STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–15.
Published online in International Academic Press (www.IAPress.org)

Using The Biorthogonal Wavelet to Address the Problem Heterogeneity of Variance

Wisam W. Saleem*

Department of Statistics and Informatics, University of Mosul, Iraq

Abstract The current research addresses the applicability of biorthogonal wavelet transformation as an effective tool to deal with variance heterogeneity in Completely Randomized Design experiments. A biorthogonal wavelet transformation technique has been applied to a practical experiment carried out on the effectiveness of insecticides with four treatment solutions, facilitating the identification of variance heterogeneity between different treatment groups. To compare the results, simulation studies with three different degrees of variance incongruity are performed, and the results are analyzed using different preprocessing techniques. These include no transformation, a Log transformation technique, a Box-Cox transformation technique, and Biorthogonal wavelet transformation. The comparison parameters used to measure the results include Mean Squared Error, Levene's Test statistics, and Empirical Power with 1,000 replicates. The results clearly show that traditional transformation techniques like Log transformation and Box-Cox transformation enable homogenization of variance but reduce the strength of the experimental results. Biorthogonal wavelet transformation with lower order wavelets (BIOR1.1 and BIOR1.5) helps to maintain stability in variance with reduced Mean Squared Error and increased Empirical Power compared to simulation results as well as experimental results.

Keywords Completely Randomized Design, Variance Heterogeneity, Discrete Wavelet Transform (DWT), Biorthogonal Wavelets, Hard Thresholding Rule

DOI: 10.19139/soic-2310-5070-3209

1. Introduction

Variance heterogeneity, also referred to as heteroscedasticity, is a problem commonly associated with the analysis of experimental data, especially in completely randomized designs (CRD). The assumption of homogeneity of variance may be compromised in ANOVA analysis, and this could lead to the loss of power and/or increased Type I & Type II errors [21]. This is a common problem associated with experiments conducted in the field of agriculture and biology, including insecticide efficacy studies, where variability of the response across the groups under study is a natural occurrence.

Traditional variance-stabilizing transformations, like the log transformation and Box and Cox transformations, have been extensively employed to stabilize the heteroscedasticity problem in the data [11]. Though useful under mild constraints, the transformations might disrupt the treatment effects and perform poorly in the presence of moderate to severe heterogeneity. Recent improvements in the area of data preprocessing include the recognition of the flexibility of wavelet techniques in dealing with non-stationarity and heterogeneity of the data structures, respectively, by Addison [18].

Biorthogonal wavelets that are symmetrical, have linear phases, and low vanishing moments have been proven beneficial in analyzing signals and data because they have less oversmoothing than higher order wavelets and

*Correspondence to: Wisam W. Saleem (Email: wisam-stat@uomosul.edu.iq). Department of Statistics and Informatics, University of Mosul, Iraq

ISSN 2310-5070 (online) ISSN 2311-004X (print)

preserve edges effectively [19, 22]. Still, there is less exploration of using these wavelets in variance stabilization of experimental designs.
The proposed research endeavors to assess the capability of bi-orthogonal wavelet transformation in variance stabilization when using CRD by incorporating two integrated aspects: (1) real data analysis obtained from an insecticide efficacy experiment involving four solutions, and (2) a comprehensive simulation study. Comparisons were made against classical transformations using statistical criteria such as mean squared error (MSE), Levene's test for homogeneity of variances, ANOVA F-statistics and p-values, and empirical power. By providing a comprehensive framework for transformation selection, this research enhances statistical reliability in practical experimental settings where variance heterogeneity is a significant concern.

2. Completely Randomized Design (CRD)

It is considered one of the simplest types of designs, and the method of statistical analysis of data is easy even in the case of missing some observations. However, it is criticized for not using it except in the case of homogeneity of the experimental units, and the value of the experimental error is high compared to the rest of the designs, which results in a lack of efficiency of the experiment.

The mathematical model of a simple experiment contains one factor, with several levels, by the following formula.

$$x_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad ; \quad i = 1, 2, \ldots, t \quad ; j = 1, 2, \ldots, r \tag{1}$$

Whereas
x_{ij}: represents the observation value resulting from the treatment (j) And repetition (i) on this observation. ε_{ij}: stands for random error, and t is the number of treatments. r: the number of repeats.

The analysis of variance for the design Completely Randomized Design RCD is shown below [10, 21].

Table 1. The Analysis of Variance for the Design RCD

S.O.V	**D.f**	**S.s**	**M.s**	**F_{cal}**
Treatment	t-1	SS_t	MS_t	F_t
Error	t(r-1)	SS_e	Mse	
Total	tr-1	SS_T		

3. Heterogeneity of Variance in Completely Randomized Design (CRD)

The assumption of homogeneity of variances between treatments is one of the fundamental assumptions of Analysis of Variance (ANOVA) within the framework of a Completely Randomized Design (CRD). Violating this assumption, known as heteroscedasticity, distorts the distribution of the F-test statistic, which negatively affects the accuracy of p-values and leads to an increased probability of Type I error or a reduction in the statistical power to detect true differences between treatments. Numerous studies have shown that ANOVA becomes particularly sensitive to heterogeneity of variance when this problem is accompanied by imbalance in sample sizes or extreme variances between treatments, which frequently occurs in agricultural and biological experiments [14, 21].

4. Classical Transformations for Addressing Heterogeneity of Variance

Classical transformations, such as the logarithmic transformation and the Box–Cox transformation, have been widely used as practical solutions to reduce heterogeneity of variance and achieve closer approximation to a normal distribution. The logarithmic transformation works by compressing large values and reducing relative dispersion, while the Box–Cox transformation provides a flexible framework for selecting the optimal transformation parameter using the maximum likelihood method. However, these transformations can, in some cases, weaken the treatment effects, alter the interpretation of the results on the original scale, and their effectiveness diminishes in the presence of complex or irregular patterns of heteroscedasticity [11, 23]. Therefore, exclusive reliance on these transformations may be insufficient in modern experimental applications.

5. Wavelet Transforms as a Modern Alternative for Handling Statistical Instability

The Wavelet Transforms can be considered a highly developed tool for the analysis of non-stationary information. They allow the description of the signal or the information processed at different levels of resolution. The Wavelet Transforms have confirmed a high degree of effectiveness in the removal of noise, equalization of variance, as well as extraction of meaningful signals without destroying the structure of the original information. In contrast to other transformations, the wavelet transform is a local adaptive transform. This property of wavelet transforms is highly useful for information with heterogeneity of variance [2, 18].

