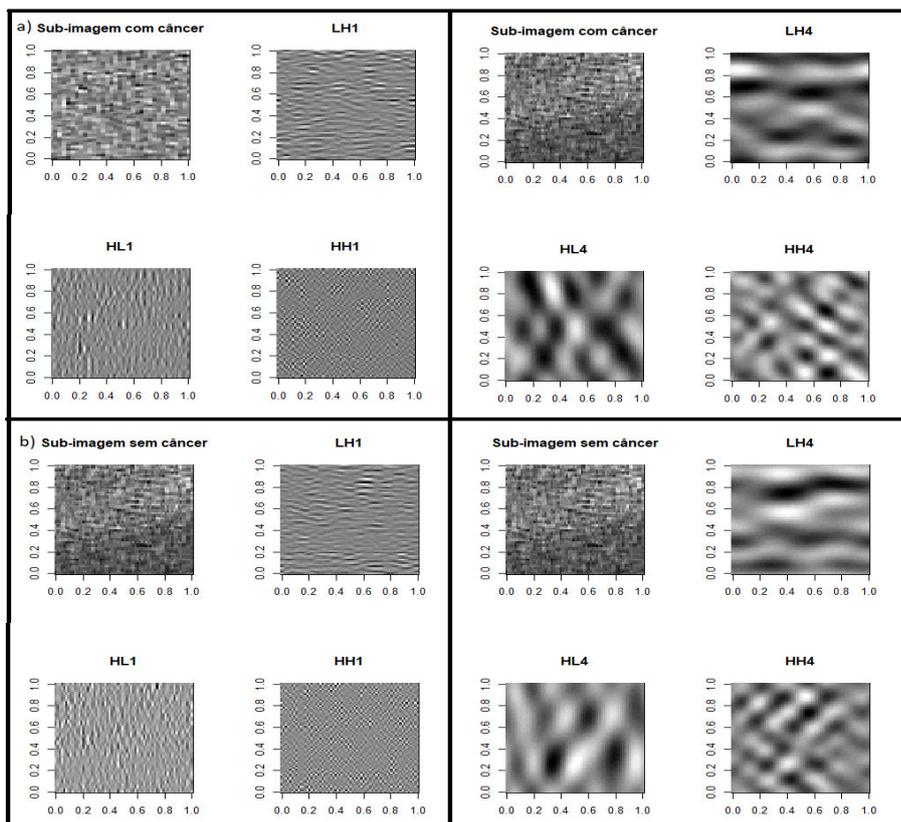


Nas próximas seções, serão apresentados os cálculos do expoente de Hurst utilizando a transformada não decimada de *wavelet*, que de acordo com Kang e Vidakovic (2017), diminuem a variância dos coeficientes *wavelet*, mas em contrapartida aumenta a correlação entre eles.

5.2 Análise temporal dos expoentes de Hurst com utilização da NDWT

Nesta seção, serão retratados os resultados utilizando, para a estimação dos expoentes de Hurst, a média dos coeficientes *wavelet* ao quadrado (MQCW), Mid energias, o estimador *Trimean* e os métodos MEDL e MEDLA. As subimagens foram decompostas com implementação das transformadas não decimada de *wavelet* (NDWT), em quatro níveis de resolução. Na Figura 5.17, apresenta-se uma subimagem do tecido cancerígeno em (a) e do tecido saudável em (b) decomposta com utilização da transformada de *wavelet* não decimada no nível 1 e no nível 4 de resolução. Nota-se que, com aumento da resolução, também aumenta o nível de detalhes.

Figura 5.7 – Decomposição de uma subimagem do tecido cancerígeno em (a) e do tecido saudável em (b), no nível 1 e 4 de resolução com utilização NDWT



Para cada método de estimação dos expoentes de Hurst direcionais, foram analisados os gráficos das séries temporais, as estatísticas descritivas e os gráficos das distribuições empíricas em subimagens do tecido cancerígeno e do saudável.

5.2.1 Estimação dos expoentes de Hurst com MQCW

Na Tabela 5.6, são apresentadas as estatísticas descritivas das séries de expoentes de Hurst direcionais, ao longo do tempo, utilizando a transformada não decimada de *wavelet* em subimagens do tecido cancerígeno e saudável, com uso da média dos coeficientes *wavelet* ao quadrado para sua estimação.

Pode-se destacar a existência de variação dos expoentes, ao longo do tempo, com a utilização da transformada não decimada de *wavelet*, ou seja, imagens tiradas em tempos diferentes geram expoentes distintos. Percebe-se que subimagens do tecido saudável apresentam médias maiores e variâncias menores dos expoentes de Hurst nas direções horizontal e vertical, comparado às médias de tecidos cancerígenos.

Além disso, observa-se também uma diminuição nos valores médios das séries nas três direções e também da variância, na direções horizontal e diagonal comparada ao uso da transformada decimada de *wavelet*, conseqüentemente valores mais baixos dos coeficientes de variação. Segundo Kang e Vidakovic (2017), o uso da transformada não decimada diminui a variância dos coeficientes *wavelet*, porém, aumenta a correlação entre eles e o viés no expoente de Hurst. Por fim, as medidas de assimetria e curtose mostram que as distribuições das séries seguem com características assimétricas e leptocúrticas.

Tabela 5.6 – Estatísticas descritivas dos expoentes de Hurst direcionais com utilização da NDWT com a MQCW

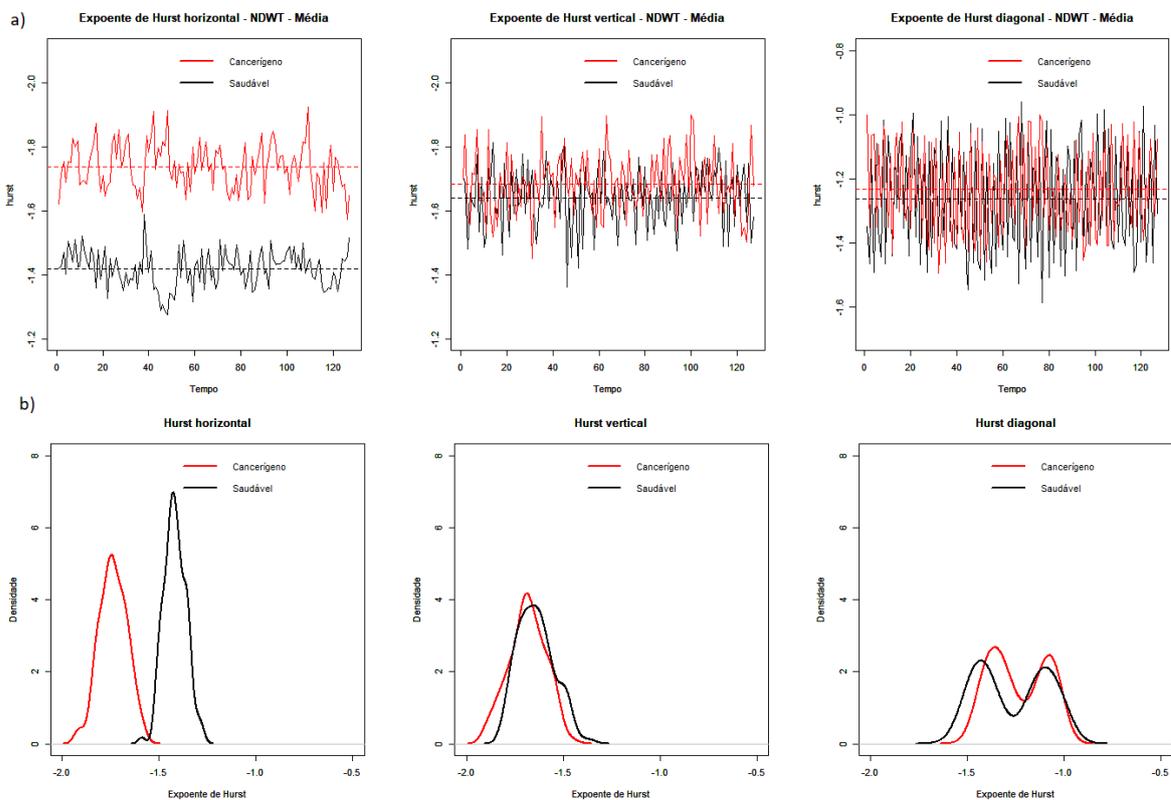
Expoentes (Tumor)	Média	Mediana	Variância	CV	Assimetria	Curtose
Horizontal	-1,7368	-1,7399	0,0050	0,0409	-0,0929	-0,3134
Vertical	-1,6842	-1,6852	0,0091	0,0567	-0,1372	-0,4822
Diagonal	-1,2344	-1,2671	0,0210	0,1174	0,0897	-1,4984
Expoentes (Saudável)	Média	Mediana	Variância	CV	Assimetria	Curtose
Horizontal	-1,4178	-1,4210	0,0031	0,0395	0,0595	-0,1200
Vertical	-1,6401	-1,6434	0,0088	0,0573	0,4158	-0,4086
Diagonal	-1,2641	-1,2915	0,0320	0,1415	0,0501	-1,5652

Na Figura 5.8b, apresentam-se os gráficos das distribuições empíricas dos expoentes de Hurst, ao longo do tempo, nas três direções com aplicação da transformada não decimada de *wavelet*. Nesses gráficos foram feitas sobreposições das distribuições dos expoentes de Hurst em subimagens do tecido cancerígeno (em vermelho) e saudável (em preto). É possível identificar uma sobreposição nas curvas principalmente nas direções vertical e diagonal, que evidencia

um comportamento semelhante dos expoentes de Hurst, nessas direções, em subimagens do tecido saudável e cancerígeno .

O comportamento semelhante nessas direções também pode ser verificado pela análise dos gráficos das séries (Figura 5.8a), em que mostra a capacidade de diferenciar subimagens do tecido cancerígeno do saudável com análises temporais dos expoentes de Hurst, já que não apresentaram comportamento semelhante em todas as direções.

