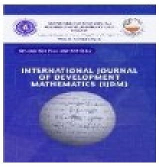
International Journal of Development Mathematics Vol 3 Issue 2, 022–042

INTERNATIONAL JOURNAL OF DEVELOPMENT MATHEMATICS

ISSN: 3026-8656 (Print) | 3026-8699 (Online)

journal homepage: https://ijdm.org.ng/index.php/Journals



A Zero Inflated Power Sine Sine Dagum Distribution for Modeling Heavy-Tailed Data with Excess Zeros

Ibrahim Abdullahi[a]* and Musa C. Dalah[b]

[a]Department of Mathematics and Statistics, Faculty of Science, Yobe State University, Damaturu, Nigeria.
[b]Department of Mathematics and Statistics, Faculty of Science, Yobe State University, Damaturu, Nigeria

ARTICLE INFO

Article history:

Received 30 December 2025
Received in Revised 20 February 2026
Accepted 25 February 2026

Keywords

Zero-inflated distribution, Power sine–sine transformation Dagum distribution, Heavy-tailed data, Maximum likelihood estimation.

MSC 2020 Subject classification: 62E15

Abstract

Zero-inflated and heavy-tailed data commonly arise in environmental, hydrological, biomedical, and reliability studies, where standard continuous distributions often fail to capture excess zeros and extreme positive observations simultaneously. This study aims to develop a flexible distributional framework for modeling such data with improved tail adaptability and zero-inflation capability. The study introduce Zero-Inflated Power Sine Sine Dagum (ZIPSSD) distribution, obtained by extending the classical Dagum distribution through a power sine–sine transformation and a zero-inflation mechanism. The resulting mixed discrete–continuous model provides greater flexibility for modeling skewed data with excess zeros and heavy-tailed behavior. The main statistical properties of the ZIPSSD distribution are derived, including the cumulative distribution function, probability density function, survival function, hazard rate, quantile function, and moments. Model parameters are es-

---

*Corresponding author Tel: +234 8066389147
Email address: ibworld82@yahoo.com

`https://doi.org/10.62054/ijdm/0302.02`

timated using maximum likelihood, and a Monte Carlo simulation study is conducted to evaluate estimator performance. An application to rainfall data demonstrates that the proposed ZIPSSD distribution provides a better fit than competing zero-inflated models based on likelihood and goodness-of-fit measures. The proposed model therefore offers a flexible and effective alternative for analyzing zero-inflated continuous data.

1 Introduction

Heavy-tailed distributions are indispensable tools for modeling real-world phenomena characterized by pronounced right skewness, extreme observations, and non-standard hazard-rate patterns. Such data structures arise frequently in income modeling, actuarial science, finance, environmental studies, survival analysis, and reliability engineering. Among classical heavy-tailed models, the Dagum distribution has attracted considerable attention because of its ability to capture income inequality, scale heterogeneity, and heavy-tail behavior (Dagum, 1977; Kleiber and Kotz, 2003). Its analytical tractability and flexible tail structure make it a valuable baseline distribution for modeling positive continuous data.

In recent years, considerable efforts have been devoted to extending the Dagum distribution to improve its adaptability to complex empirical data. Notable developments include Bayesian inference for censored Dagum data, extended Dagum distributions, odd Lomax–Dagum models, Dagum-X families, and inverted Dagum constructions (Alotaibi et al., 2021; Silva et al., 2015; Koleoso, 2023; Alghamdi et al., 2023; Osi et al., 2024). These extensions demonstrate that, although the classical Dagum distribution is useful, additional shape flexibility is often required to capture more complex distributional characteristics, including peak–tail interaction, multimodality, strong skewness, and non-monotonic hazard-rate behavior.

One major approach to enhancing distributional flexibility is the use of generator-based families of distributions. Such families provide systematic mechanisms for constructing new probability models by transforming a baseline distribution. In particular, trigonometric generators, especially sine-based transformations, have gained increasing attention because they introduce additional shape flexibility while preserving mathematical tractability. Recent contributions in this direction include the odd Lomax trigonometric generalized family, the alpha-sine-G family, and sine alpha power-G models (Bakr et al., 2022; Benchiha et al., 2023; Alghamdi et al., 2025). These models have been shown to improve goodness-of-fit and generate richer density and hazard-rate shapes. Similarly, sine-generalized odd log-logistic models have demonstrated the effectiveness of trigonometric generators in capturing complex tail behavior and reliability patterns (Osi et al., 2024).

Despite these advances, most Dagum-based extensions and sine-generated distributions remain purely continuous models defined on the positive real line. Consequently, they are not directly suitable for datasets containing a substantial proportion of structural zeros. This limitation is important in many applied settings, including environmental studies, biomedical applications, insurance claims, and reliability analysis, where the observed data may consist of both a degenerate mass at zero and a continuous positive component. In such cases, applying a purely continuous distribution may result in biased parameter estimates, poor goodness-of-fit, and misleading inferential conclusions.

Zero-inflated models provide a natural framework for addressing this limitation by combining a point mass at zero with a discrete or continuous distribution for the non-zero observations. The classical zero-inflated Poisson model of Lambert (1992) provided an important foundation for modeling excess zeros in count data. More recently, the zero-inflation concept has been extended to continuous distributions, including the zero-inflated Rayleigh distribution Liu et al. (2023), zero-inflated inverse Gaussian models Thangjai et al. (2025), and zero-inflated generalized Pareto-type models (Abbas et al., 2025). In particular, the zero-inflated generalized Pareto distribution provides a relevant example of partial integration between zero-inflation and heavy-tailed modeling. However, such models do not incorporate the Dagum baseline distribution or the additional flexibility induced by sine-based generator mechanisms.

Therefore, an important gap remains in the existing literature. Current models typically address heavy-tailed behavior, generator-based flexibility, or zero-inflation separately. Although some studies have combined zero-inflation with heavy-tailed continuous distributions, the simultaneous integration of a heavy-tailed Dagum-type baseline, a power sine–sine transformation mechanism, and a zero-inflation component has not been adequately investigated. This gap is particularly relevant for rainfall and environmental datasets, where excess zeros, strong skewness, and heavy-tailed positive observations often occur simultaneously.

Motivated by this gap, this paper introduces a new flexible distribution termed the Zero-Inflated Power Sine–Sine Dagum (ZIPSSD) distribution. The proposed model combines the heavy-tailed Dagum baseline distribution with a power sine–sine transformation and a zero-inflation mechanism. Consequently, the ZIPSSD distribution provides a mixed discrete–continuous framework capable of accommodating structural zeros, strong skewness, heavy-tailed positive observations, and diverse hazard-rate shapes within a mathematically tractable setting.