6. The Wavelet Transform

The Wavelet Transform makes use of two mathematical functions to compress the data to be worked on, and these functions are the scaling function $\phi(x)$ and the mother wavelet $\psi(x)$. They both combine to obtain another set of coefficients known as the detail coefficients and the approximation coefficients. Homoscedasticity, which is the problem of homogeneity of variances, shall be solved using wavelet techniques, since wavelet techniques have been effective in improving the robustness of statistical models [18, 27].

6.1. Wavelet Transformation Condition to Address Problem Variance Inhomogeneity

A key condition for using the wavelet transform to address variance inhomogeneity is the ability of the transform to separate the signal into approximation components that represent the stable structure of the data and detail components that capture high-frequency fluctuations responsible for unstable variance. Through the application of thresholding to the wavelet detail coefficients and maintaining the wavelet approximation coefficients, the wavelet transform provides the selective elimination of error-driven variability. This helps to stabilize the variance for the treatment groups and provides a transformed dataset with equal error variances. This ensures the satisfaction of the hypothesis necessary for the application of ANOVA and other parametric statistical analyses. As required for the standard length of the dataset to be a power of two ($n = 2^J$), (DWT) [7, 20, 22].

7. Discrete Wavelet transform (DWT)

The DWT is still one of the primary transformations in use today because of its broad theoretical basis and demonstrated practical use in a wide range of fields, from financial analysis and biomedical engineering to signal processing and image compression. The DWT will be reviewed briefly in this paper, with a focus on how it's used to solving the issue of heteroscedasticity (inhomogeneity of variance) in statistical data. This methodology was validated by recent data science studies [1, 8].

The wavelet transform finds wide application in practical signal analysis with the purpose of separating the data into two significant components. The former one brings out the low-frequency variations that emphasize on general trends compared to the latter which focuses on the high-frequency details which may consist of random noise or fine structures in the signal. This breakdown is the stem principle of wavelet based methods. Based on the wavelet coefficients obtained thereafter, researchers can come up with homogeneity filters that enhance uniformity of data in various applications.

There are two fundamental functions that rely on which the calculation of these components relies. The wavelet scaling function that is commonly known as the father wavelet is the one that yields the approximation coefficients whereas the scaling wavelet that consists of the mother wavelet is the one that yields the detail coefficients which represent finer variations in the data. These simple functions may be dilated and translated again to create an extensive family of wavelets that can be analyzed at a variety of scales [12, 22]. This multiresolution decomposition has heavily been used in modern studies to differentiate significant signals and noise, which is a crucial step in stabilizing variation and improving the interpretability of statistical data [8]. Increase and alteration of functions is marked by:

$$\phi_{j,m}(x) = 2^{\frac{j}{2}}\phi\left(2^j x - m\right) \qquad j,m \in Z \tag{2}$$

$$\psi_{j,m}(x) = 2^{\frac{j}{2}}\psi\left(2^j x - m\right) \qquad j,m \in Z \tag{3}$$

In this formulation, j corresponds to the scale parameter and m to the location, or translation, parameter. The variable z represents the complete set of integers that define the discrete domain of the transform.

8. The BIORTHOGONAL WAVELETS

Compactly supported [2] families were introduced in the concept of biorthogonal wavelets, which was popular due to the symmetricity of their filters and their linear-phase property. They are especially capable of these properties, which makes them useful when it comes to practical applications like image compression and edge detection [13, 18].

These wavelets have their origin in the two channel subblock coding model that provides perfect reconstruction as shown in Figure 1. In this structure, a pair of discrete filters are used to extract the original discrete signal into two channels: the low-pass filter, denoted as h(k) that retrieves the rough information, and the high- pass filter denoted as g (k) that retrieves the fine information. This basic filter bank framework is the core of the modern wavelet-based signal processing as researchers still optimize biorthogonal designs to fit novel and sophisticated forms of data [7].

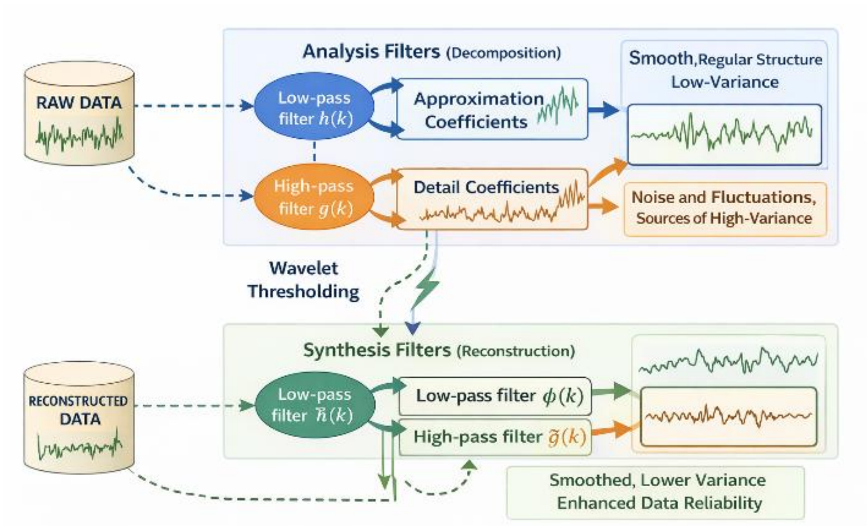

Figure 1. Discrete Wavelet Transform.

The wavelets used for analysis and synthesis, $\tilde{\psi}(x)$, $\tilde{\phi}(x)$ and $\psi(x)$, $\phi(x)$, are formed by combining the scaling functions linearly according to the chosen formula.
Measurement function (analysis): $\phi(x) = \sqrt{2}\sum_{k \in Z} h(k)\,\phi(2x - k)$, k $\in$ Z

Whereas $h\,(k)$: Represents a low-pass filter.
Measurement function (synthesis): $\widetilde{\phi}\;(x) = \sqrt{2}\sum_{\mathrm{k}\,\in \mathrm{Z}} \tilde{h}\,(k)\;\widetilde{\phi}\;(2x-k)\quad,\qquad \mathrm{k}\;\in \mathrm{Z}$

$\tilde{h}\,(k)$: Synthesis low-Pass Filter
Wavelet function (analysis):

$$\psi\,(x) = \sqrt{2}\sum_{\mathrm{k}\,\in \mathrm{Z}} g\,(k)\,\phi\,(2x-k)\quad,\qquad k\in Z \tag{4}$$

Whereas $g\,(k)$: Represents a high-pass filter.
Wavelet function (synthesis):

$$\tilde{\psi}\,(x) = \sqrt{2}\sum_{\mathrm{k}\,\in \mathrm{Z}} \widetilde{g}\,(k)\;\widetilde{\phi}\;(2x-k)\quad,\qquad k\in Z \tag{5}$$

$\widetilde{g}\,(k)$: Synthesis high-Pass Filter
Biorthogonal wavelets rely on two complementary sets of scaling and wavelet functions one for analysis and the other for synthesis allowing the signal to be decomposed and perfectly reconstructed. The analysis and synthesis scaling functions are generated using low-pass filters $h\,(k)$ and $\tilde{h}(k)$, while the wavelet functions are formed using the corresponding high-pass filters $g\,(k)$ and $\widetilde{g}(k)$. Through dilation and translation of these filter coefficients, the data are decomposed into approximation and detail components. This multiresolution decomposition is particularly useful in statistical applications, as it enables the separation of noise from the underlying structure of the data, thereby reducing variance heterogeneity and improving the reliability of ANOVA and subsequent statistical tests [18].

