Journal of Data Science 18 (4), 581–605
October 2020
DOI: 10.6339/JDS.202010_18(4).0001

Wavelet-Based Robust Estimation of Hurst Exponent with Application in Visual Impairment Classification

Chen Feng*[1], Yajun Mei[1], and Brani Vidakovic[2]

[1]*H. Milton Stewart School of Industrial & Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA*
[2]*Department of Statistics, Texas A&M University, College Station, TX, USA*

Abstract

Pupillary response behavior (PRB) refers to changes in pupil diameter in response to simple or complex stimuli. There are underlying, unique patterns hidden within complex, high-frequency PRB data that can be utilized to classify visual impairment, but those patterns cannot be described by traditional summary statistics. For those complex high-frequency data, Hurst exponent, as a measure of long-term memory of time series, becomes a powerful tool to detect the muted or irregular change patterns. In this paper, we proposed robust estimators of Hurst exponent based on non-decimated wavelet transforms. The properties of the proposed estimators were studied both theoretically and numerically. We applied our methods to PRB data to extract the Hurst exponent and then used it as a predictor to classify individuals with different degrees of visual impairment. Compared with other standard wavelet-based methods, our methods reduce the variance of the estimators and increase the classification accuracy.

Keywords *pupillary response behavior; high-frequency data; non-decimated wavelet transforms.*

1 Introduction

Visual impairment is defined as a functional limitation of the eyes or visual system. It can cause disabilities by significantly interfering with one's ability to function independently, to perform activities of daily living, and to travel safely through the environment, see Geraci et al. (1997); West et al. (2002). Many causes of severe visual impairment are hard to cure, however, there are conditions for which medical or surgical treatment will lessen the severity or progression of the vision loss, for example, recent advances in the treatment of age-related macular degeneration (AMD), see Rosenfeld et al. (2006); Fletcher and Schuchard (2006); Avery et al. (2006). Precise classification of different degrees of visual impairment for AMD patients becomes increasingly important for the sake of early intervention. It has been suggested by Moloney et al. (2006) that the high-frequency pupillary response behavior (PRB) data can be useful in visual impairment classification. PRB refers to changes in pupil diameter in response to simple or complex stimuli, and the PRB data used in our study was measured from older adults, including two groups diagnosed with AMD maintaining different ranges of visual acuities, and one visually healthy control group. Examples of PRB data from those three groups are shown in Figure 1. Moloney et al. (2006) indicated that there may be underlying unique patterns hidden within complex PRB data, and these patterns reveal the intrinsic individual differences in cognitive, sensory and motor functions. However, the proper description and interpretation of PRB are not straightforward, since it is affected by a variety of factors, including the ambient light, fatigue, and medication use.

*Corresponding author. Email: chelsea.fengchen1993@gmail.com.

Received March, 2019; Accepted December, 2019

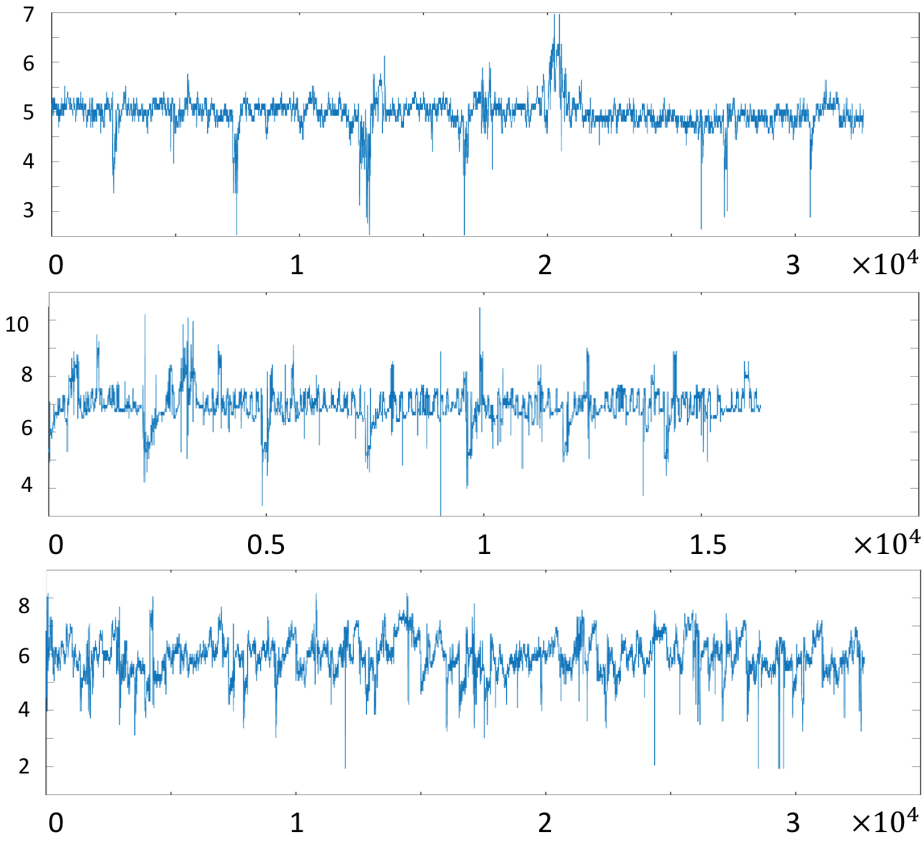

Figure 1: Examples of PRB data from three groups: Up (healthy, control), middle (AMD group I, mild case), down (AMD group II, severe case).

In fact, high-frequency, time series data from various sources often possess hidden patterns that reveal the effects of underlying functional differences, but such patterns cannot be explained by basic descriptive statistics, traditional statistical models, or global trends. Thus powerful analytical tools are needed to detect these muted or irregular change patterns for those complex high-frequency data, like PRB.

One powerful tool is the Hurst exponent, denoted by H in the sequel. It quantifies the long memory, regularity, self-similarity, and scaling in a time series, and has been used as an important feature in many applications, see Engel et al. (2009); Gregoriou et al. (2009); Katul et al. (2001); Park and Willinger (2000); Woods et al. (2016); Zhou (1996). To be more concrete, a stochastic process, $\{X(t), t \in \mathbb{R}\}$ is self-similar with Hurst exponent H if $X(t) \overset{\mathrm{d}}{=} \lambda^{-H} X(\lambda t)$, for any $\lambda \in \mathbb{R}^{+}$. Here the notation $\overset{\mathrm{d}}{=}$ means the equality in all finite-dimensional distributions. Hurst exponent 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.

One widely used example of self-similar Gaussian process is the fractional Brownian motion

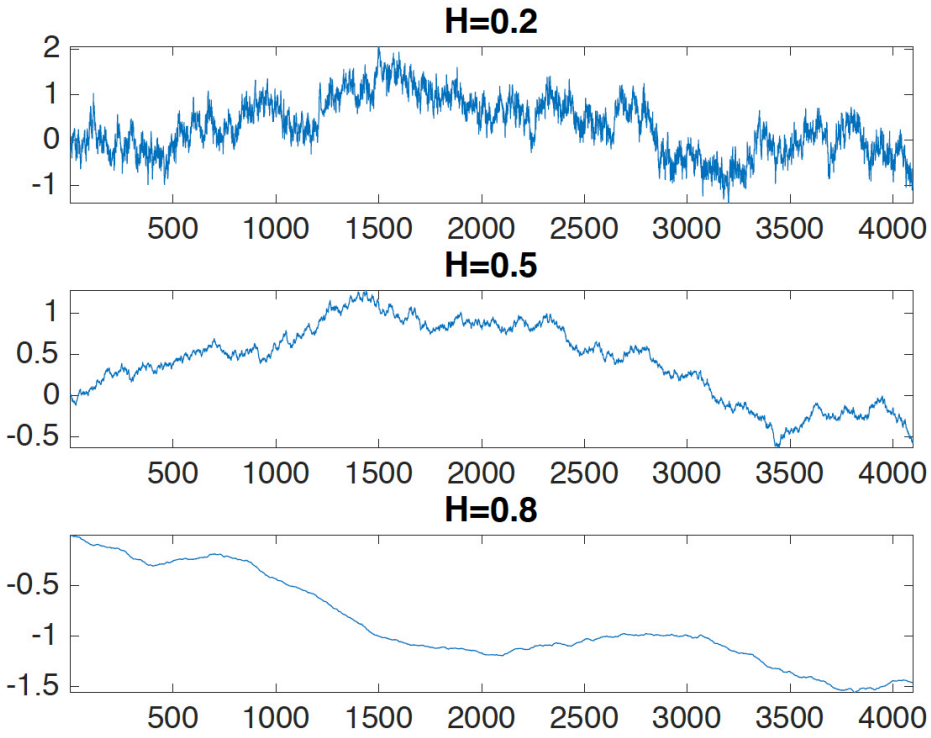

Figure 2: Examples of 1-D fBm with different Hurst exponent H.

(fBm), which was first described by Kolmogorov (1940) and formalized by Mandelbrot and Van Ness (1968), also see Abry et al. (2003); Abry (2003). The fBm is a continuous-time Gaussian process $X(t)$, which starts at zero, has expectation zero for all t, and has the following covariance function:

$$E[X(t)X(s)] = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}).$$

The 1-D fBm is our proposed model for PRB data. Figure 2 shows three examples of 1-D fBm with different Hurst exponents.