The main contributions of this paper are summarized as follows:

1. A new distribution, called the Zero-Inflated Power Sine–Sine Dagum (ZIPSSD) distribution, is proposed for modeling zero-inflated heavy-tailed continuous data.
2. Fundamental statistical properties of the proposed model are derived, including the cumulative distribution function, probability density function, survival function, hazard rate function, quantile function, raw moments, moment generating function, mode, and Shannon entropy.
3. A maximum likelihood estimation procedure is developed for estimating the unknown model parameters, and the associated inferential framework is discussed.
4. A Monte Carlo simulation study is conducted to evaluate the finite-sample performance of the maximum likelihood estimators.

5. The practical usefulness of the proposed model is demonstrated using rainfall data, and its performance is compared with competing zero-inflated models using likelihood-based and goodness-of-fit criteria.

The remainder of the paper is organized as follows. Section 2 presents the basic definitions and preliminary results. Section 3 develops the ZIPSSD distribution and its statistical properties. Section 4 presents the empirical finding of the study. Section 5 provides a detailed discussion of the empirical findings and their statistical applications. Finally, Section 6 concludes the paper and suggests possible directions for future research.

1.1 Literature Review

The development of flexible probability distributions has been an active area of research, particularly in response to increasingly complex data structures characterized by skewness, heavy tails, and non-standard hazard-rate behavior. The Dagum distribution remains a prominent model in this context due to its ability to capture strong right-tail behavior and scale variability (Dagum, 1977; Kleiber and Kotz, 2003).

Several extensions of the Dagum distribution have been proposed to improve its flexibility. These include the extended Dagum distribution, the Dagum-X family, and inverted and odd-type Dagum models, which allow for more diverse tail behavior and improved goodness-of-fit in empirical applications (Silva et al., 2015; Koleoso, 2023; Alghamdi et al., 2023; Osi et al., 2024). These developments highlight the continued relevance of Dagum-based models in statistical modeling.

In parallel, generator-based approaches have become a powerful methodology for constructing new families of distributions. Among these, trigonometric generators—particularly sine-based transformations—have proven especially effective in introducing additional flexibility. The alpha-sine-G family and related sine-generated models have demonstrated superior performance in modeling skewed and heavy-tailed data (Benchiha et al., 2023). More recent developments, including sine Type-II Topp-Leone-G and other sine-transformed families, further emphasize the versatility of trigonometric constructions (Osi et al., 2024; Isa et al., 2023).

In addition, transformation-based methods continue to play a significant role in distribution theory. These approaches enable the construction of new models with improved tail behavior, increased flexibility, and enhanced fitting capability for real-world data (Tang et al., 2025).

Another important line of research focuses on zero-inflated models. Originally developed for count data Lambert (1992), the concept has been extended to continuous distributions to address datasets with structural zeros. Recent contributions include zero-inflated Rayleigh and generalized Pareto-type models, which have demonstrated strong performance in reliability and environmental studies (Liu et al., 2023; Abbas et al., 2025).

Despite these advances, existing studies largely treat heavy-tailed modeling, generator-based transformations, and zero-inflation independently. There remains a lack of unified models that simultaneously incorporate these three important features. In particular, the integration of sine–sine transformation mechanisms with power structures and zero-inflation within a Dagum-type framework has not been adequately addressed in the literature.

2 Preliminaries

This section introduces the basic definitions and notations required for the construction of the proposed model.

Definition 2.1. *Let X be a positive continuous random variable following the Dagum distribution. Its cumulative distribution function is given by*

$$G(x) = \left[1 + \left(\frac{x}{b}\right)^{-a}\right]^{-p}, \qquad x > 0, \tag{1}$$

where $a > 0$ and $p > 0$ are shape parameters controlling the skewness and tail behavior of the distribution, while $b > 0$ is a scale parameter controlling the spread of the distribution. In particular, the parameter a regulates the rate of decay in the lower-to-middle range of the distribution, b determines the scale of the positive observations, and p provides additional power-type shape flexibility for modeling heavy-tailed behavior.

Definition 2.2. *The Power Sine Sine Dagum (PSSDagum) distribution is defined by*

$$F_{PSS}(x) = \left[\sin\left(\frac{\pi}{2}G(x)\right)\right]^{\alpha}, \qquad x > 0, \tag{2}$$

where $\alpha > 0$ is a shape parameter.

Definition 2.3. *A random variable Y is said to follow the Zero-Inflated Power Sine Sine Dagum (ZIPSSD) distribution if*

$$P(Y = 0) = \pi_0, \qquad P(Y > 0) = 1 - \pi_0, \quad 0 < \pi_0 < 1,$$

and for $x > 0$,

$$F_Y(x) = \pi_0 + (1 - \pi_0)F_{PSS}(x), \tag{3}$$

where $F_{PSS}(x)$ is given in (2).

3 Methods

This section presents the methodological framework for developing the proposed Zero-Inflated Power Sine–Sine Dagum (ZIPSSD) distribution. The method involves constructing the model by combining the Dagum baseline distribution with a power sine–sine transformation and a zero-inflation mechanism. The main statistical properties of the proposed distribution are then derived, including the density, distribution, survival, hazard, quantile, moment, and entropy functions. Finally, the model parameters are estimated using the maximum likelihood method, while simulation and real-data analyses are used to evaluate estimator performance and practical applicability.

3.1 Model Construction

3.1.1 Probability Density Function

Proposition 3.1. *The ZIPSSD distribution has a mixed density given by*

$$f_Y(y) = \pi_0\,\mathbf{1}_{\{0\}}(y) + (1-\pi_0) f_{PSS}(y)\mathbf{1}_{(0,\infty)}(y), \tag{4}$$

where

$$\begin{aligned} f_{PSS}(x) = {} & \frac{\alpha\pi}{2}\,\frac{ap}{b}\left(\frac{x}{b}\right)^{-a-1}\left(1+\left(\frac{x}{b}\right)^{-a}\right)^{-p-1} \\ & \times \left[\sin\!\left(\frac{\pi}{2}\left(1+\left(\frac{x}{b}\right)^{-a}\right)^{-p}\right)\right]^{\alpha-1} \\ & \times \cos\!\left(\frac{\pi}{2}\left(1+\left(\frac{x}{b}\right)^{-a}\right)^{-p}\right). \end{aligned} \tag{5}$$

Proof. From (2), define

$$u(x) = \sin\!\left(\frac{\pi}{2}\left(1+\left(\frac{x}{b}\right)^{-a}\right)^{-p}\right).$$

