

Chapter 5

Mammogram Diagnostics Using Robust Wavelet-Based Estimator of Hurst Exponent



Chen Feng, Yajun Mei, and Brani Vidakovic

5.1 Introduction

Breast cancer is one of the major health concerns among women. It has been estimated by the National Cancer Institute that 1 in 8 women will be diagnosed with breast cancer during their lifetime. Early detection is proven to be the best strategy for improving prognosis. Most of the references dealing with automated breast cancer detection are based on microcalcifications (El-Naqa et al. 2002; Kestener et al. 2011; Bala and Audithan 2014; Netsch and Peitgen 1999; Wang and Karayiannis 1998). Recently, predicting disease using image data becomes an active research area in statistics and machine learning (Reiss and Ogden 2010; Zhou et al. 2013; Zipunnikov et al. 2011; Reiss et al. 2005). For example, Reiss and Ogden proposed a functional generalized linear regression model with images as predictors (Reiss and Ogden 2010). However, predicting breast cancer based on the tissue images directly is like a black-box. Physicians will have a hard time to summarize the common features from the cancerous images, and the prediction results are not easily interpreted. In this paper, we study the scaling information from the tissue image and then predict breast cancer based on the estimated scaling parameter. It has been found in literatures that the scaling information is efficient and accurate in early detection of breast cancer (Hamilton et al. 2011; Nicolis et al. 2011; Ramírez-Cobo and Vidakovic 2013; Jeon et al. 2014). In fact, regular scaling is a common

C. Feng (✉) · Y. Mei

H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA

e-mail: cfeng@gatech.edu; yimei@isye.gatech.edu

B. Vidakovic

H. Milton Stewart School of Industrial and Systems Engineering and Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, USA

e-mail: brani@gatech.edu

phenomenon in high-frequency signals and high-resolution digital images collected in real life. Examples can be found in a variety of fields including economics, telecommunications, physics, geosciences, as well as in biology and medicine (Feng and Vidakovic 2017; Engel Jr et al. 2009; Gregoriou et al. 2009; Katul et al. 2001; Park and Willinger 2000; Woods et al. 2016; Zhou 1996).

The standard measure of regular scaling is the Hurst exponent, denoted by H in the sequel. Recall that a stochastic process $\{X(t), t \in \mathbb{R}^d\}$ is self-similar with Hurst exponent H if, for any $\lambda \in \mathbb{R}^+$, $X(t) \stackrel{d}{=} \lambda^{-H} X(\lambda t)$. Here the notation $\stackrel{d}{=}$ means the equality in all finite-dimensional distributions. The Hurst exponent quantifies the self-similarity and describes the rate at which autocorrelations decrease as the lag between two realizations in a time series increases. A value H in the range 0–0.5 indicates a zig-zagging intermittent time series with long-term switching between high and low values in adjacent pairs. A value H in the range 0.5 to 1 indicates a time series with long-term positive autocorrelations, which preserves trends on a longer time horizon and gives a time series more regular appearance.

Multiresolution analysis is one of the many methods to estimate the Hurst exponent. An overview can be found in Abry et al. (2000, 1995, 2013). In particular, the non-decimated wavelet transforms (NDWT) (Nason and Silverman 1995; Vidakovic 2009; Percival and Walden 2006) has several potential advantages when employed for Hurst exponent estimation. Input signals and images of arbitrary size can be transformed in a straightforward manner due to the absence of decimation. As a redundant transform, the NDWT can decrease variance in the scaling estimation (Kang and Vidakovic 2017). Least square regression can be fitted to estimate H instead of weighted least square regression since the variances of the level-wise derived distributions based on log NDWT coefficients do not depend on level. Local scaling can be assessed due to the time-invariance property. Of course, the dependence of coefficients in NDWT is much more pronounced. Similar to Soltani et al. (2004), we will control this dependence by systematic sampling of coefficients on which the estimator is based.

Different wavelet-based methods for estimation of H have been proposed in the literature for the one-dimensional case. Abry et al. (2000) suggested the estimation of H by weighted least square regression using the level-wise $\log_2(\overline{d_j^2})$. In addition, the authors corrected for the bias caused by the order of taking the logarithm and the average in $\log_2(\overline{d_j^2})$, where d_j indicates any detail coefficient at level j . We use $d_{j,k}$ to denote the k th coefficient at level j in the sequel. Soltani et al. (2004) defined a mid-energy as $D_{j,k} = (d_{j,k}^2 + d_{j,k+N_j/2}^2)/2$, and showed that the level-wise averages of $\log_2 D_{j,k}$ are asymptotically normal and more stable, which is used to estimate H by regression. The estimators in Soltani et al. (2004) consistently outperform the estimators in Abry et al. (2000). Shen et al. (2007) showed that the method of Soltani et al. (2004) yields more accurate estimators since it takes the logarithm of the mid-energy first and then averages.

The robust estimation of H has recently become a topic of interest due to the presence of outlier coefficients and outlier multiresolution levels, inter and within

level dependences, and distributional contaminations (Franzke et al. 2012; Park and Park 2009; Shen et al. 2007; Sheng et al. 2011). Hamilton et al. (2011) came up with a robust approach based on Theil-type weighted regression (Theil 1992), a method for robust linear regression that selects the weighted average of all slopes defined by different pairs of regression points. Like the VA method, they regress the level-wise $\log_2(\overline{d_j^2})$ against the level indices, but instead of weighted least square regression, they use the Theil-type weighted regression to make it less sensitive to outlier levels. Kang and Vidakovic (2017) proposed MEDL and MEDLA methods based on non-decimated wavelets to estimate H . MEDL estimates H by regressing the medians of $\log d_j^2$ on level j , while MEDLA uses the level-wise medians of $\log\left(\left(d_{j,k_1}^2 + d_{j,k_2}^2\right)/2\right)$ to estimate H , where k_1 and k_2 are properly selected locations at level j to approximate the independence.

Both MEDL and MEDLA use the median of the derived distribution instead of the mean, because the medians are more robust to potential outliers that can occur when logarithmic transform of a squared wavelet coefficient is taken and the magnitude of coefficient is close to zero. Although median is outlier-resistant, it can behave unexpectedly as a result of its non-smooth character. The fact that the median is not “universally the best outlier-resistant estimator” motivates us to develop the general trimean estimators of the level-wise derived distributions to estimate H , where the general trimean estimator was derived as a weighted average of the distribution’s median and two quantiles symmetric about the median, combining the median’s emphasis on center values with the quantiles’ attention to the tails. Tukey’s trimean estimator (Tukey 1977; Andrews and Hampel 2015) and Gastwirth estimator (Gastwirth 1966; Gastwirth and Cohen 1970; Gastwirth and Rubin 1969) are two special cases under such general framework.

In this paper, we are concerned with the robust estimation of Hurst exponent in self-similar signals. Here, the focus is on images, but the methodology applies to multiscale context of arbitrary dimension. The properties of the proposed Hurst exponent estimators are studied both theoretically and numerically. The performance of the robust approach is compared with other standard wavelet-based methods (Veitch and Abry (VA) method, Soltani, Simard, and Boichu (SSB) method, median based estimators MEDL and MEDLA, and Theil-type (TT) weighted regression method).

The rest of the paper consists of six additional sections and an Appendix. Section 5.2 discusses background of non-decimated wavelet transforms and wavelet-based spectrum in the context of estimating the Hurst exponent for fractional Brownian motion (fBm). Section 5.3 introduces the general trimean estimators and discusses two special estimators following that general framework; Sect. 5.4 describes estimation of Hurst exponent using the general trimean estimators, presents distributional results on which the proposed methods are based, and derives optimal weights that minimize the variances of the estimators. Section 5.5 provides the simulation results and compares the performance of the proposed methods to other standardly used, wavelet-based methods. The proposed methods are applied

to classify the digitized mammogram images as cancerous or non-cancerous in Sect. 5.6. The paper is concluded with a summary and discussion in Sect. 5.7.

5.2 Background

5.2.1 Non-decimated Wavelet Transforms

The non-decimated wavelet transforms (NDWT) (Nason and Silverman 1995; Vidakovic 2009; Percival and Walden 2006) are redundant transforms because they are performed by repeated filtering with a minimal shift, or a maximal sampling rate, at all dyadic scales. Subsequently, the transformed signal contains the same number of coefficients as the original signal at each multiresolution level. We start by describing algorithmic procedure of 1-D NDWT and then expand to 2-D NDWT. Traditionally, we perform a wavelet transformation as a convolution of an input data with wavelet and scaling filters. A principal difference between NDWT and DWT is the sampling rate.

Any square integrable function $f(x) \in L_2(\mathbb{R})$ can be expressed in the wavelet domain as

$$f(x) = \sum_k c_{J_0,k} \phi_{J_0,k}(x) + \sum_{j \geq J_0} \sum_k d_{j,k} \psi_{j,k}(x),$$

where $c_{J_0,k}$ denote coarse coefficients, $d_{j,k}$ indicate detail coefficients, $\phi_{J_0,k}(x)$ represent scaling functions, and $\psi_{j,k}(x)$ signify wavelet functions. For specific choices of scaling and wavelet functions, the basis for NDWT can be formed from the atoms

$$\phi_{J_0,k}(x) = 2^{J_0/2} \phi \left(2^{J_0} (x - k) \right) \text{ and}$$

$$\psi_{j,k}(x) = 2^{j/2} \psi \left(2^j (x - k) \right),$$

where $x \in \mathbb{R}$, j is a resolution level, J_0 is the coarsest level, and k is the location of an atom. Notice that atoms for NDWT have the constant location shift k at all levels, yielding the finest sampling rate on any level. The coarse coefficients $c_{J_0,k}$ and detail coefficients $d_{j,k}$ can be obtained via

$$c_{J_0,k} = \int f(x) \phi_{J_0,k}(x) dx \text{ and } d_{j,k} = \int f(x) \psi_{j,k}(x) dx. \quad (5.1)$$

In a J -level decomposition of an 1-D input signal of size N , an NDWT will yield $N \times (J + 1)$ wavelet coefficients, including $N \times 1$ coarse coefficients and $N \times J$ detail coefficients.

