A NOVEL WAVELET-BASED APPROACH FOR ANOVA IN LONGITUDINAL DATA ANALYSIS USING THE FIRST DERIVATIVE OF THE GAUSSIAN FUNCTION

Naziru Isah Muhammad*[1], Musa Tasi'u[1], Umar Kabir Abdullahi[1] and Yakubu Aliyu[1]

[1]Department of Statistics, Ahmadu Bello University, Zaria
[*]Correcponding author Email: *nazeerdabest01@gmail.com

ABSTRACT

Longitudinal data, characterized by repeated observations on the same subjects over time, present significant challenges for traditional statistical Analysis of Variance (ANOVA) methods, such as Repeated Measures ANOVA due to inherent intra-subject correlation and high dimensionality. This study introduces a novel wavelet-based ANOVA approach using the first derivative of the Gaussian function to transform high-dimensional repeated measures data into a simplified Completely Randomized Design (CRD) structure, thereby facilitating classical ANOVA. The mathematical formulation and key properties (admissibility, energy localization, vanishing moments, compact support) of this wavelet function were detailed. Through a comprehensive simulation study, we demonstrate its effectiveness in preserving treatment differences, maintaining acceptable Type I error rates, exhibiting high statistical power, and showing robustness to variations in the number of time points. The results validate the Gaussian derivative wavelet as a powerful and reliable tool for dimensionality reduction and valid inference in longitudinal data analysis, offering a robust alternative to conventional repeated measures techniques.

Keywords: Wavelet ANOVA, Longitudinal Data, Gaussian Function, Repeated Measures, Dimension Reduction.

1. INTRODUCTION

Longitudinal data, where observations are collected over time for the same subjects, are prevalent across diverse fields such as medicine, psychology, and environmental sciences. These data structures are crucial for understanding dynamic processes and temporal trends. However, their inherent characteristics, including intra-subject correlation and high dimensionality, often complicate traditional statistical methods like repeated measures ANOVA (RM-ANOVA). RM-ANOVA relies on stringent assumptions, such as sphericity, whose violations can lead to inflated Type I error rates and necessitate complex corrections (Greenhouse & Geisser, 1959). These limitations become particularly pronounced in modern studies involving high-dimensional or irregularly spaced observations.

Wavelet analysis, a powerful mathematical tool for analyzing non-stationary signals, offers a promising alternative. Unlike Fourier analysis, wavelets provide both time-domain and frequency-domain information, making them adept at detecting localized events, trends, and discontinuities (Daubechies, 1992; Nason, 2008). Wavelet transformation provides an alternative approach to analyzing longitudinal data by capturing local and global features while reducing dimensionality (Percival & Walden, 2006). The integration of wavelet techniques into statistical models for repeated measures data has led to Wavelet ANOVA (WANOVA), which transforms complex correlated data into simpler representations while retaining meaningful temporal patterns (Ramakrishnan, 2024; Kevlahan, 2021).

Despite the advancements in WANOVA, most existing methods rely on standard wavelet bases that may not optimally capture the specific variability patterns in certain longitudinal settings. This research addresses this gap by proposing and validating a customized wavelet function derived from the first derivative of the Gaussian distribution. This novel approach aims to effectively aggregate repeated measures data, reduce dimensionality, mitigate random time effects, and facilitate powerful and valid inference through a simplified CRD-style ANOVA. The primary objective is to develop and evaluate this specific wavelet-based aggregation method, comparing its performance in terms of Type I error control, statistical power, and robustness to time point variation.

2. RELATED WORK

Traditional methods for analyzing repeated measures data, such as RM-ANOVA, often struggle with the complex structures of longitudinal data, particularly when assumptions like sphericity are violated, leading to inaccurate statistical inferences (Kirk, 2012). The need for more sophisticated approaches becomes evident as data complexity increases with longitudinal studies. While wavelet functions have been successfully applied in signal processing to decompose data into various frequency components, their application to aggregate repeated measures data in experimental designs for improved data analysis and inference remains underexplored.

To address these challenges, advanced statistical methods have emerged. Functional Analysis of Variance (FANOVA) extends classical ANOVA to situations where observations are not single points, but entire functions or curves (Acal & Aguilera, 2023; Centofanti *et al*., 2023). WANOVA is a specific subset of FANOVA combines wavelet analysis with ANOVA principles to address challenges posed by non-stationary and multi-resolution data (Williams *et al*., 2022; Girimurugan & Chicken, 2013; Angelini & Vidakovic, 2003). These methods aim to transform high-dimensional correlated data into simplified forms suitable for hypothesis testing. However, the development of customized wavelet functions for specific data characteristics, particularly for aggregating repeated measures into a single value for CRD-based ANOVA, represents a significant research gap. This study focuses on filling this gap by proposing a novel wavelet derived from the Gaussian function.