Then $F_{PSS}(x) = u(x)^{\alpha}$. Differentiating gives

$$f_{PSS}(x) = \alpha u(x)^{\alpha-1} u'(x).$$

Using

$$u'(x) = \cos\!\left(\frac{\pi}{2}\left(1+\left(\frac{x}{b}\right)^{-a}\right)^{-p}\right) \cdot \frac{\pi}{2}\frac{ap}{b}\left(\frac{x}{b}\right)^{-a-1}\left(1+\left(\frac{x}{b}\right)^{-a}\right)^{-p-1},$$

yields (5). □

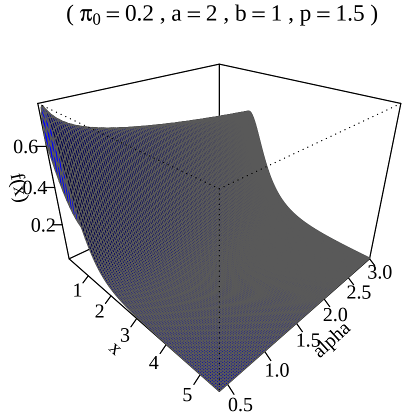


Figure 1. Probability density function of the Zero-Inflated Power Sine–Sine Dagum (ZIPSSD) distribution for selected values of the shape parameter α, with $\pi_0 = 0.2$, $a = 2$, $b = 1$, and $p = 1.5$.

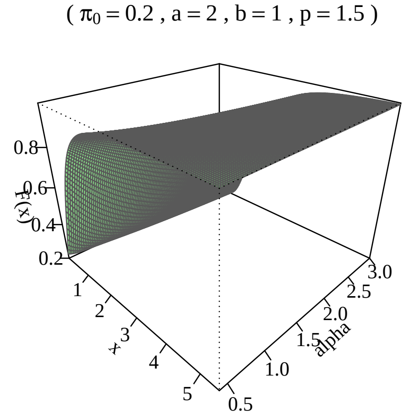


Figure 2. Cumulative distribution function of the Zero-Inflated Power Sine–Sine Dagum (ZIPSSD) distribution for selected values of the shape parameter α, with $\pi_0 = 0.2$, $a = 2$, $b = 1$, and $p = 1.5$.

3.1.2 Survival Function

Proposition 3.2. *Let* $Y \sim \mathrm{ZIPSSD}(\pi_0, \alpha, a, b, p)$. *Then the survival function is given by*

$$S_Y(x) = P(Y > x) = \begin{cases} 1, & x < 0, \\ 1 - \pi_0, & x = 0, \\ (1 - \pi_0)\left\{1 - \left[\sin\left(\frac{\pi}{2} G(x)\right)\right]^{\alpha}\right\}, & x > 0, \end{cases}$$

where $G(x)$ *is the Dagum cumulative distribution function.*

The discontinuity of $S_Y(x)$ at the origin is a direct consequence of the mixed discrete–continuous structure of the ZIPSSD model. The point mass π_0 assigned to $Y = 0$ produces an immediate downward jump in the survival function from $S_Y(x) = 1$ for $x < 0$ to $S_Y(0) = 1 - \pi_0$. Hence, only the remaining probability mass $1 - \pi_0$ is allocated to the continuous positive component of the distribution. This feature allows the model to explicitly distinguish structural zero observations from positive continuous outcomes.

Proof. By definition,

$$S_Y(x) = 1 - F_Y(x).$$

For $x < 0$, clearly $S_Y(x) = 1$.

At $x = 0$,

$$S_Y(0) = P(Y > 0) = 1 - \pi_0.$$

For $x > 0$, using (3), we obtain

$$\begin{aligned} S_Y(x) &= 1 - \pi_0 - (1-\pi_0)\left[\sin\left(\frac{\pi}{2}\left(1+\left(\frac{x}{b}\right)^{-a}\right)^{-p}\right)\right]^{\alpha} \\ &= (1-\pi_0)\left\{1 - \left[\sin\left(\frac{\pi}{2}\left(1+\left(\frac{x}{b}\right)^{-a}\right)^{-p}\right)\right]^{\alpha}\right\}. \end{aligned} \tag{6}$$

□

3.1.3 Hazard Function

Theorem 3.3. *Let* $h_D(x) = \frac{\frac{ap}{b}\left(\frac{x}{b}\right)^{-a-1}\left(1+\left(\frac{x}{b}\right)^{-a}\right)^{-p-1}}{1-\left(1+\left(\frac{x}{b}\right)^{-a}\right)^{-p}}$ *denote the hazard rate of the Dagum distribution. Then, for $x > 0$, the hazard rate of the ZIPSSD distribution can be written as*

$$h_Y(x) = h_D(x)\,R_\alpha(x),$$

where

$$R_\alpha(x) = \frac{\alpha\pi}{2}\left\{1 - \left(1+\left(\frac{x}{b}\right)^{-a}\right)^{-p}\right\}\frac{\left[\sin\left(\frac{\pi}{2}\left(1+\left(\frac{x}{b}\right)^{-a}\right)^{-p}\right)\right]^{\alpha-1}\cos\left(\frac{\pi}{2}\left(1+\left(\frac{x}{b}\right)^{-a}\right)^{-p}\right)}{1-\left[\sin\left(\frac{\pi}{2}\left(1+\left(\frac{x}{b}\right)^{-a}\right)^{-p}\right)\right]^{\alpha}}.$$

Proof. The hazard function is defined as

$$h_Y(x) = \frac{f_Y(x)}{S_Y(x)}.$$

Substituting from (4) and the (6), and simplifying, yields the result. □

3.1.4 Quantile Function

Theorem 3.4. *Let* $Y \sim \mathrm{ZIPSSD}(\pi_0, \alpha, a, b, p)$ *with* $0 < \pi_0 < 1$. *Then the quantile function of* Y *is given by*

$$Q_Y(u) = \begin{cases} 0, & 0 < u \le \pi_0, \\ b\left(\left[\dfrac{2}{\pi}\arcsin\left(\left(\dfrac{u-\pi_0}{1-\pi_0}\right)^{1/\alpha}\right)\right]^{-1/p} - 1\right)^{-1/a}, & \pi_0 < u < 1. \end{cases}$$

Proof. From (3), it follows immediately that

$$Q_Y(u) = 0, \qquad 0 < u \le \pi_0.$$

Now let $\pi_0 < u < 1$. Using (3), we have

$$u = \pi_0 + (1-\pi_0)F_{PSS}(x),$$

so that

$$F_{PSS}(x) = \frac{u-\pi_0}{1-\pi_0}.$$

By (2),

$$\left[\sin\left(\frac{\pi}{2}G(x)\right)\right]^{\alpha} = \frac{u-\pi_0}{1-\pi_0}.$$