Figura 5.8 – Gráficos dos expoentes de Hurst com utilização da NDWT com a MQCW: (a) Gráficos das séries ao longo do tempo e (b) Densidades



Na próxima seção, serão apresentados os resultados utilizando a transformada não decimada de *wavelet* e Mid-energia, proposto por Soltani et al. (2004), para reduzir o viés de ponderação no expoente de Hurst, que pode ocorrer pela não normalidade dos coeficientes *wavelet* (d_j) e também pela correlação existente entre eles.

5.2.2 Estimação dos expoentes de Hurst com Mid-energias

São apresentadas na Tabela 5.7, as estatísticas descritivas das séries de expoentes de Hurst direcionais, ao longo do tempo, com a utilização da transformada não decimada de *wavelet* em subimagens do tecido cancerígeno e saudável, com implementação de Mid-energias para sua estimação. Assim, como ocorrido com o uso da média dos coeficientes *wavelet* ao quadrado, persiste a variabilidade dos expoentes de Hurst em todas as direções. Percebe-se um aumento dos valores médios e também das variâncias dos expoentes de Hurst em subimagens dos tecidos cancerígenos e saudáveis em todas as direções, comparado ao uso da MQCW como estimativa.

Nota-se, também, que o uso da transformada não decimada de *wavelet* produz séries de expoentes de Hurst com variâncias menores, portanto mais robustas. Além disso, as medidas de assimetria e curtose mostram que as distribuições mantêm as características assimétricas e leptocúrticas com implementação da NDWT.

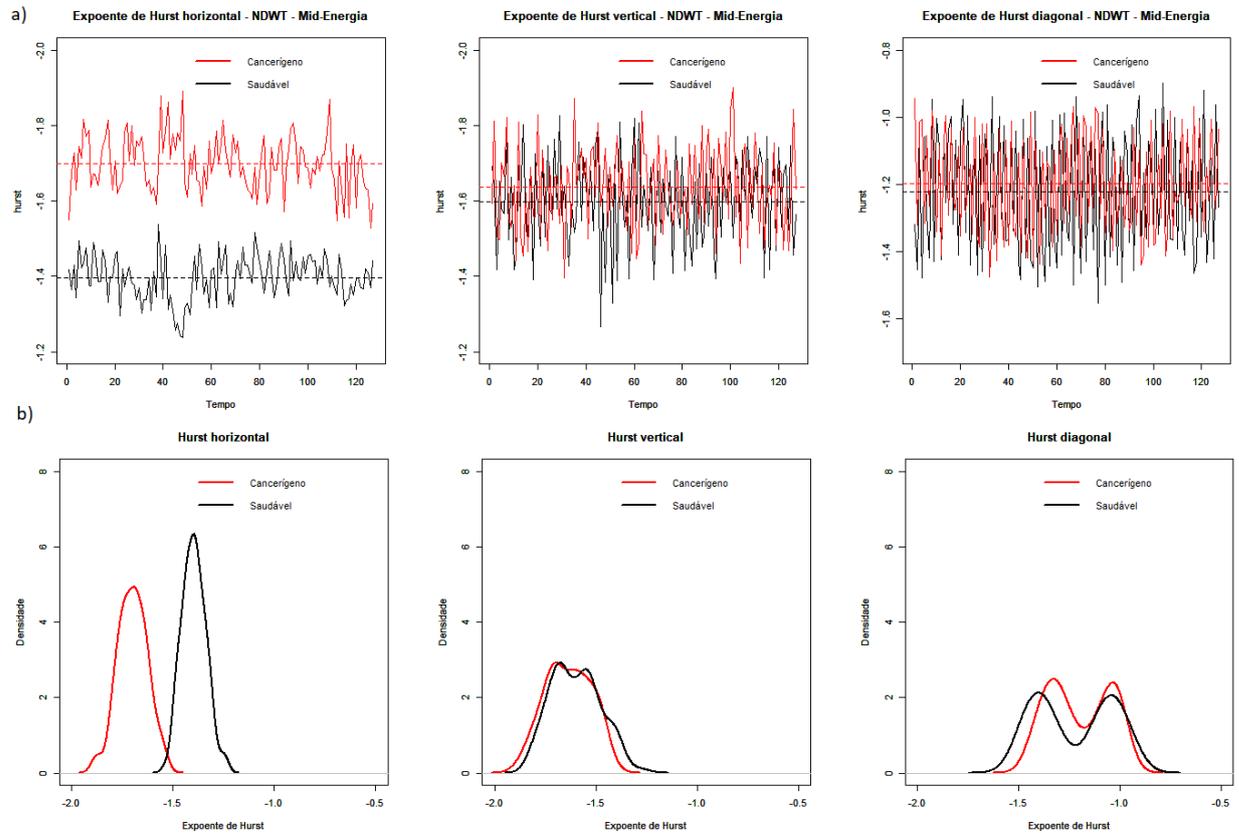
Tabela 5.7 – Estatísticas descritivas dos expoentes de Hurst direcionais com utilização da NDWT com Mid-energia

Expoentes (Tumor)	Média	Mediana	Variância	CV	Assimetria	Curtose
Horizontal	-1,6991	-1,6941	0,0053	0,0429	-0,0883	-0,2069
Vertical	-1,6352	-1,6325	0,0128	0,0691	-0,0647	-0,8648
Diagonal	-1,1991	-1,2197	0,0230	0,1265	-0,0418	-1,4938
Expoentes (Saudável)	Média	Mediana	Variância	CV	Assimetria	Curtose
Horizontal	-1,3962	-1,3957	0,0034	0,0421	0,2177	-0,2209
Vertical	-1,5984	-1,5958	0,0145	0,0753	0,2565	-0,6415
Diagonal	-1,2237	-1,2552	0,0361	0,1554	0,0286	-1,5801

Na Figura 5.9b, são apresentados os gráficos das distribuições empíricas dos expoentes de Hurst, ao longo do tempo em subimagens do tecido cancerígeno e saudável. Observa-se uma sobreposição das distribuições nas direções vertical e diagonal, reforçando a tese de que análise temporal dos expoentes de Hurst pode diferenciar imagens do tecido cancerígeno de tecido saudável, já que na direção horizontal as distribuições não se sobrepõem.

Essa informação também pode ser observada pela análise dos gráficos das séries (Figura 5.9a) que mostram médias individuais afastadas na direção horizontal e próximas nas direções vertical e diagonal. Para mais, constata-se uma variância maior e uma mistura de distribuições na direção diagonal.

Figura 5.9 – Gráficos dos expoentes de Hurst com utilização da NDWT com Mid-energia: (a) Gráficos das séries ao longo do tempo e (b) Densidades



Na próxima seção, serão apresentados os resultados utilizando a transformada não decimada de *wavelet* com implementação de método MEDL, que aplica a mediana do logaritmo dos coeficientes *wavelet* ao quadrado, proposto por Kang e Vidakovic (2017), para tentar reduzir a correlação existente entre eles.

5.2.3 Estimação dos expoentes de Hurst com método MEDL

Na Tabela 5.8, são apresentadas as estatísticas descritivas das séries de expoentes de Hurst direcionais, ao longo do tempo, com a utilização da transformada não decimada de *wavelet* em subimagens do tecido cancerígeno e com a aplicação do método MEDL para sua estimação.

Nota-se, como nos casos anteriores, uma variabilidade dos expoentes de Hurst, ao longo do tempo, para as duas regiões. Outra observação importante é em relação à variância das séries de expoentes de Hurst quando implementado o método MEDL, que se apresentam menores em subimagens dos tecidos cancerígeno e saudável do que quando se aplica a MQCW e Mid-energia para sua estimação. Além disso, percebe-se que subimagens do tecido cancerígeno produzem expoentes de Hurst menores do que subimagens do tecido saudável.

Em relação às medidas de assimetria e curtose, constata-se que as distribuições empíricas apresentam características assimétricas e leptocúrticas.

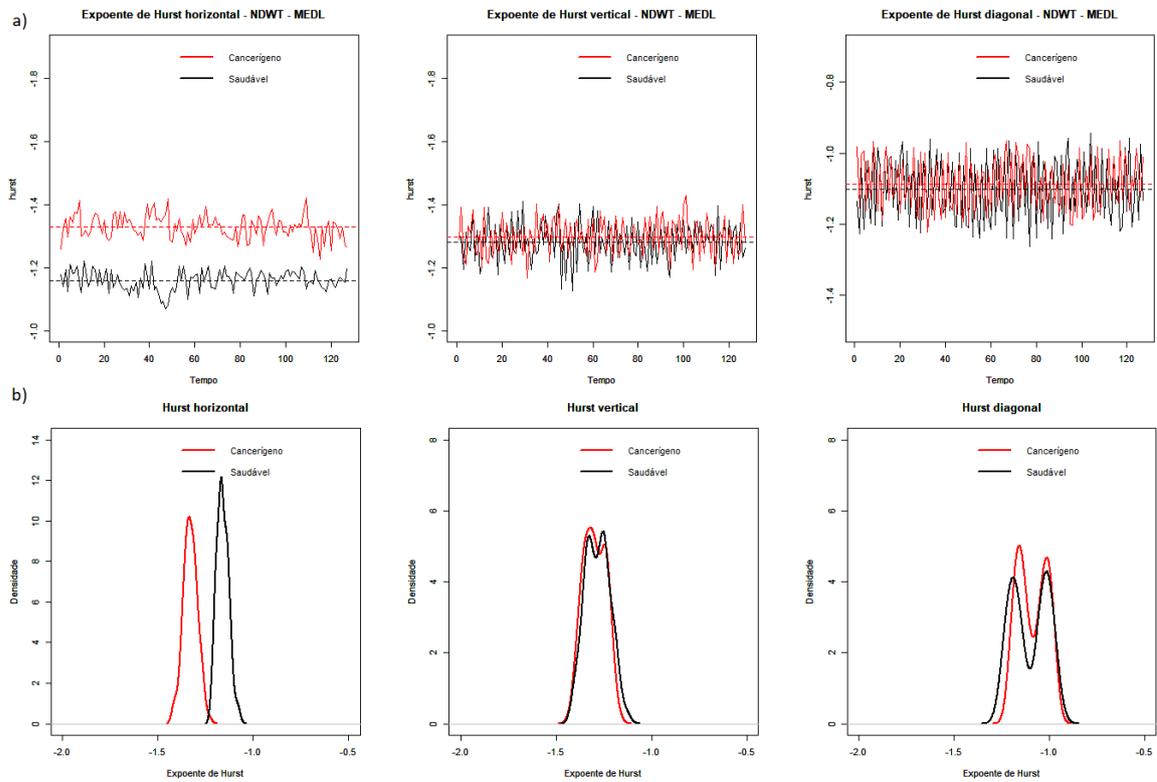
Tabela 5.8 – Estatísticas descritivas dos expoentes de Hurst direcionais com utilização da NDWT com método MEDL