3. PROPOSED METHODOLOGY

This study proposes a wavelet function derived from the first derivative of the Gaussian probability density function (PDF). The choice of the Gaussian function is motivated by its optimal time-frequency localization and smoothness (Shafi et al., 2009; Chui & Wang, 1996), making it well-suited for detecting continuous and subtle transitions common in longitudinal data. The use of the first derivative is crucial as it captures the rate of change in the signal, emphasizing transitions between measurements that often hold significant analytical value in repeated measures data.

3.1. Wavelet Function using Gaussian Distribution

The Gaussian pdf is given by

$$f_G(y;\mu,\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \tag{1}$$

where $y \in \Re$ is the observation variable, $\mu \in \Re$ is the location parameter, and $\sigma > 0$ is the scale parameter.

The proposed wavelet function, $\psi_G(y)$, is obtained by taking the first derivative of the Gaussian function

$$\psi_G(y)=-\frac{(y-\mu)}{\sigma^3\sqrt{2\pi}}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \tag{2}$$

3.2. Aggregation Process

The proposed wavelet function is applied to each subject's repeated measurements to aggregate them into a single value. For each subject's time series $y_{ij}(t)$, the aggregation process involves:

1. Computing a location parameter for each pair of consecutive observations
2. Shifting this location parameter using the average of consecutive observations to obtain the wavelet function $\psi_G(y)$
3. Summing the wavelet transformation across all time points for each subject to obtain a single aggregated observation W_{ij}

$$W_{ij}=\sum_{t=1}^{T}\psi_G(y) \tag{3}$$

This transforms the original high-dimensional repeated measures dataset into a Completely Randomized Design (CRD) structure, where each subject is represented by one aggregated value.

3.3.Mathematical Properties of the First Derivative of the Gaussian Function

For a wavelet function to be valid for analysis, it must satisfy certain mathematical properties. These were evaluated for the proposed Gaussian derivative wavelet.

3.2.1. Admissibility Condition

The admissibility condition ensures that a wavelet function can be used for signal reconstruction and is defined as

$$C_\psi=\int_0^\infty \frac{|\hat{\psi}(\omega)|^2}{|\omega|}d\omega<\infty \tag{4}$$

where $\hat{\psi}(\omega)$ is the Fourier transform of the wavelet function $\psi(y)$, and ω is the angular frequency parameter of the Fourier transform.

Proposition: The first derivative of the Gaussian function, satisfies the admissibility condition.

Proof:

The Fourier transform of the first derivative Gaussian function is given by:

$$\hat{\psi}_G(\omega) = \int_{-\infty}^{\infty} e^{-i\omega y}\psi_G(y)dy = \frac{1}{\sigma^2} i\omega e^{-\frac{\omega^2}{2}} \tag{5}$$

Substituting this into the admissibility integral in equation (4)

$$C_{\psi_G} = \int_0^{\infty} \frac{\left|\frac{1}{\sigma^2} i\omega e^{-\frac{\omega^2}{2}}\right|^2}{|\omega|} d\omega = \frac{1}{\sigma^4}\int_0^{\infty} \frac{\omega^2 e^{-\omega^2}}{\omega} d\omega = \frac{1}{\sigma^4}\int_0^{\infty} \omega e^{-\omega^2} d\omega \tag{6}$$

Let $z = \omega^2 \Rightarrow dz = 2\omega d\omega.$ then as $z \to 0$ as $\omega \to 0$ and $z \to \infty$ as $\omega \to \infty$.
Therefore

$$C_{\psi_G} = \frac{1}{2\sigma^4}\int_0^{\infty} e^{-z} dz = -\frac{1}{2\sigma^4} e^{-z}\Big|_0^{\infty} = \frac{1}{2\sigma^4} \tag{7}$$

This confirmed that the Gaussian derivative wavelet satisfies the admissibility condition.