Taking power $1/\alpha$ gives

$$\sin\left(\frac{\pi}{2}G(x)\right) = \left(\frac{u-\pi_0}{1-\pi_0}\right)^{1/\alpha}.$$

Applying the inverse sine function,

$$G(x) = \frac{2}{\pi}\arcsin\left(\left(\frac{u-\pi_0}{1-\pi_0}\right)^{1/\alpha}\right).$$

Using (1), we obtain

$$\left(1+\left(\frac{x}{b}\right)^{-a}\right)^{-p} = \frac{2}{\pi}\arcsin\left(\left(\frac{u-\pi_0}{1-\pi_0}\right)^{1/\alpha}\right).$$

Hence,

$$1+\left(\frac{x}{b}\right)^{-a} = \left[\frac{2}{\pi}\arcsin\left(\left(\frac{u-\pi_0}{1-\pi_0}\right)^{1/\alpha}\right)\right]^{-1/p},$$

and therefore

$$\left(\frac{x}{b}\right)^{-a} = \left[\frac{2}{\pi}\arcsin\left(\left(\frac{u-\pi_0}{1-\pi_0}\right)^{1/\alpha}\right)\right]^{-1/p} - 1.$$

Taking power $-1/a$ yields the stated result. □

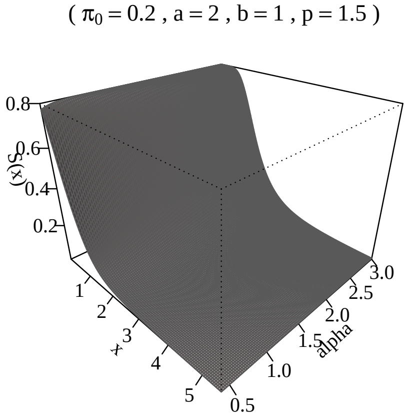


Figure 3. Survival function of the ZIPSSD distribution for varying α, with $\pi_0 = 0.2$, $a = 2$, $b = 1$, and $p = 1.5$.

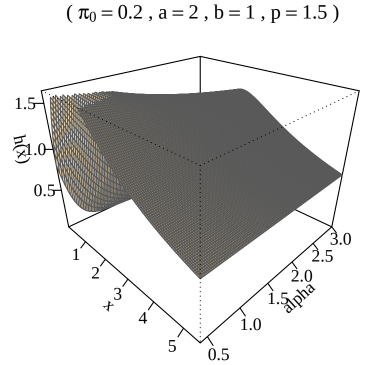


Figure 4. Hazard rate function of the ZIPSSD distribution for varying α, with $\pi_0 = 0.2$, $a = 2$, $b = 1$, and $p = 1.5$.

3.1.5 Raw Moment Representation

Proposition 3.5. *Let $Y \sim \text{ZIPSSD}(\pi_0, \alpha, a, b, p)$. Then, for $r > 0$, the rth raw moment of Y is given by*

$$E(Y^r) = (1-\pi_0)\frac{\alpha\pi}{2} b^r \int_0^1 \left(u^{-1/p} - 1\right)^{-r/a} \left[\sin\left(\frac{\pi}{2}u\right)\right]^{\alpha-1} \cos\left(\frac{\pi}{2}u\right)\, du.$$

Proof. Since $P(Y = 0) = \pi_0$ and $r > 0$, we have

$$E(Y^r) = (1-\pi_0) \int_0^\infty x^r f_{PSS}(x)\, dx,$$

where $f_{PSS}(x)$ is given in (5).

Substituting (5) into the above expression gives

$$E(Y^r) = (1-\pi_0)\frac{\alpha\pi}{2}\int_0^\infty x^r \left[\sin\Big(\frac{\pi}{2}G(x)\Big)\right]^{\alpha-1} \cos\Big(\frac{\pi}{2}G(x)\Big)\, g(x)\, dx.$$

Now let

$$u = G(x),$$

where $G(x)$ is defined in (1). Then

$$du = g(x)\, dx.$$

Also, since $G(0^+) = 0$ and $G(\infty) = 1$, the limits become $u \in (0,1)$. From (1),

$$x = b\left(u^{-1/p} - 1\right)^{-1/a},$$

and hence

$$x^r = b^r \left(u^{-1/p} - 1\right)^{-r/a}.$$

Therefore,

$$E(Y^r) = (1-\pi_0)\frac{\alpha\pi}{2} b^r \int_0^1 \left(u^{-1/p} - 1\right)^{-r/a} \left[\sin\Big(\frac{\pi}{2}u\Big)\right]^{\alpha-1} \cos\Big(\frac{\pi}{2}u\Big)\, du.$$

This completes the proof. □

3.1.6 Moment Generating Function

Proposition 3.6. *Let* $Y \sim \mathrm{ZIPSSD}(\pi_0, \alpha, a, b, p)$*. Whenever it exists, the moment generating function of* Y *is*

$$M_Y(t) = E(e^{tY}) = \pi_0 + (1-\pi_0) M_{PSS}(t),$$

where

$$M_{PSS}(t) = \int_0^\infty e^{tx} f_{PSS}(x)\, dx.$$

Equivalently,

$$M_Y(t) = \pi_0 + (1-\pi_0)\frac{\alpha\pi}{2}\int_0^1 \exp\Big\{tb\left(u^{-1/p} - 1\right)^{-1/a}\Big\} \left[\sin\Big(\frac{\pi}{2}u\Big)\right]^{\alpha-1} \cos\Big(\frac{\pi}{2}u\Big)\, du.$$

Proof. By definition,

$$M_Y(t) = E(e^{tY}).$$

Since Y has a point mass π_0 at zero and a continuous PSS–Dagum component on $(0,\infty)$, we have

$$M_Y(t) = \pi_0 + (1-\pi_0)\int_0^\infty e^{tx} f_{PSS}(x)\, dx.$$

Using (5), this becomes

$$M_Y(t) = \pi_0 + (1-\pi_0)\frac{\alpha\pi}{2}\int_0^\infty e^{tx}\left[\sin\Big(\frac{\pi}{2}G(x)\Big)\right]^{\alpha-1}\cos\Big(\frac{\pi}{2}G(x)\Big)\,g(x)\,dx.$$

Now let $u = G(x)$, where $G(x)$ is defined in (1). Then

$$du = g(x)\,dx,$$

and, from (1),

$$x = b\left(u^{-1/p}-1\right)^{-1/a}.$$

Substituting into the integral yields

$$M_Y(t) = \pi_0 + (1-\pi_0)\frac{\alpha\pi}{2}\int_0^1 \exp\Big\{tb\left(u^{-1/p}-1\right)^{-1/a}\Big\}\left[\sin\Big(\frac{\pi}{2}u\Big)\right]^{\alpha-1}\cos\Big(\frac{\pi}{2}u\Big)\,du.$$

□

3.1.7 Mode

Proposition 3.7. *Let* $Y \sim \mathrm{ZIPSSD}(\pi_0, \alpha, a, b, p)$ *with* $0 < \pi_0 < 1$. *Then the distribution has an atom at* $y = 0$ *of size* π_0. *Any interior mode of the continuous component on* $(0,\infty)$ *is obtained by solving*

$$\frac{d}{dx}\log f_{PSS}(x) = 0, \tag{7}$$

where $f_{PSS}(x)$ *is given in* (5).