One popular method to estimate Hurst exponent from fBm is the multiresolution analysis through wavelet transformations, see Abry et al. (2000, 1995, 2009). The idea is to explore the fact that H is linearly correlated to wavelet coefficients d_j's at level j on the log-scale, and the following two estimation methods of H have been proposed: 1) Veitch & Abry (VA) method in Abry et al. (2000) by weighted least square regression on the level-wise $\log_2\left(\bar{d}_j^2\right)$; 2) Soltani, Simard, & Boichu (SSB) method in Soltani et al. (2004) by first defining a mid-energy as $D_{j,k} = \left(d_{j,k}^2 + d_{j,k+N_j/2}^2\right)/2$, then taking the mean of the logarithm of mid-energies, and last applying weighted least square regression. Later Shen et al. (2007) demonstrated that the SSB method yielded more accurate estimators than the VA method since it took the logarithm first and then averaged. Unfortunately, both methods are sensitive to outlier coefficients and outlier multiresolution levels, inter and within level dependences, and distributional contaminations, thus, it is important to robustify them.

The robust estimation of Hurst exponent H has recently become a topic of interest, see Franzke et al. (2012); Park and Park (2009); Shen et al. (2007); Sheng et al. (2011). Three

robust estimation methods have been developed. The first one is the Theil-type regression (TT) method in Hamilton et al. (2011), which modified the VA method by replacing the weighted least square regression with the Theil-type weighted regression in Theil (1992) to make it less sensitive to outlier levels. The second and third methods are MEDL and MEDLA proposed in Kang and Vidakovic (2017), where the median, not mean, was used for level-wise wavelet coefficients. To be concrete, the MEDL method took median of $\log d_j^2$, and then applied linear regression to estimate H. The MEDLA method is very similar to MEDL, but it took median of sampled $\log\left(\left(d_{jk_1}^2 + d_{jk_2}^2\right)/2\right)$. The difference between these two methods is that MEDL took logarithm on squared wavelet coefficients, while MEDLA was close to the concept in SSB method that paired and averaged wavelet coefficients prior to taking logarithm. 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" provides a practical motivation for examining alternatives that are intermediate in behavior between the very smooth but outlier-sensitive mean and the very outlier-insensitive but non-smooth median.

In this article, we propose to robustly estimate the Hurst exponent from the fBm model, where the mean or median of the wavelet coefficients is replaced by a general trimean estimator that is inspired by Tukey (1977) and Gastwirth and Cohen (1970). Here the general trimean estimator is defined as a weighted average of the median and two quantiles symmetric about the median, which balances the tradeoff between median value and extreme values. It turns out that this will yield a robust estimator of the Hurst exponent from PRB data, which in turn allows us to efficiently classify PRB into groups with different degrees of visual impairment.

The remaining structure is as follows. Section 2 introduces the general trimean estimators; Section 3 describes estimation of Hurst exponent using the general trimean estimators and derives the asymptotic distributions of the proposed estimators. Section 4 provides the simulation results and compares the performance of the proposed methods to other standardly used, wavelet-based methods. The proposed methods are illustrated using the real PRB data for visual impairment classification in Section 5. The paper is concluded with a summary and discussion in Section 6. The detailed proofs of all theorems are provided in the Appendix.

2 General Trimean Estimators

In this section, we propose a general trimean estimator under the point estimation context, which will be used later for robust estimation of Hurst exponent.

Let $X_1, ..., X_n$ be i.i.d. continuous random variables with pdf $f(x)$ and cdf $F(x)$ with mean μ. For $0 < p < 1$, let $Y_p = X_{\lfloor np \rfloor:n}$ denote a sample pth quantile, where $\lfloor np \rfloor$ denotes the greatest integer that is less than or equal to np. In the context of outliers, a remarkably efficient robust estimator of population mean μ is Tukey's trimean estimator in Tukey (1977), defined as

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

Another robust estimator is the Gastwirth's estimator in Gastwirth and Cohen (1970),

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

Here we propose a general trimean estimator, which is defined as a weighted average of the distribution's median and its two quantiles Y_p and Y_{1-p}, for $p \in (0, 1/2)$:

$$\hat{\mu} = \frac{\alpha}{2}\, Y_p + (1-\alpha)\; Y_{1/2} + \frac{\alpha}{2}\, Y_{1-p}. \tag{3}$$

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 is more robust then mean but smoother than the median. It turns out that Tuckey's trimean estimator and Gastwirth's estimator are two special cases. To be specific, $\alpha = 1/2,\ p = 1/4$ in Tuckey's trimean estimator, and $\alpha = 0.6,\ p = 1/3$ in Gastwirth's estimator.

To derive its asymptotic distribution, we need to first define some notations for the population distribution. Let $0 < p < 1$ and ξ_p denotes the pth quantile of F, so that $\xi_p = \inf\{x|F(x) \geq p\}$. If F is monotone, the pth quantile is simply defined as $F(\xi_p) = p$. Moreover, define

$$A = \left[\frac{\alpha}{2} \quad 1-\alpha \quad \frac{\alpha}{2}\right], \tag{4}$$

$$\boldsymbol{\xi} = \left[\xi_p \quad \xi_{1/2} \quad \xi_{1-p}\right]^T, \tag{5}$$

and the asymptotic covariance matrix of $\boldsymbol{y} = \left[Y_p \quad Y_{1/2} \quad Y_{1-p}\right]^T$ is

$$\Sigma = (\sigma_{ij})_{r\times r}, \qquad \text{with} \qquad \sigma_{ij} = \frac{p_i\,(1-p_j)}{f\,(x_{p_i})\,f\,(x_{p_j})},\ i \leq j, \tag{6}$$

see DasGupta (2008).

Now we are ready to present the asymptotic properties of our proposed general trimean estimator $\hat{\mu}$ in (3).

Lemma 1. *As $n \to \infty$, for $\hat{\mu}$ in (3),*

$$\sqrt{n}(\hat{\mu} - A \cdot \xi) \overset{approx}{\sim} \mathcal{N}\left(0, A\Sigma A^{-1}\right) \tag{7}$$

The proof of Lemma 1 follows directly from the asymptotic joint distribution of the order statistics, and thus are omitted.

3 Robust Estimations of Hurst Exponent

In this section, we propose two different robust methods for estimating Hurst exponent H under the fBm model through wavelet transformations.

For that purpose, let us first provide a brief background on non-decimated wavelet transforms (NDWT), also see Nason and Silverman (1995); Vidakovic (2009); Percival and Walden (2006). The NDWT 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. To be more specific, any square integrable function $f(x) \in \boldsymbol{L}_2(\mathbb{R})$ can be expressed in the wavelet domain as

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

where $c_{J_0,k}$ denotes coarse coefficients, $d_{j,k}$ indicates detail coefficients, $\phi_{J_0,k}(x)$ represents scaling functions, and $\psi_{j,k}(x)$ signifies wavelet functions. For specific choices of scaling and wavelet

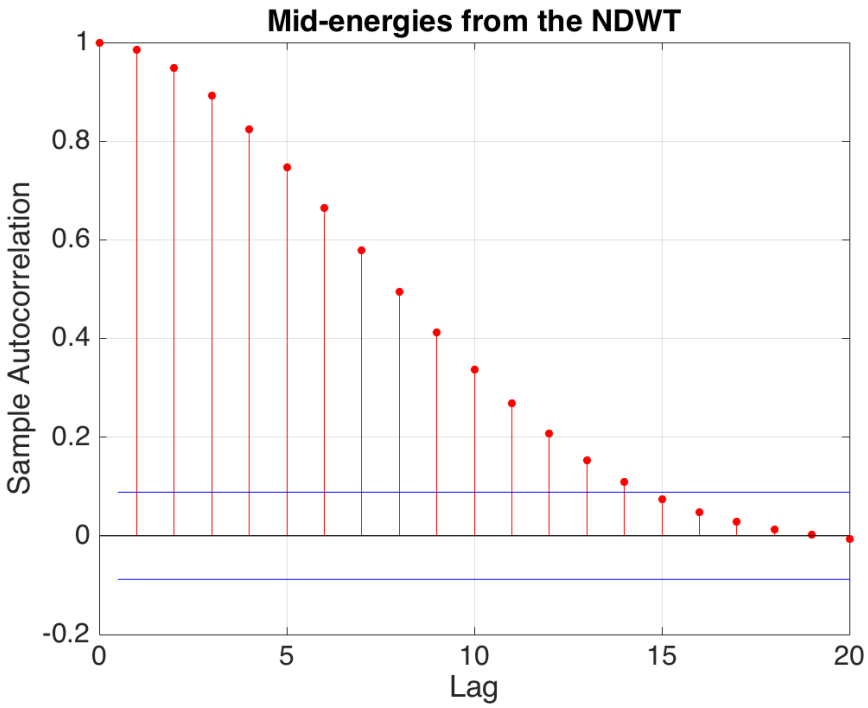

Figure 3: Autocorrelation plot of mid-energies.

functions, the basis for NDWT can be formed from the atoms $\phi_{J_0,k}(x) = 2^{J_0/2}\phi\left(2^{J_0}(x-k)\right)$ 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. A principal difference between NDWT and orthogonal discrete wavelet transforms (DWT) is the sampling rate, and notice that atoms for NDWT have the constant location shift k at all levels, yielding the finest sampling rate on any level. The coefficients are $c_{J_0,k} = \int f(x)\ \phi_{J_0,k}(x)dx$ and $d_{j,k} = \int f(x)\ \psi_{j,k}(x)dx$. In a J-depth decomposition of a fBm of size N, a NDWT generates J detail levels and one smooth level, therefore containing $N \times (J+1)$ wavelet coefficients, N in each level.

In the Hurst exponent estimation literature, most research was based on the standard orthogonal discrete wavelet transforms (DWT), but NDWT turns out to have several advantages when employed for Hurst exponent estimation: 1) Input signals and images of arbitrary size can be processed due to the absence of decimation; 2) as a redundant transform, the NDWT increases the accuracy of the scaling estimation; 3) 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 logged NDWT coefficients do not depend on level; 4) local scaling can be assessed due to the time-invariance property. As we will discuss later, the price we pay is that the dependence of coefficients in NDWT is more profound than in DWT.