3.2.2. Energy Localization

A wavelet has energy localization if its squared magnitude integrates to a finite, well-defined total energy. This property is crucial as it ensures that the wavelet's energy is concentrated in a specific region, making it effective for signal analysis. The condition for this is:

$$E = \int_{-\infty}^{\infty} |\psi(y)|^2 dy < \infty \tag{8}$$

Proposition 2: The first derivative of the Gaussian function has a finite, well-defined total energy
Proof:

$$E = \int_{-\infty}^{\infty} \left| \frac{-(y-\mu)}{\sigma^3\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \right|^2 dy = \int_{-\infty}^{\infty} \frac{(y-\mu)^2}{\sigma^6(2\pi)} e^{-\frac{(y-\mu)^2}{\sigma^2}} dy \tag{9}$$

Substituting $t = \dfrac{y-\mu}{\sigma} \Rightarrow dy = \sigma dt$. And as $y \to \pm\infty$, $t \to \pm\infty$, so,

$$E = \frac{1}{2\pi\sigma^3}\int_{-\infty}^{\infty} t^2 e^{-t^2} dt \tag{10}$$

Let $u = t^2 \quad \Rightarrow \quad t = u^{1/2},$ and $dt = \dfrac{1}{2}u^{-1/2} du.$

Therefore,

$$\begin{aligned} E &= \frac{1}{2\pi\sigma^3} 2\int_0^{\infty} ue^{-u} \frac{1}{2} u^{-1/2} du = \frac{1}{2\pi\sigma^3}\int_0^{\infty} u^{1/2} e^{-u} du \\ &= \frac{1}{2\pi\sigma^3}\Gamma\left(\frac{3}{2}\right) = \frac{1}{2\pi\sigma^3}\frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{1}{4\pi\sigma^3}\sqrt{\pi} \\ &= \frac{1}{4\sigma^3\sqrt{\pi}} \end{aligned} \tag{11}$$

This indicates that the function has good energy concentration.

3.2.3. Vanishing Moments

A wavelet has vanishing moments if its integral against any polynomial of degree k is zero. This allows the wavelet to ignore smooth background trends and focus on abrupt changes. For $k=0$

$$\int_{-\infty}^{\infty} y^k \psi_G(y)dy = \int_{-\infty}^{\infty} -\frac{y-\mu}{\sigma^3\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy = 0 \tag{12}$$

Since $\psi_G(y)$ is an odd function (symmetric around μ) and the integral is over a symmetric interval, the integral evaluates to zero. Thus, it satisfies the first vanishing moment property.

3.2.4. Compact Support

A wavelet is said to have compact support if it is non-zero only within a finite interval. While the Gaussian function's derivative is not strictly compactly supported, it is effectively compactly supported because its values rapidly decay towards zero as $|y-\mu|$ increases. For 95% of the energy (from the normal distribution properties), the function is confined to the interval $[\mu-2\sigma, \mu+2\sigma]$.

3.3. ANOVA Model for Transformed Data

After the wavelet transformation, the dataset of aggregated observations W_{ij} for each subject is obtained. To compare treatment effects, classical one-way (ANOVA) is applied to this CRD-structured data. The model is given by:

$$W_{ij} = \eta + \kappa_i + \varepsilon_{ij} \tag{13}$$

where η is the grand mean, κ_i is the treatment effect for group i, and ε_{ij} is the random error, assumed to be normally distributed with mean zero and constant variance $\varepsilon_{ij} \sim N\left(0, \sigma^2\right)$. The standard ANOVA F-statistic is then used to test the null hypothesis of no treatment effect $H_0 : \kappa_1 = \kappa_2 = ... = \kappa_m = 0$.

4. SIMULATION STUDY

To evaluate the performance of the proposed Gaussian derivative wavelet, a synthetic dataset was generated to simulate an experimental study with treatment groups.

- **Experimental Design:** Four treatment groups $(T1, T2, T3, T4)$, with five subjects per treatment group, totaling 20 subjects.
- **Time Points:** Ten repeated measurements were generated for each subject, resulting in a balanced design.
- **Data Generation:** Each subject's signal was generated by combining a fixed treatment effect (linearly increasing: 10, 20, 30, 40 for $T1-T4$, respectively) with normal random noise (mean zero, equal variance).

The first derivative of the Gaussian function was applied to aggregate the repeated measurements of each subject into a single value. Subsequently, a one-way ANOVA was conducted on these aggregated values to assess the preservation of treatment differences. The performance was further evaluated by comparing Type I error rates, statistical power, and robustness to varying numbers of time points.

5. RESULTS AND DISCUSSION

5.1. ANOVA on Aggregated Data

The one-way ANOVA conducted on the data aggregated using the Gaussian derivative wavelet yielded the following results

Table 1: ANOVA Result

Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	3	1.235	0.4117	19.04	1.58e-05
Residuals	16	0.346	0.0216		

The ANOVA results show a highly significant $F-\text{value}$ of 19.04 with an extremely low $p-\text{value}\ (1.58e-05)$, strongly indicating significant differences among the treatment means. This confirms that the Gaussian derivative wavelet effectively preserves the underlying treatment-related information during the aggregation process, allowing for robust detection of group differences. Compared to the Mexican Hat benchmark $(F=151.50)$, while the F-value is lower, the Gaussian derivative still demonstrates strong statistical significance, validating its ability to extract meaningful signals.