Proof. Since

$$P(Y=0) = \pi_0,$$

the ZIPSSD distribution has a point mass at zero.

For $x > 0$, the density of the continuous component is

$$f_Y(x) = (1-\pi_0)f_{PSS}(x).$$

Because $(1-\pi_0)$ is a positive constant, maximizing $f_Y(x)$ over $(0,\infty)$ is equivalent to maximizing $f_{PSS}(x)$ over $(0,\infty)$. Hence any interior mode of the continuous component satisfies (7). □

Proposition 3.8. *Any interior mode* $x_m > 0$ *of the continuous component satisfies*

$$\begin{aligned} 0 = &-\frac{a+1}{x_m} + \frac{a(p+1)b^a x_m^{-a-1}}{1+b^a x_m^{-a}} \\ &+ \frac{\pi}{2}g(x_m)\left[(\alpha-1)\cot\Big(\frac{\pi}{2}G(x_m)\Big) - \tan\Big(\frac{\pi}{2}G(x_m)\Big)\right], \end{aligned} \tag{8}$$

where $G(x)$ *and* $g(x)$ *are the Dagum cumulative distribution function and density, respectively.*

Proof. From (5),

$$f_{PSS}(x) = \frac{\alpha\pi}{2}\, g(x) \left[\sin\Big(\frac{\pi}{2}G(x)\Big)\right]^{\alpha-1} \cos\Big(\frac{\pi}{2}G(x)\Big).$$

Hence,

$$\log f_{PSS}(x) = \log\Big(\frac{\alpha\pi}{2}\Big) + \log g(x) + (\alpha-1)\log\Big[\sin\Big(\frac{\pi}{2}G(x)\Big)\Big] + \log\Big[\cos\Big(\frac{\pi}{2}G(x)\Big)\Big].$$

Differentiating with respect to x gives

$$\frac{d}{dx}\log f_{PSS}(x) = \frac{g'(x)}{g(x)} + \frac{\pi}{2}g(x)\left[(\alpha-1)\cot\Big(\frac{\pi}{2}G(x)\Big) - \tan\Big(\frac{\pi}{2}G(x)\Big)\right],$$

since $G'(x) = g(x)$. Now,

$$g(x) = \frac{ap}{b}\left(\frac{x}{b}\right)^{-a-1}\left(1+\left(\frac{x}{b}\right)^{-a}\right)^{-p-1},$$

so that

$$\log g(x) = \log\Big(\frac{ap}{b}\Big) - (a+1)\log\Big(\frac{x}{b}\Big) - (p+1)\log\Big(1+\Big(\frac{x}{b}\Big)^{-a}\Big).$$

Therefore,

$$\begin{aligned}\frac{g'(x)}{g(x)} &= -\frac{a+1}{x} - (p+1)\frac{d}{dx}\log\Big(1+\Big(\frac{x}{b}\Big)^{-a}\Big)\\ &= -\frac{a+1}{x} + \frac{a(p+1)b^a x^{-a-1}}{1+b^a x^{-a}}.\end{aligned}$$

Substituting this expression into (7) and evaluating at $x = x_m$ yields (8). □

3.1.8 Shannon Entropy

Proposition 3.9. *Let $Y \sim \mathrm{ZIPSSD}(\pi_0, \alpha, a, b, p)$ with $0 < \pi_0 < 1$. Then the Shannon entropy of Y is given by*

$$H(Y) = -\pi_0\log\pi_0 - (1-\pi_0)\log(1-\pi_0) + (1-\pi_0)H_{PSS}, \tag{9}$$

where

$$H_{PSS} = -\int_0^\infty f_{PSS}(x)\log f_{PSS}(x)\,dx \tag{10}$$

is the differential entropy of the continuous PSS–Dagum component.

Proof. By definition,

$$H(Y) = E[-\log f_Y(Y)].$$

Since the ZIPSSD distribution has a point mass π_0 at zero and a continuous component

$(1-\pi_0)f_{PSS}(x)$ on $(0,\infty)$, we write

$$H(Y) = -\pi_0 \log \pi_0 - \int_0^\infty (1-\pi_0)f_{PSS}(x)\log[(1-\pi_0)f_{PSS}(x)]\,dx.$$

Expanding the logarithm,

$$\log[(1-\pi_0)f_{PSS}(x)] = \log(1-\pi_0) + \log f_{PSS}(x),$$

gives

$$\begin{aligned} H(Y) = -\pi_0 \log \pi_0 - (1-\pi_0)\log(1-\pi_0)\int_0^\infty f_{PSS}(x)\,dx \\ - (1-\pi_0)\int_0^\infty f_{PSS}(x)\log f_{PSS}(x)\,dx. \end{aligned}$$

Since $\int_0^\infty f_{PSS}(x)\,dx = 1$, we obtain

$$H(Y) = -\pi_0 \log \pi_0 - (1-\pi_0)\log(1-\pi_0) + (1-\pi_0)H_{PSS}.$$

□

3.2 Parameter Estimation

Proposition 3.10. *Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from the ZIPSSD distribution with parameter vector*

$$\mathbf{\Theta} = (\pi_0, \alpha, a, b, p)^\top,$$

where $0 < \pi_0 < 1$ and $\alpha, a, b, p > 0$. Then the model parameters can be estimated by the method of maximum likelihood.