Expanding on the 1-D definitions, we could easily describe 2-D NDWT of $f(x, y)$ with $(x, y) \in \mathbb{R}^2$. Several versions of 2-D NDWT exist, but we only focus on the scale-mixing version based on which our methods are proposed. For the scale-mixing 2-D NDWT, the wavelet atoms are

$$\begin{aligned}\phi_{J_{01}, J_{02}; \mathbf{k}}(x, y) &= 2^{(J_{01}+J_{02})/2} \phi(2^{J_{01}}(x - k_1)) \phi(2^{J_{02}}(y - k_2)), \\ \psi_{J_{01}, j_2; \mathbf{k}}(x, y) &= 2^{(J_{01}+j_2)/2} \phi(2^{J_{01}}(x - k_1)) \psi(2^{j_2}(y - k_2)), \\ \psi_{j_1, J_{02}; \mathbf{k}}(x, y) &= 2^{(j_1+J_{02})/2} \psi(2^{j_1}(x - k_1)) \phi(2^{J_{02}}(y - k_2)), \\ \psi_{j_1, j_2; \mathbf{k}}(x, y) &= 2^{(j_1+j_2)/2} \psi(2^{j_1}(x - k_1)) \psi(2^{j_2}(y - k_2)),\end{aligned}$$

where $\mathbf{k} = (k_1, k_2)$ is the location index, J_{01} and J_{02} are coarsest levels, $j_1 > J_{01}$, and $j_2 > J_{02}$. The wavelet coefficients for $f(x, y)$ after the scale-mixing NDWT can be obtained as

$$\begin{aligned}c_{J_{01}, J_{02}; \mathbf{k}} &= \iint f(x, y) \phi_{J_{01}, J_{02}; \mathbf{k}}(x, y) dx dy, \\ h_{J_{01}, j_2; \mathbf{k}} &= \iint f(x, y) \psi_{J_{01}, j_2; \mathbf{k}}(x, y) dx dy, \\ v_{j_1, J_{02}; \mathbf{k}} &= \iint f(x, y) \psi_{j_1, J_{02}; \mathbf{k}}(x, y) dx dy, \\ d_{j_1, j_2; \mathbf{k}} &= \iint f(x, y) \psi_{j_1, j_2; \mathbf{k}}(x, y) dx dy.\end{aligned}\tag{5.2}$$

Note that $c_{J_{01}, J_{02}; \mathbf{k}}$ are coarse coefficients and represent the coarsest approximation, $h_{J_{01}, j_2; \mathbf{k}}$ and $v_{j_1, J_{02}; \mathbf{k}}$ represent the mix of coarse and detail information, and $d_{j_1, j_2; \mathbf{k}}$ carry information about details only. In our methods, only detail coefficients $d_{j_1, j_2; \mathbf{k}}$ are used to estimate H .

5.2.2 The fBm: Wavelet Coefficients and Spectra

Among models having been proposed for analyzing the self-similar phenomena, arguably the most popular is the fractional Brownian motion (fBm) first described by Kolmogorov (1940) and formalized by Mandelbrot and Van Ness (1968).

In this section, an overview of 1-D fBm and its extension to 2-D fBm is provided. Consider a stochastic process $\{X(t), t \in \mathbb{R}\}$ is self-similar with Hurst exponent H , then the 1-D detail coefficients defined in (5.1) satisfy

$$d_{jk} \stackrel{d}{=} 2^{-j(H+1/2)} d_{0k},$$

for a fixed level j (Abry et al. 2003). If the process has stationary increments, i.e., $X(t+h) - X(t)$ is independent of t , then $\mathbb{E}(d_{0k}) = 0$ and $\mathbb{E}(d_{0k}^2) = \mathbb{E}(d_{00}^2)$. We obtain

$$\mathbb{E}(d_{jk}^2) \propto 2^{-j(2H+1)}. \quad (5.3)$$

The Hurst exponent can be estimated by taking logarithms on both sides of Eq. (5.3). The wavelet spectrum is defined by the sequence $\{S(j) = \log \mathbb{E}(d_{jk}^2), j \in \mathbb{Z}\}$. Fractional Brownian motion (fBm), denoted as $B_H(t)$ is the unique Gaussian process with stationary increments that is self-similar (Abry et al. 2003; Abry 2003). The definition of the one-dimensional fBm can be extended to the multivariate case. In particular, a two-dimensional fBm, $B_H(t)$, for $t \in [0, 1] \times [0, 1]$ and $H \in (0, 1)$, is a Gaussian process with stationary zero-mean increments, satisfying

$$B_H(at) \stackrel{d}{=} a^H B_H(t).$$

It can be shown that the detail coefficients $d_{j_1, j_2; k}$ defined in Eq. (5.2) satisfy

$$\log_2 \mathbb{E}(|d_{j_1, j_2; k}|^2) = -(2H + 2)j + C,$$

which defines the two-dimensional wavelet spectrum, from which the Hurst exponent can be estimated. Our proposed methods in next sections are based on but improve from this spectrum.

5.3 General Trimean Estimators

Let X_1, X_2, \dots, X_n be i.i.d. continuous random variables with pdf $f(x)$ and cdf $F(x)$. Let $0 < p < 1$, and let ξ_p denote the p th quantile of F , so that $\xi_p = \inf\{x | F(x) \geq p\}$. If F is monotone, the p th quantile is simply defined as $F(\xi_p) = p$.

Let $Y_p = X_{[np]:n}$ denote a sample p th quantile. Here $[np]$ denotes the greatest integer that is less than or equal to np . The general trimean estimator is defined as a weighted average of the distribution's median and its two quantiles Y_p and Y_{1-p} , for $p \in (0, 1/2)$:

$$\hat{\mu} = \frac{\alpha}{2} Y_p + (1 - \alpha) Y_{1/2} + \frac{\alpha}{2} Y_{1-p}. \quad (5.4)$$

The weights for the two quantiles are the same for Y_p and Y_{1-p} , and $\alpha \in [0, 1]$. This is equivalent to the weighted sum of the median and the average of Y_p and Y_{1-p} with weights $1 - \alpha$ and α :

$$\hat{\mu} = (1 - \alpha) Y_{1/2} + \alpha \left(\frac{Y_p + Y_{1-p}}{2} \right).$$

This general trimean estimator turns out to be more robust than mean but smoother than the median. To derive its asymptotic distribution, the asymptotic joint distribution of sample quantiles is needed, as shown in Lemma 5.1; detailed proof can be found in DasGupta (2008).

Lemma 5.1 Consider r sample quantiles, $Y_{p_1}, Y_{p_2}, \dots, Y_{p_r}$, where $1 \leq p_1 < p_2 < \dots < p_r \leq n$. If for any $1 \leq i \leq r$, $\sqrt{n} (\lfloor np_i \rfloor / n - p_i) \rightarrow 0$ is satisfied, then the asymptotic joint distribution of $Y_{p_1}, Y_{p_2}, \dots, Y_{p_r}$ is:

$$\sqrt{n} ((Y_{p_1}, Y_{p_2}, \dots, Y_{p_r}) - (\xi_{p_1}, \xi_{p_2}, \dots, \xi_{p_r})) \overset{\text{approx}}{\sim} \mathcal{MVN}(0, \Sigma),$$

where

$$\Sigma = (\sigma_{ij})_{r \times r},$$

and

$$\sigma_{ij} = \frac{p_i (1 - p_j)}{f(x_{p_i}) f(x_{p_j})}, \quad i \leq j. \quad (5.5)$$

From Lemma 5.1, the asymptotic distribution of general trimean estimator will be normal as a linear combination of the components each with an asymptotic normal distribution. The general trimean estimator itself may be defined in terms of order statistics as

$$\hat{\mu} = A \cdot \mathbf{y},$$

where

$$A = \left[\frac{\alpha}{2} \quad 1 - \alpha \quad \frac{\alpha}{2} \right], \quad \text{and } \mathbf{y} = [Y_p \quad Y_{1/2} \quad Y_{1-p}]^T.$$

It can be easily verified that $\sqrt{n} (\lfloor pn \rfloor / n - p) \rightarrow 0$ for $p \in (0, 1/2]$. If we denote $\xi = [\xi_p \quad \xi_{1/2} \quad \xi_{1-p}]^T$ the population quantiles, the asymptotic distribution of \mathbf{y} is

$$\sqrt{n} (\mathbf{y} - \xi) \overset{\text{approx}}{\sim} \mathcal{MVN}(0, \Sigma),$$

where $\Sigma = (\sigma_{ij})_{3 \times 3}$, and σ_{ij} follows Eq. (5.5) for $p_1 = p$, $p_2 = 1/2$, and $p_3 = 1 - p$. Therefore

$$\hat{\mu} \overset{\text{approx}}{\sim} \mathcal{N}(\mathbb{E}(\hat{\mu}), \text{Var}(\hat{\mu})),$$

with the theoretical expectation and variance being

$$\mathbb{E}(\hat{\mu}) = \mathbb{E}(A \cdot \mathbf{y}) = A \cdot \mathbb{E}(\mathbf{y}) = A \cdot \boldsymbol{\xi}, \quad (5.6)$$

and

$$\text{Var}(\hat{\mu}) = \text{Var}(A \cdot \mathbf{y}) = A \text{Var}(\mathbf{y}) A^T = \frac{1}{n} A \Sigma A^T. \quad (5.7)$$

5.3.1 Tukey's Trimean Estimator

Tukey's trimean estimator is a special case of the general trimean estimators, with $\alpha = 1/2$ and $p = 1/4$ in Eq. (5.4). To compute this estimator, we first sort the data in ascending order. Next, we take the values that are one-fourth of the way up this sequence (the first quartile), half way up the sequence (i.e., the median), and three-fourths of the way up the sequence (the third quartile). Given these three values, we then form the weighted average, giving the central (median) value a weight of $1/2$ and the two quartiles a weight of $1/4$ each.

If we denote Tukey's trimean estimator as $\hat{\mu}_T$, then

$$\hat{\mu}_T = \frac{1}{4} Y_{1/4} + \frac{1}{2} Y_{1/2} + \frac{1}{4} Y_{3/4}.$$

The asymptotic distribution is

$$\hat{\mu}_T \overset{\text{approx}}{\sim} \mathcal{N}\left(A_T \cdot \boldsymbol{\xi}_T, \frac{1}{n} A_T \Sigma_T A_T^T\right),$$

where $A_T = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}$, $\boldsymbol{\xi}_T = [\xi_{1/4} \ \xi_{1/2} \ \xi_{3/4}]^T$, $\Sigma_T = (\sigma_{ij})_{3 \times 3}$ is the covariance matrix of the asymptotic multivariate normal distribution, and σ_{ij} follows Eq. (5.5) with $p_1 = 1/4$, $p_2 = 1/2$, and $p_3 = 3/4$.

5.3.2 Gastwirth Estimator

As Tukey's estimator, the Gastwirth estimator is another special case of the general trimean estimators, with $\alpha = 0.6$ and $p = 1/3$ in Eq. (5.4).

If we denote this estimator as $\hat{\mu}_G$, then

$$\hat{\mu}_G = 0.3 Y_{1/3} + 0.4 Y_{1/2} + 0.3 Y_{2/3}.$$

The asymptotic distribution can be derived as

$$\hat{\mu}_G \overset{\text{approx}}{\sim} \mathcal{N} \left(A_G \cdot \xi_G, \frac{1}{n} A_G \Sigma_G A_G^T \right),$$

where $A_G = [0.3 \ 0.4 \ 0.3]$, $\xi_G = [\xi_{1/3} \ \xi_{1/2} \ \xi_{2/3}]^T$, $\Sigma_G = (\sigma_{ij})_{3 \times 3}$, and σ_{ij} follows Eq. (5.5) with $p_1 = 1/3$, $p_2 = 1/2$, and $p_3 = 2/3$.