Expoentes (Tumor)	Média	Mediana	Variância	CV	Assimetria	Curtose
Horizontal	-1,3286	-1,3289	0,0013	0,0279	0,0261	-1,1329
Vertical	-1,2962	-1,2981	0,0034	0,0450	0,0036	-0,9536
Diagonal	-1,0875	-1,0973	0,0056	0,0692	0,0455	-1,4958
Expoentes (Saudável)	Média	Mediana	Variância	CV	Assimetria	Curtose
Horizontal	-1,1592	-1,1631	0,0009	0,0266	0,3642	-0,2404
Vertical	-1,2809	-1,2797	0,0040	0,0494	0,1053	-0,7027
Diagonal	-1,1009	-1,1048	0,0086	0,0845	-0,0123	-1,5631

Os gráficos das distribuições empíricas dos expoentes de Hurst direcionais, com utilização do método MEDL para sua estimação, são apresentados na Figura 5.10b. Observa-se uma sobreposição das curvas nas direções vertical e diagonal, como ocorrido quando utilizou-se a média dos coeficientes *wavelet* ao quadrado e Mid-energia, o que evidencia o comportamento semelhante dos expoentes em subimagens do tecido saudável e cancerígeno nessas direções.

Persiste o afastamento entre as curvas na direção horizontal, com a implementação da transformada não decimada, indicador de que essa metodologia também pode diferenciar subimagens do tecido cancerígeno do saudável (Figura 5.10a).

Figura 5.10 – Gráficos dos expoentes de Hurst com utilização da NDWT com o método MEDL: (a) Gráficos das séries ao longo do tempo e (b) Densidades.



Na próxima seção, serão apresentados os cálculos com a utilização da transformada não decimada de *wavelet*, com implementação do método MEDLA para estimar os expoentes de Hurst. Esse método aplica a mediana do logaritmo da média de pares de coeficientes *wavelet* ao quadrado, selecionados aleatoriamente em cada nível de resolução.

5.2.4 Estimação dos expoentes de Hurst com método MEDLA

Para gerar os resultados com utilização do método MEDLA, foram estimados 100 expoentes de Hurst direcionais para cada subimagem e calculada sua média. Na Tabela 5.9, são apresentadas as estatísticas descritivas das séries de expoentes de Hurst direcionais, ao longo do tempo, com a utilização da transformada não decimada e o método MEDLA, em subimagens do tecido cancerígeno e saudável.

Nota-se, como ocorrido nos casos anteriores com utilização da transformada não decimada, que persiste a variabilidade nas três direções para as duas regiões. Comparado ao método MEDL, que também utiliza a mediana, percebe-se valores médios das séries próximos e uma diminuição da variância, além de uma aproximação maior entre os valores médios e medianos.

Em relação às medidas de assimetria e curtose, verifica-se que, como nos métodos anteriores, persiste o comportamento assimétrico e leptocúrtico das distribuições em todas as direções.

Tabela 5.9 – Estatísticas descritivas dos expoentes de Hurst direcionais com utilização da NDWT com método MEDLA

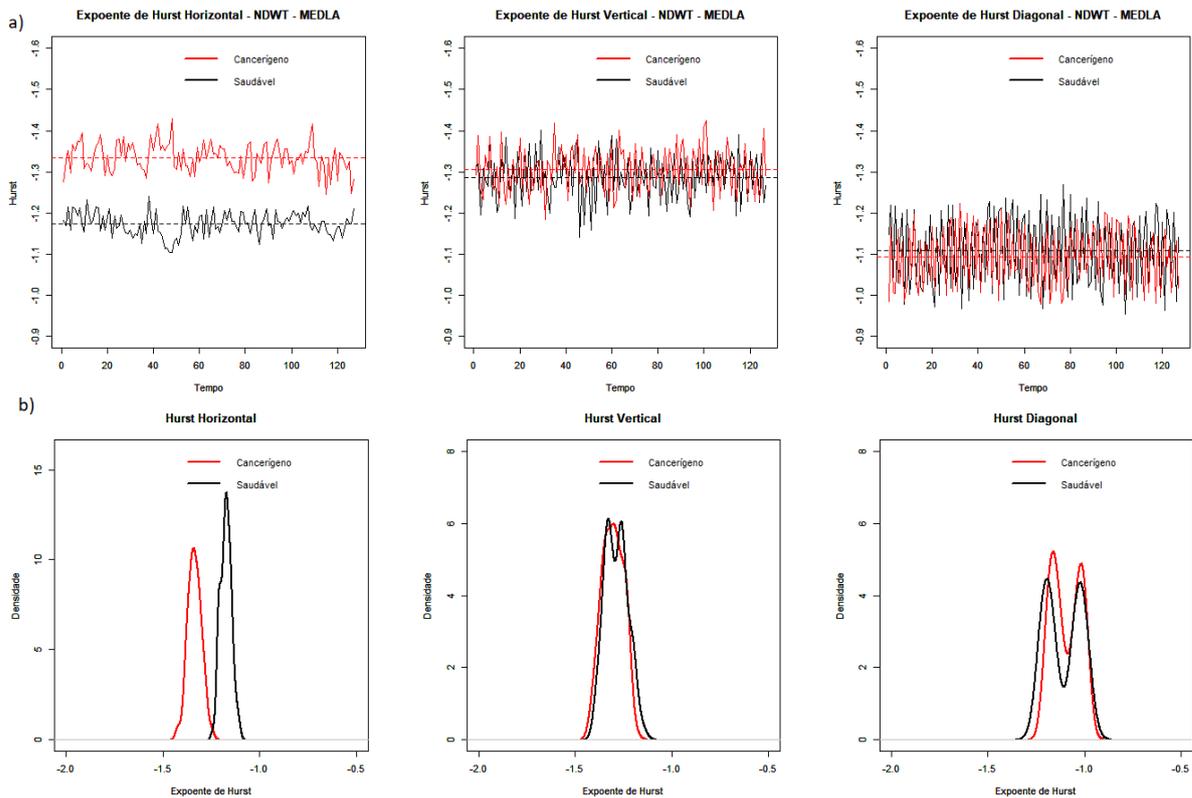
Expoentes (Tumor)	Média	Mediana	Variância	CV	Assimetria	Curtose
Horizontal	-1,3347	-1,3373	0,0012	0,0350	0,0460	-0,1506
Vertical	-1,3049	-1,3024	0,0029	0,0415	-0,0524	-0,8747
Diagonal	-1,0936	-1,1054	0,0054	0,0673	0,0565	-1,5276
Expoentes (Saudável)	Média	Mediana	Variância	CV	Assimetria	Curtose
Horizontal	-1,1728	-1,1732	0,0007	0,0237	0,1017	-0,4111
Vertical	-1,2867	-1,2898	0,0032	0,0445	0,1984	-0,7386
Diagonal	-1,1073	-1,1272	0,0083	0,0822	0,0224	-1,5907

São apresentados, na Figura 5.11b, os gráficos das distribuições empíricas dos expoentes de Hurst direcionais, em subimagens do tecido cancerígeno e saudável, com implementação do método MEDLA para sua estimação. Como verificado nos casos anteriores, com aplicação da transformada não decimada, tem-se uma sobreposição das curvas nas direções vertical e diagonal e um afastamento na direção horizontal. Esse fato corrobora a comprovação de que essa metodologia de análise de imagens interferométricas, podem diferenciar imagens do tecido cancerígeno do saudável.

Percebe-se também, com a análise dos gráficos das séries (Figura 5.11a), o afastamento das médias individuais dos expoentes de Hurst na direção horizontal, o que não acontece nas

direções vertical e diagonal. Além disso, nota-se uma mistura de distribuições na direção diagonal, independentemente da transformada de *wavelet* e do método utilizado para estimação dos expoentes de Hurst. Por fim, nota-se uma mistura de distribuições na direção diagonal.

Figura 5.11 – Gráficos dos expoentes de Hurst com utilização da NDWT com método MEDLA: (a) Gráficos das séries ao longo do tempo e (b) Densidades.



Na próxima seção, serão apresentados os resultados utilizando a transformada não decimada de *wavelet*, com a implementação do estimador *Trimean* para o cálculo dos expoentes de Hurst.

5.2.5 Estimação dos expoentes de Hurst com estimador Trimean

Para gerar os resultados, com a utilização do estimador *Trimean*, foram considerados 8 subgrupos de coeficientes *wavelet* em cada nível de resolução e calculada a média *Trimean* em cada subgrupo para estimação do expoente de Hurst. Na Tabela 5.10, são apresentadas as estatísticas descritivas das séries de expoentes de Hurst direcionais, ao longo do tempo, utilizando a transformada não decimada de *wavelet* em subimagens do tecido cancerígeno e saudável. Destaca-se, assim como ocorrido com os casos anteriores, que os expoentes apresentam variação ao longo do tempo nas duas regiões.

Observa-se uma diminuição nos valores médios e aumento das variâncias, nas três direções, comparado aos métodos que utilizam MEDL e MEDLA para estimação do expoente de Hurst, tanto na região cancerígena quanto na saudável. Também a confirmação de que subimagens do tecido saudável tendem a ter expoentes de Hurst maiores nas três direções do que subimagens de tecido cancerígeno.

As medidas de assimetria e curtose (Tabela 5.19) mostraram que as distribuições possuem características assimétrica e leptocúrtica.