Proof. Let

$$n_0 = \sum_{i=1}^{n} \mathbf{1}_{\{0\}}(y_i)$$

denote the number of zero observations and let $n_1 = n - n_0$ denote the number of positive observations. Without loss of generality, let $y_1, \ldots, y_{n_1}$ represent the positive observations. From (4), the likelihood function is given by

$$L(\mathbf{\Theta}) = \prod_{i=1}^{n} f_Y(y_i) = \pi_0^{n_0}(1-\pi_0)^{n_1}\prod_{i=1}^{n_1} f_{PSS}(y_i).$$

Hence, the log-likelihood function becomes

$$\ell(\mathbf{\Theta}) = n_0 \log \pi_0 + n_1 \log(1-\pi_0) + \sum_{i=1}^{n_1} \log f_{PSS}(y_i).$$

Substituting (5) into the above expression gives

$$\begin{aligned}\ell(\boldsymbol{\Theta}) = {} & n_0 \log \pi_0 + n_1 \log(1-\pi_0) + n_1 \log\left(\frac{\alpha\pi}{2}\right) + n_1 \log\left(\frac{ap}{b}\right) \\ & + (-a-1)\sum_{i=1}^{n_1} \log\left(\frac{y_i}{b}\right) + (-p-1)\sum_{i=1}^{n_1} \log\left[1+\left(\frac{y_i}{b}\right)^{-a}\right] \\ & + (\alpha-1)\sum_{i=1}^{n_1} \log\left[\sin\left\{\frac{\pi}{2}G(y_i)\right\}\right] + \sum_{i=1}^{n_1} \log\left[\cos\left\{\frac{\pi}{2}G(y_i)\right\}\right], \end{aligned} \tag{11}$$

where

$$G(y_i) = \left[1+\left(\frac{y_i}{b}\right)^{-a}\right]^{-p}.$$

let

$$z_i = \left(\frac{y_i}{b}\right)^{-a}, \qquad w_i = 1 + z_i, \qquad G_i = w_i^{-p},$$

$$t_i = \frac{\pi}{2}G_i, \qquad q_i = \log\left(\frac{y_i}{b}\right),$$

and

$$A_i = \frac{\pi}{2}\left[(\alpha-1)\cot(t_i) - \tan(t_i)\right], \qquad i = 1, \ldots, n_1.$$

Differentiating (11) with respect to the model parameters gives the following score functions:

$$\frac{\partial \ell}{\partial \pi_0} = \frac{n_0}{\pi_0} - \frac{n_1}{1-\pi_0},$$

$$\frac{\partial \ell}{\partial \alpha} = \frac{n_1}{\alpha} + \sum_{i=1}^{n_1} \log[\sin(t_i)],$$

$$\frac{\partial \ell}{\partial a} = \frac{n_1}{a} - \sum_{i=1}^{n_1} q_i + (p+1)\sum_{i=1}^{n_1} \frac{q_i z_i}{w_i} + \sum_{i=1}^{n_1} A_i \left(\frac{p q_i z_i}{w_i} G_i\right),$$

$$\frac{\partial \ell}{\partial b} = \frac{n_1 a}{b} - \frac{a(p+1)}{b}\sum_{i=1}^{n_1} \frac{z_i}{w_i} - \frac{ap}{b}\sum_{i=1}^{n_1} A_i \left(\frac{z_i}{w_i} G_i\right),$$

and

$$\frac{\partial \ell}{\partial p} = \frac{n_1}{p} - \sum_{i=1}^{n_1} \log(w_i) - \sum_{i=1}^{n_1} A_i G_i \log(w_i).$$

The maximum likelihood estimates are obtained by solving the system of score equations

$$\frac{\partial \ell}{\partial \pi_0} = 0, \qquad \frac{\partial \ell}{\partial \alpha} = 0, \qquad \frac{\partial \ell}{\partial a} = 0, \qquad \frac{\partial \ell}{\partial b} = 0, \qquad \frac{\partial \ell}{\partial p} = 0.$$

In particular, solving the score equation for π_0 yields the closed-form estimator

$$\widehat{\pi}_0 = \frac{n_0}{n}.$$

The remaining parameters (α, a, b, p) do not admit closed-form solutions and are therefore obtained by numerically maximizing the log-likelihood function. Suitable iterative methods include Newton–Raphson, quasi-Newton, or other gradient-based optimization procedures.

Furthermore, the observed Fisher information matrix is obtained from the negative Hessian matrix,

$$\mathcal{I}(\widehat{\boldsymbol{\Theta}}) = - \left. \frac{\partial^2 \ell(\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}\, \partial \boldsymbol{\Theta}^\top} \right|_{\boldsymbol{\Theta} = \widehat{\boldsymbol{\Theta}}},$$

which may be used to compute approximate standard errors and confidence intervals for the model parameters. □

4 Results

This section presents the empirical findings obtained from the Monte Carlo simulation study and the real data application. The simulation study evaluates the finite-sample performance of the maximum likelihood estimators, while the real data analysis assesses the practical applicability of the proposed ZIPSSD distribution in comparison with competing zero-inflated models.

4.1 Simulation Study

The performance of the ZIPSSD model was evaluated through a Monte Carlo simulation study. Random samples were generated using the inverse transformation method based on the quantile function derived in Section 3.1.4. The simulation considered sample sizes $n = 50, 100, 250, 500$, with fixed parameter values. For each configuration, $N = 1000$ replications were performed. Parameter estimation was carried out using the maximum likelihood estimation (MLE) method. The performance of the estimators was assessed using bias and mean squared error (MSE), defined as

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta, \qquad \text{MSE}(\hat{\theta}) = E\big[(\hat{\theta} - \theta)^2\big].$$

The simulation results indicate that the estimators are consistent, as the mean estimates converge to the true values and both bias and variance decrease with increasing sample size. The accuracy and stability of the estimators improve significantly for larger samples, although the shape parameters require relatively larger sample sizes for reliable estimation.

Table 1. Monte Carlo simulation results for the MLEs of the ZIPSSD model under scenario I considers the true parameter setting $(\pi_0, \alpha, a, b, p) = (0.2, 1.5, 2.0, 1.0, 1.2)$

n	Parameter	True	Mean	Bias	Variance	MSE	RMSE
50	π_0	0.2	0.2015	0.0015	0.003170	0.003172	0.056324
	α	1.5	7.8927	6.3927	268.710	309.576	17.5948
	a	2.0	5.7533	3.7533	151.050	165.137	12.8506
	b	1.0	0.9661	-0.0339	0.539102	0.540251	0.735017
	p	1.2	13.0532	11.8532	319.812	460.312	21.4549
100	π_0	0.2	0.1996	-0.0004	0.001620	0.001620	0.040249
	α	1.5	4.4264	2.9264	143.768	152.332	12.3423
	a	2.0	2.7304	0.7304	18.873	19.4062	4.40525
	b	1.0	0.9772	-0.0228	0.288850	0.289372	0.537933
	p	1.2	8.1144	6.9144	160.927	208.735	14.4477
250	π_0	0.2	0.2006	0.0006	0.000594	0.000594	0.024371
	α	1.5	2.5308	1.0308	56.869	57.9314	7.61127
	a	2.0	2.2064	0.2064	0.159447	0.202056	0.449507
	b	1.0	0.9595	-0.0405	0.105980	0.107623	0.328060
	p	1.2	5.2268	4.0268	51.896	68.1113	8.25296
500	π_0	0.2	0.2011	0.0011	0.000330	0.000331	0.018201
	α	1.5	1.5161	0.0161	11.349	11.3491	3.36885
	a	2.0	2.1337	0.1337	0.073213	0.091076	0.301788
	b	1.0	0.9539	-0.0461	0.043197	0.045324	0.212895
	p	1.2	3.6688	2.4688	21.877	27.9720	5.28886