5.4 Methods

Our proposal for robust estimation of Hurst exponent H is based on non-decimated wavelet transforms (NDWT). In a J -depth decomposition of a 2-D fBm of size $N \times N$, a scale-mixing 2-D NDWT generates $(J + 1) \times (J + 1)$ blocks of coefficients, with each block the same size as original image, i.e., $N \times N$. The tessellation of coefficients of scale-mixing 2-D NDWT is shown in Fig. 5.1a. From the 2-D NDWT wavelets coefficients, our methods use the diagonal blocks ($j_1 = j_2 = j$) of the detail coefficients $d_{j_1, j_2; k}$ to predict H , as is shown in Fig. 5.1b.

At each detail level j , the corresponding level- j diagonal block is of size $N \times N$, the same size as original image. Note that those coefficients $d_{j, j; k}$ in level- j diagonal block are not independent, however, their autocorrelations decay exponentially, that is, they possess only the short memory. We reduce such within block dependency by dividing the block into $M \times M$ equal grids and then random sampling one coefficient from each grid, therefore increasing the distance between two consecutive coefficients. To improve the efficiency, here we apply symmetric sampling. To be specific, we partition the level- j diagonal block into four equal parts (top left, top right, bottom left, and bottom right), only sample from the $M^2/4$

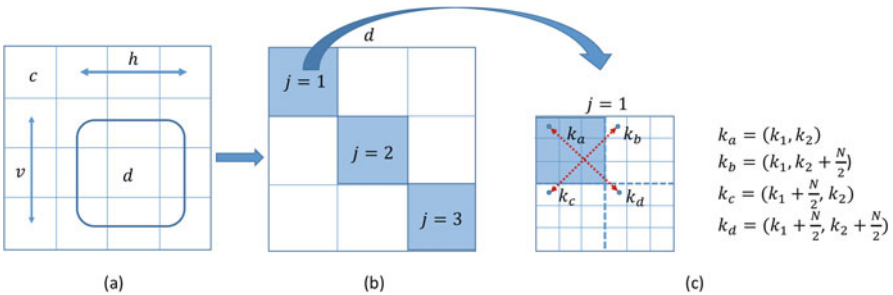


Fig. 5.1 (a) Four types of wavelet coefficients with their locations in the tessellation of a 2-D scale mixing NDWT of depth of 3 ($J = 3$), with each block the size of $N \times N$. Coefficients c represent the coarsest approximation, h and v are the mix of coarse and detail information, and d carry detail information only. (b) Detail coefficients d and its diagonal blocks corresponding to 3 ($J = 3$) levels. (c) Symmetric random sampling from level-1 ($j = 1$) diagonal block divided into 6×6 ($M = 6$) grids

grids at the top left, and then get the corresponding coefficients that have the same location in other parts, which is shown in Fig. 5.1c.

If assuming the coefficient $d_{j,j;(k_{i1},k_{i2})}$ is randomly sampled from grid $i \in \{1, \dots, \frac{M^2}{4}\}$ at the top left part of level- j diagonal block, and $k_{i1}, k_{i2} \in \{1, 2, \dots, \frac{N}{2}\}$ being the corresponding location indexes, then we can extract corresponding coefficients $d_{j,j;(k_{i1},k_{i2}+\frac{N}{2})}$, $d_{j,j;(k_{i1}+\frac{N}{2},k_{i2})}$, and $d_{j,j;(k_{i1}+\frac{N}{2},k_{i2}+\frac{N}{2})}$ from the top right, bottom left, and bottom right parts, respectively. From the set

$$\{d_{j,j;(k_{i1},k_{i2})}, d_{j,j;(k_{i1},k_{i2}+\frac{N}{2})}, d_{j,j;(k_{i1}+\frac{N}{2},k_{i2})}, d_{j,j;(k_{i1}+\frac{N}{2},k_{i2}+\frac{N}{2})}\},$$

we could generate two mid-energies as

$$\begin{aligned} D_{i,j} &= \frac{d_{j,j;(k_{i1},k_{i2})}^2 + d_{j,j;(k_{i1}+\frac{N}{2},k_{i2}+\frac{N}{2})}^2}{2} \\ D'_{i,j} &= \frac{d_{j,j;(k_{i1},k_{i2}+\frac{N}{2})}^2 + d_{j,j;(k_{i1}+\frac{N}{2},k_{i2})}^2}{2}, \quad i \in \{1, \dots, \frac{M^2}{4}\}, \end{aligned} \quad (5.8)$$

where $D_{i,j}$ and $D'_{i,j}$ denote the two mid-energies corresponding to grid i at level j . If we denote D_j as the set of all mid-energies at level j , then

$$D_j = \{D_{1,j}, D'_{1,j}, D_{2,j}, D'_{2,j}, \dots, D_{\frac{M^2}{4},j}, D'_{\frac{M^2}{4},j}\}. \quad (5.9)$$

The $M^2/2$ mid-energies at each level j are treated as if they are independent. Note that M must be divisible by 2.

Our methods have two different versions, one is based on mid-energies D_j , while the other is using logged mid-energies $\log D_j$ (in bracket). First, the distribution of D_j ($\log D_j$) is derived under the independence approximation between $d_{j,j;(k_{i1},k_{i2})}$, $d_{j,j;(k_{i1},k_{i2}+\frac{N}{2})}$, $d_{j,j;(k_{i1}+\frac{N}{2},k_{i2})}$, and $d_{j,j;(k_{i1}+\frac{N}{2},k_{i2}+\frac{N}{2})}$. Next, we calculate the general trimean estimators from the level-wise derived distributions to estimate H .

5.4.1 General Trimean of the Mid-energy (GTME) Method

At each decomposition level j , the asymptotic distribution of the general trimean estimator on $M^2/2$ mid-energies in D_j is derived, from which we find the relationship between the general trimean estimators and H . The general trimean of the mid-energy (GTME) method is described in the following theorem:

Theorem 5.1 *Let $\hat{\mu}_j$ be the general trimean estimator based on the $M^2/2$ mid-energies in D_j defined by (5.9) at level j in a J -level NDWT of a 2-D fBm of size $N \times N$ with Hurst exponent H . Then, the asymptotic distribution of $\hat{\mu}_j$ is normal,*

$$\hat{\mu}_j \stackrel{\text{approx}}{\sim} \mathcal{N} \left(c(\alpha, p) \lambda_j, \frac{2}{M^2} f(\alpha, p) \lambda_j^2 \right), \quad (5.10)$$

where

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

$$f(\alpha, p) = \frac{\alpha(1-2p)(\alpha-4p)}{4p(1-p)} + 1,$$

$$\lambda_j = \sigma^2 \cdot 2^{-(2H+2)j},$$

and σ^2 is the variance of wavelet coefficients from level 0, the Hurst exponent can be estimated as

$$\hat{H} = -\frac{\hat{\beta}}{2} - 1, \quad (5.11)$$

where $\hat{\beta}$ is the regression slope in the least square linear regression on pairs $(j, \log_2(\hat{\mu}_j))$ from level J_1 to J_2 , $J_1 \leq j \leq J_2$. The estimator \hat{H} follows the asymptotic normal distribution

$$\hat{H} \stackrel{\text{approx}}{\sim} \mathcal{N}(H, V_1), \quad (5.12)$$

where the asymptotic variance V_1 is a constant number independent of sample size N and level j ,

$$V_1 = \frac{6f(\alpha, p)}{(\log 2)^2 M^2 c^2(\alpha, p) q(J_1, J_2)},$$

and

$$q(J_1, J_2) = (J_2 - J_1)(J_2 - J_1 + 1)(J_2 - J_1 + 2). \quad (5.13)$$

The proof of Theorem 5.1 is deferred to the Appendix.

To find the optimal α and p by minimizing the asymptotic variance of $\hat{\mu}_j$, we take partial derivatives of $f(\alpha, p)$ with respect to α and p and set them to 0. The optimal $\hat{\alpha}$ and \hat{p} can be obtained by solving

$$\begin{aligned} \frac{\partial f(\alpha, p)}{\partial \alpha} &= -\frac{2p-1}{2p(1-p)}\alpha + \frac{1+p}{2(1-p)} - \frac{3}{2} = 0, \\ \frac{\partial f(\alpha, p)}{\partial p} &= \frac{\alpha(2-\alpha)}{2(1-p)^2} + \frac{\alpha^2(2p-1)}{4p^2(1-p)^2} = 0. \end{aligned} \quad (5.14)$$

Since $\alpha \in [0, 1]$ and $p \in (0, 1/2)$, we get the unique solution $\alpha = 2p \approx 0.6$ and $p = 1 - \sqrt{2}/2 \approx 0.3$. The Hessian matrix of $f(\alpha, p)$ is

$$\begin{bmatrix} \frac{\partial^2 f(\alpha, p)}{\partial \alpha^2} & \frac{\partial^2 f(\alpha, p)}{\partial \alpha \partial p} \\ \frac{\partial^2 f(\alpha, p)}{\partial \alpha \partial p} & \frac{\partial^2 f(\alpha, p)}{\partial p^2} \end{bmatrix} = \begin{bmatrix} -\frac{2p-1}{2p(1-p)} & \frac{2p^2-2\alpha p^2+\alpha(2p-1)}{2p^2(1-p)^2} \\ \frac{2p^2-2\alpha p^2+\alpha(2p-1)}{2p^2(1-p)^2} & \frac{2p^3\alpha(2-\alpha)+\alpha^2 p(1-p)+\alpha^2(2p-1)^2}{2p^3(1-p)^3} \end{bmatrix}.$$

Since $-\frac{2p-1}{2p(1-p)} > 0$ and the determinant is $5.66 > 0$ when $\alpha = 2p \approx 0.6$ and $p = 1 - \sqrt{2}/2 \approx 0.3$, the above Hessian matrix is positive definite. Therefore, $\hat{\alpha} = 2 - \sqrt{2}$ and $\hat{p} = 1 - \sqrt{2}/2$ provide the global minima of $f(\alpha, p)$, minimizing also the asymptotic variance of $\hat{\mu}_{j,i}$. In comparing these optimal $\hat{\alpha} \approx 0.6$ and $\hat{p} \approx 0.3$ with $\alpha = 0.6$ and $p = 1/3$ from the Gastwirth estimator, curiously, we find that the optimal general trimean estimator is very close to the Gastwirth estimator.

5.4.2 General Trimean of the Logarithm of Mid-energy (GTLME) Method

Previously discussed the GTME method calculates the general trimean estimator of the mid-energy first and then takes the logarithm. In this section, we will calculate the general trimean estimator of the logged mid-energies at each level j . The following theorem describes the general trimean of the logarithm of mid-energy, the GTLME method.