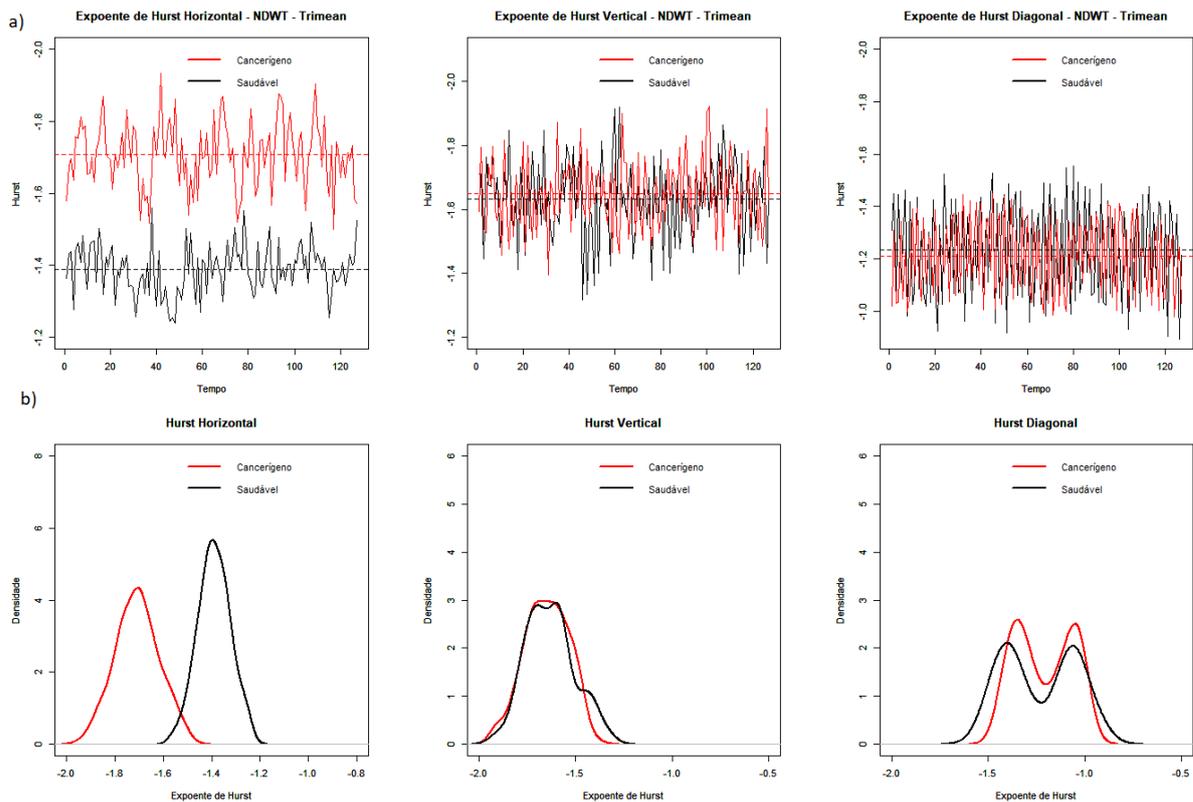
Tabela 5.10 – Estatísticas descritivas dos expoentes de Hurst direcionais com utilização da NDWT com estimador Trimean

Expoentes (Tumor)	Média	Mediana	Variância	CV	Assimetria	Curtose
Horizontal	-1,7003	-1,7059	0,0051	0,0422	-0,0623	-0,2044
Vertical	-1,6378	-1,6380	0,0123	0,0677	-0,0983	-0,8236
Diagonal	-1,2005	-1,2294	0,0231	0,1268	0,0617	-1,5110
Expoentes (Saudável)	Média	Mediana	Variância	CV	Assimetria	Curtose
Horizontal	-1,3936	-1,3982	0,0031	0,0404	0,0672	-0,1484
Vertical	-1,6007	-1,6128	0,0137	0,0732	0,3223	-0,6088
Diagonal	-1,2268	-1,2595	0,0357	0,1540	0,0419	-1,5882

Na Figura 5.12b, são apresentados os gráficos das distribuições empíricas dos expoentes de Hurst, ao longo do tempo, nas três direções, com aplicação da transformada não decimada de *wavelet*. Como mostrado nos métodos anteriores com utilização da NDWT, os expoentes de Hurst horizontais são capazes de diferenciar subimagem de tecido saudável do cancerígeno, já que houve sobreposição entre as curvas nas direções vertical e diagonal e um afastamento na direção horizontal.

Esse comportamento também foi verificado na análise das séries ao longo do tempo (Figura 5.12a) que mostram médias individuais próximas nas direções vertical e diagonal e afastadas na direção horizontal.

Figura 5.12 – Gráficos dos expoentes de Hurst com utilização da NDWT com estimador Trimean: (a) Gráficos das séries ao longo do tempo e (b) Densidades.



Com a finalidade de verificar o comportamento dos expoentes de Hurst direcionais, ao longo do tempo, em subimagens interferométricas de um carcinoma mamário anaplástico canino, utilizaram-se as transformadas decimada e não decimada de *wavelet* com a implementação da *wavelet* de Daubechies com oito momentos nulos e quatro níveis de resolução. Ademais, verificaram-se os efeitos do uso da média dos coeficientes *wavelet* ao quadrado, Mid-energia, estimador *Trimean* e dos métodos MEDL e MEDLA para estimar os expoentes de Hurst direcionais, nas séries temporais.

Constatou-se que, independentemente do método adotado e da transformada utilizada para estimação dos expoentes de Hurst, existe uma variabilidade nos expoentes tanto em subimagens do tecido cancerígeno quanto no saudável, consequência de um erro de estimação. Portanto, mesmo que sejam imagens de uma mesma região, se tiradas em tempos distintos podem gerar diferentes expoentes de Hurst.

Os métodos analisados para estimar os expoentes de Hurst foram propostos com o objetivo de aumentar sua robustez. Neste estudo, constatou-se que esses métodos contribuem para a diminuição da variância nas séries. Portanto, a utilização do método MEDLA seria o mais indicado para diferenciação entre subimagens de tecidos cancerígenos e saudável, por apresentar variabilidade setenta por cento menor (em média) nas séries de expoentes de Hurst, comparado ao uso da MQCW. Tanto para a transformada decimada quanto para a transformada não decimada, além de um afastamento maior das distribuições empíricas na direção horizontal.

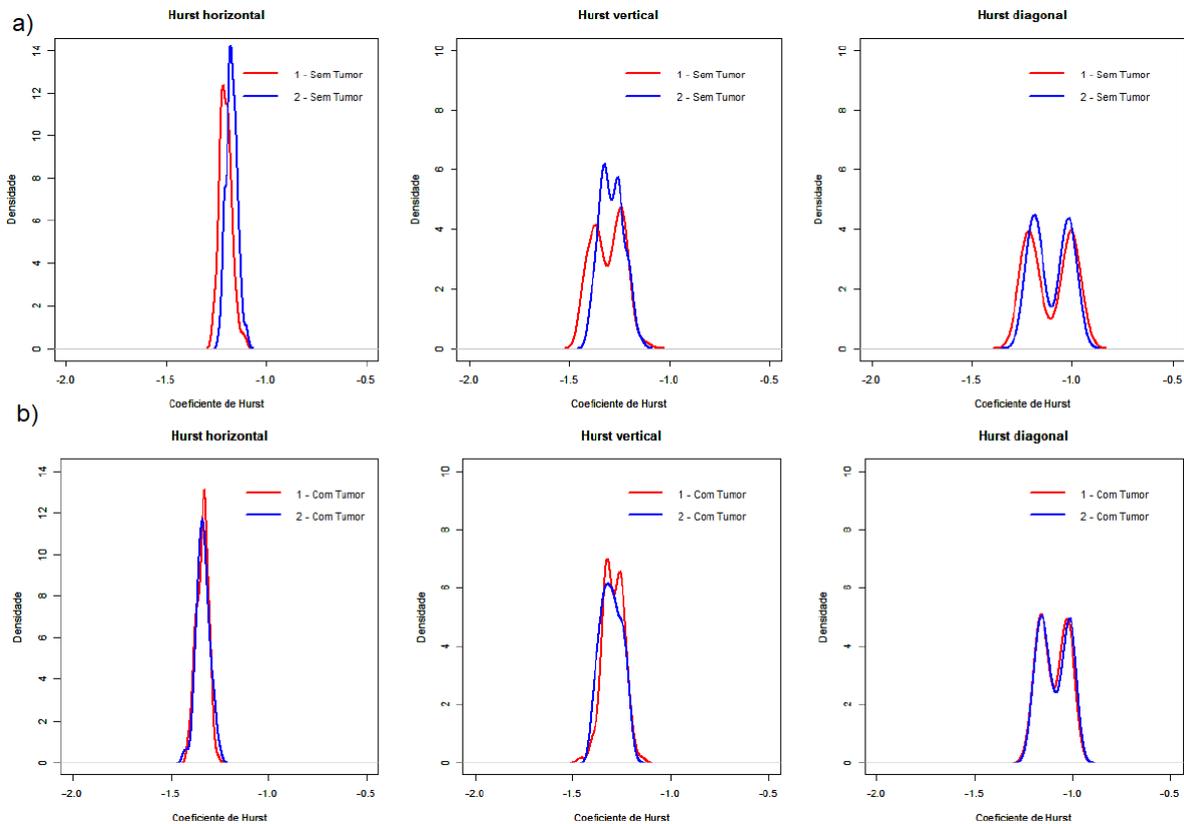
Outra importante observação a ser destacada é sobre as variâncias e os valores médios das séries de expoentes de Hurst direcionais, que tendem a ser menores nas duas regiões com implementação da transformada não decimada. Esse fato contribuiu para menores valores dos coeficientes de variação. Além disso, são pequenas as distâncias entre médias e medianas, a partir da segunda casa decimal, em todos os métodos utilizados para estimar o expoente de Hurst e também com a implementação das duas transformadas (DWT e NDWT).

Foi constatado também, com análises das distribuições empíricas, que a diferenciação entre subimagens do tecido saudável e cancerígeno, com implementação da transformada decimada, foi detectada pelo expoente de Hurst vertical. No entanto, com aplicação da transformada não decimada, a distinção é percebida pelo expoente de Hurst horizontal, independentemente do método utilizado para estimá-lo.