The simulation results presented in Tables 1 and 2 indicate that the estimators improve as the sample size increases, confirming their consistency. The parameter π_0 is accurately estimated across all sample sizes with negligible bias and very low variance. The parameters a and b show clear improvement as n increases, with reduced bias and variability. The parameter α exhibits large bias for small sample sizes but improves substantially for larger n, suggesting asymptotic reliability. However, the parameter p remains difficult to estimate, with relatively high bias and variance even at larger sample sizes. Overall, the results demonstrate that the ZIPSSD model performs well for most parameters, particularly as the sample size increases.

4.2 Application to Real Data

We analyzed a real rainfall dataset obtained from the NOAA GHCN-Daily database NOAA (2024) using the `GHCNr` package Chadwick et al. (2023) to demonstrate the applicability of the proposed Zero-Inflated Power Sine–Sine Dagum (ZIPSSD) distribution. After preprocessing, the dataset contained $n = 6206$ observations, comprising 3835 zero observations corresponding to dry days and 2371 positive rainfall observations. This yielded an observed zero proportion of 0.618, indicating a pronounced zero-inflated structure. The positive rainfall observations exhibited strong right skewness and heavy-tailed behavior. Therefore, we fitted three competing zero-inflated models: the zero-inflated Weibull, zero-inflated Dagum, and proposed ZIPSSD models. The parameters were estimated using maximum likelihood. For the ZIPSSD model, the estimated zero-inflation parameter was $\widehat{\pi}_0 = 0.618$, which closely matches the observed zero proportion and indicates that the model adequately captures the structural zero component of the data.

Table 2. Monte Carlo simulation results for the MLEs of the ZIPSSD model under scenario II considers the true parameter setting $(\pi_0, \alpha, a, b, p) = (0.2, 1.0, 1.5, 1.0, 1.0)$.

n	Parameter	True	Mean	Bias	Variance	MSE	RMSE
50	π_0	0.2	0.201500	0.001500	0.003170	0.003172	0.0563
	α	1.0	7.494739	6.494739	250.6706	292.8502	17.1129
	a	1.5	8.720262	7.220262	287.5245	339.6757	18.4303
	b	1.0	1.253888	0.253888	1.1840	1.2485	1.1174
	p	1.0	6.026664	5.026664	134.8008	160.0652	12.6517
100	π_0	0.2	0.199600	-0.000400	0.001620	0.001620	0.0402
	α	1.0	5.953165	4.953165	194.0004	219.5092	14.8158
	a	1.5	3.928296	2.428296	100.5060	106.4007	10.3151
	b	1.0	1.205568	0.205568	0.7089	0.7512	0.8667
	p	1.0	3.437701	2.437701	38.3527	44.2973	6.6556
250	π_0	0.2	0.200564	0.000564	0.000594	0.000594	0.0244
	α	1.0	4.724278	3.724278	139.6343	153.8504	12.4036
	a	1.5	1.855141	0.355141	7.2910	7.4171	2.7234
	b	1.0	1.180033	0.180033	0.4460	0.4790	0.6921
	p	1.0	2.236884	1.236884	11.1767	12.7055	3.5645
500	π_0	0.2	0.201096	0.001096	0.000330	0.000331	0.0182
	α	1.0	2.578460	1.578460	45.6892	48.1807	6.9412
	a	1.5	1.602617	0.102617	0.0623	0.0728	0.2699
	b	1.0	1.117178	0.117178	0.2301	0.2438	0.4938
	p	1.0	1.659505	0.659506	4.9562	5.3912	2.3219

Table 3 presents the model comparison results based on the log-likelihood and AIC values, together with the corresponding parameter estimates. The ZIPSSD model achieves the highest log-likelihood value and the lowest Akaike information criterion (AIC) value among the competing zero-inflated models, indicating the best likelihood-based fit to the rainfall data. The ZIP-Dagum model also improves upon the ZIP-Weibull model, suggesting that the Dagum baseline provides greater flexibility for capturing the heavy-tailed nature of the rainfall distribution. However, the further improvement achieved by the ZIPSSD model shows that the additional power sine–sine transformation enhances the flexibility of the Dagum baseline. Therefore, based on the likelihood-based model selection criteria, the ZIPSSD model is selected as the preferred model for modeling the zero-inflated and heavy-tailed rainfall data.

Table 3. Model comparison and parameter estimates for zero-inflated rainfall data

Model	**Model selection**		**Parameter estimates**				
	LogLik	**AIC**	$\widehat{\pi}_0$	$\widehat{\alpha}$	$\widehat{a}$	$\widehat{b}$	$\widehat{p}$
ZIP-Weibull	-8725.31	17456.62	0.618	–	1.247	2.135	–
ZIP-Dagum	-8705.88	17419.76	0.618	–	2.431	1.782	1.965
ZIPSSD	**−8698.75**	**17407.51**	0.618	1.284	2.215	1.695	1.843

Note: – indicates that the parameter is not applicable to the corresponding model specification.

5 Discussion

The simulation results provide evidence that the maximum likelihood estimators of the ZIPSSD parameters generally improve as the sample size increases. The zero-inflation parameter π_0 is estimated with high accuracy in both scenarios, even for small sample sizes. This suggests that the discrete mass at zero is well identified in the zero-inflated framework. The continuous component parameters show different levels of finite-sample stability. The parameters a and b generally become more stable as the sample size increases, with reductions in bias, variance, mean squared error (MSE), and root mean squared error (RMSE). The shape parameters α and p, however, exhibit larger finite-sample bias and variability, especially when $n = 50$ and $n = 100$. This behaviour is expected because the ZIPSSD model contains several shape parameters, and such parameters often require larger samples for stable estimation. Nevertheless, the improvement observed as n increases supports the asymptotic reliability of the estimation procedure. The real data application further demonstrates the usefulness of the proposed ZIPSSD model for rainfall data with excess zeros and heavy-tailed positive observations. The estimated value $\widehat{\pi}_0 = 0.618$ confirms the presence of a large proportion of dry days, supporting the use of a zero-inflated model rather than a purely continuous distribution. In addition, the improvement of ZIP-Dagum over ZIP-Weibull suggests that the Dagum baseline is more suitable for capturing heavy-tailed rainfall behaviour. Most importantly, the ZIPSSD model achieved the best likelihood-based performance among the fitted models. Its higher log-likelihood and lower AIC indicate that the additional power sine–sine transformation improves the flexibility of the Dagum baseline distribution. Therefore, the proposed ZIPSSD distribution provides a more flexible and empirically suitable model for zero-inflated rainfall data than the ZIP-Weibull and ZIP-Dagum alternatives. Overall, the findings confirm that the ZIPSSD model is effective for modeling datasets characterized by a mixture of structural zeros, skewness, and heavy-tailed positive observations.