At high level, we propose to estimate Hurst exponent from NDWT as follows. At each detail level j, we generate $N/2$ mid-energies as $D_{j,k} = \left(d_{j,k}^2 + d_{j,k+N/2}^2\right)/2$, for $k = 1, 2, ..., N/2$. Then we have two different approaches to robustly estimate Hurst exponent: One is based on mid-energies $D_{j,k}$ themselves, and the other is based on the logarithm of mid-energies $\log D_{j,k}$. In each approach, we first calculate the general trimean estimator on $D_{j,k}$ or $\log D_{j,k}$, and then derive its asymptotic distribution, which depends on Hurst exponent H and allows us to provide

a robust estimation of H.

Here, we use the mid-energies instead of the raw wavelet coefficients. The concept of mid-energies was first introduced by Soltani et al. (2004) to reduce the estimation bias caused by the non-normality of $\log_2 \bar{d}_j^2$ and correlation among detail wavelet coefficients. Soltani et al. (2004) showed that level-wise averages of $\log_2 D_{j,k}$ were asymptotically normal with the mean $-(2H+1)j+C$, which could be used to estimate H by regression. However, the asymptotic distribution was conducted under the independent assumption between mid-energies. Unfortunately, for a fixed detail level j, these mid-energies or the logarithm versions are generally dependent. The good news is that their autocorrelations decay exponentially as their distance increases, see Figure 3. Thus for the practice purpose, we will be able to reduce such dependency by increasing the distance between two consecutive points.

To be specific, we sample every M points from the original $N/2$ mid-energies or their logarithm versions, resulting in M groups in each level j. Note that at each level j, the M groups are generated by switching the starting point from $D_{j,1}$ $(\log D_{j,1})$ to $D_{j,M}$ $(\log D_{j,M})$. Since the distances are large, the $(N/2)/M$ sampled values within each group can be thought of as independent for the practice purpose. The general trimean estimators are then calculated from each of the M groups. Note that M must be divisible by $N/2$.

Group 1: $\{D_{j,1}, D_{j,1+M}, D_{j,1+2M}, ..., D_{j,(N/2-M+1)}\}$
$(\{\log(D_{j,1}), \log(D_{j,1+M}), ..., \log(D_{j,(N/2-M+1)})\})$
Group 2: $\{D_{j,2}, D_{j,2+M}, D_{j,2+2M}, ..., D_{j,(N/2-M+2)}\}$
$(\{\log(D_{j,2}), \log(D_{j,2+M}), ..., \log(D_{j,(N/2-M+2)})\})$
$\vdots$

Group M: $\{D_{j,M}, D_{j,2M}, D_{j,3M}, ..., D_{j,N/2}\}$
$(\{\log(D_{j,M}), \log(D_{j,2M}), ..., \log(D_{j,N/2})\})$

The parameter M affects the efficiency of dependency reductions. Figure 4 shows the autocorrelation plot of grouped mid-energies following the above procedure, with $M = 8$. As can be seen, our procedure efficiently reduces the effect of the correlation among original mid-energies.

In practice, parameter M could be either user specified or determined via grid search based on different needs and criterion. In our application to classify individuals with different degrees of visual impairment, M is chosen to minimize the misclassification rate on the testing data.

Below we will present our proposed two methods in two subsections. Section 3.1 introduces the general trimean of the mid-energy (GTME) method, and Section 3.2 discusses the general trimean of the logarithm of mid-energy (GTLME) method. These two methods are closely related, except switching the order of logarithm and general trimean estimator.

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

Our proposed GTME method involves the following three steps:

1) Compute the general trimean estimators $\hat{\mu}_{j,i}$ on

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

where $D(j,i)$, for $1 \le j \le J$ and $1 \le i \le M$, is the ith group of mid-energies at level j in a J-level NDWT of a fBm of size N with Hurst exponent H.

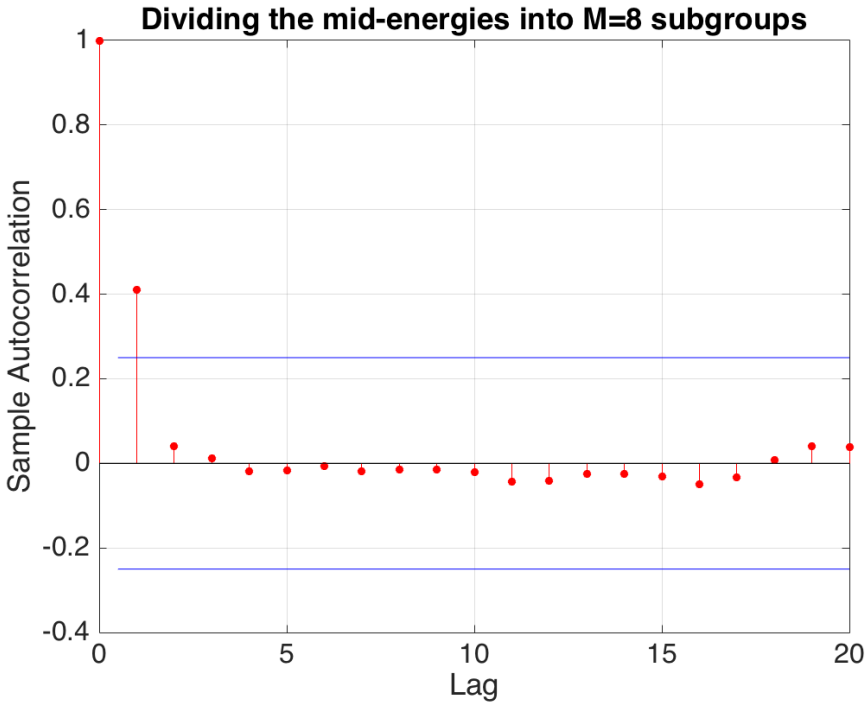

Figure 4: Autocorrelation plot of grouped mid-energies.

2) For each $i = 1, 2, ...M$, calculate the regression slope $\hat{\beta}_i$ in the least square linear regression on pairs $(j, \log_2(\hat{\mu}_{j,i}))$, for $1 \leq J_1 \leq j \leq J_2 \leq J$.

3) Estimate the Hurst exponent by

$$\hat{H}_1 = \frac{1}{M}\sum_{i=1}^{M}\left(-\frac{\hat{\beta}_i}{2} - \frac{1}{2}\right) = -\frac{\bar{\beta}}{2} - \frac{1}{2}, \tag{8}$$

where $\bar{\beta} = \frac{1}{M}\sum_{i=1}^{M}\hat{\beta}_i$ is the average of regression slopes over the M groups for $i = 1, 2, ...M$.

The motivation of GTME method is based on the asymptotic distribution of $\hat{\mu}_{j,i}$ from Lemma 1:

$$\sqrt{N}\left(\hat{\mu}_{j,i} - c(\alpha, p)\lambda_j\right) \overset{\text{approx}}{\sim} \mathcal{N}\left(0, 2Mf(\alpha, p)\lambda_j^2\right), \tag{9}$$

where

$$\begin{aligned} c(\alpha, p) &= \frac{\alpha}{2}\log\left(\frac{1}{p(1-p)}\right) + (1-\alpha)\log 2, \\ f(\alpha, p) &= \frac{\alpha(1-2p)(\alpha-4p)}{4p(1-p)} + 1, \text{and} \\ \lambda_j &= \sigma^2 \cdot 2^{-(2H+1)j}. \end{aligned} \tag{10}$$

Here σ is the standard deviation of wavelet coefficients from level 0. Hence, $\log_2(\hat{\mu}_{j,i})$ is linearly related to $(2H+1)j$, which allows us to use the slopes $\hat{\beta}_i$ to estimate $2H+1$ and leads to the proposed estimator in (8).

For our proposed GTME method in (8), its asymptotic properties are established in the following theorem, whose proof is postponed in the Appendix:

Theorem 3.1. *The estimator $\hat{H}_1$ follows the asymptotic normal distribution*

$$\sqrt{N}\left(\hat{H}_1 - H\right) \overset{approx}{\sim} \mathcal{N}\left(0, V_1\right). \tag{11}$$

The asymptotic variance V_1 is a constant number,

$$\begin{aligned} V_1 &= \frac{6f(\alpha,p)}{(\log 2)^2 (c\,(\alpha,p))^2 q(J_1,J_2)}, \\ q(J_1,J_2) &= (J_2-J_1)(J_2-J_1+1)(J_2-J_1+2), \end{aligned} \tag{12}$$

where $c(\alpha,p)$ and $f(\alpha,p)$ are given in (10).

There are different ways to determine the tuning parameters α and p in general trimean estimator. One could use the settings in Tukey's trimean estimator ($\alpha = 1/2$, $p = 1/4$) or in Gastwirth's estimator ($\alpha = 0.6$, $p = 1/3$). Alternatively, we could find the optimal α and p in the sense of minimizing the asymptotic variance of general trimean estimators $\hat{\mu}_{j,i}$ in (9). To see this, we take partial derivatives of $f\,(\alpha,p)$ with respect to α and p and set them to 0. The optimal α and p can be obtained by solving

$$\begin{aligned} \frac{\partial f\,(\alpha,p)}{\partial \alpha} &= -\frac{2p-1}{2p\,(1-p)}\alpha + \frac{1+p}{2\,(1-p)} - \frac{3}{2} = 0, \\ \frac{\partial f\,(\alpha,p)}{\partial p} &= \frac{\alpha\,(2-\alpha)}{2\,(1-p)^2} + \frac{\alpha^2\,(2p-1)}{4p^2\,(1-p)^2} = 0. \end{aligned} \tag{13}$$

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

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

In this section, we propose our second method, the general trimean of the logarithm of mid-energy (GTLME) method, which takes logarithm first and then calculates the general trimean estimators. The GTLME method involves the following three steps:

1) Calculate the general trimean estimators $\hat{\mu}'_{j,i}$ on

$$\left\{\log\left(D_{j,i}\right), \log\left(D_{j,i+M}\right), ..., \log\left(D_{j,(N/2-M+i)}\right)\right\} := L(j,i),$$

where $L(j,i)$ is the ith group of logarithm of mid-energies at level j in a J-level NDWT of a fBm of size N with Hurst exponent H, $1 \le i \le M$ and $1 \le j \le J$.