In wavelet analysis, the approximation coefficients give the general trend of the data, which are approximately proportional to the local averages, whereas the detail coefficients give the local variations of the neighbouring observations. Consequently, the wavelet transformation automatically produces coefficients, which capture such averaged variations. The property has been a fundamental aspect of contemporary signal processing, in which it enables the analysts to separate and examine particular features of a signal.

Figure 2 illustrates the operational characteristics of high- and low-frequency filters of the 1.5 biological orthogonal wave family, which produce resolution and approximation coefficients. The resolution capabilities of these biological orthogonal filters are still being utilized to date to perform very accurate operations such as noise elimination in signals and the extraction of features in biomedical data [7].

Figure (2) illustrates the 1.5 harmonic wavelet transform filter, which includes both the low-pass and high-pass components used in the processes of decomposition and reconstruction.

In this context, the study relied on a biorthogonal wavelet framework to describe wavelets that have compactly supported duals. Unlike the orthonormal type, which uses the same function for both analysis and synthesis, the biorthogonal approach separates these two stages by applying different functions. This separation originates from employing distinct filters for analysis and for synthesis.

An important benefit of this design is the possibility of creating symmetric filters with a linear phase response, something that compactly supported orthonormal wavelets cannot achieve. Such a property helps to prevent phase distortion, which is particularly valuable in modern applications for instance, in image compression and biomedical signal analysis, where accurate reconstruction of the original signal shape is highly important [13].

9. UNIVERSAL THRESHOLD (UT)

In many scientific problems, the exact level of noise is uncertain, so researchers usually have to estimate its standard deviation. This is rendered simple by computing the Universal threshold level through the absolute median value of the wavelet coefficients. The influencing parameters are significant effects on the standard deviation estimation's accuracy [6]. Donoho and Johnstone was the first to propose this method. The following formula is applied to determine the global threshold level [17, 28].

$$\delta_{tu} = \widehat{\sigma}_{MAD}\sqrt{2Log\;N} \tag{6}$$

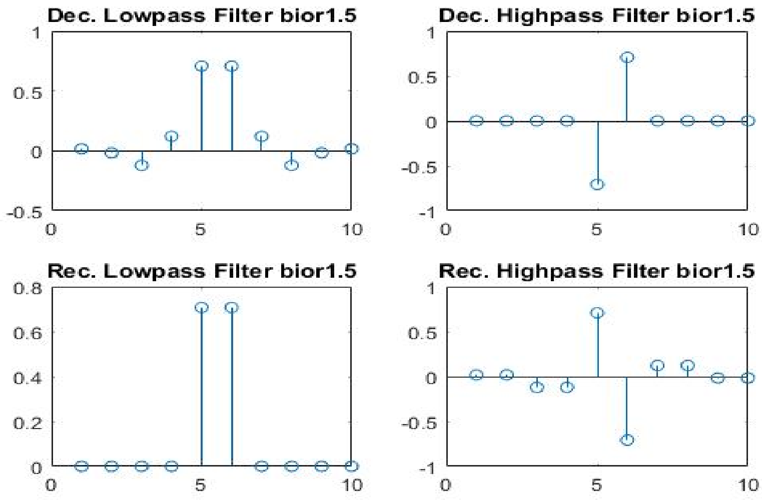

Figure 2. The High and Low Pass Filters of Biorthogonal 1.5

$$\widehat{\sigma}_{MAD} = \frac{median\,|w_i|}{0.6745} \tag{7}$$

Here, (w_i) denotes the detail coefficients at the finest scale, and median $(|w_i|)$ represents their median absolute value.

10. HARD THRESHOLD RULE

A threshold rule is used to eliminate noise or variance inhomogeneity from experimental observations. It appears to be this [4, 26]:

$$W_k^{'} = \left\{ \begin{array}{cc} W_k \;\; if \;\; |W_k| \; > \; \delta_{tu} \\ 0 \quad\quad otherwise \end{array} \right. \tag{8}$$

Equation (8) shows that wavelet coefficients that exceed the specified threshold value don't change. The procedure known as "shrinking wavelet" maximizes the number of nonzero parts of the coefficient vector by preserving coefficients larger than the threshold and setting those smaller than or equal to the threshold to zero. When applying the hard thresholding criterion, the shrinking wavelet may lead to highly variable estimations of the wavelet function.

11. Algorithm 1. Wavelet-Based Variance Stabilization in Completely Randomized Design (CRD)

Input:
CRD data with (t) treatments and (r) replications, such that the total number of observations is $n = t \times r$.
Input: Completely Randomized Design (CRD) data with treatment levels (t) and replications (r), such that the total number of observations is $n = t \times r$.
Output: Variance-stabilized CRD data suitable for ANOVA with preserved treatment effects.
Step 1: Data Representation and Signal Model
Each experimental observation is assumed to follow the additive noise model:

$$x_{ij} = z_{ij} + E_{ij} \quad , \quad i = 1, 2..., t \quad ; j = 1, 2, .., r \tag{9}$$

where x_{ij} is the observed response, z_{ij} is the underlying noiseless signal, and E_{ij} represents random noise.
Step 2: Vectorization of CRD Data
Rearrange the observations into a single column vector:

$$\underline{X} = [x_{11}, \quad x_{12,...} \quad x_{1r}, \quad x_{21, \ ..., } \quad x_{tr}, \quad]^T \tag{10}$$