Theorem 5.2 *Let $\hat{\mu}_j$ be the general trimean estimator based on $\log(D_j)$, which is the set of $M^2/2$ logged mid-energies at level j in a J -level NDWT of a 2-D fBm of size $N \times N$ with Hurst exponent H , and $1 \leq j \leq J$. Then, the asymptotic distribution of $\hat{\mu}_j$ is normal,*

$$\hat{\mu}_j \overset{\text{approx}}{\sim} \mathcal{N} \left(c(\alpha, p) + \log(\lambda_j), \frac{2}{M^2} f(\alpha, p) \right), \quad (5.15)$$

where

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

$$f(\alpha, p) = \frac{\alpha^2}{4g_1(p)} + \frac{\alpha(1-\alpha)}{2g_2(p)} + \frac{(1-\alpha)^2}{(\log 2)^2},$$

$g_1(p)$ and $g_2(p)$ are two functions of p given in the Appendix,

$$\lambda_j = \sigma^2 \cdot 2^{-(2H+2)j},$$

and σ^2 is the variance of wavelet coefficients from level 0. The Hurst exponent can be estimated as

$$\hat{H} = -\frac{1}{2 \log 2} \hat{\beta} - 1, \tag{5.16}$$

where $\hat{\beta}$ is the regression slope in the least square linear regressions on pairs $(j, \hat{\mu}_j)$ from level J_1 to J_2 , $J_1 \leq j \leq J_2$. The estimator \hat{H} follows the asymptotic normal distribution

$$\hat{H} \overset{\text{approx}}{\sim} \mathcal{N}(H, V_2), \tag{5.17}$$

where the asymptotic variance V_2 is a constant number independent of simple size N and level j ,

$$V_2 = \frac{6f(\alpha, p)}{(\log 2)^2 M^2 q(J_1, J_2)},$$

and $q(J_1, J_2)$ is given in Eq. (5.13).

The proof of Theorem 5.2 is provided in the Appendix. Similarly, as for the GTME, the optimal α and p which minimize the asymptotic variance of $\hat{\mu}_j$ can be obtained by solving

$$\frac{\partial f(\alpha, p)}{\partial \alpha} = 0, \text{ and } \frac{\partial f(\alpha, p)}{\partial p} = 0. \tag{5.18}$$

From the first equation in (5.18) it can be derived that

$$\alpha = \frac{\frac{2}{\log(2)^2} - \frac{1}{2}g_2(p)}{\frac{1}{2}g_1(p) - g_2(p) + \frac{2}{(\log 2)^2}}.$$

The second equation in (5.18) cannot be simplified to a finite form. As an illustration, we plot the $f(\alpha, p)$ with p ranging from 0 to 0.5 and α being a function of p . The plot of α against p is also shown in Fig. 5.2. Numerical computation gives $\hat{\alpha} = 0.5965$ and $\hat{p} = 0.24$. These optimal parameters are close to $\alpha = 0.5$ and $p = 0.25$ in the Tukey’s trimean estimator, but put some more weight on the median.

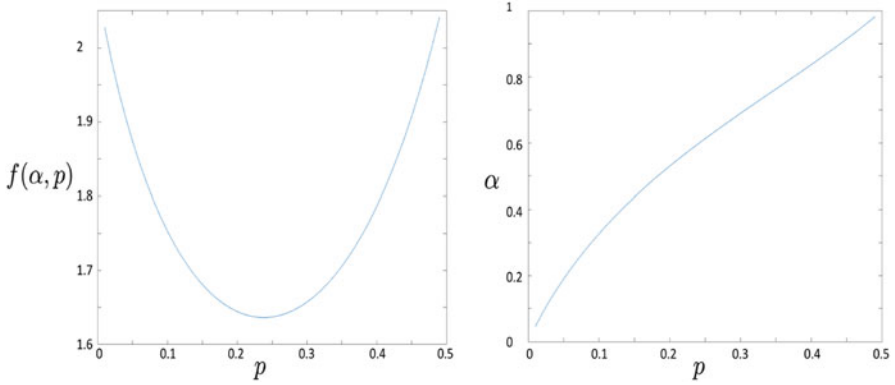


Fig. 5.2 Plot of $f(\alpha, p)$ against p on the left; plot of α against p on the right

5.4.3 Special Cases: Tukey's Trimean and Gastwirth Estimators

The Tukey's trimean of the mid-energy (TTME) method and Gastwirth of the mid-energy (GME) method are described in the following Lemma.

Lemma 5.2 Let $\hat{\mu}_j^T$ and $\hat{\mu}_j^G$ be the Tukey's trimean and Gastwirth estimators based on D_j defined in (5.9). Then the asymptotic distributions of $\hat{\mu}_j^T$ and $\hat{\mu}_j^G$ are normal:

$$\hat{\mu}_j^T \overset{\text{approx}}{\sim} \mathcal{N}\left(c_1 \lambda_j, \frac{5}{3M^2} \lambda_j^2\right), \quad (5.19)$$

$$\hat{\mu}_j^G \overset{\text{approx}}{\sim} \mathcal{N}\left(c_2 \lambda_j, \frac{1.67}{M^2} \lambda_j^2\right), \quad (5.20)$$

where c_1 and c_2 are constant numbers and can be found in the Appendix, $\lambda_j = \sigma^2 \cdot 2^{-(2H+2)j}$, and σ^2 is the variance of wavelet coefficients from level 0. The Hurst exponent can be estimated as

$$\hat{H}^T = -\frac{\hat{\beta}^T}{2} - 1, \text{ and } \hat{H}^G = -\frac{\hat{\beta}^G}{2} - 1, \quad (5.21)$$

where $\hat{\beta}^T$ and $\hat{\beta}^G$ are the regression slopes in the least square linear regression on pairs $(j, \log_2(\hat{\mu}_j^T))$ and pairs $(j, \log_2(\hat{\mu}_j^G))$ from level J_1 to J_2 , $J_1 \leq j \leq J_2$. The estimators \hat{H}^T and \hat{H}^G follow the asymptotic normal distributions

$$\hat{H}^T \overset{\text{approx}}{\sim} \mathcal{N}\left(H, V_1^T\right), \text{ and } \hat{H}^G \overset{\text{approx}}{\sim} \mathcal{N}\left(H, V_1^G\right), \quad (5.22)$$

where the asymptotic variances V_1^T and V_1^G are constant numbers,

$$V_1^T = \frac{5}{(\log 2)^2 M^2 c_1^2 q(J_1, J_2)},$$

$$V_1^G = \frac{5.01}{(\log 2)^2 M^2 c_2^2 q(J_1, J_2)}.$$

The function $q(J_1, J_2)$ is the same as Eq. (5.13) in Theorem 5.1.

The following Lemma describes the Tukey's trimean (TTLME) and Gastwirth (GLME) of the logarithm of mid-energy method.

Lemma 5.3 Let $\hat{\mu}_j^T$ and $\hat{\mu}_j^G$ be the Tukey's trimean estimator and Gastwirth estimator based on $\log(D_j)$ defined in the Theorem 5.2. The asymptotic distributions of $\hat{\mu}_j^T$ and $\hat{\mu}_j^G$ are normal,

$$\hat{\mu}_j^T \overset{\text{approx}}{\sim} \mathcal{N}(- (2H + 2) \log 2j + c_3, V_T), \quad (5.23)$$

$$\hat{\mu}_j^G \overset{\text{approx}}{\sim} \mathcal{N}(- (2H + 2) \log 2j + c_4, V_G), \quad (5.24)$$

where $c_3, V_T, c_4,$ and V_G are constant numbers and can be found in the Appendix. The Hurst exponent can be estimated as

$$\hat{H}^T = -\frac{\hat{\beta}^T}{2 \log 2} - 1, \text{ and } \hat{H}^G = -\frac{\hat{\beta}^G}{2 \log 2} - 1, \quad (5.25)$$

where $\hat{\beta}^T$ and $\hat{\beta}^G$ are the regression slopes in the least square linear regression on pairs $(j, \hat{\mu}_j^T)$ and pairs $(j, \hat{\mu}_j^G)$ from level J_1 to J_2 , $J_1 \leq j \leq J_2$. The estimators \hat{H}^T and \hat{H}^G follow the asymptotic normal distributions

$$\hat{H}^T \overset{\text{approx}}{\sim} \mathcal{N}(H, V_2^T), \text{ and } \hat{H}^G \overset{\text{approx}}{\sim} \mathcal{N}(H, V_2^G), \quad (5.26)$$

where the asymptotic variances V_2^T and V_2^G are constant numbers,

$$V_2^T = \frac{3V_T}{(\log 2)^2 q(J_1, J_2)},$$

$$V_2^G = \frac{3V_G}{(\log 2)^2 q(J_1, J_2)}.$$

The function $q(J_1, J_2)$ is provided in Eq. (5.13).

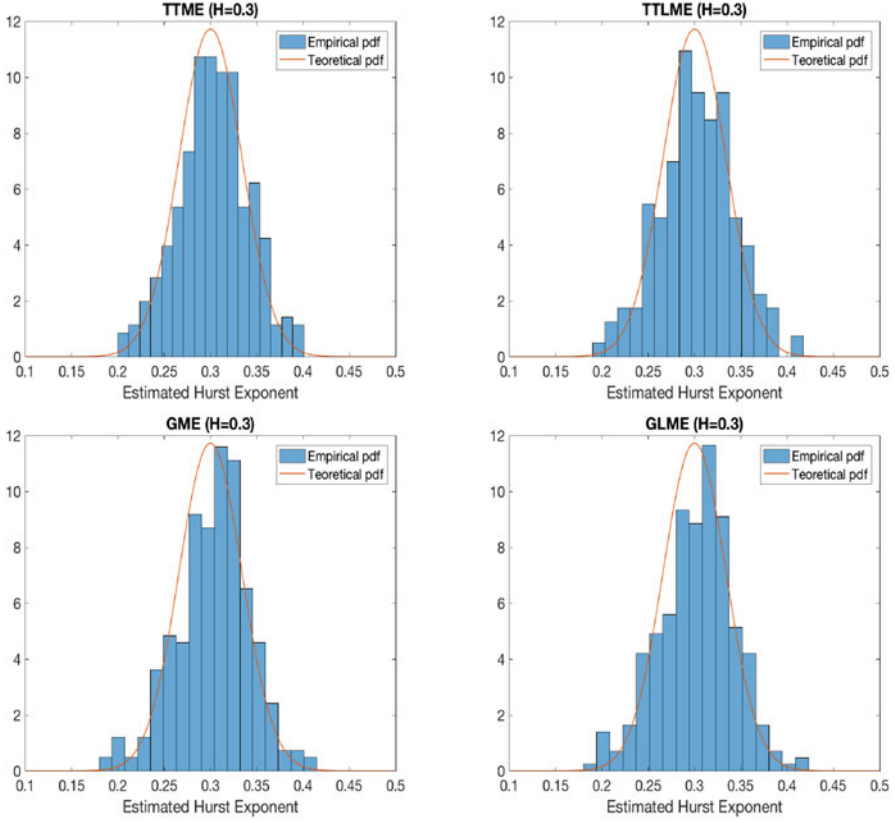


Fig. 5.3 Histograms and theoretical distributions of \hat{H}

The proofs of Lemmas 5.2 and 5.3 are provided in the Appendix. To verify the asymptotic normal distributions of predictors in Lemmas 5.2 and 5.3, we perform an NDWT of depth 10 on 300 simulated fBm's with $H = 0.3$. We use resulting wavelet coefficients from levels 4 to 10 inclusive to estimate H . Figure 5.3 shows the histograms and theoretical distributions of \hat{H} using TTME, TTLME, GME, and GLME methods, respectively.