2) Obtain the regression slope $\hat{\beta}'_i$ in the least square linear regressions on pairs $\left(j, \hat{\mu}'_{j,i}\right)$ for $1 \leq J_1 \leq j \leq J_2 \leq J$.

3) Estimate the Hurst exponent by

$$\hat{H}_2 = \frac{1}{M}\sum_{i=1}^{M}\left(-\frac{\hat{\beta}'_i}{2\log 2} - \frac{1}{2}\right) = -\frac{1}{2\log 2}\bar{\beta}' - \frac{1}{2}, \tag{14}$$

where $\bar{\beta}' = \frac{1}{M}\sum_{i=1}^{M}\hat{\beta}'_i$ is the average of regression slopes over the M groups, $i = 1, 2, ...M$.

The motivation of GTLME method is from the asymptotic distribution of general trimean estimator $\hat{\mu}'_{j,i}$ that is derived from Lemma 1:

$$\sqrt{N}\left(\hat{\mu}'_{j,i} - (g(\alpha, p) + \log(\lambda_j))\right) \overset{\text{approx}}{\sim} \mathcal{N}(0, 2Mh(\alpha, p)), \tag{15}$$

where

$$\begin{aligned} g(\alpha, p) &= \frac{\alpha}{2}\log\left(\log\frac{1}{1-p}\cdot\log\frac{1}{p}\right) + (1-\alpha)\log(\log 2), \\ h(\alpha, p) &= \frac{\alpha^2}{4h_1(p)} + \frac{\alpha(1-\alpha)}{2h_2(p)} + \frac{(1-\alpha)^2}{(\log 2)^2}. \end{aligned} \tag{16}$$

The $h_1(p)$ and $h_2(p)$ are two functions of p that are provided in the Appendix, $\lambda_j = \sigma^2 \cdot 2^{-(2H+1)j}$, and σ^2 is the variance of wavelet coefficients from level 0.

It is interesting to compare our two proposed methods, GTME and GTLME. The main difference is whether to calculate general trimean estimators before or after taking the logarithm. In GTLME method, the general trimean estimator $\hat{\mu}'_{j,i}$, not the $\log_2(\hat{\mu}'_{j,i})$, is linearly related to $(2H+1)j$, and we use the slopes $\bar{\beta}'_i$ to estimate $2H+1$, therefore leading to the proposed estimator in (14). Based on our extensive experience, the GTME seems more efficient in terms of smaller variance, whereas the GTLME method is more robust to outliers.

The asymptotic distribution of Hurst exponent estimator $\hat{H}_2$ in GTLME method is provided in the following theorem, whose proof is postponed in the Appendix.

Theorem 3.2. *The estimator $\hat{H}_2$ follows the asymptotic normal distribution*

$$\sqrt{N}\left(\hat{H}_2 - H\right) \overset{approx}{\sim} \mathcal{N}(0, V_2). \tag{17}$$

The asymptotic variance V_2 is a constant number,

$$V_2 = \frac{6h(\alpha, p)}{(\log 2)^2 q(J_1, J_2)}, \tag{18}$$

where $q(J_1, J_2)$ is given in (12) and $h(\alpha, p)$ is in (16).

Now we want to determine the optimal tuning parameters α and p. Again, we can set $\alpha = 1/2$ and $p = 1/4$ from Tukey's trimean estimator or $\alpha = 0.6$ and $p = 1/3$ from Gastwirth's estimator. Here, we will find the optimal α and p by minimizing the asymptotic variance of general trimean estimator $\hat{\mu}'_{j,i}$ in (15), and the corresponding results also lead to the smallest asymptotic variance V_2 in (18). They can be obtained by solving

$$\frac{\partial h(\alpha, p)}{\partial \alpha} = 0, \text{ and } \frac{\partial h(\alpha, p)}{\partial p} = 0. \tag{19}$$

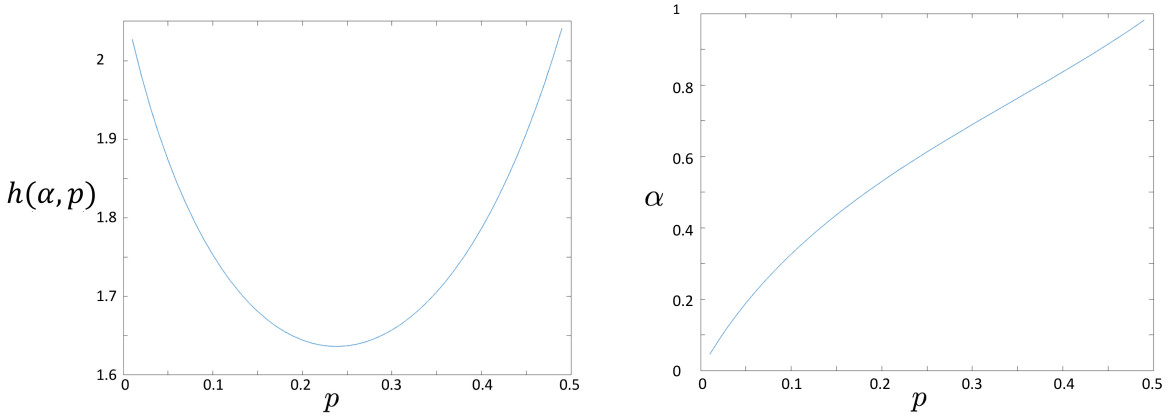

Figure 5: Left: Plot of $h(\alpha, p)$ against p; Right: Plot of α against p.

From the first equation in (19), it can be derived that

$$\alpha = \frac{\frac{2}{\log(2)^2} - \frac{1}{2}h_2(p)}{\frac{1}{2}h_1(p) - h_2(p) + \frac{2}{(\log 2)^2}}. \tag{20}$$

The second equation in (19) cannot be simplified to a finite form. As an illustration, we plot $h(\alpha, p)$ with p ranging from 0 to 0.5, and α is a function of p in (20). The plot of α against p is also shown in Figure 5. Numerical calculation gives $\alpha = 0.5965$ and $p = 0.24$. 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.

Even though we have proved that both the asymptotic distributions of $\hat{H}_1$ and $\hat{H}_2$ do not depend on parameter M, M affects the estimation accuracy. To illustrate this, we simulate 300 fBm(0.7) of length 2^{10} and apply our two methods, GTME and GTLME, with different M's to obtain the estimated Hurst exponent H. The parameters α and p are fixed at the optimal values, in other words, $\alpha = 2 - \sqrt{2}$ and $p = 1 - \sqrt{2}/2$ for GTME, and $\alpha = 0.5965$ and $p = 0.24$ for GTLME. Figure 6 plots the estimation mean square errors (MSEs) against different M's for our two methods. We notice that small M cannot efficiently reduce the dependency among the mid-energies, while large M yields too small number of mid-energies within each group, although the independency can be guaranteed, the estimation accuracy will be sacrificed.

4 Simulation

In this section, we will illustrate our proposed methods via simulation. We simulate one dimensional fBm signals of sizes $N = 2^{10}, N = 2^{11}$, and $N = 2^{12}$ with Hurst exponent $H = 0.3, 0.5, 0.7, 0.8, 0.9$, respectively. NDWT of depth $J = 10$ using Pollen wavelets with angles $\pi/6$ (Daubechies 2), $\pi/4$, $\pi/3$, and $\pi/2$ (Haar) are performed on each simulated signal to obtain wavelet coefficients. Pollen wavelets with different angles generates a family possessing continuum many wavelet bases of various degrees of regularity, see Vidakovic (2002). Special cases of Pollen's representation for $\pi/6$ and $\pi/2$ give Daubechies 2 filter and Haar filter, respectively.

Our proposed methods, GTME and GTLME, are then applied on the NDWT coefficients to estimate Hurst exponent H. We select different combinations of parameters α and p in each

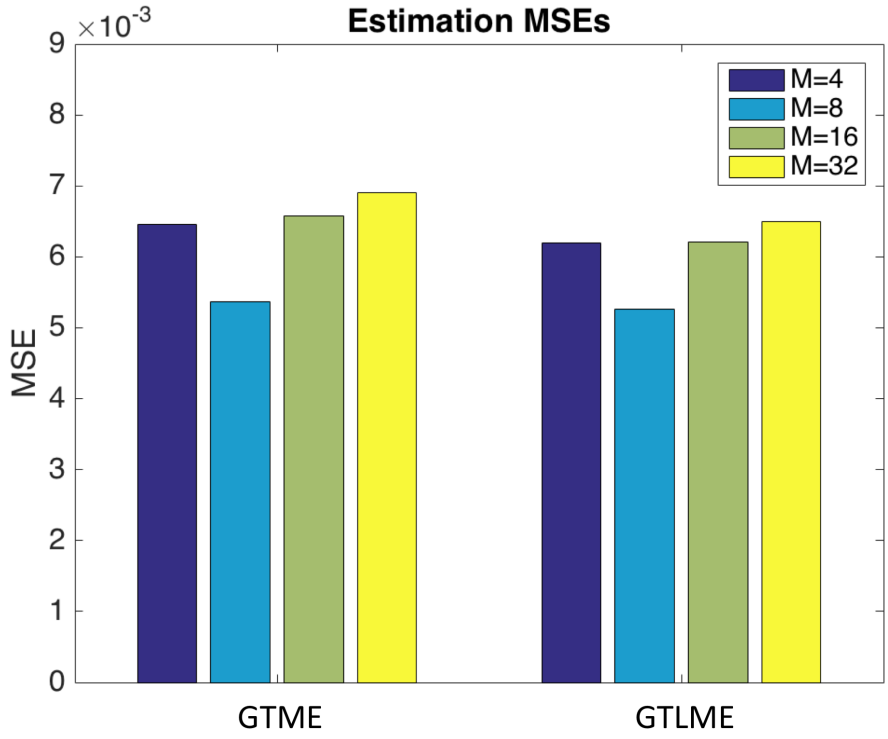

Figure 6: Estimation MSE's against parameter M.