Nota-se, também, a presença de expoentes de Hurst negativos, tanto para subimagens do tecido cancerígeno como em saudável independentemente da transformada ou do método utilizado para estimação do expoente de Hurst, que de acordo com trabalhos de Jeon, Nicolis e Vidakovic (2014), estes podem ocorrer devido a efeitos da correlação existente entre os coeficientes *wavelet* em cada nível de resolução.

Para a validação do método, foram consideradas duas regiões saudáveis distintas e duas regiões cancerígenas distintas, com aplicação da transformada não decimada de *wavelet* e o método MEDLA, no qual foi verificado a sobreposição das distribuições empíricas nas três direções. Como pode ser observado nos gráficos das Figuras 5.13a (Regiões saudáveis) e 5.13b (Regiões cancerígenas). A análise completa é apresentada no Anexo E.

Figura 5.13 – Gráficos dos expoentes de Hurst com utilização da NDWT - MEDLA: (a) Densidade regiões saudáveis (b) Densidades regiões cancerígenas.



Na prática, a metodologia proposta para diferenciar subimagens de tecido cancerígeno do saudável com análise das séries de expoentes de Hurst, pode ser aplicada da seguinte maneira: se na imagem analisada pelo profissional existir uma região suspeita de ser tecido cancerígena, pode-se selecionar uma subimagem da região suspeita e compará-la com outra subimagem de tecido saudável. Assim, caso as séries de expoentes de Hurst se sobrepuserem em todas as direções, essa região suspeita pode ser considerada uma região saudável. Caso as séries não se sobrepuserem em, pelo menos, uma direção, suspeita-se de uma região cancerígena. Vale salientar que essa análise deve ser feita junto às análises clínicas.

Encerram-se, aqui, as análises e discussões dos resultados obtidos e, na próxima seção, será apresentada a conclusão final da presente pesquisa.

6 CONCLUSÃO

Conclui-se que, a análise temporal dos expoentes de Hurst em imagens interferométricas, obtidas pela técnica Biospeckle laser, é uma ferramenta eficaz para distinguir imagens de tecido saudável do cancerígeno. Constatou-se a existência de variabilidade dos expoentes de Hurst direcionais, ao longo do tempo, em torno de um valor médio tanto para regiões cancerígenas como para regiões saudáveis. A diferenciação entre áreas cancerígenas e saudáveis foi detectada pela direção vertical com utilização da DWT, enquanto com a utilização da NDWT foi a direção horizontal. De acordo com os resultados obtidos, o melhor método de estimação do expoente de Hurst para esses dados foi o método MEDLA com implementação da transformada não decimada de wavelet, por produzir menor variância nas séries de expoentes de Hurst além de uma diferenciação maior entre imagens do tecido cancerígeno e saudável.

REFERÊNCIAS

- ABRY, P. et al. Wavelet for the analysis, estimation and synthesis of scaling data. **Self-Similar Network Traffic and Performance Evaluation**, [S.l.], p. 39–88, 2002.
- ALTEKRUSE, S. et al. Seer cancer statistics review: 1975-2014. **National Cancer Institute**, 2010. Disponível em: <https://seer.cancer.gov/archive/csr/1975_2014/>. Acesso em: 10 out. 2018.
- ARIÑO, M. A.; MORETTIN, P. A.; VIDA KOVIC, B. On wavelet scalograms and their application in economic time series. **Brazilian Journal of Probability and Statistics**, São Paulo, v. 18, p. 37–51, 2004.
- BRAGA, R. et al. Biospeckle numerical values over spectral image maps of activity. **Optics Communications**, [S.l.], v. 285, p. 553–561, 2012.
- DAUBECHIES, I. Orthonormal basis of compactly supported wavelets. **Communications on pure and applied mathematics**, [S.l.], v. 41, n. 7, p. 909–996, 1989.
- DONOH O D.L.; JOHNSTONE, I. M. Minimax estimation via wavelet shrinkage. **The Annals of Statistics**, Stanford, v. 26, n. 3, p. 879–921, 1998.
- FEDER, J. **Fractals**. 4. ed. New York: Springer Science & Business Media, 1989. 243 p.
- FENG C.; VIDA KOVIC, B. Estimation of the Hurst exponent using trimean estimators on nondecimated wavelet coefficients. **Cornell University Library**, [S.l.], v. 14, p. 738–751, 2017.
- FRANCIS I.; KIM, S. **An introduction to Wavelet Theory in Finance: A Wavelet Multiscale Approach**. Singapore: World Scientific Publishing Co.Pte.Ltd, 2012.
- GABOR, D. Theory of communications. **Journal of the Institution of Electrical Engineers**, [S.l.], v. 94, p. 429–457, 1946.
- GROSSMANN A.; MOLET, J. Decomposition of hardy functions into square integrable wavelets of constant shape. **SIAM Journal on Mathematical Analysis**, [S.l.], v. 15, p. 723–736, 1984.
- HAAR, A. Zur theorie der orthogonalen funktionensysteme. **Mathematische Annalen**, [S.l.], v. 69, n. 3, p. 331–371, 1910. Disponível em: <<https://doi.org/10.1007/BF01456927>>. Acesso em: 15 abr. 2018.
- HAROLD, H. E.; BLACK, R. P.; SIMAIKA, Y. M. **Long-term storage: an experimental study**. London: Constable, 1965. Disponível em: <<http://books.google.com/books?id=X-7cAAAAIAAJ>>. Acesso em: 15 abr. 2018.
- INCA. Todos contra os erros. **Instituto Nacional do Câncer**, Rio de Janeiro, 2013. Disponível em: <<http://www2.inca.gov.br>>. Acesso em: 05 mar. 2018.
- JEFFERIS, G. **bmp: Read Windows Bitmap (BMP) Images**. [S.l.], 2017. R package version 0.3. Disponível em: <<https://CRAN.R-project.org/package=bmp>>. Acesso em: 10 jun. 2018.
- JEON, S.; NICOLIS, O.; VIDA KOVIC, B. Mammogram diagnostics via 2-d complex wavelet-based self-similarity measures. **São Paulo Journal of Mathematical Sciences**, São Paulo, v. 8, p. 265–284, 2014.

- KAMPSTRA, P. Beanplot: A boxplot alternative for visual comparison of distributions. **Journal of Statistical Software, Code Snippets**, [S.l.], v. 28, n. 1, p. 1–9, 2008. Disponível em: <<http://www.jstatsoft.org/v28/c01/>>. Acesso em: 15 jun. 2018.
- KANG, M.; VIDAKOVIC, B. Medl and medla: Methods for assessment of scaling by medians of log-squared nondecimated wavelet coefficients. **arXiv**, [S.l.], 2017. Disponível em: <<https://arxiv.org/pdf/1703.04180.pdf>>. Acesso em: 03 set. 2018.
- KOZAKEVICIUS A. J.; BAYER, F. M. Filtragem de sinais via limiarização de coeficientes wavelet. **Centro de Ciências Naturais e Exatas-UFSM**, Santa Maria, v. 36, p. 37–51, 2014.
- LIMA, P. C. Wavelets: Uma introdução. **Revista Universitária**, Belo Horizonte, v. 33, p. 13–44, 2002.
- MALLAT, S. A theory for multiresolution signal decomposition: the wavelet representation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, [S.l.], v. 11, n. 7, p. 674–693, 1989.
- MALLAT, S. **A Wavelet tour of signal processing: The Sparse Way**. 3. ed. Burlington: Elsevier, 2009. 805 p.
- MANDELBROT B.B.; VAN NESS, J. Fractional brownian motion, fractional noises and applications. **Society for Industrial and Applied Mathematics**, [S.l.], v. 10, n. 4, p. 422–437, 1968.
- MARTIN, J.; MOSKOWITZ, M.; MILBRATH, J. W. Breast cancer missed by mammography. **American Journal of Roentgenology**, [S.l.], v. 132, n. 5, p. 737–739, 1979.
- MEYER, Y. **Ondelettes et opérateurs**. Paris: Hermann, 1990. 264 p.
- MORETTIN, P. A. **Ondas e Ondaletas: Da Análise de Fourier à Análise de Ondaletas de Séries Temporais**. 1. ed. São Paulo: Editora Universitária de São Paulo, 2014. 320 p.
- NASON, G. **Wavelet Methods in Statistics with R**. New York: Springer Science, 2008. 268 p.
- NASON, G. **wavethresh: Wavelets Statistics and Transforms**. [S.l.], 2016. R package version 4.6.8. Disponível em: <<https://CRAN.R-project.org/package=wavethresh>>. Acesso em: 10 jun. 2018.
- NICOLIS, O.; RAMÍREZ-COBO, P.; VIDAKOVIC, B. 2d wavelet-based spectral with applications. **Computational Statistics and Data Analysis**, [S.l.], v. 55, p. 738–751, 2011.
- OLIVEIRA, H. M. **Análise de sinais para Engenheiros: Uma abordagem via Wavelets**. Rio de Janeiro: Brasport, 2007(a). 244 p.
- OLIVEIRA, H. M. **Análise de Fourier e Wavelets: Sinais Estacionários e Não Estacionários**. Recife: Editora Universitária UFPE, 2007(b). 342 p.
- PERCIVAL D. B.; CONSTANTINE, W. Wavelet-based simulation of stochastic processes. **Technical Report, University of Washington**, Seattle, 2000.
- PISKAREV, D. Cálculo do coeficiente de hurst. **Metatrader 5**, [S.l.], 2017. Disponível em : <<https://www.mql5.com/pt/articles/2930>>. Acesso em: 15 jun. 2018.