However, the simulation results also suggest that some shape parameters may require moderately large samples for stable estimation. This should be considered when applying the model to small datasets.

6 Conclusion

This study introduced the Zero-Inflated Power Sine–Sine Dagum (ZIPSSD) distribution, a flexible mixed discrete–continuous model designed to capture excess zeros and heavy-tailed behavior within a unified framework. By integrating a sine-based transformation with the classical Dagum distribution and a zero-inflation mechanism, the proposed model significantly enhances distributional flexibility. Key structural properties of the model were derived, and parameter estimation was performed using the maximum likelihood approach. A Monte Carlo simulation study confirmed that the estimators exhibit satisfactory performance and improve with increasing sample size. The practical relevance of the ZIPSSD distribution was demonstrated through an application to rainfall data, where it outperformed competing zero-inflated models based on likelihood-based criteria and goodness-of-fit measures. These results highlight the capability of the model to effectively capture complex data features involving both structural zeros and heavy-tailed positive observations. Overall, the ZIPSSD distribution provides a robust and versatile tool for modeling zero-inflated continuous data. Future research may extend the model to regression frameworks, Bayesian inference, and multivariate settings.

Acknowledgements

The authors acknowledge the National Oceanic and Atmospheric Administration (NOAA) for providing access to the Global Historical Climatology Network-Daily (GHCN-Daily) rainfall data used in this study. The authors also acknowledge the developers of the `GHCNr` package for facilitating access to the rainfall data.

References

Dagum, C. A new model of personal income distribution. *Econometrica* **1977**, *45*, 1259–1276.

Kleiber, C.; Kotz, S. *Statistical Size Distributions in Economics and Actuarial Sciences*; John Wiley & Sons: Hoboken, NJ, USA, 2003.

Alotaibi, R.; Rezk, H.; Dey, S.; Okasha, H. Bayesian estimation for Dagum distribution based on progressive type I interval censoring. *PLOS ONE* **2021**, *16*(6), e0252556. https://doi.org/10.1371/journal.pone.0252556

Silva, A.O.; da Silva, L.C.M.; Cordeiro, G.M. The extended Dagum distribution: Properties and application. *Journal of Data Science* **2015**, *13*, 53–72.

Bakr, M.E.; Al-Babtain, A.A.; Mahmood, Z.; Aldallal, R.A.; Khosa, S.K.; Abd El-Raouf, M.M.; Hussam, E.; Gemeay, A.M. Statistical modelling for a new family of generalized distributions with real data applications. *Mathematical Biosciences and Engineering* **2022**, *19*, 8705–8740.

Benchiha, S.A.; Sapkota, L.P.; Al Mutairi, A.; Kumar, V.; Khashab, R.H.; Gemeay, A.M.; Elgarhy, M.; Nassr, S.G. A new sine family of generalized distributions: Statistical inference with applications. *Mathematical and Computational Applications* **2023**, *28*(4), 83. https://doi.org/10.3390/mca28040083

Koleoso, P.O. The properties of odd Lomax–Dagum distribution and its application. *Scientific African* **2023**, *19*, e01555. https://doi.org/10.1016/j.sciaf.2023.e01555

Alghamdi, A.S.; Alghamdi, H.; Fayomi, A. The new Dagum-X family of distributions: Properties and applications. *International Journal of Analysis and Applications* **2023**, *21*, 21–45.

Osi, A.A.; Sabo, S.A.; Musa, I.Z. Inverted Dagum distribution: Properties and application to lifetime dataset. *Reliability: Theory & Applications* **2024**, *19*(3), 595–604.

Osi, A.A.; Abubakar, U.; Isma'il, L.A. Sine generalized odd log-logistic family of distributions: Properties and application to real data. *Reliability: Theory & Applications* **2024**, *19*(4), 190–200.

Liu, F.X.; Liu, J.N.; Xu, P. A novel zero-inflated Rayleigh distribution and its properties. *Results in Physics* **2023**, *51*, 106634.

Alghamdi, A.S.; ALoufi, S.F.; Baharith, L.A. The Sine Alpha Power-G family of distributions: Characterizations, regression modeling, and applications. *Symmetry* **2025**, *17*(3), 468. https://doi.org/10.3390/sym17030468

Tang, B.; Frye, H.A.; Silander, J.A., Jr.; Gelfand, A.E. Zero-inflated multivariate Tobit regression modeling. *Journal of Statistical Planning and Inference* **2025**, *236*, 106229. https://doi.org/10.1016/j.jspi.2024.106229

Thangjai, W.; Niwitpong, S.-A.; Niwitpong, S. Estimation of the common mean of zero-inflated inverse Gaussian distributions: Application to road accident fatalities in Thailand. *Symmetry* **2025**, *17*(11), 1944. https://doi.org/10.3390/sym17111944

Isa, A.; Yahaya, A.; Suleiman, A. The sine Type II Topp–Leone-G family of distributions: Properties and applications. *Journal of Applied Statistics* **2023**, *50*, 2501–2523.

Abbas, M.; Khalil, M.; Ibrahim, S. Zero-inflated generalized Pareto distribution with applications. *Communications in Statistics – Simulation and Computation* **2025**, *54*, 112–130.

Lambert, D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **1992**, *34*, 1–14.

National Oceanic and Atmospheric Administration (NOAA). *Global Historical Climatology Network Daily (GHCN-Daily)*; 2024. Available online: https://www.ncei.noaa.gov/products/land-based-station/global-historical-climatology-network-daily (accessed on 11 April 2026).

Chadwick, R.; Goodess, C.; Harris, I.; Osborn, T. *GHCNr: Download and Process NOAA GHCN Daily Data*; R package, 2023. Available online: https://CRAN.R-project.org/package=GHCNr (accessed on 11 April 2026).