method, leading to the following 6 variations:

$$\text{I:} \quad \text{GTME with } \alpha = \frac{1}{2},\ p = \frac{1}{4}; \tag{21}$$

$$\text{II:} \quad \text{GTLME with } \alpha = \frac{1}{2},\ p = \frac{1}{4}; \tag{22}$$

$$\text{III:} \quad \text{GTME with } \alpha = 0.6,\ p = \frac{1}{3}; \tag{23}$$

$$\text{IV:} \quad \text{GTLME with } \alpha = 0.6,\ p = \frac{1}{3}; \tag{24}$$

$$\text{V:} \quad \text{GTME with } \alpha = 2 - \sqrt{2},\ p = 1 - \frac{\sqrt{2}}{2}; \tag{25}$$

$$\text{VI:} \quad \text{GTLME with } \alpha = 0.5965,\ p = 0.24. \tag{26}$$

Variations I and II are based on Tuckey's trimean estimator in (1), and variations III and IV use Gastwirth's estimator in (2). The α and p in variations V and VI are the optimal values obtained in Section 3 to minimize the corresponding asymptotic variance of general trimean estimator.

Wavelet coefficients on each level are divided into eight groups ($M = 8$), and we use wavelet coefficients from levels 4 ($J_1 = 4$) to 9 ($J_2 = 9$) for the least square linear regression. Note that the performances of our proposed methods depend on the choices of those parameters, and here they are properly chosen by grid search to guarantee good estimations of Hurst exponent. The estimation performance of the proposed methods are compared to five other existing methods: VA method, SSB method, MEDL method, MEDLA method, and TT method, all in the context of NDWT. The main idea of all those existing methods is fitting linear regression and then utilizing

Table 1: Simulation results for $N = 2^{10}$ fBm using Haar wavelet.

	Existing Methods					Proposed Methods					
H	VA	SSB	MEDL	MEDLA	TT	I	II	III	IV	V	VI
					$\hat{H}$						
0.3	0.2759	0.2795	0.2711	0.2720	0.2699	0.2759	0.2746	0.2743	0.2739	0.2748	0.2751
0.5	0.4663	0.5243	0.5116	0.5111	0.4709	0.5179	0.5160	0.5157	0.5148	0.5168	0.5166
0.7	0.5408	0.7320	0.7120	0.7111	0.5451	0.7187	0.7163	0.7172	0.7157	0.7183	0.7168
0.8	0.5393	0.8459	0.8132	0.8128	0.5406	0.8216	0.8188	0.8192	0.8175	0.8206	0.8196
0.9	0.5222	0.9595	0.9155	0.9064	0.5232	0.9114	0.9130	0.9102	0.9109	0.9108	0.9143
					Variances						
0.3	0.0031	0.0032	0.0036	0.0030	**0.0023**	0.0030	0.0031	0.0031	0.0032	0.0030	0.0031
0.5	0.0022	0.0049	0.0050	0.0038	**0.0014**	0.0040	0.0042	0.0041	0.0043	0.0040	0.0043
0.7	0.0052	0.0067	0.0059	**0.0045**	0.0048	0.0048	0.0050	0.0049	0.0050	0.0049	0.0049
0.8	0.0072	0.0083	0.0072	0.0052	0.0068	**0.0050**	0.0054	0.0052	0.0055	**0.0050**	0.0054
0.9	0.0076	0.0112	0.0087	0.0058	0.0075	**0.0056**	0.0062	0.0059	0.0062	0.0058	0.0063
					MSEs						
0.3	0.0037	0.0036	0.0044	0.0037	**0.0032**	0.0035	0.0037	0.0037	0.0038	0.0036	0.0037
0.5	0.0033	0.0055	0.0051	0.0039	**0.0022**	0.0043	0.0045	0.0043	0.0045	0.0043	0.0045
0.7	0.0306	0.0077	0.0060	**0.0046**	0.0288	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052
0.8	0.0751	0.0103	0.0073	**0.0054**	0.0741	**0.0054**	0.0058	0.0056	0.0058	**0.0054**	0.0058
0.9	0.1503	0.0147	0.0089	0.0058	0.1494	**0.0057**	0.0064	0.0060	0.0063	0.0059	0.0065

Note. Different variations of our proposed methods can be found in (21)-(26).

the slope to estimate Hurst exponent H: 1) For VA method, the weighted least square regression is fitted on $(j, \log_2\left(\bar{d}_j^2\right))$, and d_j indicates the wavelet coefficients at level j; 2) the SSB method is fitting a weighted least square regression on $(j, \overline{\log_2 D_j})$, where D_j represents the mid-energies at level j; 3) the MEDL method is to fit an ordinary linear regression on $(j, \text{median}\{\log d_j^2\})$; 4) for the MEDLA method, a simple linear regression is fitted on $\left(j, \text{median}\left\{\log\left(\left(d_{jk_1}^2 + d_{jk_2}^2\right)/2\right)\right\}\right)$, where k_1 and k_2 are the positions of wavelet coefficients that are at least 2^{J-j} apart; 5) the TT method is similar to VA method, but it fits a Theil-type regression on $(j, \log_2\left(\bar{d}_j^2\right))$. The details of those methods have been provided and discussed in the Introduction section. Estimation performance is reported in terms of mean, variance, and mean square error (MSE) based on 300 repetitions for each case.

The proposed methods preform the best using Haar wavelet (Pollen wavelets with angle $\pi/2$), and the simulation results are shown in Table 1 to Table 3 for fBm of sizes $N = 2^{10}, N = 2^{11}$, and $N = 2^{12}$, respectively. Similar results are obtained for other wavelets. For each H (corresponding to each row in the table), the smallest variances and MSEs are highlighted in bold. Compared with those existing methods, our methods yield significantly smaller variances and MSEs when H is large ($H >= 0.7$). When $H = 0.3$ and 0.5, the proposed methods perform better than the SSB method, similar to the MEDL and the MEDLA, but worse than the VA and the TT methods,

Table 2: Simulation results for $N = 2^{11}$ fBm using Haar wavelet

	Existing Methods					Proposed Methods					
H	VA	SSB	MEDL	MEDLA	TT	I	II	III	IV	V	VI
					$\hat{H}$						
0.3	0.2808	0.2767	0.2747	0.2749	0.2707	0.2739	0.2736	0.2739	0.2738	0.2738	0.2736
0.5	0.4811	0.5137	0.5023	0.4979	0.4830	0.5038	0.5024	0.5012	0.5005	0.5028	0.5034
0.7	0.5423	0.7311	0.7036	0.7075	0.5428	0.7126	0.7106	0.7109	0.7096	0.7116	0.7110
0.8	0.5386	0.8417	0.8114	0.8095	0.5369	0.8155	0.8148	0.8130	0.8131	0.8141	0.8156
0.9	0.5258	0.9449	0.9055	0.9016	0.5229	0.9055	0.9075	0.9047	0.9062	0.9048	0.9083
					Variances						
0.3	0.0017	0.0014	0.0018	0.0013	0.0014	**0.0012**	0.0014	0.0013	0.0014	**0.0012**	0.0014
0.5	0.0011	0.0021	0.0021	0.0016	**0.0007**	0.0017	0.0019	0.0019	0.0020	0.0017	0.0018
0.7	0.0041	0.0039	0.0034	0.0026	0.0039	**0.0025**	0.0028	0.0028	0.0029	0.0026	0.0027
0.8	0.0051	0.0061	0.0044	0.0038	0.0052	**0.0036**	0.0041	0.0038	0.0041	0.0037	0.0041
0.9	0.0064	0.0046	0.0041	0.0026	0.0063	**0.0021**	0.0024	0.0023	0.0024	0.0022	0.0024
					MSEs						
0.3	0.0021	**0.0019**	0.0024	0.0020	0.0023	**0.0019**	0.0021	0.0020	0.0021	**0.0019**	0.0021
0.5	0.0014	0.0023	0.0021	0.0016	**0.0010**	0.0017	0.0019	0.0018	0.0020	0.0017	0.0018
0.7	0.0289	0.0048	0.0034	0.0027	0.0286	**0.0026**	0.0028	0.0028	0.0030	0.0027	0.0028
0.8	0.0733	0.0078	0.0045	**0.0038**	0.0744	**0.0038**	0.0043	0.0039	0.0042	**0.0038**	0.0043
0.9	0.1463	0.0065	0.0041	0.0026	0.1484	**0.0021**	0.0024	0.0023	0.0024	0.0022	0.0025

Note. Different variations of our proposed methods can be found in (21)-(26).

Table 3: Simulation results for $N = 2^{12}$ fBm using Haar wavelet.