- RAMÍREZ-COBO, P. et al. A wavelet-based spectral method for extracting self-similarity measures in time-varying two-dimensional rainfall maps. **Journal of Time Series Analysis**, [S.l.], v. 32, n. 4, p. 351–363, 2011.
- RAMÍREZ-COBO, P.; VIDAKOVIC, B. A 2d wavelet-based multiscale approach with applications to the analysis of digital mammograms. **Computational Statistics and Data Analysis**, [S.l.], v. 58, p. 71–81, 2013.
- SHANNON, C. E. A mathematical theory of communication. **The Bell System Technical Journal**, Massachusetts, v. 27, n. 3, p. 379–423, 1948.
- SOLTANI, S.; SIMARD, P.; BOICHU, D. Estimation of the self-similarity parameter using wavelet transform. **Signal Processing**, [S.l.], v. 84, n. 1, p. 117–123, 2004.
- WALNUT, D. F. **An Introduction Wavelets Analysis**. Boston: Springer Science & Business Media, 2004. 445 p.
- WANG, G.; SHIYING, Z.; ROBERTSON, D. X-ray ct metal artifact reduction using wavelets: An application for imaging total hip prostheses. **IEEE Transactions on Medical Imaging**, [S.l.], v. 19, n. 12, p. 1238–1247, 2000.
- WHITCHER, B. **waveslim: Basic wavelet routines for one, two and three-dimensional signal processing**. [S.l.], 2015. R package version 1.7.5. Disponível em: <<https://CRAN.R-project.org/package=waveslim>>. Acesso em: 10 jun. 2018.
- WUERTZ, D.; SETZ, T.; CHALABI, Y. **fBasics: Rmetrics - Markets and Basic Statistics**. [S.l.], 2017. R package version 3042.89. Disponível em: <<https://CRAN.R-project.org/package=fBasics>>. Acesso em: 10 jun. 2018.

APÊNDICE A – Mid-energias

Abordagem usando a distribuição Gumbel e o teorema do limite central

A abordagem desenvolvida combina os coeficientes wavelet de tal forma que as variáveis aleatórias $\log_2 D_{j,k}$ são assintoticamente normais. A demonstração dessa afirmação foi retirada do artigo “*Estimation of the Self-similarity Parameter Using the Wavelet Transform*”, Soltani et al. (2004)

Demonstração:

Um conjunto de coeficientes wavelets em uma NDWT de um fBm é normalmente distribuído, com variância dependendo do nível j , ou seja:

$$d_j \sim N(0, 2^{-(2H+1)j} \sigma_2).$$

Seja $D_{j,k}$ uma nova variável aleatória resultado da média de um par de quadrados de coeficientes wavelet da seguinte forma:

$$D_{j,k} = \frac{d_{j,k}^2 + d_{j,k+\frac{N_j}{2}}^2}{2},$$

$$k = 0, 1, 2, 3, \dots, \frac{N_j}{2} - 1$$

Essa combinação garante a quase decorrelação dos coeficientes wavelet, para valores suficientemente altos de j . A nova variável pode ser escrita da seguinte forma:

$$D_{j,k} = \frac{\sigma_j^2}{2} \left[\left(\frac{d_{j,k}}{\sigma_j} \right)^2 + \left(\frac{d_{j,k+\frac{N_j}{2}}}{\sigma_j} \right)^2 \right],$$

pode-se verificar que $E(D_{j,k}) = \sigma_j^2 = C 2^{-(2H+1)j}$.

São assumidos que $\frac{d_{j,k}}{\sigma_j}$ e $\frac{d_{j,k+N_j/2}}{\sigma_j}$ seguem uma distribuição normal com média zero e variância unitária. Portanto a variável:

$$y_{j,k} = \left(\frac{d_{j,k}}{\sigma_j} \right)^2 + \left(\frac{d_{j,k+N_j/2}}{\sigma_j} \right)^2,$$

segue uma distribuição qui-quadrado com dois graus de liberdade (χ_2^2). Os índices j e k podem ser removidos, uma vez que os coeficientes wavelet normalizados tem a mesma distribuição.

Portanto, a densidade de probabilidade de $y = y_{j,k}$ é dada por:

$$f(y) = \frac{1}{2} e^{-\frac{y}{2}}.$$

Tomando $-\log_2$ de $D_{j,k}$, a nova variável $D'_{j,k}$ é definida da seguinte forma:

$$D'_{j,k} = -\log_2 D_{j,k} = (2H + 1)j + \log_2 C + 1 - \log_2 y.$$

Vamos agora estudar o comportamento da variável aleatória $z = -\log_2 y$. Sua densidade de probabilidade e função distribuição acumulada são dadas por:

$$f_Z(z) = \log 2 e^{(z+1)\log 2} e^{-e^{(z+1)\log 2}}$$

e

$$F_Z(z) = 1 - F_Y(e^{-z \log 2}) = \frac{1}{2} e^{-\frac{1}{2} e^{-z \log 2}}.$$

Corresponde a uma função Gumbel com parâmetros $\alpha = -1$ e $\beta = \frac{1}{\log 2}$ com média e variâncias dadas por:

$$E(z) = -1 + \frac{\gamma}{\log 2} \quad \text{Var}(z) = \frac{\pi^2}{6 \log^2 2}$$

Nota

Uma variável aleatória Z tem uma distribuição Gumbel com parâmetros α e β quando a distribuição de probabilidade é dada por:

$$F_Z(z; \alpha, \beta) = e^{-e^{-\frac{(z-\alpha)}{\beta}}}$$

logo,

$$E(z) = \alpha + \frac{\gamma}{\beta} \quad \text{e} \quad \text{Var}(z) = \frac{\pi^2 \beta^2}{6}, \quad \gamma = 0,577$$

Depois de adicionar a parte determinística $(2H + 1)j - \log_2 C + 1$, concluímos que $D'_{j,k}$ segue uma distribuição Gumbel com:

$$E(D'_{j,k}) = (2H + 1)j - \log_2 C + \frac{\gamma}{\log 2} \quad \text{e} \quad \text{Var}(D'_{j,k}) = \frac{\pi^2}{6 \log^2 2}$$

Note que a média é linearmente dependente de H . A estimação usando o método de máxima verossimilhança não pode ser realizado diretamente usando os coeficientes $D'_{j,k}$, por causa da natureza complexa da distribuição Gumbel. Para simplificar o problema, uma nova variável é definida:

$$\begin{aligned}\Delta_j &= \frac{1}{\lfloor N_j/2 \rfloor} \sum_{k=0}^{\lfloor N_j/2 \rfloor - 1} D'_{j,k} = (2H+1)j - \log_2 C + 1 + \frac{1}{\lfloor N_j/2 \rfloor} \sum_{k=0}^{\lfloor N_j/2 \rfloor - 1} z_k = \\ &= (2H+1)j - \log_2 C + 1 - 1 \frac{\gamma}{\log 2} + \varepsilon_j\end{aligned}$$

o índice de translação foi introduzido em z_k para mostrar que $N_j/2$ realizações em z .

O Teorema do limite central implica que ε_j tem distribuição normal para valores suficientemente grandes para j com média zero e variância $\frac{\pi^2}{6 \log^2 2}$. Sendo assim, finalmente temos que:

$$\Delta_j \sim N\left((2H+1)j - \log_2 C + \frac{\gamma}{\log 2}, \frac{\pi^2}{6 \log^2 2} \frac{1}{N_j/2}\right)$$

Portanto para estimar o expoente de Hurst \hat{H} usamos a seguinte relação:

$$\overline{-\log_2 D_{j,k}} = (2H+1)j - \log_2 C + \frac{\gamma}{\log 2}$$

Sendo $\hat{\beta}$ a inclinação da reta resultado da regressão linear entre os pares $(j, \overline{-\log_2 D_{j,k}})$, então o expoente de Hurst pode ser estimado como:

$$\hat{\beta} = -(2H+1)$$

$$\hat{H} = -\frac{\hat{\beta}}{2} - \frac{1}{2}$$

APÊNDICE B – Teorema 3.1

A demonstração desse teorema foi retirado do artigo “*MEDL and MEDLA: Methods for Assessment of Scaling by Medians of Log-Squared Nondecimated Wavelet Coefficients*”, Kang e Vidakovic (2017).

Seja y_j^* a mediana do logarítimo de d_j^2 , onde d_j é um arbitrário coeficiente wavelet no nível j em um transformação wavelet de um fBm, com expoente de Hurst H . Então a mediana da população é dada por:

$$y_j^* = -\ln 2(2H + 1)j + k ,$$

onde k é uma constante independente de j . O expoente de Hurst pode ser estimado como:

$$\hat{H} = -\frac{\hat{\beta}}{2\ln 2} - \frac{1}{2} ,$$

onde $\hat{\beta}$ é a inclinação da reta resultado da regressão linear entre os pares (j, y_j^*)

Demonstração:

Um conjunto de coeficientes wavelets em uma NDWT de um fBm é normalmente distribuído, com variância dependendo do nível j , ou seja:

$$d_j \sim N(0, 2^{-(2H+1)j} \sigma_2).$$

Sua energia d_j^2 segue uma distribuição χ^2 com um grau de liberdade. Definimos δ como:

$$\delta = \frac{(d_j - 0)^2}{var} = \frac{d_j^2}{2^{-(2H+1)j} \sigma_2^2} = \frac{2^{(2H+1)j}}{\sigma_2^2} d_j^2 \cong \chi_1^2 ,$$

com função densidade dada por:

$$f(\sigma) = \frac{1}{\sqrt{2\sigma}\Gamma(\frac{1}{2})} e^{-\frac{\sigma}{2}} .$$