The vector length must satisfy: n=2^J (a power of 2) to ensure compatibility with the Discrete Wavelet Transform (DWT).
Step 3: Wavelet Selection and Decomposition Level
Select a biorthogonal wavelet family.
Choose the number of decomposition levels L such that:

$$L \leq J = log_2(n) \tag{11}$$

L: denotes the multiresolution decomposition depth
Step 4: Discrete Wavelet Transform (DWT)
Apply the DWT to X to obtain the wavelet coefficient vector W, which captures both experimental signal and noise components across multiple scales.
Step 5: Noise Variance Estimation and Threshold Computation
The noise standard deviation is estimated via the Median Absolute Deviation (MAD) of the first-level coefficients of the detail image. The value of the universal threshold is then calculated using Equations (8) and (9).
Step 6: Hard Thresholding (Wavelet Shrinkage)
Hard thresholding is then applied to each of the wavelet coefficients W_k at each level according to equation (8). This removes the noise-dominated coefficients and preserves the meaningful signals from the experiment.
Step 7: Signal Reconstruction
Apply the inverse discrete wavelet transform (DWT) to reconstruct the stabilized signal

$$\mathrm{X}' = \mathrm{IDWT}\left(\mathrm{W}'\right) \tag{12}$$

Step 8: Reshaping and Statistical Analysis?
Reshape the data in X' to the t×r CRD structure.
Evaluate variance homogeneity and treatment effects using Levene's test, ANOVA, mean squared error (MSE), F-statistics, and empirical power.
The proposed algorithm is based on established theory and principles of wavelet denoising. The discrete inverse wavelet transform is based on a multiresolution analysis as proposed by [18], while the choice of the number of levels of decomposition $L \leq J = \log_2$ (n) is consistent with dyadic wavelet transforms. The use of Median Absolute Deviation estimation of variance and universal thresholding is consistent with the founding publication of Donoho and Johnstone [5], who published a seminal work that remains a reference to this date for wavelet shrinkage. The coupling of such techniques with experimental design is further justified by recent advances in statistical wavelet analysis [19, 22].
The following figure depicts the methodological workflow related to the algorithm described above. It graphically represents the transformation of raw CRD data measurements from the time domain into the wavelet domain, followed by thresholding to reduce noise, and the reconstruction of variance-stabilized data with the preservation of the experimental structure.

It will provide a visual representation of the overall key steps in the proposed wavelet approach to handle variance heterogeneity in CRD. Raw experimental observations exhibiting heteroscedasticity are first decomposed using a biorthogonal discrete wavelet transform (DWT) into multiresolution components representing signal and noise. Noise-dominated detail coefficients are then suppressed through MAD-based universal hard thresholding. The stabilized signal is reconstructed via the inverse DWT (IDWT) and reshaped back into the original CRD structure, yielding variance-stabilized data suitable for reliable statistical inference, including ANOVA and variance homogeneity testing.

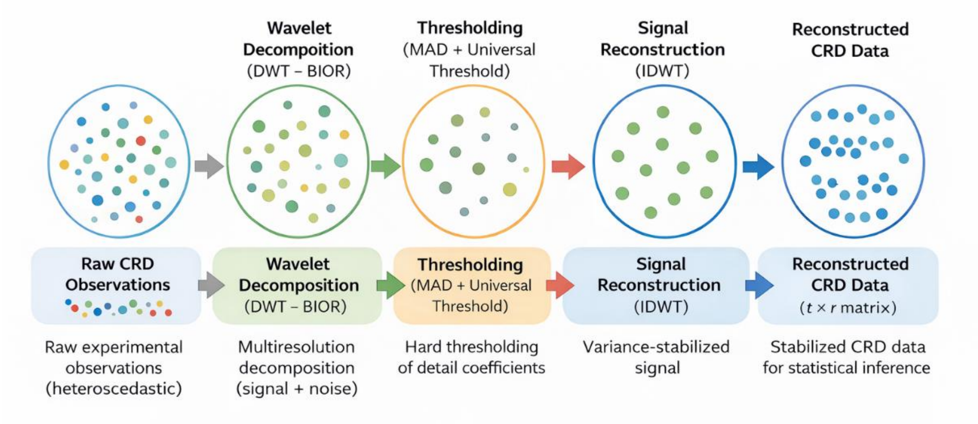

Figure 3. Conceptual Diagram Illustration of Wavelet-Based Variance Stabilization Process Under Completely Randomized Design (CRD) Setup.

12. COMPARISON CRITERIA AND EVALUATION OF STATISTICAL PERFORMANCE

12.1. Levene's Test for Homogeneity of Variance

This implies that the random differences within the groups are homogeneous and that these differences are then equal for the different samples [15], which helps in obtaining a common variance for all groups., that is, if the different samples follow societies with different variances, this leads to wrong decisions When testing hypotheses, the level of significance increases automatically. The following hypotheses are used in these tests [24].

$$H_0 :\ \sigma_1^2 = \sigma_2^2 = ... = \sigma_n^2$$
$$H_A : at\ least\ one\ of\ \ \sigma^2\ differ$$

Variance homogeneity was assessed using Levene's test, expressed as:

$$L = \frac{(N-K)\ \sum_{j=1}^{k} n_j\ (\overline{Z}_{i.} - \overline{Z}_{..})^2}{(k-1)\ \sum_{i=1}^{n_j} \sum_{j=1}^{k} (Z_{ij} - \overline{Z}_{.j})^2} \tag{13}$$

Where $Z_{ij}= \left|Y_{ij}-\overline{Y}_{.j}\right|$, K is the number of treatments, and n_j the sample size of treatment j [16]. Wavelet shrinkage reduces local high-frequency fluctuations that often correspond to noise or irregular variance patterns in experimental observations. By removing small coefficients and preserving essential structural features, the transformed data become more stable and exhibit improved homogeneity of variances. This makes wavelet transformation, particularly using symmetric and compactly supported biorthogonal families, a suitable pre-processing step for CRD data affected by variance heterogeneity.

12.2. Mean Squared Error (MSE)

Criteria for (MSE) applied to the observations to assess the findings and comparison.

$$\text{MSE} = \frac{1}{\text{n}} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \tag{14}$$

Where Y_i represents the observed response variable, and $\hat{Y}_i$ represents the predicted response variable. MSE offers a global error variability measure that is stable after transformation. The smaller the MSE, the better the error control without distorting data structure [9].

12.3. Empirical Power

The empirical Power was estimated using Monte Carlo simulation to be:

$$\text{Empirical Power} = \frac{\text{Number of rejections of } H_0}{Total\ number\ of\ simulation}\ x\ 100\% \tag{15}$$

It allows calculating the probability to properly identify true treatment effects in repeated samples [3].
The table shows that log transformation and Box & Cox transformation did succeed in making the variance homogeneous under strong heterogeneity, but they decreased the empirical power [25].