	Existing Methods					Proposed Methods					
H	VA	SSB	MEDL	MEDLA	TT	I	II	III	IV	V	VI
					$\hat{H}$						
0.3	0.2773	0.2736	0.2693	0.2708	0.2682	0.2710	0.2707	0.2708	0.2705	0.2709	0.2710
0.5	0.4781	0.5008	0.4891	0.4896	0.4790	0.4932	0.4923	0.4920	0.4915	0.4925	0.4926
0.7	0.5482	0.7170	0.6964	0.6998	0.5463	0.7052	0.7030	0.7045	0.7031	0.7052	0.7031
0.8	0.5352	0.8186	0.7948	0.7970	0.5319	0.8021	0.8004	0.8009	0.7999	0.8014	0.8005
0.9	0.5357	0.9200	0.8933	0.8913	0.5317	0.8933	0.8943	0.8933	0.8939	0.8932	0.8946
					Variances						
0.3	0.0010	0.0008	0.0009	0.0008	**0.0007**	0.0008	0.0008	0.0008	0.0008	0.0008	0.0008
0.5	0.0005	0.0011	0.0014	0.0011	**0.0004**	0.0010	0.0011	0.0010	0.0011	0.0010	0.0010
0.7	0.0045	0.0018	0.0021	0.0014	0.0046	**0.0013**	0.0015	0.0014	0.0015	0.0014	0.0015
0.8	0.0049	0.0021	0.0022	0.0016	0.0047	**0.0014**	0.0017	0.0015	0.0017	0.0015	0.0016
0.9	0.0079	0.0038	0.0038	0.0023	0.0075	**0.0020**	0.0024	0.0022	0.0024	0.0021	0.0024
					MSEs						
0.3	**0.0015**	**0.0015**	0.0019	0.0016	0.0017	0.0016	0.0016	0.0016	0.0016	0.0016	0.0016
0.5	**0.0010**	0.0011	0.0015	0.0012	0.0008	**0.0010**	0.0011	0.0011	0.0012	**0.0010**	0.0011
0.7	0.0275	0.0021	0.0021	0.0014	0.0282	**0.0013**	0.0015	0.0014	0.0015	0.0014	0.0015
0.8	0.0750	0.0024	0.0022	0.0016	0.0765	**0.0014**	0.0016	0.0015	0.0017	**0.0014**	0.0016
0.9	0.1405	0.0042	0.0038	0.0023	0.1431	**0.0021**	0.0024	0.0023	0.0024	0.0022	0.0024

Note. Different variations of our proposed methods can be found in (21)-(26).

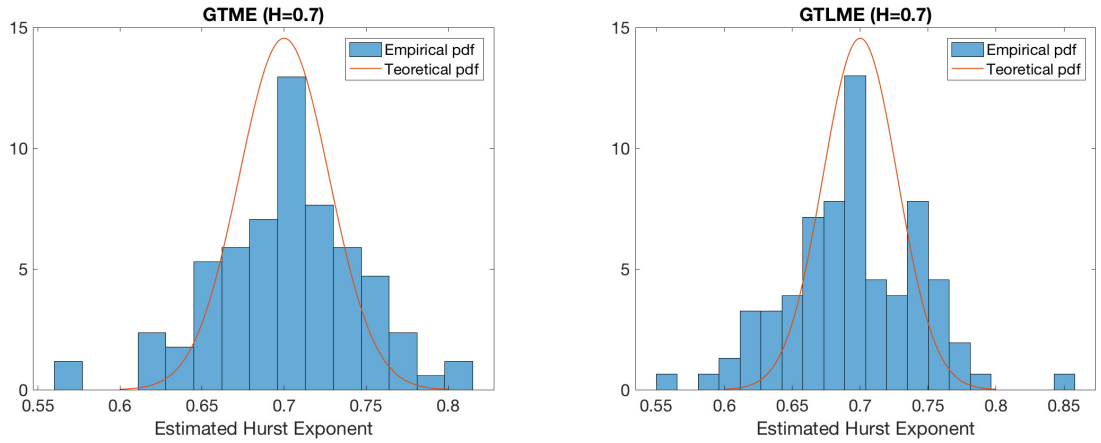

Figure 7: Histograms and theoretical distributions of $\hat{H}$.

however, both VA and TT methods have bad performances when H increases. Besides, the estimation variances of our proposed methods decrease as the length of the fBm signal increases. The performances of our 6 variations are different, but still very similar, regarding variances and MSEs. It is due to the fact that they use different combinations of α and p in the general trimean estimator, however, the values used in those variations are very similar, as is shown in (21)-(26). Besides, the variances of variation V and VI that are based on the optimal α's and p's are not always the smallest, compared to other variations. The optimal values that we obtained through theoretical analysis provide useful suggestions in reality, however, since the independency constraints can be easily violated, it is still beneficial to empirically check other possible values of α and p to obtain the most accurate results. Based on our simulation study, we notice that here variation I based on Tukey's trimean estimator of the mid-energy has the best performance.

Figure 7 shows the histograms of $\hat{H}$ using the GTME method with optimal parameters given in (25) and the GTLME method with optimal parameters in (26), respectively. To generate these plots, we simulate 300 fBm(0.7) with size 2^{10} and apply the proposed two methods to obtain the estimated $\hat{H}$. The solid curves in the plots are the theoretical distributions of $\hat{H}$ given in theorems 3.1 and 3.2. As can be seen, the empirical distribution of $\hat{H}$ is close to the theoretical results given in the theorems.

5 Application

In this section, we apply the proposed methods to PRB data in order to classify individuals according to their visual impairment. Participants in this study consists of 24 older adults, solicited from the patient pool of the Bascom Palmer Eye Institute of the University of Miami School of Medicine. Participants were selected on the basis of having either no ocular disease or only Age-related Macular Degeneration (AMD), as assessed by patient history and clinical testing.

Participants were assigned to three groups: one control group, and two experimental groups I and II. The control group is a set of individuals with healthy, unaffected vision and no evidence of any ocular disease or trauma. Individuals in two experimental groups had varying visual acuity and were diagnosed with AMD. Patients in group II had more severe visual impairment

Table 4: Group characterization summary.

Group	N	Visual Acuity	AMD	Number of original data	Number of cleaned data
Control	6	20/20-20/40	No	60	49
I	8	20/20-20/50	Yes	100	92
II	10	20/60-20/100	Yes	96	262

Note. N represents the number of individuals in the group; Visual Acuity signifies the range of Snellen acuity scores for the individuals in the given group; AMD indicates whether the individuals were diagnosed with age-related macular degeneration or not; Number of original data and Number of cleaned data show the number of 2048 length original data and cleaned data, respectively.

than those in group I. The number of participants is 6 in control group, 8 in group I, and 10 in group II. In Introduction part, we have already shown in Figure 1 the examples of raw PRB data of three different individuals from control group, group I, and group II, respectively.

Researchers have utilized simple statistical methods for analyzing PRB, for example, comparing the relative mean or variance of pupil size deviation in response to stimuli; some sophisticated techniques have also been utilized, like power, frequency and spectral analysis using mathematical tools. However, they failed to characterize the underlying patterns within time series PRB data. Wavelet analysis to estimate the Hurst exponent of the high-frequency, time series physiological data is a useful tool for detecting these hidden patterns and differentiating individuals based on these unique patterns in their physiological behavior.

In this section, we propose to use 1-D fractional Brownian motion (fBm) to model the PRB data. We have provided brief introduction to fBm in Section 1, and the examples of 1-D fBm are plotted in Figure 2. Intuitively, the PRB data, as is shown in Figure 1, looks close to 1-D fBm, except that it is noisy due to the human blinks and the noise caused by equipment. To reduce such noise, we manually remove the blinks and equipment artifacts from the original signals. In order to illustrate the robustness of our methods, the proposed methods and other existing methods will be applied to both the original, noisy data and the cleaned data with blink and equipment artifacts removed.

Like in many human-subject studies, the limited number of participants is a major disadvantage, but in PRB data set, each subject has enough measurements to segment into multiple pieces with a length of 2048 observations. The number of 2048 length original data and cleaned data within each group are shown in Table 4. Although this induces dependence between the data, we will use hierarchical models to accommodate for the subject induced dependence later. Here we use the same 6 variations of our proposed methods as those in Simulation section, and the corresponding parameter settings can be found in (21)-(26). The parameters $M = 8$, $J_1 = 3$, and $J_2 = 7$ in our methods are selected by grid search to minimize the misclassification rate on testing data. The classification performances of the proposed methods are compared to five other methods: VA, SSB, MEDL, MEDLA, and TT. They are implemented as described in the Simulation section.

Table 5 and Table 6 provide descriptive statistics of the estimated Hurst exponent $\hat{H}$ from original data and cleaned data, respectively. As can be seen from Table 6, for the cleaned data, the control group exhibited the smallest value for $\hat{H}$ in both the mean and median. In fact, signals with smaller Hurst exponent H tend to be more disordered and unsystematic, therefore individuals without visual impairment tend to have more disordered pupil diameter signals. However, for the original data, control group did not exhibit the smallest $\hat{H}$ due to the noise

Table 5: Descriptive statistics group summary (original noisy data).

	Existing Methods					Proposed Methods					
H	VA	SSB	MEDL	MEDLA	TT	I	II	III	IV	V	VI
					Mean of $\hat{H}$						
Control	0.2206	0.4242	0.3583	0.3660	0.2740	0.3524	0.3602	0.3602	0.3615	0.3567	0.3590
I	0.2781	0.5698	0.4195	0.4201	0.3346	0.4229	0.4255	0.4261	0.4262	0.4254	0.4246
II	0.1949	0.4391	0.3522	0.3297	0.2306	0.3099	0.3273	0.3214	0.3292	0.3163	0.3260
					Median of $\hat{H}$						
Control	0.2336	0.4511	0.3416	0.3597	0.2795	0.3734	0.3681	0.3443	0.3540	0.3479	0.3719
I	0.2537	0.5696	0.4036	0.4227	0.3301	0.4219	0.4255	0.4197	0.4283	0.4174	0.4262
II	0.2107	0.4322	0.3686	0.3396	0.2544	0.3248	0.3370	0.3369	0.3406	0.3319	0.3377
					Variance of $\hat{H}$						
Control	0.0197	0.0358	0.0349	0.0414	0.0191	0.0327	0.0333	0.0356	0.0353	0.0340	0.0317
I	0.0190	0.0344	0.0153	0.0158	0.0177	0.0186	0.0166	0.0174	0.0168	0.0181	0.0167
II	0.0225	0.0381	0.0190	0.0164	0.0293	0.0185	0.0167	0.0173	0.0167	0.0178	0.0167

Note. Different variations of our proposed methods can be found in (21)-(26).

caused by blinks and equipment artifacts, which can be seen from Table 5.