Sendo assim, a função densidade de probabilidade de d_j^2 será:

$$f(d_j^2) = \frac{e^{-\frac{d_j^2}{2} c_j}}{\sqrt{2c_j d_j^2} \Gamma(\frac{1}{2})} c_j ,$$

onde $c_j = \frac{2^{(2H+1)j}}{\sigma_2^2}$

Agora seja $y = \ln d_j^2$, logo $d_j^2 = e^y$ e $\frac{\partial d_j^2}{\partial y} = e^y$, então a função densidade de probabilidade de y será:

$$f(y) = \frac{e^{-\frac{e^y c_j}{2}}}{\sqrt{2c_j e^y} \Gamma(\frac{1}{2})} e^y = \frac{\sqrt{c_j} e^{-\frac{e^y c_j}{2}}}{\sqrt{2} \Gamma(\frac{1}{2})} e^{\frac{y}{2}} = \sqrt{\frac{c_j}{2\pi}} e^{-\frac{c_j e^y}{2}} e^{\frac{y}{2}} = \sqrt{\frac{c_j}{2\pi}} e^{-\frac{c_j e^y + y}{2}},$$

com isso, a função distribuição de y será:

$$F(y) = \int_{-\infty}^y f(t) dt = 2\Phi(\sqrt{c_j} e^{\frac{y}{2}}) - 1,$$

onde Φ é a função distribuição cumulativa da normal padrão. Agora seja y_j^* a mediana da distribuição de y_j no nível j . Esse resultado é obtido fazendo $F(y_j^*) = \frac{1}{2}$, ou seja:

$$2\Phi(\sqrt{c_j} e^{\frac{y_j^*}{2}}) - 1 = \frac{1}{2},$$

logo

$$y_j^* = 2 \ln \left(\frac{\Phi^{-1}(\frac{3}{4})}{\sqrt{c_j}} \right).$$

Desta equação pode-se encontrar uma ligação entre y_j^* e o expoente de Hurst H , substituindo $c_j = \frac{2^{(2H+1)j}}{\sigma^2}$, temos:

$$\begin{aligned} y_j^* &= 2 \ln [\Phi^{-1}(\frac{3}{4})] - \ln c_j = -\ln \frac{2^{(2H+1)j}}{\sigma^2} + 2 \ln [\Phi^{-1}(\frac{3}{4})] = -(2H+1)j \ln 2 + \ln \sigma^2 + 2 \ln [\Phi^{-1}(\frac{3}{4})] \\ &= -\ln 2(2H+1)j + k, \end{aligned}$$

com $k = \ln \sigma^2 + 2 \ln [\Phi^{-1}(\frac{3}{4})]$

Sendo $\hat{\beta}$ a inclinação da reta resultado da regressão linear entre os pares (j, y_j^*) , então o expoente de Hurst pode ser estimado como:

$$\hat{\beta} = -\ln 2(2H+1)$$

$$\hat{H} = -\frac{\hat{\beta}}{2 \ln 2} - \frac{1}{2}$$

APÊNDICE C – Teorema 3.2

A demonstração desse teorema foi retirado do artigo “*MEDL and MEDLA: Methods for Assessment of Scaling by Medians of Log-Squared Nondecimated Wavelet Coefficients*”, Kang e Vidakovic (2017).

Sejam d_{j,k_1} e d_{j,k_2} dois coeficientes wavelet no nível j , em posições k_1 e k_2 respectivamente, de uma NDWT de um fBm com expoente de Hurst H . Suponha que $|k_1 - k_2| > q_j$, onde q_j é a distância mínima de separação que depende do nível j e da base wavelet selecionada. Seja y_j^* a mediana do logaritmo de $[\frac{d_{j,k_1}^2 + d_{j,k_2}^2}{2}]$. Então a mediana da população é dada por:

$$y_j^* = -\ln 2 (2H + 1) j + k ,$$

onde k é uma constante independente de j , logo expoente de Hurst pode ser estimado como:

$$\hat{H} = -\frac{\hat{\beta}}{2\ln 2} - \frac{1}{2} ,$$

onde $\hat{\beta}$ é a inclinação da reta resultado da regressão linear entre os pares (j, y_j^*)

Demonstração:

Começamos selecionando o par de coeficientes wavelet que seguem uma distribuição normal com média zero e um variância populacional dependente do nível j , a partir do qual os coeficientes wavelet são amostrados.

$$d_{j,k_1}, d_{j,k_2} \sim N(0, 2^{-(2H+1)j} \sigma^2) ,$$

onde σ é o desvio padrão dos coeficientes wavelet no nível 0, k_1 e k_2 são posições dos coeficientes wavelet no nível j e H é o expoente de Hurst.

Também assumimos que os coeficientes d_{j,k_1} e d_{j,k_2} são independentes, o que é uma razoável suposição quando $|k_1 - k_2| > q_j = 2^{J-j}$.

Definimos δ como:

$$\delta = \frac{(d_{j,k_1} - 0)^2}{var} + \frac{(d_{j,k_2} - 0)^2}{var} = \frac{(d_{j,k_1}^2 + d_{j,k_2}^2)}{2^{-(2H+1)j} \sigma^2} = \frac{2^{(2H+1)j}}{\sigma^2} (d_{j,k_1}^2 + d_{j,k_2}^2) = C_j a ,$$

onde:

$$C_j = \frac{2 \cdot 2^{(2H+1)j}}{\sigma^2} \quad e \quad a = \frac{d_{j,k_1}^2 + d_{j,k_2}^2}{2} .$$

δ segue uma distribuição χ_2^2 , a função densidade probabilidade da média de dois coeficientes wavelet ao quadrado (a) será:

$$f(a) = \frac{C_j}{2} e^{-\frac{C_j a}{2}} .$$

Denotando $y = \ln a$, então $a = e^y$ e $\frac{\partial a}{\partial y} = e^y$. As funções densidade de probabilidade e distribuição acumulada são dada por:

$$f(y) = \frac{C_j}{2} e^{-\frac{C_j e^y}{2}} e^y$$

$$F(y) = 1 - e^{-\frac{C_j e^y}{2}} .$$

Seja y_j^* a mediana de y_j no nível j . A expressão para y_j^* é obtida fazendo $F(y_j^*) = \frac{1}{2}$, sendo assim:

$$1 - e^{-\frac{C_j e^{y_j^*}}{2}} = \frac{1}{2}$$

$$-\frac{C_j e^{y_j^*}}{2} = -\ln 2$$

$$C_j e^{y_j^*} = 2 \ln 2$$

$$\ln C_j + y_j^* = \ln (\ln 4)$$

$$y_j^* = \ln (\ln 4) - \ln C_j ,$$

substituindo $C_j = \frac{2 \cdot 2^{(2H+1)j}}{\sigma^2}$, a mediana torna-se:

$$y_j^* = \ln (\ln 4) - \ln \frac{2}{\sigma^2} - (2H + 1)j \ln 2 ,$$

ou seja,

$$y_j^* = -\ln 2(2H + 1)j + k,$$

onde $k = \ln(\ln 4) - \ln \frac{2}{\sigma^2}$

Sendo $\hat{\beta}$ a inclinação da reta resultado da regressão linear entre os pares (j, y_j^*) , então o expoente de Hurst pode ser estimado como:

$$\hat{\beta} = -\ln 2(2H + 1)$$

$$\hat{H} = -\frac{\hat{\beta}}{2 \ln 2} - \frac{1}{2}.$$

APÊNDICE D – Teorema 3.3

A demonstração desse teorema foi retirada do artigo “ *Estimation of the Hurst Exponent Using Trimean Estimators on Nondecimated Wavelet Coefficients*”, Feng e Vidakovic (2017).

Seja $\hat{\mu}_{j,i}$ o estimador trimean geral baseado em

$$\{D_{j,i}, D_{j,i+M}, D_{j,i+2M}, \dots, D_{j,N/2-M+i}\} = D(i, j),$$

onde $D(j, i)$ para $1 \leq i \leq M$ e $1 \leq j \leq J$ o i -ésimo subgrupo de Mid-energias no nível j em uma NDWT de um fBm de tamanho N com expoente de Hurst H . Então a distribuição assintótica de $\hat{\mu}_{j,i}$ é normal, ou seja:

$$\hat{\mu}_{j,i} \sim N(c(\alpha, p)\lambda_j, \frac{2M}{N}f(\alpha, p)\lambda_j^2),$$

onde

$$c(\alpha, p) = \frac{\alpha}{2} \ln \left(\frac{1}{p(1-p)} \right) + (1-\alpha) \ln 2,$$

$$f(\alpha, p) = \frac{\alpha(1-2p)(\alpha-4p)}{4p(1-p)} + 1 \quad e \quad \lambda_j = \sigma^2 \cdot 2^{-(2H+1)j}$$

em que σ é o desvio padrão dos coeficientes wavelet no nível zero. O expoente de Hurst pode ser estimado:

$$\hat{H} = \frac{1}{M} \sum_{i=1}^M \hat{H}_i = \frac{1}{M} \sum_{i=1}^M \left(-\frac{\hat{\beta}_i}{2} - \frac{d}{2} \right) = -\frac{\bar{\beta}}{2} - \frac{d}{2}$$

onde $\bar{\beta} = \frac{1}{M} \sum_{i=1}^M \hat{\beta}_i$ é a média das inclinações da regressão linear dos pares $(j, \log_2(\hat{\mu}_{i,j}))$ nos grupos $i = 1, 2, 3, \dots, M$ e d relacionado a dimensão dos dados.

Demonstração:

Um coeficiente wavelet em uma NDWT de um fBm de tamanho N com expoente de Hurst H é normalmente distribuído com variância dependente do nível j . Portanto, cada $d_{j,k}$ e $d_{j,k+N/2}$ em Mid-energia $D_{j,k}$ são assumidos independentes e seguem uma distribuição normal, ou seja:

$$d_{j,k}, d_{j,k+N/2} \sim N(0, 2^{-(2H+1)j} \sigma^2),$$

então Mid-energia é definida como:

$$D_{j,k} = \frac{d_{j,k}^2 + d_{j,k+N/2}}{2}, \quad j = 1, 2, 3, \dots, J \text{ e } k = 1, 2, 3, \dots, N/2,$$

pode ser verificado que $D_{j,k}$ tem distribuição exponencial com parâmetro $\lambda_j = \sigma^2 \cdot 2^{-(2H+1)j}$, isto é:

$$f(D_{j,k}) = \frac{1}{\lambda_j} e^{-\frac{1}{\lambda_j} D_{j,k}}, \quad \forall k = 1, 2, 3, \dots, N/2.$$

Portanto, o i -ésimo subgrupo $\{D_{j,i}, D_{j,i+M}, D_{j,i+2M}, \dots, D_{j,N/2-M+i}\}$ são iid. Ao aplicar o estimador trimean em $\{D_{j,i}, D_{j,i+M}, D_{j,i+2M}, \dots, D_{j,N/2-M+i}\}$, temos:

$$\varepsilon = \left[\ln \left(\frac{1}{1-p} \right) \lambda_j \quad \lambda_j \ln 2 \quad \ln \left(\frac{1}{p} \right) \lambda_j \right]^T,$$

e

$$\Sigma = \begin{bmatrix} \frac{p}{(1-p)} & \frac{p}{(1-p)} & \frac{p}{(1-p)} \\ \frac{p}{(1-p)} & 1 & 1 \\ \frac{p}{(1-p)} & 1 & \frac{(1-p)}{p} \end{bmatrix} \frac{1}{\lambda_j^2},$$

portanto, a distribuição assintótica de $\hat{\mu}_{j,i}$ é normal com:

$$E(\hat{\mu}_{j,i}) = A.x = \left(\frac{\alpha}{2} \ln \left(\frac{1}{p(1-p)} \right) + (1 - \alpha \ln 2) \right) \lambda_j = c(\alpha, p) \lambda_j$$

e variância

$$Var(\hat{\mu}_{j,i}) = \frac{2M}{N} A \Sigma A^T = \frac{2M}{N} \left(\frac{\alpha(1-2p)(\alpha-4p)}{4p(1-p)} + 1 \right) \lambda_j^2 = \frac{2M}{N} f(\alpha, p) \lambda_j^2.$$

APÊNDICE E – Regiões similares - NDWT - MEDLA

Para reafirmar o que foi dito por Jeon, Nicolis e Vidakovic (2014), de que estimativas robustas dos expoentes de Hurst podem diferenciar tecido saudável de tecido cancerígeno, foram selecionadas duas regiões do tecido cancerígeno e duas regiões do tecido saudável de tamanho 64 x 64 pixel, estimados e calculada a média de 100 expoentes de Hurst direcionais em cada subimagem com utilização do método MEDLA e da transformada não decimada de wavelet.