5.5 Simulation

We simulate 2-D fBm of sizes $2^{10} \times 2^{10}$ ($N = 2^{10}$) with Hurst exponent $H = 0.3, 0.5, 0.7, 0.8, 0.9$, respectively. NDWT of depth $J = 10$ using Haar wavelet is performed on the simulated signal to obtain wavelet coefficients. The two-

Table 5.1 Simulation results for $2^{10} \times 2^{10}$ fBm using Haar wavelet (300 replications)

H	Existing methods					Proposed methods					
	VA	SSB	MEDL	MEDLA	TT	TTME	TTLME	GME	GLME	GTME	GTLME
\hat{H}											
0.3	0.3103	0.3055	0.3018	0.3031	0.3054	0.3032	0.3028	0.3032	0.3034	0.3028	0.3030
0.5	0.5220	0.5132	0.5095	0.5102	0.5151	0.5126	0.5111	0.5108	0.5100	0.5118	0.5116
0.7	0.7382	0.7235	0.7175	0.7165	0.7326	0.7193	0.7179	0.7193	0.7184	0.7199	0.7181
0.8	0.8458	0.8261	0.8200	0.8204	0.8398	0.8222	0.8214	0.8208	0.8206	0.8212	0.8221
0.9	0.9593	0.9328	0.9241	0.9274	0.9641	0.9303	0.9282	0.9287	0.9278	0.9295	0.9287
Variances											
0.3	0.0014	0.0016	0.0026	0.0020	0.0017	0.0015	0.0016	0.0016	0.0016	0.0015	0.0016
0.5	0.0020	0.0017	0.0027	0.0018	0.0034	0.0013	0.0016	0.0014	0.0016	0.0014	0.0016
0.7	0.0037	0.0019	0.0030	0.0026	0.0086	0.0018	0.0021	0.0020	0.0021	0.0019	0.0020
0.8	0.0050	0.0021	0.0027	0.0023	0.0095	0.0018	0.0020	0.0020	0.0021	0.0019	0.0020
0.9	0.0073	0.0021	0.0028	0.0022	0.0168	0.0018	0.0019	0.0019	0.0020	0.0018	0.0019
MSEs											
0.3	0.0015	0.0016	0.0026	0.0020	0.0017	0.0015	0.0016	0.0016	0.0016	0.0015	0.0016
0.5	0.0025	0.0019	0.0027	0.0019	0.0037	0.0015	0.0017	0.0016	0.0017	0.0015	0.0017
0.7	0.0052	0.0025	0.0033	0.0028	0.0097	0.0022	0.0024	0.0024	0.0025	0.0023	0.0024
0.8	0.0070	0.0027	0.0031	0.0028	0.0110	0.0023	0.0024	0.0024	0.0025	0.0023	0.0025
0.9	0.0108	0.0032	0.0033	0.0030	0.0208	0.0027	0.0027	0.0027	0.0028	0.0027	0.0027

dimensional fBm signals were simulated based on the method of Wood and Chan (1994).

The proposed methods (with six variations) are applied on the NDWT detail coefficients to estimate Hurst exponent H . Each level diagonal block is divided into 16×16 grids ($M = 16$) for all proposed methods, and we use wavelet coefficients from levels 4 to 10 for the least square linear regression. The estimation performance of the proposed methods is compared to five other existing methods: Veitch and Abry (VA) method, Soltani, Simard, and Boichu (SSB) method, MEDL method, MEDLA method, and Theil-type regression (TT) method. The GTME and GTLME methods are based on the optimal parameters which minimize the variances. Estimation performance is reported in terms of mean, variance, and mean square error (MSE) based on 300 repetitions for each case.

The simulation results are shown in Table 5.1. For each H (corresponding to each row in the table), the smallest variances and MSEs are highlighted in bold. From simulations results, all our six variations outperform SSB, MEDL, MEDLA, and TT methods for all H 's regarding variances and MSEs. Compared with VA method, our methods yield significantly smaller variances and MSEs when $H > 0.5$. When $H = 0.3$, our methods are still comparable to VA. Although the performances of

our six variations are very similar regarding variances and MSEs, the TTME method based on Tukey's trimean estimator of the mid-energy has the best performance among all of them. The variances of GTME based on the optimal parameters are very close or equal to those of GME and TTME methods in most cases. Besides, in most cases the optimized GTLME method is superior to other logged mid-energy methods TTLME and GLME with respect to variances; however, such superiority is not significant, since the variances are close to each other.

5.6 Application

In this section, we apply the proposed methodology to classification of digitized mammogram images. The digitized mammograms were obtained from the University of South Florida's Digital Database for Screening Mammography (DDSM) (Heath et al. 2000). All cases examined had biopsy results which served as ground truth. Researchers used the HOWTEK scanner at the full 43.5-micron per pixel spatial resolution to scan 45 mammograms from patients with normal studies (control group) and 79 from patients with confirmed breast cancer (study group). Figure 5.4 shows an example of mammograms from study group, and it is almost impossible for physicians to distinguish a cancerous mammogram with a non-cancerous mammogram just by eyes. Each subject contains two mammograms from a screening exam, one craniocaudal projection for each side breast. We only keep one projection for each subject, either right side or left side breast image. A sub-image of size 1024×1024 was taken manually from each mammogram.

Our methods were then applied on each sub-image to estimate the Hurst exponent parameter for each subject. To be specific, the NDWT of depth $J = 10$ using Haar wavelet was performed on each sub-image to obtain wavelet coefficients. The proposed methods (with six variations) are applied on the NDWT detail coefficients

Fig. 5.4 An example of mammograms with breast cancer

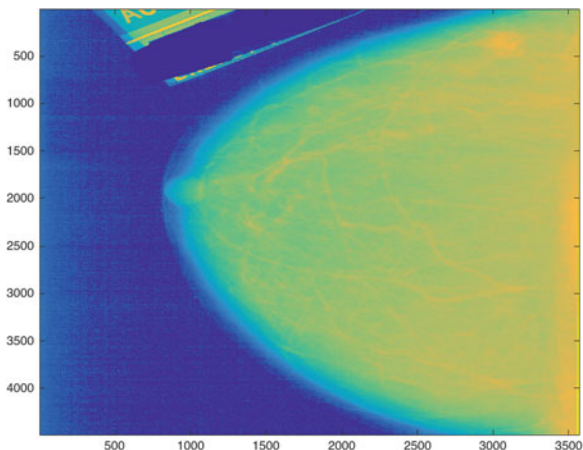


Table 5.2 Descriptive statistics group summary

Group	Existing methods					Proposed methods					
	VA	SSB	MEDL	MEDLA	TT	TTME	TTLME	GME	GLME	GTME	GTLME
Mean of \hat{H}											
Control	0.3570	0.3457	0.3323	0.3403	0.3716	0.3454	0.3422	0.3444	0.3420	0.3450	0.3430
Study	0.4310	0.4038	0.3935	0.4023	0.4203	0.4061	0.4026	0.4031	0.4019	0.4053	0.4027
Median of \hat{H}											
Control	0.3368	0.3339	0.3326	0.3140	0.3871	0.3248	0.3188	0.3198	0.3240	0.3263	0.3278
Study	0.4286	0.4147	0.3865	0.4165	0.4204	0.4194	0.4211	0.4178	0.4150	0.4168	0.4209
Variance of \hat{H}											
Control	0.0267	0.0270	0.0268	0.0298	0.0305	0.0284	0.0279	0.0285	0.0279	0.0281	0.0277
Study	0.0159	0.0172	0.0198	0.0175	0.0128	0.0169	0.0173	0.0174	0.0175	0.0175	0.0174

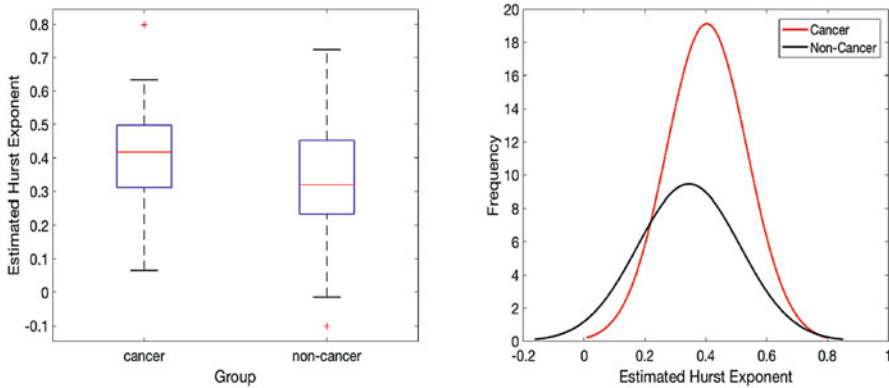


Fig. 5.5 Using GME method to estimate Hurst exponent, boxplots in cancer and non-cancer groups on the left; normal density curves fitted in cancer and non-cancer groups on the right

to estimate Hurst exponent H . Each level diagonal block is divided into 16×16 grids ($M = 16$) for all proposed methods, and we use levels 4 to 10 for the least square linear regression. Veitch and Abry (VA) method, Soltani, Simard, and Boichu (SSB) method, MEDL method, MEDLA method, and Theil-type regression (TT) method were applied, as well, to compare with our methods.

Table 5.2 provides descriptive statistics of the estimated Hurst exponent \hat{H} in each group using our proposed methods and other standard methods to compare with. To visualize the difference in \hat{H} across cancer and non-cancer groups, we present in Fig. 5.5 the boxplots of estimated H and fitted normal density curves in two groups based on proposed GME method. As can be seen, the non-cancer group exhibited a smaller value for \hat{H} in both the mean and median, and the variance of

Table 5.3 Results of classification by logistic regression

	Existing methods					Proposed methods					
	VA	SSB	MEDL	MEDLA	TT	TTME	TTLME	GME	GLME	GTME	GTLME
Overall accuracy	0.629	0.597	0.645	0.589	0.547	0.622	0.613	0.654	0.605	0.628	0.645
Sensitivity	0.695	0.659	0.709	0.623	0.543	0.659	0.659	0.722	0.647	0.684	0.708
Specificity	0.511	0.491	0.532	0.534	0.553	0.553	0.534	0.536	0.532	0.528	0.534

\hat{H} is slightly larger. In fact, images with smaller Hurst exponent tend to be more disordered and unsystematic, therefore healthy individuals tend to have more rough breast tissue images.