The objective is to classify the visual impairment groups based on the estimated Hurst exponent for a given 2048 length pupil diameter data. Before doing the classification, we need to first deal with subject induced dependence through the following hierarchical model. If we denote i to be the group index where the piece of observations is from, with $i = 0$ for control group, $i = 1$ for group I, $i = 2$ for group II, and n_j as the number of pieces generated from subject j (j=1,2,..,24), the estimated Hurst exponent $\hat{H}_{ijk}$ for the kth piece of subject j nested in group i can be expressed in the following model:

$$\hat{H}_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{ijk}, \tag{27}$$

where μ is the overall mean, α_i is the effect for ith group, $\beta_{j(i)}$ is the effect for jth participant within ith group, and ϵ_{ijk} is the random error. In avoid of dependency between data due to subject effects, the estimated $\hat{\beta}_{j(i)}$ is first subtracted from $\hat{H}_{ijk}$, and then multinomial logistic regression model is fitted on the data $\left\{\left(\hat{H}_{ijk} - \hat{\beta}_{j(i)}, i\right), i = 0, 1, 2, j = 1, ..., 24, k = 1, ..., n_j\right\}$. To test the model performance, we randomly choose 80% of the data points to form a training set, and the remaining 20% forms the testing set. Model is developed on the training set and applied on the testing set.

Misclassification rates are reported in Table 7. When doing the classification on the original noised data, our robust methods performed the best, with the minimal misclassification error 37.21%. On the blinks-removed data, our methods outperformed or were comparable to other methods. In general, our methods provide a robust tool to classify different degrees of visual impairment for AMD patients.

Table 6: Descriptive statistics group summary (cleaned data).

	Existing Methods					Proposed Methods					
H	VA	SSB	MEDL	MEDLA	TT	I	II	III	IV	V	VI
					Mean of $\hat{H}$						
Control	0.1811	0.3650	0.3025	0.2804	0.2466	0.2722	0.2826	0.2806	0.2851	0.2772	0.2821
I	0.2751	0.5601	0.4048	0.4092	0.3311	0.3926	0.4040	0.4013	0.4076	0.3975	0.4031
II	0.2088	0.4356	0.3494	0.3246	0.2489	0.3028	0.3198	0.3140	0.3219	0.3085	0.3186
					Median of $\hat{H}$						
Control	0.1775	0.3801	0.3311	0.3010	0.2387	0.3012	0.3065	0.3043	0.3070	0.3004	0.3040
I	0.2729	0.5210	0.4168	0.4117	0.3354	0.3959	0.4122	0.4033	0.4113	0.4011	0.4105
II	0.2301	0.4227	0.3580	0.3380	0.2865	0.3121	0.3329	0.3237	0.3345	0.3171	0.3312
					Variance of $\hat{H}$						
Control	0.0094	0.0238	0.0126	0.0116	0.0091	0.0122	0.0122	0.0123	0.0122	0.0124	0.0121
I	0.0077	0.0310	0.0097	0.0131	0.0057	0.0121	0.0124	0.0124	0.0126	0.0122	0.0123
II	0.0149	0.0390	0.0182	0.0153	0.0195	0.0148	0.0152	0.0149	0.0152	0.0147	0.0152

Note. Different variations of our proposed methods can be found in (21)-(26).

Table 7: Classification error.

	Existing Methods					Proposed Methods					
	VA	SSB	MEDL	MEDLA	TT	I	II	III	IV	V	VI
Blinks removed	0.4568	0.4074	0.4691	0.3704	0.4444	0.3951	0.3827	0.3827	0.3827	0.3951	0.3827
Original data	0.4651	0.3953	0.4535	0.3837	0.4419	0.3721	0.3837	0.3721	0.3721	0.3837	0.3721

Note. Different variations of our proposed methods can be found in (21)-(26).

6 Conclusions

In this paper, we proposed two methods, GTME and GTLME, to improve the robust estimation of Hurst exponent from the fractional Brownian motion through wavelet transformations. The three key ideas in our proposed methods are: 1) We define a general trimean estimator that is a weighted average of median and two quantiles, and it turns out that the well known Tukey's trimean estimator and Gastwirth estimator are two special cases under this framework; 2) When utilizing non-decimated wavelet transforms (NDWT) wavelet coefficients to obtain Hurst exponent estimators, we reduce the dependency of NDWT wavelet coefficients by rearranging each level coefficients into groups, so that the distance between any two points within the same group are large enough; 3) Instead of using mean or median, we apply the general trimean estimator to wavelet coefficients (GTME) or the logarithm of wavelet coefficients (GTLME), and then derive its asymptotic distribution, which depends on Hurst exponent H and leads to the robust estimation of H.

The estimation performance of the proposed methods were compared to five other existing methods: Veitch & Abry (VA) method, Soltani, Simard, & Boichu (SSB) method, MEDL method, MEDLA method, and Theil-type regression (TT) method. Simulation results indicated our proposed methods yielded smaller variance and MSEs when estimating Hurst exponent H, in particular for large H's. For fBm with small to moderate Hurst exponent, for example $H = 0.3$ or 0.5, our methods still outperformed SSB, MEDL and MEDLA, and were comparable to VA and TT.

Our proposed two methods have been applied to a real pupillary response behavior (PRB) data set for visual impairment classification. The unique pattern of PRB data cannot be efficiently represented by the trends or traditional statistical summaries of the signal, and our proposed methods helped to detect those unique patterns by estimating the Hurst exponent from the data. The estimated Hurst exponent was then used as a predictor in the multinomial logistic regression model to classify individuals with different degrees of visual impairment. It turns out that our robust methods yielded the smallest three-class misclassification rate 37.21% on the noisy PRB data. Besides, we noticed that the healthy group exhibited the smallest value for estimated Hurst exponent H, which indicated individuals without visual impairment had more disordered signals. This is common for many other biometric signals: EEG, EKG, high frequency protein mass-spectra, high resolution medical images of tissue, to list a few.

Acknowledgment

The authors were supported in part by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR000454 as well as NSF awards DMS-1613258 and DMS-1830344 at Georgia Institute of Technology.

Appendix

In the Appendix, we include the detailed proof of Theorem 3.1 and Theorem 3.2.

Proof of Theorem 3.1

Proof. A single wavelet coefficient in a non-decimated wavelet transform of a fBm of size N with Hurst exponent H is normally distributed, with variance depending on its level j, therefore, each

pair $d_{j,k}$ and $d_{j,k+N/2}$ in mid-energy $D_{j,k}$ are assumed to be independent and follow the same normal distribution.

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

Then the mid-energy is defined as

$$D_{j,k} = \frac{\left(d_{j,k}^2 + d_{j,k+N/2}^2\right)}{2},\ j = 1, .., J, \text{and } k = 1, ..., N/2,$$

and it can be readily shown that $D_{j,k}$ has exponential distribution with scale parameter $\lambda_j = \sigma^2 \cdot 2^{-(2H+1)j}$, i.e.,

$$f\left(D_{j,k}\right) = \lambda_j^{-1} e^{-\lambda_j^{-1} D_{j,k}},\ \text{for any } k = 1, .., N/2.$$

Therefore the ith subgroup $\left\{D_{j,i}, D_{j,i+M}, D_{j,i+2M}, ..., D_{j,(N/2-M+i)}\right\}$ are i.i.d. $\exp\left(\lambda_j^{-1}\right)$, and when applying general trimean estimator $\hat{\mu}_{j,i}$ on it, following the derivation in Section 2, we have

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

and

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

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

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

and variance

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

Since the Hurst exponent can be estimated as

$$\hat{H}_1 = -\frac{\bar{\beta}}{2} - \frac{1}{2}, \tag{28}$$

where $\bar{\beta} = 1/M \sum_{i=1}^{M} \hat{\beta}_i$ is the average regression slope in the least square linear regression on pairs $\left(j, \log_2\left(\hat{\mu}_{j,i}\right)\right)$ from level J_1 to J_2, $J_1 \le j \le J_2$. It can be easily derived that each $\hat{\beta}_i$ is a linear combination of $\log_2\left(\hat{\mu}_{j,i}\right)$,

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

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

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

based on which we calculate

$$\begin{aligned} &\mathbb{E}\left(\log_2\left(\hat{\mu}_{j,i}\right)\right) = -(2H+1)j + \text{ Constant, and} \\ &\text{Var}\left(\log_2\left(\hat{\mu}_{j,i}\right)\right) = \frac{\frac{2M}{N} f(\alpha,p)}{(\log 2)^2 c^2(\alpha,p)} \end{aligned}$$

Therefore

$$\begin{aligned} &\mathbb{E}\left(\hat{\beta}_i\right) = \sum_{j=J_1}^{n} a_j \mathbb{E}\left(\log_2\left(\hat{\mu}_{j,i}\right)\right) = -(2H+1), \text{ and} \\ &\text{Var}\left(\hat{\beta}_i\right) = \sum_{j=J_1}^{J_2} a_j^2 \,\text{Var}\left(\log_2\left(\hat{\mu}_{j,i}\right)\right) := 4M \times V_1 \end{aligned}$$

and

$$\mathbb{E}\left(\hat{H}_1\right) = H, \text{ and } \text{Var}\left(\hat{H}_1\right) = \frac{1}{N} \cdot V_1 \tag{29}$$

where the asymptotic variance V_1 is a constant number independent of group number M and level j,