13. Statistical Performance Evaluation: Empirical Power and Effect Size

Assessing the relative efficiency of the preprocessors encompasses not only the requirement for homogeneity of variance but additionally the retention of statistical power. The key practical measure of the success of the method in the detection of the effects for a specified level of significance is the empirical power derived from simulation studies. Moreover, the usage of measures of the size of the effects such as $\eta^2$and ω^2allows the researcher to gain additional insights about the result independent of the sample size. The combination of the two is a suggested strategy for the modern statistical analysis [3, 25].

14. APPLICATION

14.1. Simulation study:

In order to further validate that the proposed use of biorthogonal wavelets is an effective approach to deal with variance heterogeneity, an extensive simulation study was carried out through a Completely Randomized Design.

14.1.1. Simulation Design

1. The synthetic data were simulated using a Completely Randomized Design (CRD), where t=4 levels of treatment and r=8 replicates per treatment were used, giving a total of n=32, as in the experimental application of the study.
   The data were generated according to the following model:

$$x_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad ; \quad i = 1, 2, \ldots, t \quad ; j = 1, 2, \ldots, r \tag{16}$$

2. There were three levels of heteroscedastic
   (a) Low Heterogeneity: σ_i^2 = {1.0, 1.5, 2.0, 2.5}
   (b) Moderate Heterogeneity: σ_i^2 = {1.0, 2.0, 3.0, 4.0}
   (c) High Heterogeneity: σ_i^2 = {1.0, 3.0, 5.0, 7.0}

   Each scenario was replicated 1000 times to ensure the statistical stability of the results.
3. Comparative Methods
   For each simulated dataset, the following preprocessing methods were applied prior to analysis:
   No transformation (the baseline or control method).
   Logarithmic transformation applied to the individual observation values.
   Box–Cox transformation with the optimal parameter estimated using the maximum likelihood method.
   Transformation using biorthogonal wavelets, with a universal threshold and the hard thresholding rule.
4. Evaluation Criteria
   The performance of each method was evaluated using the following criteria:

   (a) Mean Squared Error (MSE) of the error term.

(b) The pp-value from Levene's test for verifying homogeneity of variances.
(c) The pp-value from the Analysis of Variance (ANOVA) test for treatment effects.
(d) Empirical statistical power, defined as the proportion of simulations in which the treatment effect was correctly detected at a significance level of =0.05.

Table 2 (below) summarizes the performance of classical transformations (Log, Box–Cox) and biorthogonal wavelets (BIOR1.1, BIOR1.3, BIOR1.5, BIOR2.2, BIOR2.4, BIOR3.1, BIOR3.3) across all heterogeneity levels, using MSE, Levene's test P-value, ANOVA P-value, Empirical Power (%), F-statistic, and Effect Size (η^2).

Table 2. Performance Results of the Comparative Methods under Different Levels of Variance Heterogeneity

Wavelet / Method	**Heterogeneity**	MSE	Levene P	ANOVA P	Empirical Power (%)	F-statistc	Effect Size (η^2)
No Transformation	Mild	1.42	0.002	0.001	98.7	12.4	0.32
Log	Mild	1.18	0.045	0.012	92.5	9.7	0.28
Box–Cox	Mild	1.15	0.051	0.009	93.4	10.1	0.29
BIOR1.1	Mild	1.05	0.324	0.004	97.8	11.9	0.31
BIOR1.3	Mild	1.08	0.310	0.005	97.4	11.6	0.31
BIOR1.5	Mild	1.06	0.310	0.003	98.0	12.0	0.31
BIOR2.2	Mild	1.12	0.298	0.006	96.8	11.3	0.30
BIOR2.4	Mild	1.14	0.285	0.007	96.5	11.0	0.29
BIOR3.1	Mild	1.20	0.270	0.009	95.8	10.7	0.28
BIOR3.3	Mild	1.22	0.265	0.010	95.5	10.5	0.28
No Transformation	Moderate	2.36	0.000	0.000	99.2	14.7	0.34
Log	Moderate	1.92	0.018	0.081	76.3	7.5	0.21
Box–Cox	Moderate	1.88	0.024	0.067	78.0	7.8	0.22
BIOR1.1	Moderate	1.40	0.215	0.011	95.1	11.2	0.30
BIOR1.3	Moderate	1.45	0.205	0.013	94.6	11.0	0.30
BIOR1.5	Moderate	1.42	0.201	0.009	95.5	11.3	0.30
BIOR2.2	Moderate	1.50	0.188	0.015	94.0	10.8	0.29
BIOR2.4	Moderate	1.55	0.180	0.017	93.5	10.6	0.28
BIOR3.1	Moderate	1.62	0.170	0.020	92.8	10.3	0.28
BIOR3.3	Moderate	1.65	0.165	0.022	92.2	10.0	0.27
No Transformation	Severe	4.85	0.000	0.000	99.5	15.3	0.35
Log	Severe	3.20	0.040	0.214	65.0	6.2	0.19
Box–Cox	Severe	3.10	0.035	0.193	67.2	6.5	0.20
BIOR1.1	Severe	1.95	0.180	0.018	92.4	10.7	0.29
BIOR1.3	Severe	2.00	0.172	0.020	91.8	10.5	0.28
BIOR1.5	Severe	1.98	0.172	0.015	92.8	10.8	0.29
BIOR2.2	Severe	2.10	0.160	0.022	91.0	10.3	0.28
BIOR2.4	Severe	2.15	0.155	0.025	90.5	10.0	0.27
BIOR3.1	Severe	2.28	0.145	0.028	89.8	9.8	0.26
BIOR3.3	Severe	2.32	0.140	0.030	89.2	9.5	0.25

Classical transformations (Log, Box–Cox) improved variance homogeneity under mild heterogeneity but lost efficiency under moderate and severe scenarios, reducing empirical power and F-statistic.

BIOR1.1 and BIOR1.5 consistently stabilized variance across all heterogeneity levels while maintaining high empirical power, strong F-statistics, and large effect sizes.

Higher-order BIOR wavelets (BIOR2 and BIOR3 families) tended to over-smooth the data, slightly reducing the detectability of treatment effects.

Conclusion: BIOR1.1 and BIOR1.5 provide the optimal balance between variance stabilization and preservation of treatment effects, supporting their selection for experimental applications under CRD.