For subject i , we generated the data $\{Y_i, H_i\}$, where H_i represents the estimated Hurst exponent, and Y_i is the indicator of the disease status with 1 and 0 signifying cancer and non-cancer, respectively. The subjects were classified using a logistic regression model by treating H_i as the predictor and Y_i as the response. The overall classification accuracy, true positive rate (sensitivity), and true negative rate (specificity) were obtained by using a fourfold-cross validation. Instead of the constant 0.5 threshold, we used a training-data-determined adaptive threshold, i.e., each time the threshold of the logistic regression was first chosen to maximize Youden index on the training set and then applied to the testing set to classify.

Table 5.3 summarizes the results of the classification for each estimation method. The best classification rate (0.6538) and sensitivity (0.7217) were both achieved using GME estimator, and the best specificity (0.5530) was achieved using TT or TTME estimator (highlighted in bold). In general, the six variations of our robust method performed better as compared to other methods in classification of breast cancers using mammograms.

Real-world images like mammograms may be characterized by non-stationary conditions such as extreme values, causing outlier coefficients in multiresolution levels after NDWT. VA method estimates H by weighted least square regression using the level-wise $\log_2 \left(\overline{d_{j,j}^2} \right)$, and SSB method uses $\overline{\log_2 D_j}$, with D_j defined in (5.9), they are easily affected by those within level outliers, in that they both use mean of derived distributions on level-wise detail coefficients to estimate H . Besides, potential outliers can also occur when logarithmic transform is taken and the magnitude of coefficient is close to zero. Like the VA method, TT method regress the level-wise $\log_2 \left(\overline{d_{j,j}^2} \right)$ against the level indices, but instead of weighted least square regression, they use the Theil-type weighted regression, the weighted average of all slopes between different pairs of regression points, to make it less sensitive to outlier levels. However, it is still not robust to within level outlier coefficients. MEDL and MEDLA use the median of the derived distribution instead of the mean. Although median is outlier-resistant, it can behave unexpectedly as a result of its non-smooth character. To improve, our methods (six derivations) use the general trimean estimator on non-decimated wavelet detail coefficients of the transformed data, combining the median's emphasis on central values with the

quantiles' attention to the extremes. Besides, in the context of our scenario, Theil-type regression is equivalent to least square regression, since the variance of our pair-wise slope is independent of levels and sample size. Those explain why our robust methods performed the best in classification of mammograms.

5.7 Conclusions

In this paper, we proposed methodologies and derived six variations to improve the robustness of estimation of Hurst exponent H in two-dimensional setting. Non-decimated wavelet transforms (NDWT) are utilized for its redundancy and time-invariance. Instead of using mean or median of the derived distribution on level-wise wavelet coefficients, we defined the general trimean estimators that combine the median's emphasis on center values with the quantiles' attention to the extremes and used them on the level-wise derived distributions to estimate H .

The proposed variations were: (1) Tukey's trimean of the mid-energy (TTME) method; (2) Tukey's trimean of the logged mid-energy (TTLME) method; (3) Gastwirth of the mid-energy (GME) method; (4) Gastwirth of the logged mid-energy (GLME) method; (5) general trimean of the mid-energy (GTME) method; (6) general trimean of the logarithm of mid-energy (GTLME) method. The GTME and GTLME methods are based on the derived optimal parameters in general trimean estimators to minimize the asymptotic variances. Tukey's trimean and Gastwirth estimators are two special cases following the general trimean estimators' framework. These estimators are applied on both mid-energy (as defined by Soltani et al. 2004) and logarithm of the mid-energy at each NDWT level detail coefficient diagonal block. The estimation performance of the proposed methods is compared to five other existing methods: Veitch and Abry (VA) method, Soltani, Simard, and Boichu (SSB) method, MEDL method, MEDLA method, and Theil-type regression (TT) method.

Simulation results indicate all our six variations outperform SSB, MEDL, MEDLA, and TT methods for all H 's regarding variances and MSEs. Compared with VA method, our methods yield significantly smaller variances and MSEs when $H > 0.5$. When $H = 0.3$, our methods are still comparable to VA. Although the performances of our six variations are very similar regarding variances and MSEs, the TTME method based on Tukey's trimean estimator of the mid-energy has the best performance among all of them.

The proposed methods have been applied to digitized mammograms to classify patients with and without breast cancer. Our methods helped to differentiate individuals based on the estimated Hurst parameters \hat{H} . Higher values for \hat{H} have been found in cancer group, and individuals with breast cancer have smoother breast tissue images. This increase of regularity with increase of the degree of pathology is common for many other biometric signals: EEG, EKG, high frequency protein mass-spectra, high resolution medical images of tissue, to list a few.

Appendix

Proof of Theorem 5.1

Proof A single wavelet coefficient in a non-decimated wavelet transform of a 2-D fBm of size $N \times N$ with Hurst exponent H is normally distributed, with variance depending on its level j . The four coefficients in each set

$$\{d_{j,j;(k_{i1},k_{i2})}, d_{j,j;(k_{i1},k_{i2}+\frac{N}{2})}, d_{j,j;(k_{i1}+\frac{N}{2},k_{i2})}, d_{j,j;(k_{i1}+\frac{N}{2},k_{i2}+\frac{N}{2})}\}$$

are assumed to be independent and follow the same normal distribution.

$$\begin{aligned} & d_{j,j;(k_{i1},k_{i2})}, d_{j,j;(k_{i1},k_{i2}+\frac{N}{2})}, d_{j,j;(k_{i1}+\frac{N}{2},k_{i2})}, d_{j,j;(k_{i1}+\frac{N}{2},k_{i2}+\frac{N}{2})} \\ & \sim \mathcal{N}\left(0, 2^{-(2H+2)j} \sigma^2\right). \end{aligned}$$

Then the mid-energies in D_j defined in (5.9) and (5.8) can be readily shown to have exponential distribution with scale parameter $\lambda_j = \sigma^2 \cdot 2^{-(2H+2)j}$. Therefore at each detail level j , the mid-energies in D_j are i.i.d. $\mathcal{E}xp(\lambda_j^{-1})$, and when applying general trimean estimator $\hat{\mu}_j$ on D_j , following the derivation in Sect. 5.3, we have

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

and

$$\boldsymbol{\Sigma} = \begin{bmatrix} \frac{p}{(1-p)} \lambda_j^2 & \frac{p}{(1-p)} \lambda_j^2 & \frac{p}{(1-p)} \lambda_j^2 \\ \frac{p}{(1-p)} \lambda_j^2 & \lambda_j^2 & \lambda_j^2 \\ \frac{p}{(1-p)} \lambda_j^2 & \lambda_j^2 & \frac{1-p}{p} \lambda_j^2 \end{bmatrix}_{3 \times 3},$$

therefore, the asymptotic distribution of $\hat{\mu}_{j,i}$ is normal with mean

$$\begin{aligned} \mathbb{E}(\hat{\mu}_{j,i}) &= A \cdot \boldsymbol{x} \\ &= \left(\frac{\alpha}{2} \log\left(\frac{1}{p(1-p)}\right) + (1-\alpha) \log 2 \right) \lambda_j \\ &\triangleq c(\alpha, p) \lambda_j, \end{aligned}$$

and variance

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

Since the Hurst exponent can be estimated as

$$\hat{H} = -\frac{\hat{\beta}}{2} - 1, \quad (5.27)$$

where $\hat{\beta}$ is the regression slope in the least square linear regression on pairs $(j, \log_2(\hat{\mu}_j))$ from level J_1 to J_2 , $J_1 \leq j \leq J_2$. It can be easily derived that $\hat{\beta}$ is a linear combination of $\log_2(\hat{\mu}_j)$,

$$\hat{\beta} = \sum_{j=J_1}^{J_2} a_j \log_2(\hat{\mu}_j), \quad a_j = \frac{j - (J_1 + J_2)/2}{\sum_{j=J_1}^{J_2} (j - (J_1 + J_2)/2)^2}.$$

We can check that $\sum_{j=J_1}^{J_2} a_j = 0$ and $\sum_{j=J_1}^{J_2} a_j j = 1$. Also, if $X \sim \mathcal{N}(\mu, \sigma^2)$, the approximate expectation and variance of $g(X)$ are

$$\mathbb{E}(g(X)) = g(\mu) + \frac{g''(\mu)\sigma^2}{2}, \quad \text{and} \quad \text{Var}(g(X)) = (g'(\mu))^2 \sigma^2,$$

based on which we calculate

$$\mathbb{E}(\log_2(\hat{\mu}_j)) = -(2H+2)j + \text{Constant}, \quad \text{and} \quad \text{Var}(\log_2(\hat{\mu}_j)) = \frac{\frac{2}{M^2} f(\alpha, p)}{(\log 2)^2 c^2(\alpha, p)}.$$

Therefore

$$\begin{aligned}
\mathbb{E}(\hat{\beta}) &= \sum_{j=J_1}^{J_2} a_j \mathbb{E}(\log_2(\hat{\mu}_j)) = -(2H+2), \quad \text{and} \quad \text{Var}(\hat{\beta}) \\
&= \sum_{j=J_1}^{J_2} a_j^2 \text{Var}(\log_2(\hat{\mu}_j)) := 4V1,
\end{aligned}$$

and

$$\mathbb{E}(\hat{H}) = H, \quad \text{and} \quad \text{Var}(\hat{H}) = V1, \quad (5.28)$$

where the asymptotic variance V_1 is a constant number independent of simple size N and level j ,

$$V_1 = \frac{6f(\alpha, p)}{(\log 2)^2 M^2 c^2(\alpha, p) q(J_1, J_2)},$$

and

$$q(J_1, J_2) = (J_2 - J_1)(J_2 - J_1 + 1)(J_2 - J_1 + 2).$$

Proof of Theorem 5.2

Proof We have stated that each mid-energy in D_j follows $\mathcal{E}_{xp}(\lambda_j^{-1})$ with scale parameter $\lambda_j = \sigma^2 \cdot 2^{-(2H+2)j}$. If we denote the k th element in $\log(D_j)$ as $y_{j,k}$ for $k = 1, \dots, \frac{M^2}{2}$ and $j = 1, \dots, J$, the pdf and cdf of $y_{j,k}$ are