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

and

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

□

Proof of Theorem 3.2

Proof. We have stated that $D_{j,k} \sim \mathcal{E}xp\left(\lambda_j^{-1}\right)$ with scale parameter $\lambda_j = \sigma^2 \cdot 2^{-(2H+1)j}$, so that

$$f\left(D_{j,k}\right) = \lambda_j^{-1} e^{-\lambda_j^{-1} D_{j,k}}, \text{ for any } k = 1, .., N/2.$$

Let $y_{j,k} = \log\left(D_{j,k}\right)$ for any $j = 1, ..., J$ and $k = 1, ..., N/2$. The pdf and cdf of $y_{j,k}$ are

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

and

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

The p-quantile can be obtained by solving $F\left(y_p\right) = 1 - e^{-\lambda_j^{-1} e^{y_p}} = p$, and $y_p = \log\left(-\lambda_j \log(1-p)\right)$. Then it can be shown that $f\left(y_p\right) = -(1-p)\log(1-p)$. When applying the general trimean estimator $\hat{\mu}'_{j,i}$ on

$$\left\{\log\left(D_{j,i}\right), \log\left(D_{j,i+M}\right), ..., \log\left(D_{j,(N/2-M+i)}\right)\right\},$$

following the derivation in Section 2, we get

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

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

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

and variance

$$\begin{aligned} \mathrm{Var}\left(\hat{\mu}'_{j,i}\right) &= \frac{1}{N/16} A\Sigma A^T \\ &= \frac{2M}{N}\left(\frac{\alpha^2}{4}h_1(p) + \frac{\alpha(1-\alpha)}{2}h_2(p) + \frac{(1-\alpha)^2}{(\log 2)^2}\right) \\ &\triangleq \frac{2M}{N}h(\alpha, p) \end{aligned}$$

where

$$\begin{aligned} h_1\left(p\right) =& \frac{p}{(1-p)\left(\log\left(1-p\right)\right)^2} + \\ & \frac{1-p}{p\left(\log p\right)^2} + \frac{2p}{(1-p)\log\left(1-p\right)\log p}, \end{aligned}$$

and

$$h_2\left(p\right) = \frac{2p}{(1-p)\log\left(1-p\right)\log\frac{1}{2}} + \frac{2}{\log\frac{1}{2}\log p}.$$

Since the Hurst exponent can be estimated as

$$\hat{H}_2 = -\frac{1}{2\log 2}\bar{\beta}' - \frac{1}{2}, \tag{30}$$

where $\bar{\beta}' = 1/M\sum_{i=1}^{M}\hat{\beta}'_i$ is the avearge regression slope in the least square linear regressions on pairs $\left(j, \hat{\mu}'_{j,i}\right)$ from level J_1 to J_2, $J_1 \le j \le J_2$. It can be easily derived that $\hat{\beta}'_i$ is a linear combination of $\hat{\mu}'_{j,i}$,

$$\hat{\beta}'_i = \sum_{j=J_1}^{J_2} a_j \hat{\mu}'_{j,i}, \quad a_j = \frac{j - (J_1+J_2)/2}{\sum_{j=J_1}^{J_2}\left(j-(J_1+J_2)/2\right)^2}.$$

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

$$\begin{aligned} \mathbb{E}\left(\hat{\beta}'_i\right) &= \textstyle\sum_{j=J_1}^{J_2} a_j \mathbb{E}\left(\hat{\mu}'_{j,i}\right) = -(2H+1)\log 2, \text{ and} \\ \mathrm{Var}\left(\hat{\beta}'_i\right) &= \textstyle\sum_{j=J_1}^{J_2} a_j^2 \,\mathrm{Var}\left(\hat{\mu}'_{j,i}\right) := 4(\log 2)^2 M \times V_2 \end{aligned}$$

and

$$V_2 = \frac{6f(\alpha, p)}{(\log 2)^2 q\left(J_1, J_2\right)}, \tag{31}$$

where the asymptotic variance V_2 is a constant number independent of group number M and level j,

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

and

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

□

References

Abry P (2003). Scaling and wavelets: An introductory walk. In: *Processes with long-range correlations* (G Rangarajan, M Ding, eds.), 34–60. Springer.

Abry P, Flandrin P, Taqqu MS, Veitch D (2000). Wavelets for the analysis, estimation, and synthesis of scaling data. In: *Self-Similar Network Traffic and Performance Evaluation* (K Park, W Willinger, eds.), 39–88. Wiley Online Library.

Abry P, Flandrin P, Taqqu MS, Veitch D, et al. (2003). Self-similarity and long-range dependence through the wavelet lens. In: *Theory and applications of long-range dependence* (G Rangarajan, M Ding, eds.), 527–556. Birkhäuser,.

Abry P, Gonçalvés P, Flandrin P (1995). Wavelets, spectrum analysis and $1/f$ processes. In: *Wavelets and Statistics* (A Antoniadis, G Oppenheim, eds.), 15–29. Springer.

Abry P, Goncalves P, Véhel JL (2009). *Scaling, Fractals and Wavelets*. John Wiley & Sons, Ltd.

Avery RL, Pieramici DJ, Rabena MD, Castellarin AA, Ma'an AN, Giust MJ (2006). Intravitreal Bevacizumab (Avastin) for Neovascular Age-related macular degeneration. *Ophthalmology*, 113(3): 363–372.

DasGupta A (2008). *Asymptotic Theory of Statistics and Probability*. Springer Science & Business Media.

Engel J Jr, Bragin A, Staba R, Mody I (2009). High-frequency oscillations: What is normal and what is not? *Epilepsia*, 50(4): 598–604.

Fletcher DC, Schuchard RA (2006). Visual function in patients with choroidal neovascularization resulting from age-related macular degeneration: The importance of looking beyond visual acuity. *Optometry and Vision Science*, 83(3): 178–189.

Franzke CL, Graves T, Watkins NW, Gramacy RB, Hughes C (2012). Robustness of estimators of long-range dependence and self-similarity under on-Gaussianity. *Philosophical Transactions of the Royal Society A*, 370(1962): 1250–1267.

Gastwirth JL, Cohen ML (1970). Small sample behavior of some robust linear estimators of location. *Journal of the American Statistical Association*, 65(330): 946–973.

Geraci JM, Ashton CM, Kuykendall DH, Johnson ML, Wu L (1997). International classification of diseases, 9th revision, clinical modification codes in discharge abstracts are poor measures of complication occurrence in medical inpatients. *Medical Care*, 35(6): 589–602.

Gregoriou GG, Gotts SJ, Zhou H, Desimone R (2009). High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science*, 324(5931): 1207–1210.

Hamilton EK, Jeon S, Cobo PR, Lee KS, 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*, 384–389.

Kang M, Vidakovic B (2017). MEDL and MEDLA: Methods for Assessment of Scaling by

Medians of Log-squared Nondecimated Wavelet Coefficients. ArXiv preprint https://arxiv.org/abs/1703.04180.

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.

Kolmogorov AN (1940). Wienersche spiralen und einige andere interessante Kurven in Hilbertscen Raum, C. R. (doklady). *Acad. Sci. URSS (NS)*, 26: 115–118.

Mandelbrot BB, Van Ness JW (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10(4): 422–437.

Moloney KP, Jacko JA, Vidakovic B, Sainfort F, Leonard VK, Shi B (2006). Leveraging data complexity: Pupillary behavior of older adults with visual impairment during HCI. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 13(3): 376–402.

Nason GP, Silverman BW (1995). The stationary wavelet transform and some statistical applications. In: *Wavelets and Statistics* (A Antoniadis, G Oppenheim, eds.), 281–299. Springer.

Park J, Park C (2009). Robust estimation of the Hurst parameter and selection of an onset scaling. *Statistica Sinica*, 19(4): 1531–1555.

Park K, Willinger W (2000). *Self-Similar Network Traffic and Performance Evaluation*. John Wiley & Sons.

Percival DB, Walden AT (2006). *Wavelet Methods for Time Series Analysis*, volume 4. Cambridge University Press.

Rosenfeld PJ, Brown DM, Heier JS, Boyer DS, Kaiser PK, Chung CY, et al. (2006). Ranibizumab for Neovascular Age-related macular degeneration. *New England Journal of Medicine*, 355(14): 1419–1431.

Shen H, Zhu Z, Lee TC (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 Theil's Contributions to Economics and Econometrics* (B Raj, J Koerts, eds.), 345–381. Springer.

Tukey JW (1977). *Exploratory Data Analysis*, volume 2. Reading, MA.

Vidakovic B (2002). Pollen bases and daubechies-lagarias algorithm in Matlab. *Technical report*, Wallace H. Coulter Department of Biomedical Engineering at Georgia Tech and Emory University. https://www2.isye.gatech.edu/~brani/datasoft/DL.pdf.

Vidakovic B (2009). *Statistical Modeling by Wavelets*, volume 503. John Wiley & Sons.

West SK, Rubin GS, Broman AT, Munoz B, Bandeen-Roche K, Turano K (2002). How does visual impairment affect performance on tasks of everyday life? The SEE project. *Archives of Ophthalmology*, 120(6): 774–780.

Woods T, Preeprem T, Lee K, Chang W, Vidakovic B (2016). Characterizing exons and introns by regularity of nucleotide strings. *Biology Direct*, 11(6): 6.

Zhou B (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business & Economic Statistics*, 14(1): 45–52.