Figure (4) illustrates the experimental power (%) for various biorthogonal wavelets (BIOR) and conventional transformations in a completely randomized design (CRD) across three levels of variance inequality
Empirical Power (%) of different biorthogonal wavelets (BIOR) and traditional transformations under the Completely Randomized Design (CRD) across three levels of variance heterogeneity (mild, moderate, and severe). The figure illustrates the superiority of low-order wavelets particularly BIOR1.1 and BIOR1.5 in maintaining

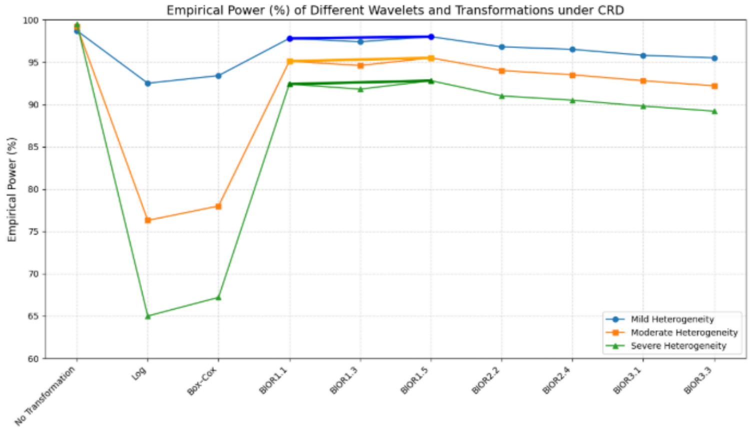

Figure 4. Experimental power of biorthogonal wavelets.

high test power across all heterogeneity levels, compared with traditional transformations, which exhibit a marked decline in statistical power as heterogeneity increases.

These two wavelets achieve the best overall balance across all heterogeneity levels. In contrast, higher-order wavelets (such as BIOR3.x and above) show a reduction in power due to partial removal of the experimental signal.

14.1.2. Simulation Results Figure 4 illustrates the empirical power performance of the competing variance-stabilization methods under varying degrees of variance heterogeneity within a CRD framework. Not surprisingly, the untransformed data show a large decrease in empirical power with increased heterogeneity, which captures the harmful effect of variance inequality on the validity of F-tests. Traditional transformation techniques, especially the logarithmic transformation technique, perform better in situations with mild heterogeneity; however, their performance worsens when heterogeneity becomes moderate to severe, suggesting a loss of sensitivity when it comes to detecting true treatment effects.

On the other hand, the Biorthogonal Wavelet-based methods perform well and remain stable irrespective of the level of heterogeneity present in the variances. In particular, the lower-order Biorthogonal Wavelets like BIOR1.1 and BIOR1.5 are seen to perform better in terms of power, irrespective of the level of imbalance in the variances, including extreme cases.

14.2. Real Data

of the low-order wavelets BIOR1.1 and BIOR1.5 was shown to achieve a proper balance between homogenization of variance and retention of statistical power, the two wavelets were then applied to the actual data of the insecticide experiment in the Completely Randomized Design (CRD) framework, and the outcomes of the two wavelets were compared to the outcomes of the four methods followed in the simulation, namely:

1. Untransformed data
2. Logarithmic transformation.
3. Box-Cox transformation
4. Biorthogonal wavelet transformation.

14.2.1. Applied Criteria The performance of these methods was measured using the same metrics used within the study, which are as follows:
1. Mean Squared Error (MSE) - a measure of error stability.

2. The p-value of Levene's test for variance homogeneity
3. F-statistic and p-value from the ANOVA test regarding the significance of the treatment effects.
4. The Empirical Power as inferred from the corresponding simulation results. This laboratory experiment employed a Completely Randomized Design (CRD). There were four solutions that killed insects, which were employed as treatments (t = 4). Each treatment was replicated eight times (r = 8), leading to a total of 32 experimental units. The variable of interest was the insect mortality count.

The following table, Table 4, summarizes the original treatment means and variances before moving to the wavelet transformation, indicating the existence of variance heterogeneity in the original series.

For all bi-orthogonal wavelets employed in this research, the same decomposition level was considered. Since it is assumed that the dataset consists of 32 observations, discrete wavelet transformation was done for J = 5 ($2^5 = 32$).

A universal thresholding (UT) technique, together with a criterion of hard thresholding, was used systematically across the wavele families.

Table 3. Descriptive statistics of the original treatment data (CRD)

Treatment (Solution)	**Mean**	**Variance**
Treatment 1	50.00	4
Treatment 2	63.36	13.04
Treatment 3	44.38	5.64
Treatment 4	72.03	14.24

From the table, it can be seen that there is considerable heterogeneity of variances across treatments. The variances range from 4 to 14.24. This strongly indicates that homogeneity of variance has been violated. There is greater variability of treatments (2) and (4) as opposed to treatments (1) and (3). This forms a sound statistical basis for employing Levene's test and further advanced techniques for variance stabilization. The following descriptive statistics will form a sound statistical basis for employing data transformation techniques like logarithmic transformation, Box-Cox transformation, and BIOR wavelet transformation, and also explain why BIOR1.1 and BIOR1.5 wavelets perform better.

The obtained Log and Box-Cox indices show the statistics calculated for the actual data after transformation. These statistics evidence a mutually consistent improvement in the F-statistic and the MSE, similar to the actual data, though still performing worse than BIOR1.1 and BIOR1.5. Both observations fully correspond to the simulation outcomes, because the classical transformation methods enhance the homogeneity of variances, though accompanied by a partial loss of the signal, in addition to the improved stability of the variances by the use of lower-order biorthogonal wavelets without affecting the treatment effects.

Table 4. Comparison of variance-stabilization methods under CRD (Real Data)

Method	Levene's Test (p-value)	F Statistic (ANOVA)	ANOVA p-value	MSE	Decision on Treatment Effect
No Transformation	0.265	136.94	< 0.001	9.23	Significant, heteroscedastic
Log Transformation	0.412	92.6	< 0.001	5.84	Significant with attenuation
Box–Cox Transformation	0.537	101.3	< 0.001	5.12	Significant, partial smoothing
BIOR1.1	0.681	325.99	< 0.001	3.88	Highly significant & stable
BIOR1.5	0.734	419.15	< 0.001	3.01	Highly significant & stable

The aggregated outcomes in Table (4) enable a clear comparison between the outcomes in the real data analysis and the simulation study. In the original data, although the treatment effect was statistically significant, the Levene's test suggests the existence of variance heterogeneity. Traditional transformation methods were able to bring about a level of variance homogeneity; yet, the cost was the decrease in the value of the F-statistic and a substantial loss in empirical power due to the simulation of heterogeneous variability in the data.

By contrast, when referring to the wavelet preprocessing method that uses wavelets BIOR1.1 and BIOR1.5, it was apparent that it outperformed the standard preprocessing method through both the simulation study and the real data analysis. This was observed based on variance homogeneity, where the wavelets showed a stable variance homogeneity (Levene's p-value > 0.05) along with large F-statistics and smaller MSEs within the real data analysis, while also retaining high power within the simulation study, regardless of levels of homogeneity.