$$f(y_{j,k}) = \lambda_j^{-1} e^{-\lambda_j^{-1} e^{y_{j,k}}} e^{y_{j,k}},$$

and

$$F(y_{j,k}) = 1 - e^{-\lambda_j^{-1} e^{y_{j,k}}}.$$

The p -quantile can be obtained by solving $F(y_p) = 1 - e^{-\lambda_j^{-1} e^{y_p}} = p$, and $y_p = \log(-\lambda_j \log(1 - p))$. Then it can be shown that $f(y_p) = -(1 - p) \log(1 - p)$. When applying the general trimean estimator $\hat{\mu}_j$ on $\log(D_j)$, following the derivation in Sect. 5.3, we get

$$\xi = \begin{bmatrix} \log\left(\log\left(\frac{1}{1-p}\right)\right) + \log(\lambda_j) \\ \log(\log 2) + \log(\lambda_j) \\ \log\left(\log\left(\frac{1}{p}\right)\right) + \log(\lambda_j) \end{bmatrix},$$

and

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

thus, the asymptotic distribution of $\hat{\mu}_{j,i}$ is normal with mean

$$\begin{aligned} \mathbb{E}(\hat{\mu}_{j,i}) &= A \cdot \xi \\ &= \frac{\alpha}{2} \log \left(\log \frac{1}{1-p} \cdot \log \frac{1}{p} \right) + (1-\alpha) \log(\log 2) + \log(\lambda_j) \\ &\triangleq c(\alpha, p) + \log(\lambda_j), \end{aligned}$$

and variance

$$\begin{aligned} \text{Var}(\hat{\mu}_{j,i}) &= \frac{2}{M^2} A \Sigma A^T \\ &= \frac{2}{M^2} \left(\frac{\alpha^2}{4} g_1(p) + \frac{\alpha(1-\alpha)}{2} g_2(p) + \frac{(1-\alpha)^2}{(\log 2)^2} \right) \\ &\triangleq \frac{2}{M^2} f(\alpha, p), \end{aligned}$$

where

$$\begin{aligned} g_1(p) &= \frac{p}{(1-p)(\log(1-p))^2} + \\ &\quad \frac{1-p}{p(\log p)^2} + \frac{2p}{(1-p)\log(1-p)\log p}, \end{aligned}$$

and

$$g_2(p) = \frac{2p}{(1-p)\log(1-p)\log \frac{1}{2}} + \frac{2}{\log \frac{1}{2} \log p}.$$

Since the Hurst exponent can be estimated as

$$\hat{H} = -\frac{1}{2 \log 2} \hat{\beta} - 1, \tag{5.29}$$

where $\hat{\beta}$ is the regression slope in the least square linear regressions on pairs $(j, \hat{\mu}_j)$ from level J_1 to J_2 , $J_1 \leq j \leq J_2$. It can be easily derived that $\hat{\beta}$ is a linear combination of $\hat{\mu}_j$,

$$\hat{\beta} = \sum_{j=J_1}^{J_2} a_j \hat{\mu}_j, \quad a_j = \frac{j - (J_1 + J_2)/2}{\sum_{j=J_1}^{J_2} (j - (J_1 + J_2)/2)^2}.$$

Again, we can check that $\sum_{j=J_1}^{J_2} a_j = 0$ and $\sum_{j=J_1}^{J_2} a_j j = 1$. Therefore

$$\begin{aligned} \mathbb{E}(\hat{\beta}) &= \sum_{j=J_1}^{J_2} a_j \mathbb{E}(\hat{\mu}_{j,i}) = -(2H+2)\log 2, \text{ and } \text{Var}(\hat{\beta}) \\ &= \sum_{j=J_1}^{J_2} a_j^2 \text{Var}(\hat{\mu}_{j,i}) := 4(\log 2)^2 V_2, \end{aligned}$$

and

$$\mathbb{E}(\hat{H}) = H, \text{ and } \text{Var}(\hat{H}) = V_2, \quad (5.30)$$

where the asymptotic variance V_2 is a constant number independent of simple size N and level j ,

$$V_2 = \frac{6f(\alpha, p)}{(\log 2)^2 M^2 q(J_1, J_2)},$$

and $q(J_1, J_2)$ is given in Eq. (5.13).

Proof of Lemma 5.2

Proof When applying Tukey's trimean estimator $\hat{\mu}_j^T$ on D_j , following the derivation in Sect. 5.3.1, we have

$$\xi_T = \begin{bmatrix} \log\left(\frac{4}{3}\right)\lambda_j \\ \log(2)\lambda_j \\ \log(4)\lambda_j \end{bmatrix},$$

and

$$\Sigma_T = \begin{bmatrix} \frac{1}{3}\lambda_j^2 & \frac{1}{3}\lambda_j^2 & \frac{1}{3}\lambda_j^2 \\ \frac{1}{3}\lambda_j^2 & \lambda_j^2 & \lambda_j^2 \\ \frac{1}{3}\lambda_j^2 & \lambda_j^2 & \frac{1}{3}\lambda_j^2 \end{bmatrix}_{3 \times 3},$$

therefore, the asymptotic distribution of $\hat{\mu}_j^T$ is normal with mean

$$\mathbb{E}(\hat{\mu}_j^T) = A_T \cdot \xi_T = \frac{1}{4} \log\left(\frac{64}{3}\right)\lambda_j \triangleq c_1 \lambda_j,$$

and variance

$$\text{Var}(\hat{\mu}_{j,i}^T) = \frac{2}{M^2} A_T \Sigma_T A_T^T = \frac{5}{3M^2} \lambda_j^2.$$

When applying Gastwirth estimator $\hat{\mu}_j^G$ on D_j , following the derivation in Sect. 5.3.2, we have

$$\xi_G = \begin{bmatrix} \log\left(\frac{3}{2}\right) \lambda_j \\ \log(2) \lambda_j \\ \log(3) \lambda_j \end{bmatrix},$$

and

$$\Sigma_G = \begin{bmatrix} \frac{1}{2} \lambda_j^2 & \frac{1}{2} \lambda_j^2 & \frac{1}{2} \lambda_j^2 \\ \frac{1}{2} \lambda_j^2 & \lambda_j^2 & \lambda_j^2 \\ \frac{1}{2} \lambda_j^2 & \lambda_j^2 & \frac{1}{2} \lambda_j^2 \end{bmatrix},$$

therefore, the asymptotic distribution of $\hat{\mu}_j^G$ is normal with mean

$$\begin{aligned} \mathbb{E}(\hat{\mu}_{j,i}^G) &= A_G \cdot \xi_G \\ &= \left(0.3 \times \log\left(\frac{9}{2}\right) + 0.4 \times \log(2)\right) \lambda_j \\ &\triangleq c_2 \lambda_j, \end{aligned}$$

and variance

$$\text{Var}(\hat{\mu}_{j,i}^G) = \frac{2}{M^2} A_G \Sigma_G A_G^T = \frac{1.67}{M^2} \lambda_j^2.$$

Based on Eq. (5.28), we have

$$\hat{H}^T \overset{\text{approx}}{\rightsquigarrow} \mathcal{N}(H, V_1^T), \text{ and } \hat{H}^G \overset{\text{approx}}{\rightsquigarrow} \mathcal{N}(H, V_1^G), \tag{5.31}$$

where the asymptotic variances V_1^T and V_1^G are constant numbers,

$$V_1^T = \frac{5}{(\log 2)^2 M^2 c_1^2 q(J_1, J_2)},$$

$$V_1^G = \frac{5.01}{(\log 2)^2 M^2 c_2^2 q(J_1, J_2)}.$$

The function $q(J_1, J_2)$ is the same as Eq. (5.13) in Theorem 5.1.

Proof of Lemma 5.3

Proof When applying Tukey's trimean estimator $\hat{\mu}_j^T$ on $\log(D_j)$, following the derivation in Sect. 5.3.1, we have

$$\xi_T = \begin{bmatrix} \log\left(\log\left(\frac{4}{3}\right)\right) + \log(\lambda_j) \\ \log(\log 2) + \log(\lambda_j) \\ \log(\log 4) + \log(\lambda_j) \end{bmatrix},$$

and

$$\Sigma_T = \begin{bmatrix} \frac{1}{3\left(\log\left(\frac{3}{4}\right)\right)^2} & \frac{1}{3\log\left(\frac{3}{4}\right)\log\left(\frac{1}{2}\right)} & \frac{1}{3\log\left(\frac{3}{4}\right)\log\left(\frac{1}{4}\right)} \\ \frac{1}{3\log\left(\frac{3}{4}\right)\log\left(\frac{1}{2}\right)} & \frac{1}{(\log 2)^2} & \frac{1}{\log\left(\frac{1}{2}\right)\log\left(\frac{1}{4}\right)} \\ \frac{1}{3\log\left(\frac{3}{4}\right)\log\left(\frac{1}{4}\right)} & \frac{1}{\log\left(\frac{1}{2}\right)\log\left(\frac{1}{4}\right)} & \frac{3}{(\log 4)^2} \end{bmatrix},$$

therefore, the asymptotic distribution of $\hat{\mu}_j^T$ is normal with mean

$$\begin{aligned} \mathbb{E}\left(\hat{\mu}_{j,i}^T\right) &= A_T \cdot \xi_T \\ &= -(2H + 2) \log 2 \cdot j + \log \sigma^2 + \\ &\quad \frac{1}{4} \log\left(\log\left(\frac{4}{3}\right) \cdot \log 4\right) + \frac{1}{2} \log(\log 2) \\ &\triangleq -(2H + 2) \log 2 \cdot j + c_3 \end{aligned}$$

and variance

$$\begin{aligned} \text{Var}\left(\hat{\mu}_{j,i}^T\right) &= \frac{2}{M^2} A_T \Sigma_T A_T^T \\ &= \frac{2}{M^2} \left(\frac{1}{48 \left(\log\left(\frac{3}{4}\right)\right)^2} + \frac{1}{12 \log\left(\frac{3}{4}\right) \log\left(\frac{1}{2}\right)} + \frac{1}{24 \log\left(\frac{3}{4}\right) \log\left(\frac{1}{4}\right)} + \right. \\ &\quad \left. \frac{1}{4 (\log 2)^2} + \frac{1}{4 \log\left(\frac{1}{2}\right) \log\left(\frac{1}{4}\right)} + \frac{3}{16 \left(\log\left(\frac{1}{4}\right)\right)^2} \right) \\ &\triangleq V_T. \end{aligned}$$