5.2. Type I Error Rate

The Type I error rate for the Gaussian derivative wavelet was found to be 0.044. This value is remarkably close to the nominal 5% significance level (α=0.05). This indicates excellent control over false positives, implying that the aggregation process using the Gaussian derivative wavelet does not inflate the risk of incorrectly rejecting a true null hypothesis.

5.3. Power of the Test

The statistical power of the test using the Gaussian derivative wavelet was estimated at 0.947. This high power estimate (greater than 94%) demonstrates the strong ability of the method to detect true treatment effects when they exist. This provides robust evidence that the Gaussian derivative wavelet aggregation strategy successfully retains the treatment-level signal, yielding reliable inference via standard ANOVA.

5.4. Sensitivity to Time Point Variation

The robustness of the Gaussian derivative wavelet to varying numbers of time points was assessed by evaluating power at different longitudinal lengths

Table 2: Robustness to time point

Time Points	Power
5	0.938
10	0.932
15	0.952
20	0.944

The power remains consistently high across 5, 10, 15, and 20 time points, confirming the robustness of the Gaussian derivative wavelet-based aggregation to changes in longitudinal data length. While slight fluctuations are observed, the maintained high power demonstrates its adaptability and effectiveness regardless of the density of repeated measurements.

5.5. Discussion

The simulation results collectively validate the effectiveness of the proposed wavelet-based aggregation method using the first derivative of the Gaussian function. This method successfully transforms high-dimensional longitudinal data into ANOVA-ready univariate responses while preserving essential treatment differences, maintaining acceptable false-positive control, and exhibiting high statistical power. Its robustness to variations in measurement length further highlights its practical applicability. The ability to retain treatment differences after transforming complex repeated measures into a simplified CRD structure.

REFERENCES

Acal, C., & Aguilera, A. M. (2023). Basis expansion approaches for functional analysis of variance with repeated measures. *Advances in Data Analysis and Classification*, 17(2), 291-321.

Angelini, C., & Vidakovic, B. (2003). Some novel methods in wavelet data analysis: Wavelet ANOVA, F-test shrinkage and Γ-minimax wavelet shrinkage. *Wavelets and their Applications. Allied Publishers*, 31-45.

Centofanti, F., Colosimo, B. M., Grasso, M. L., Menafoglio, A., Palumbo, B., & Vantini, S. (2023). Robust functional ANOVA with application to additive manufacturing. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 72(5), 1210-1234.

Chui, C. K., & Wang, J. Z. (1996). A study of asymptotically optimal time-frequency localization by scaling functions and wavelets. *Annals of Numerical Mathematics*, *4*, 193-216.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. SIAM.

Girimurugan, S. B., & Chicken, E. (2013). Wavelet analysis of variance (wanova). *Technical Reports of the FSU. Department of Statistics M*, *1012*, 1-7.

Greenhouse, S. W., & Geisser, S. (1959). On Methods in the Analysis of Profile Data. *Psychometrika*, 24(2), 95–112.

Kevlahan, N. K. R. (2021). Adaptive wavelet methods for Earth systems modelling. *Fluids*, 6(7), 236.

Kirk, R. E. (2012). Experimental Design: Procedures for the Behavioral Sciences. SAGE Publications.

Nason, G. P. (2008). *Wavelet Methods in Statistics with R*. Springer.

Percival, D. B., & Walden, A. T. (2006). Wavelet Methods for Time Series Analysis. Cambridge University Press.

Ramakrishnan, S. (Ed.). (2024). *Modern Applications of Wavelet Transform*. BoD–Books on Demand.

Shafi, I., Ahmad, J., Shah, S. I., & Kashif, F. M. (2009). Techniques to obtain good resolution and concentrated time-frequency distributions: a review. *EURASIP Journal on Advances in Signal processing*, *2009*(1), 673539.

Williams, J., Hill, R. R., Pignatiello Jr, J. J., & Chicken, E. (2022). Wavelet analysis of variance box plot. *Journal of Applied Statistics*, 49(14), 3536-3563.

Zhang, J. T., & Liang, X. (2014). One-way ANOVA for functional data via globalizing the pointwise F-test. *Scandinavian Journal of Statistics*, 41(1), 51-71.