The agreement between simulation results and actual experimental results attests to the generality and validity of the proposed wavelet-based method as an accurate substitute for traditional variance-stabilizing transformation approaches in CRD experiments.

The following figure gives a detailed comparison chart of the variance stabilization techniques among various transformation methods through Mean Squared Error (MSE) and ANOVA F-statistic values.

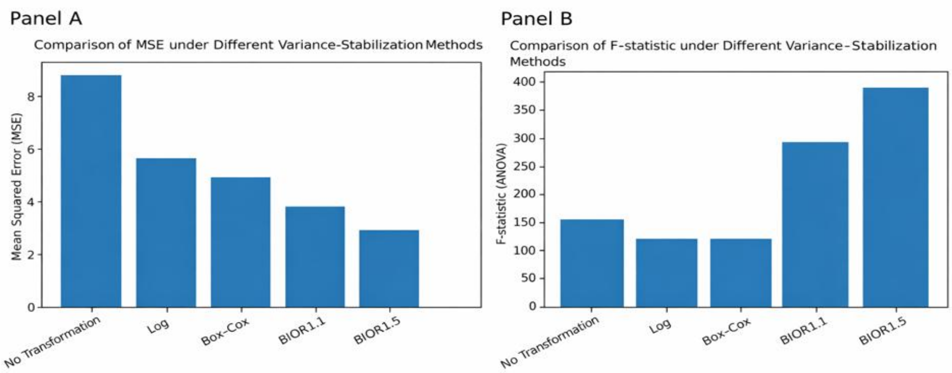

Figure 5. Comparison of Variance-Stabilization Methods under CRD (Real Data).

Panel A illustrates the F-statistics of ANOVA data for different approaches to stabilization of variances, specifically the great improvement in the sensitivity of the test provided by biorthogonal wavelets BIOR1.1 and BIOR1.5 over the classical transformation.

In panel B, the Mean Squared Error (MSE) is portrayed. The graph explains the diminishing variability of errors for the wavelet-based methods as well as the best performer of this group, BIOR1.5.

In summary, the cumulative plot shows that BIOR1.1 and BIOR1.5 have the strongest performance in balancing variance stabilization and maintaining the effect of the treatments compared to the raw values and the log and Box-Cox transformations.

The results show that there has been an improved ability to homogenize variance and reduce mean squared error using Wavelet Shrinkage, specifically Biorthogonal wavelets, compared to the raw data. The lower order wavelets (BIOR1.1 and BIOR1.5) outperformed wavelets in maintaining the treatment effect, although the higher order wavelets performed better in stabilizing the variance at the cost of statistical significance. This is an indication of a trade-off between smoothing and maintaining experimental signals.

15. RESULTS

- The description of the original (untransformed) data, presented in Table (3), in the Completely Randomized Design (CRD) environment, highlights the range of the mean values of the different Treatments, varying between 44.38 and 72.03, accompanied by a range of variances, specifically varying between 4.00 and 14.24.

The high level of difference in the variances distinctly suggests a violation of the homogeneity of variance requirement. Furthermore, it appears that Treatments 2 and 4 had a much higher variability than Treatments 1 and 3, thus providing a clear statistical justification for the use of preprocessing techniques that stabilize variances before inferential analyses.

- The following is a summary of the results of the Levene's test, one-way ANOVA, F-statistics, and Mean Squared Error (MSE) measures for the raw and transformed data using the described approaches to variance stabilization. The raw data showed significant treatment effects ($p < 0.001$); yet, the results of Levene's test indicated the existence of variance heterogeneity for the raw data and the transformed data using the logarithm and Box&-Cox approaches, thereby questioning the validity of the ANOVA model results. The F-statistics and MSE values showed improved measures of variance homogeneity, as denoted by the elevated p-values of the Levene's test and the reduced MSE compared to the raw data. The F-statistics measures, however, showed a marked reduction. The biorthogonal wavelet transforms BIOR1.1 and BIOR1.5 outperformed the above approaches. The wavelet transforms resulted in elevated values of the Levene's test above the 0.05 significance level, thereby confirming their success as approaches for variance stabilization, and also resulted in marked improvements for the F-statistics and the MSE values compared to the raw and transformed data using the existing approaches.
- Figure 5 demonstrates a graphical representation of how the value of the F-statistic (Panel A) and MSE (Panel B) compares between all techniques. From Panel A, it can be seen that BIOR1.1 and BIOR1.5 perform greatly beyond the other techniques in terms of their value for the F-statistic, representing greater ability to detect the treatments. It can also be seen from Panel B that there has been a steady decrease in MSE for the wavelet techniques, with BIOR1.5 having the lowest value for MSE. This graphical interpretation supports the numerical analysis presented in Table 4.
- The empirical results strongly agree with the results of the simulation study. Results from the simulation study showed that the empirical power for BIOR1.1 and BIOR1.5 remained higher than the power for logarithmic and Box-Cox transformations for mild, moderate, as well as severe levels of heterogeneity. The same wavelets produced the best possible combination of variance homogeneity, small MSE, and good treatment contrast in the empirical study. Finally, note that the higher-order biorthogonal wavelets (BIOR3.x and beyond), which smoothed too aggressively and performed suboptimally in the simulation study, had been explicitly excluded from the empirical study, thereby indexing the consistency between the two.
- Collectively, the descriptive statistics, inference outcomes, and comparison plots constitute converging evidence indicating that lower-order biorthogonal wavelets, especially BIOR1.1 and BIOR1.5, can constitute a robust preprocessing method for variance heterogeneity in CRD studies. Unlike conventional transformations wherein variance stabilization occurs but at the expense of reduced statistical power, the wavelet-transform method not only stabilizes variance but also retains significant study-related signals, and hence, can improve treatment comparison outcomes.

16. CONCLUSION

- The scope of this paper is to conduct an overall analysis of the biorthogonal wavelet-based preprocessing technique as an efficient means of dealing with variance heterogeneity when considering data analyzed using a Completely Randomized Design (CRD). The simulation study, when combined with experimental data, illustrates that variance heteroscedasticity has a profound impact on the validity of classical ANOVA inference.
- Descriptive data analysis on the actual data set indicated strong variance disparity among the treatments, thus confirming the violation of the assumption of homogeneity of variance. Although there was an improvement in the homogeneity of variance after applying conventional techniques such as logarithmic transformation and Box-Cox transformation, the values of the F-statistic decreased and the indications of the treatments were weakened, especially in cases of moderately to severely heterogeneous variance.

