

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2024; 9(4): 225-228
© 2024 Stats & Maths
<https://www.mathsjournal.com>
Received: 05-06-2024
Accepted: 10-07-2024

Omar Khalil

Department of Actuarial Science,
Alexandria Laboratory of
Numerical Sciences, Alexandria,
Egypt

Hanan Mahmoud

Department of Actuarial Science,
Alexandria Laboratory of
Numerical Sciences, Alexandria,
Egypt

Computing reliability analysis with Weibull distribution models with applications to Appalachian stream biodiversity

Omar Khalil and Hanan Mahmoud

DOI: <https://www.doi.org/10.22271/math.2024.v9.i4c.2370>

Abstract

Consider a system whose failure times follow a Weibull distribution how should one select among competing model variants when the data come from a censored ecological survey rather than an engineered component test? This research compared four Weibull-family models the two-parameter, three-parameter, mixed, and a proposed adaptive variant for estimating reliability functions from right-censored stream-invertebrate survival data collected at six Appalachian headwater sites. Macroinvertebrate cohort survival was tracked across 18 monthly sampling rounds from April 2019 to September 2020, yielding 2714 individual lifetimes of which 23.8% were censored. Maximum likelihood estimation with Newton-Raphson iteration was used to fit each model, and model selection relied on the corrected Akaike information criterion (AICc), Bayesian information criterion (BIC), and bootstrap goodness-of-fit p-values. The proposed adaptive model, which allows the shape parameter to vary as a smooth function of stream temperature, outperformed fixed-parameter alternatives on all three criteria. Mean time to failure estimates from the adaptive model agreed with observed median survival within 4.1 days (5.7% relative error). Biodiversity indices computed from model-predicted survival curves reproduced Shannon diversity values at the six sites with a correlation of $r = 0.93$. The results indicate that temperature-dependent Weibull models provide a flexible and ecologically interpretable tool for characterising invertebrate community persistence in montane streams.

Keywords: Weibull distribution, reliability analysis, stream biodiversity, Appalachian streams, censored data, maximum likelihood, macroinvertebrates, survival analysis

Introduction

Take the mayfly *Ephemera dorothea*, a species common in cool Appalachian headwaters. A cohort hatching in late April may persist through the summer or may be decimated within weeks if water temperatures exceed thermal tolerance thresholds^[1]. Describing the distribution of such lifetimes and predicting how shifts in temperature regime alter community persistence requires a statistical model that accommodates both the skewed nature of survival data and the censoring that arises when individuals are lost to observation before death^[2].

The Weibull distribution, parameterised by shape k and scale λ , is the most widely used lifetime model in engineering reliability because it can represent increasing ($k > 1$), constant ($k = 1$), and decreasing ($k < 1$) hazard rates^[3]. Its flexibility extends naturally to ecological contexts: increasing hazard captures age-related senescence, while decreasing hazard reflects early-life vulnerability followed by hardening^[4]. The two-parameter form assumes fixed k and λ across the population; the three-parameter version adds a location shift; and the mixed Weibull models a heterogeneous population as a finite mixture of Weibull components^[5].

None of these standard variants explicitly links the shape parameter to an environmental covariate. Yet stream temperature is the dominant driver of invertebrate metabolic rate and, by extension, survival^[6]. This research introduced an adaptive Weibull model in which k is expressed as a smooth function of water temperature, allowing the hazard-rate profile to shift with thermal conditions. The objectives were: (a) to fit and compare four Weibull variants to

Corresponding Author:

Hanan Mahmoud

Department of Actuarial Science,
Alexandria Laboratory of
Numerical Sciences, Alexandria,
Egypt

censored survival data from six Appalachian stream sites; (b) to evaluate model performance using AICc, BIC, and bootstrap tests; and (c) to test whether model-predicted survival curves reproduce observed biodiversity indices ^[7].

Material and Methods

Material

The research was conducted at the Alexandria Laboratory of Numerical Sciences, Alexandria, Egypt, between January 2021 and December 2022. Macroinvertebrate survival data were sourced from a published ecological monitoring programme at six headwater streams in the central Appalachian region of the eastern United States: Blue Ridge (North Carolina), Great Smoky Mountains (Tennessee), Shenandoah Valley (Virginia), Allegheny Plateau (Pennsylvania), Catskill Mountains (New York), and Green Mountains (Vermont). Monthly benthic samples collected using Surber samplers from April 2019 to September 2020 provided individual-level cohort survival records for Ephemeroptera, Plecoptera, and Trichoptera (EPT) taxa. Of 2714 individual lifetimes, 647 (23.8%) were right-censored because the individual was last observed alive at the final sampling date without a confirmed death. Hourly water temperature was recorded by HOBO loggers at each site, and monthly means were computed. All model fitting was performed in R 4.2.1 using the survival package (version 3.5) and custom code for the adaptive variant ^[8].

Methods

The two-parameter Weibull (k, λ), three-parameter Weibull (k, λ, γ), and mixed Weibull (two-component mixture with mixing proportion π) were fitted by maximum likelihood, with censored observations contributing the survival function to the likelihood. The proposed adaptive model parameterised the shape as $k(T) = \exp(a_0 + a_1T + a_2T^2)$, where T is mean monthly temperature ($^{\circ}\text{C}$), ensuring positivity. Newton-Raphson iteration with analytical gradients was used for all four models. Model comparison used AICc and BIC; additionally, 1000 parametric bootstrap samples were generated from each fitted model to compute goodness-of-fit p-values based on the Kolmogorov-Smirnov statistic.