When applying Gastwirth estimator $\hat{\mu}_j^G$ on $\log(D_{j,i})$, following the derivation in Sect. 5.3.2, we have

$$\xi_G = \begin{bmatrix} \log\left(\log\left(\frac{3}{2}\right)\right) + \log(\lambda_j) \\ \log(\log 2) + \log(\lambda_j) \\ \log(\log 3) + \log(\lambda_j) \end{bmatrix},$$

and

$$\Sigma_G = \begin{bmatrix} \frac{1}{2\left(\log\frac{2}{3}\right)^2} & \frac{1}{2\log\left(\frac{2}{3}\right)\log\left(\frac{1}{2}\right)} & \frac{1}{2\log\left(\frac{1}{3}\right)\log\left(\frac{2}{3}\right)} \\ \frac{1}{2\log\left(\frac{2}{3}\right)\log\left(\frac{1}{2}\right)} & \frac{1}{(\log 2)^2} & \frac{1}{\log\left(\frac{1}{2}\right)\log\left(\frac{1}{3}\right)} \\ \frac{1}{2\log\left(\frac{1}{3}\right)\log\left(\frac{2}{3}\right)} & \frac{1}{\log\left(\frac{1}{2}\right)\log\left(\frac{1}{3}\right)} & \frac{2}{(\log 3)^2} \end{bmatrix},$$

therefore, the asymptotic distribution of $\hat{\mu}_j^G$ is normal with mean

$$\begin{aligned} \mathbb{E}\left(\hat{\mu}_{j,i}^G\right) &= A_g \cdot \xi_G \\ &= -(2H + 2) \log 2 \cdot j + \log \sigma^2 + \\ &\quad 0.3 \times \log\left(\log\left(\frac{3}{2}\right) \cdot \log 3\right) + 0.4 \times \log(\log 2) \\ &\triangleq -(2H + 2) \log 2 \cdot j + c_4 \end{aligned}$$

and variance

$$\begin{aligned} \text{Var}\left(\hat{\mu}_{j,i}^G\right) &= \frac{2}{M^2} A_G \Sigma_G A_G^T \\ &= \frac{2}{M^2} \left(\frac{0.09}{2\left(\log\frac{2}{3}\right)^2} + \frac{0.12}{\log\frac{2}{3}\log\frac{1}{2}} + \frac{0.09}{\log\frac{1}{3}\log\frac{2}{3}} + \right. \\ &\quad \left. \frac{0.16}{\left(\log\frac{1}{2}\right)^2} + \frac{0.24}{\log\frac{1}{2}\log\frac{1}{3}} + \frac{0.18}{\left(\log\frac{1}{3}\right)^2} \right) \\ &\triangleq V_G. \end{aligned}$$

Based on Eq. (5.30), we can easily derive

$$\hat{H}^T \overset{\text{approx}}{\sim} \mathcal{N}\left(H, V_2^T\right), \text{ and } \hat{H}^G \overset{\text{approx}}{\sim} \mathcal{N}\left(H, V_2^G\right), \quad (5.32)$$

where the asymptotic variances V_2^T and V_2^G are constant numbers,

$$V_2^T = \frac{3V_T}{(\log 2)^2 q(J_1, J_2)},$$

$$V_2^G = \frac{3V_G}{(\log 2)^2 q(J_1, J_2)}.$$

The function $q(J_1, J_2)$ is provided in Eq. (5.13).

References

- Abry, P. (2003). Scaling and wavelets: An introductory walk. In *Processes with long-range correlations* (pp. 34–60). Berlin: Springer.
- Abry, P., Gonçalves, P., & Flandrin, P. (1995). Wavelets, spectrum analysis and 1/f processes. In: *Wavelets and statistics* (pp. 15–29). Springer: New York.
- Abry, P., Gonçalves, P., & Véhel, J. L. (2013). *Scaling, fractals and wavelets*. New York: Wiley.
- Abry, P., Flandrin, P., Taqqu, M. S., & Veitch, D. (2000). Wavelets for the analysis, estimation and synthesis of scaling data. *Self-similar network traffic and performance evaluation* (pp. 39–88). New York: Wiley.
- Abry, P., Flandrin, P., Taqqu, M. S., & Veitch, D. (2003). Self-similarity and long-range dependence through the wavelet lens. In *Theory and applications of long-range dependence* (pp. 527–556). Boston, MA: Birkhauser.
- Andrews, D. F., & Hampel, F. R. (2015). *Robust estimates of location: Survey and advances*. Princeton: Princeton University Press.
- Bala, B. K., & Audithan, S. (2014). Wavelet and curvelet analysis for the classification of microcalcification using mammogram images. In *Second International Conference on Current Trends in Engineering and Technology - ICCTET 2014* (pp. 517–521). <https://doi.org/10.1109/ICCTET.2014.6966351>
- DasGupta, A. (2008). Edgeworth expansions and cumulants. In *Asymptotic theory of statistics and probability* (pp. 185–201). New York: Springer.
- El-Naqa, I., Yang, Y., Wernick, M. N., Galatsanos, N. P., & Nishikawa, R. M. (2002). A support vector machine approach for detection of microcalcifications. *IEEE Transactions on Medical Imaging*, 21(12), 1552–1563.
- Engel, Jr J., Bragin, A., Staba, R., & Mody, I. (2009). High-frequency oscillations: What is normal and what is not? *Epilepsia*, 50(4), 598–604.
- Feng, C., & Vidakovic, B. (2017). Estimation of the hurst exponent using trimean estimators on nondecimated wavelet coefficients. arXiv preprint arXiv:170908775.
- Franzke, C. L., Graves, T., Watkins, N. W., Gramacy, R. B., & Hughes, C. (2012). Robustness of estimators of long-range dependence and self-similarity under non-gaussianity. *Philosophical Transactions of the Royal Society A*, 370(1962), 1250–1267.
- Gastwirth, J. L. (1966). On robust procedures. *Journal of the American Statistical Association*, 61(316), 929–948.
- Gastwirth, J. L., & Cohen, M.L (1970) Small sample behavior of some robust linear estimators of location. *Journal of the American Statistical Association*, 65(330), 946–973
- Gastwirth, J. L., & Rubin, H. (1969). On robust linear estimators. *The Annals of Mathematical Statistics*, 40(1), 24–39.

- Gregoriou, G. G., Gotts, S. J., Zhou, H., & Desimone, R. (2009). High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science*, *324*(5931), 1207–1210.
- Hamilton, E. K., Jeon, S., Cobo, P. R., Lee, K. S., & Vidakovic, B. (2011). Diagnostic classification of digital mammograms by wavelet-based spectral tools: A comparative study. In *2011 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE (pp. 384–389).
- Heath, M., Bowyer, K., Kopans, D., Moore, R., & Kegelmeyer, W. P. (2000). The digital database for screening mammography. In *Proceedings of the 5th International Workshop on Digital Mammography* (pp. 212–218). Medical Physics Publishing.
- Jeon, S., Nicolis, O., & Vidakovic, B. (2014). Mammogram diagnostics via 2-d complex wavelet-based self-similarity measures. *The São Paulo Journal of Mathematical Sciences*, *8*(2), 265–284.
- Kang, M., & Vidakovic, B. (2017). Medl and medla: Methods for assessment of scaling by medians of log-squared nondecimated wavelet coefficients. ArXiv Preprint ArXiv:170304180.
- Katul, G., Vidakovic, B., & Albertson, J. (2001). Estimating global and local scaling exponents in turbulent flows using discrete wavelet transformations. *Physics of Fluids*, *13*(1), 241–250.
- Kestener, P., Lina, J. M., Saint-Jean, P., & Arneodo, A. (2011). Wavelet-based multifractal formalism to assist in diagnosis in digitized mammograms. *Image Analysis & Stereology*, *20*(3), 169–174.
- Kolmogorov, A. N. (1940). Wienerische spiralen und einige andere interessante kurven in hilbertscen raum, cr (doklady). *Academy of Sciences URSS (NS)*, *26*, 115–118.
- Mandelbrot, B. B., & Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review*, *10*(4), 422–437.
- Nason, G. P., & Silverman, B. W. (1995). The stationary wavelet transform and some statistical applications. In *Wavelets and statistics* (pp. 281–299). New York: Springer.
- Netsch, T., & Peitgen, H. O. (1999). Scale-space signatures for the detection of clustered microcalcifications in digital mammograms. *IEEE Transactions on Medical Imaging*, *18*(9), 774–786.
- Nicolis, O., Ramírez-Cobo, P., & Vidakovic, B. (2011). 2d wavelet-based spectra with applications. *Computational Statistics & Data Analysis*, *55*(1), 738–751.
- Park, J., & Park, C. (2009). Robust estimation of the hurst parameter and selection of an onset scaling. *Statistica Sinica*, *19*, 1531–1555.
- Park, K., & Willinger, W. (2000). *Self-similar network traffic and performance evaluation*. Wiley Online Library. <https://doi.org/10.1002/047120644X>.
- Percival, D. B., & Walden, A. T. (2006). *Wavelet methods for time series analysis* (vol. 4). New York: Cambridge University Press.
- Ramírez-Cobo, P., & Vidakovic, B. (2013). A 2d wavelet-based multiscale approach with applications to the analysis of digital mammograms. *Computational Statistics & Data Analysis*, *58*, 71–81.
- Reiss, P. T., & Ogden, R. T. (2010). Functional generalized linear models with images as predictors. *Biometrics*, *66*(1), 61–69.
- Reiss, P. T., Ogden, R. T., Mann, J. J., & Parsey, R. V. (2005). Functional logistic regression with pet imaging data: A voxel-level clinical diagnostic tool. *Journal of Cerebral Blood Flow & Metabolism*, *25*(1_suppl), S635–S635.
- Shen, H., Zhu, Z., & Lee, T. C. (2007). Robust estimation of the self-similarity parameter in network traffic using wavelet transform. *Signal Processing*, *87*(9), 2111–2124.
- Sheng, H., Chen, Y., & Qiu, T. (2011). On the robustness of hurst estimators. *IET Signal Processing*, *5*(2), 209–225.
- Soltani, S., Simard, P., & Boichu, D. (2004). Estimation of the self-similarity parameter using the wavelet transform. *Signal Processing*, *84*(1), 117–123.
- Theil, H. (1992). A rank-invariant method of linear and polynomial regression analysis. In *Henri Theils contributions to economics and econometrics* (pp. 345–381). The Netherlands: Springer.
- Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2). Reading, MA: Addison-Wesley.
- Vidakovic, B. (2009). *Statistical modeling by wavelets* (Vol. 503). New York: Wiley.

- Wang, T. C., & Karayiannis, N. B. (1998). Detection of microcalcifications in digital mammograms using wavelets. *IEEE Transactions on Medical Imaging*, *17*(4), 498–509.
- Wood, A. T., & Chan, G. (1994). Simulation of stationary gaussian processes in $[0, 1]$ d. *Journal of Computational and Graphical Statistics*, *3*(4), 409–432.
- Woods, T., Preeprem, T., Lee, K., Chang, W., & Vidakovic, B. (2016). Characterizing exons and introns by regularity of nucleotide strings. *Biology Direct*, *11*(1), 6.
- Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business and Economic Statistics*, *14*(1), 45–52.
- Zhou, H., Li, L., & Zhu, H. (2013). Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association*, *108*(502), 540–552.
- Zipunnikov, V., Caffo, B., Yousem, D. M., Davatzikos, C., Schwartz, B. S., Crainiceanu, C. (2011). Functional principal component model for high-dimensional brain imaging. *NeuroImage*, *58*(3), 772–784.