Na Tabela 1, estão as estatísticas descritivas das regiões sem carcinoma (Região 1 e Região 2). Pode-se observar que as duas regiões saudáveis apresentaram resultados próximos em todas as direções. Além disso, baixa variância como constatado anteriormente e uma aproximação entre média e mediana em todas as direções. As medidas de assimetria e curtose reforçam as características assimétricas e leptocúrticas das distribuições.

Tabela 1 – Estatísticas descritivas dos expoentes de Hurst direcionais (Regiões sem carcinoma - NDWT - MEDLA)

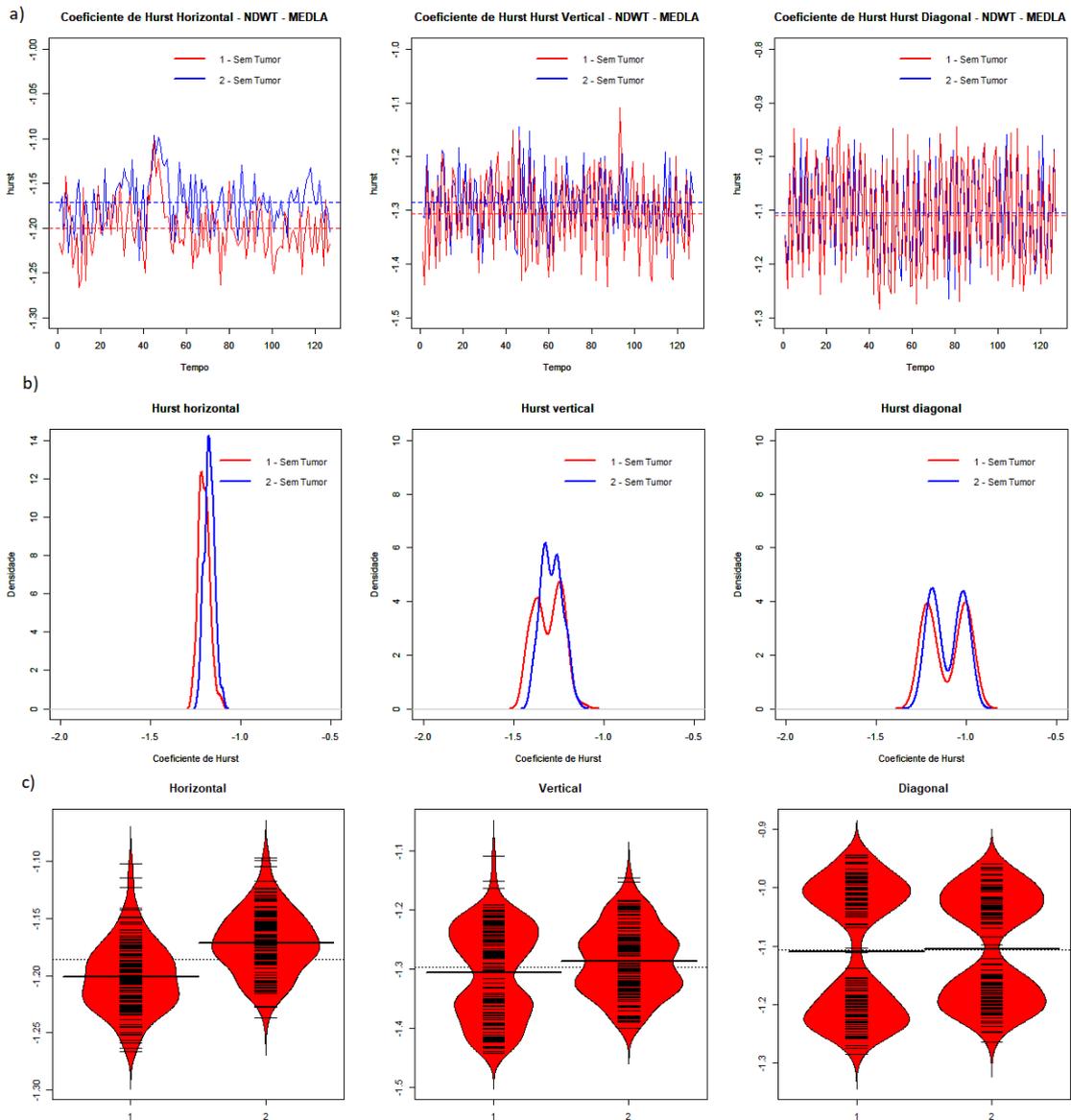
Região 1	Média	Mediana	Variância	CV	Assimetria	Curtose
Horizontal	-1,1714	-1,1727	0,0007	0,0237	0,1779	-0,2019
Vertical	-1,2861	-1,2902	0,0033	0,0450	0,1564	-0,7647
Diagonal	-1,1044	-1,2110	0,0083	0,0826	0,0166	-1,5669
Região 2	Média	Mediana	Variância	CV	Assimetria	Curtose
Horizontal	-1,2008	-1,2036	0,0009	0,0255	0,4594	0,3042
Vertical	-1,3061	-1,2987	0,0060	0,0597	0,0088	-1,1220
Diagonal	-1,1095	-1,1110	0,0119	0,0985	0,0038	-1,6911

Na Figura 1, são apresentados os gráficos ao longo do tempo em (a), os gráficos das distribuições empíricas em (b) e seus respectivos *beanplots*¹ em (c) das séries de expoentes de Hurst direcionais (Região 1 em azul e Região 2 em vermelho).

Percebe-se que as séries produzidas por subimagens distintas do tecido saudável apresenta grande sobreposição em todas as direções, reafirmando a importância dessa metodologia para diferenciar tecido saudável do cancerígeno, já que caso fossem tecidos com características distintas pelo menos uma direção não teria sobreposição relevante. Informação que também pode ser vista com análise dos *beanplots* que mostram valores próximos entre as médias individuais com a média geral das distribuições, além de uma mistura de distribuições na direção diagonal.

¹ Gráfico para comparar distribuições de diferentes grupos. Composto de um gráfico de dispersão unidimensional, as distribuições em forma de densidade, uma linha média para cada distribuição e uma linha média geral de todas distribuições

Figura 1 – Gráficos dos expoentes de Hurst das regiões sem carcinoma utilizando NDWT - MEDLA : (a) Gráficos das séries ao longo do tempo, (b) Distribuições dos expoentes direcionais e (c) beanplots.



As estatísticas descritivas das regiões com carcinoma (Região 1 e Região 2) estão na Tabela 2. Assim como ocorrido com subimagens do tecido saudável, as duas regiões com carcinoma também mostram estatísticas descritivas muito próximas. Além disso, também confirma o que foi dito por Nicolis et al. (2011), que tecidos malignos tendem a ser mais regulares e com expoentes de Hurst maiores, já que a maioria das médias e das medianas apresentaram diferença a partir da terceira casa decimal.

Tabela 2 – Estatísticas descritivas dos expoentes de Hurst direcionais (Regiões com carcinoma - NDWT - MEDLA)

Região 1	Média	Mediana	Variância	CV	Assimetria	Curtose
Horizontal	-1,3334	-1,3342	0,0011	0,0255	0,0101	0,1574
Vertical	-1,3026	-1,3016	0,0028	0,0412	0,0337	-0,9724
Diagonal	-1,0915	-1,1016	0,0055	0,0682	0,0209	-1,5037
Região 2	Média	Mediana	Variância	CV	Assimetria	Curtose
Horizontal	-1,3372	-1,3327	0,0009	0,0228	-0,2334	-0,4582
Vertical	-1,2904	-1,2888	0,0027	0,0405	-0,1829	0,0531
Diagonal	-1,0964	-1,1087	0,0052	0,0663	0,0098	-1,4030

Os gráficos das séries, ao longo do tempo, dos expoentes de Hurst direcionais em (a), as distribuições empíricas em (b) e seus respectivos beanplots são apresentados na Figura 2. Além da variância já verificada, também identifica-se o comportamento semelhante do expoente de Hurst nas três direções.

A sobreposição das curvas nas três direções confirma o comportamento semelhante do expoente de Hurst, quando considerado subimagens de regiões cancerígenas distintas, confirmando a teoria de que essa metodologia pode diferenciar subimagens do tecido cancerígeno do saudável. Para mais, esse comportamento também é evidenciado com análise dos *beanplots* (c), que mostram as médias individuais de cada distribuição próximas da média geral.

Por fim, observando os *beanplots* (c), confirma-se as características assimétricas e leptocúrticas das distribuições, além de uma mistura de distribuições na direção diagonal.

Figura 2 – Gráficos dos expoentes de Hurst das regiões com carcinoma utilizando NDWT - MEDLA :
 (a) Gráficos das séries ao longo do tempo, (b) Distribuições dos expoentes direcionais e (c) beanplots.