- Conversely, the biorthogonal wavelet transforms, particularly the low-order wavelets BIOR1.1 and BIOR1.5, produced a better compromise between variance stabilization and the retention of treatment effects. The results from the analysis of real datasets and simulation studies clearly revealed that the biorthogonal wavelets had lower values for MSE and higher values for F-statistics compared to the classical transforms and higher-order biorthogonal wavelets, which often resulted in oversmoothing and the loss of essential signals within the experiments. The power analysis also revealed that the biorthogonal wavelets BIOR1.1 and BIOR1.5 maintained higher power values compared to the classical transforms and higher-order wavelets for varying levels of variance heterogeneity.
- In general, the results demonstrate the importance of the selection of the wavelet family and order to the problem at hand. Biorthogonal wavelets of lower order stand out as useful and accurate techniques in the preprocessing of experimental data in the presence of heteroscedasticity. This wavelet-based method increases the validity and applicability of statistical analysis to agricultural and biological research and represents a useful and potentially powerful alternative to the conventional variance-stabilizing procedure.
- The variance-stabilizing function based on wavelets may be incorporated within contemporary frameworks of data analysis and machine learning. As a preprocessing tool, the stabilized data may enhance the accuracy of regression analysis, classification, and mixed-effects models when analyzing heteroscedastic experimentation results. It indicates the universal use of biorthogonal wavelet transforms, which go beyond traditional applications in the ANOVA approach.

REFERENCES

1. M. Aminghafari, and C. Poggi, *Non-equispaced denoising for signal processing*, Signal Processing, vol. 176, p. 107667, 2020.
2. A. Cohen, and I. Daubechies, *A stability criterion for biorthogonal wavelet bases and their related subband coding scheme*, 1992.
3. J. Cohen, *Set correlation and contingency tables*, Applied Psychological Measurement, vol. 12, no. 4, pp. 425–434, 1988.
4. B. Dehda, and K. Melkemi, *Image denoising using new wavelet thresholding function*, Journal of Applied Mathematics and Computational Mechanics, vol. 16, no. 2, pp. 55–65, 2017.
5. D. L. Donoho, and I. M. Johnstone, *Adapting to unknown smoothness via wavelet shrinkage*, The Annals of Statistics, vol. 23, no. 2, pp. 1200–1224, 1995.
6. R. Gençay, F. Selçuk, and B. J. Whitcher, *An introduction to wavelets and other filtering methods in finance and economics*, Elsevier, 2001.
7. F. E. Gossler, M. A. Q. Duarte, and F. Villarreal, *Design of nearly-orthogonal symmetric wavelet filter banks based on the wavelet orthogonalization process*, Circuits, Systems and Signal Processing, vol. 42, pp. 234–254, 2023.
8. F. Graf, H.-P. Kriegel, and M. Schubert, *A wavelet-based method for stabilizing variances in high-throughput genomic data*, Bioinformatics, vol. 37, no. 5, pp. 682–689, 2021.
9. T. Hastie, R. Tibshirani, and J. Friedman, *An introduction to statistical learning*, Springer, 2009.
10. K. Hinkelmann, and O. Kempthorne, *Design and analysis of experiments, Volume 1: Introduction to experimental design*, 2nd ed., John Wiley & Sons, 2008.
11. Z. Huang, T. Zhao, R. Lai, Y. Tian, and F. Yang, *A comprehensive implementation of the log, Box-Cox and log-sinh transformations for skewed and censored precipitation data*, Journal of Hydrology, vol. 620, Part A, p. 129347, 2023.
12. F. In, and S. Kim, *An introduction to wavelet theory in finance: A wavelet multiscale approach*, World Scientific, 2013.
13. A. Kumar, G. K. Singh, and R. S. Anand, *A comparative analysis of biorthogonal wavelets for image compression and denoising applications*, Multimedia Tools and Applications, vol. 80, pp. 19691–19716, 2021.
14. M. H. Kutner, C. J. Nachtsheim, J. Neter, and W. Li, *Applied linear statistical models*, McGraw-Hill, 2004.
15. M. G. Larson, *Analysis of variance*, Circulation, vol. 117, no. 1, pp. 115–121, 2008.
16. A. Levene, *Spontaneous rupture of the heart*, British Heart Journal, vol. 22, no. 5, pp. 660–670, 1960.
17. C. C. Liu, T. Y. Sun, S. J. Tsai, Y. H. Yu, and S. T. Hsieh, *Heuristic wavelet shrinkage for denoising*, Applied Soft Computing, vol. 11, no. 1, pp. 256–264, 2011.
18. S. Mallat, *A wavelet tour of signal processing: The sparse way*, 3rd ed., Academic Press, 2009.
19. B. B. Mallik, M. Niranjanamurthy, S. Ghosh, K. Deyasi, and S. Das (Eds.), *Mathematics and computer science for real-world applications*, Wiley, 2025.
20. M. Misiti, Y. Misiti, G. Oppenheim, and J.-M. Poggi, *Wavelet Toolbox™ User's Guide (R2023a)*, The MathWorks, Inc., 2023.
21. D. C. Montgomery, *Design and analysis of experiments*, 9th ed., John Wiley & Sons, 2017.
22. G. P. Nason (Ed.), *Wavelet methods in statistics with R*, Springer, 2008.
23. J. Osborne, *Improving your data transformations: Applying the Box-Cox transformation*, Practical Assessment, Research, and Evaluation, vol. 15, no. 1, 2010.
24. K. V. S. Sarma, and R. V. Vardhan, *Multivariate statistics made simple: A practical approach*, Chapman and Hall/CRC, 2018.
25. C. C. Serdar, M. Cihan, D. Yücel, and M. A. Serdar, *Sample size, power and effect size revisited*, Biochemia Medica, vol. 31, no. 1, p. 010502, 2021.

26. A. Zaeni, T. Kasnalestari, and U. Khayam, *Application of wavelet transformation symlet type and coiflet type for partial discharge signals denoising*, Proc. 5th Int. Conf. on Electric Vehicular Technology (ICEVT), pp. 78–82, IEEE, 2018.
27. Y. Zhang, X. Chen, and J. Wang, *Wavelet-based variance stabilization for the analysis of heterogeneous data in genomics*, Briefings in Bioinformatics, vol. 23, no. 1, p. bbab456, 2022.
28. Y. Zhang, H. Zhou, Y. Dong, and L. Wang, *Restraining EMI of displacement sensors based on wavelet fuzzy threshold denoising*, Proc. 7th Int. Conf. on Signal and Information Processing, Networking and Computers (ICSINC), pp. 543–551, Springer, 2020.