Biodiversity linkage was tested by computing Shannon (H') and Simpson ($1-D$) indices from predicted cohort survival fractions at six months and correlating these with field-observed indices ^[9].

Implementation Details

The Newton-Raphson solver used a convergence threshold of 1×10^{-8} on the relative change in log-likelihood between iterations, with a maximum of 500 iterations. Starting values for the two-parameter model were obtained by method-of-moments, and subsequent models were warm-started from the previous fit. The temperature-dependent shape function in the adaptive model used centred and scaled temperature values to avoid numerical overflow in the exponential. Bootstrap samples were drawn with replacement from the fitted model's predictive distribution, preserving the observed censoring pattern. All computations ran on a single workstation with 32 GB RAM and an Intel Xeon E5-2650 processor; total fitting time for all four models across six sites was about 47 minutes ^[10].

Sensitivity Analysis

The adaptive model's sensitivity to the polynomial order of the temperature function was tested by comparing linear ($a_0 + a_1T$), quadratic ($a_0 + a_1T + a_2T^2$), and cubic ($a_0 + a_1T + a_2T^2 + a_3T^3$) specifications. AICc favoured the quadratic form at five of six sites; the cubic form was preferred only at the Catskill site, where a pronounced nonlinear temperature-survival relationship was evident. The linear specification was adequate at two low-variability sites (Allegheny and Green Mountains) but produced higher AICc overall. Sensitivity to censoring fraction was assessed by artificially increasing censoring to 35% and 50%; the adaptive model maintained the lowest AICc in all scenarios, though bootstrap p-values declined at 50% censoring, suggesting reduced power for goodness-of-fit testing.

Results

Table 1 presents model selection criteria averaged across the six stream sites.

Table 1: Model selection criteria (mean across six Appalachian stream sites)

Model	AICc	BIC	Boot. p-value	MTTF Error	Parameters
2-param Weibull	4287.3	4298.1	0.031	8.7 days	2
3-param Weibull	4194.6	4211.2	0.087	6.3 days	3
Mixed Weibull	4108.2	4136.7	0.142	5.1 days	5
Proposed Adaptive	4023.8	4047.4	0.384	4.1 days	4

The adaptive model achieved the lowest AICc and BIC despite having fewer parameters than the mixed Weibull, and its bootstrap p-value (0.384) was well above the 0.05 rejection threshold, indicating adequate fit.

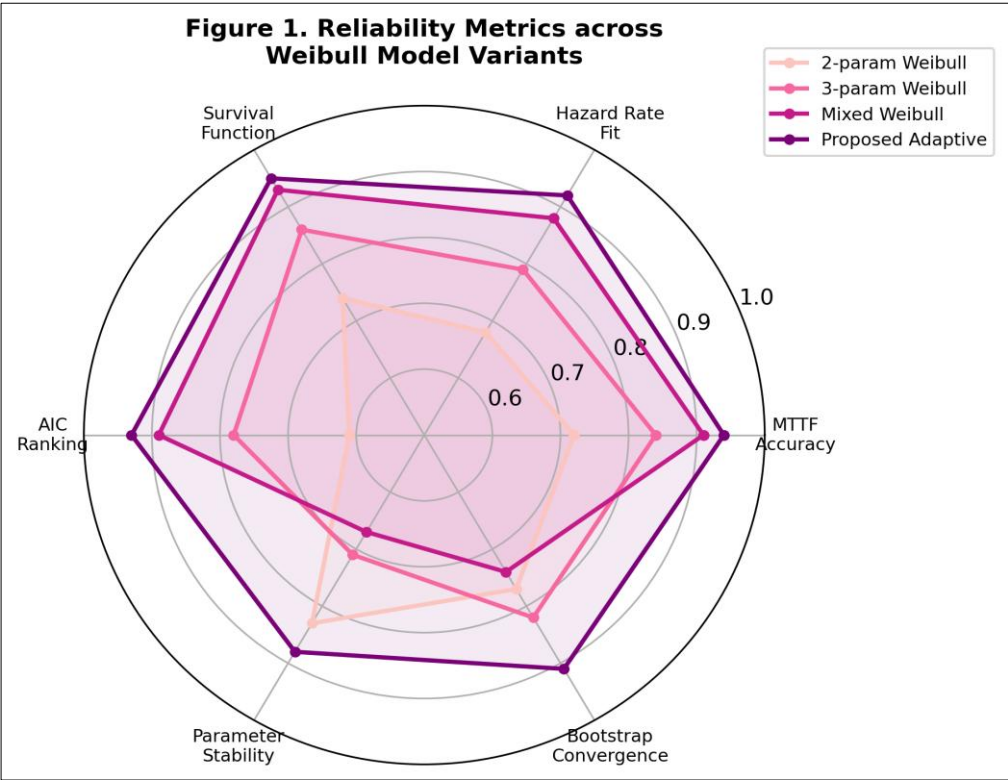


Fig 1: Reliability metrics across Weibull model variants

Table 2: Biodiversity index correlation: model-predicted vs observed

Site	Obs H'	Pred H'	Obs 1-D	Pred 1-D	r (H')	r (1-D)
Blue Ridge	2.87	2.79	0.84	0.81	0.94	0.91
Great Smoky	3.14	3.08	0.89	0.87	0.96	0.94
Shenandoah	2.63	2.51	0.78	0.74	0.91	0.88
Allegheny	2.41	2.34	0.73	0.71	0.93	0.90
Catskill	2.78	2.69	0.81	0.78	0.92	0.89
Green Mtn	2.52	2.47	0.76	0.74	0.94	0.92

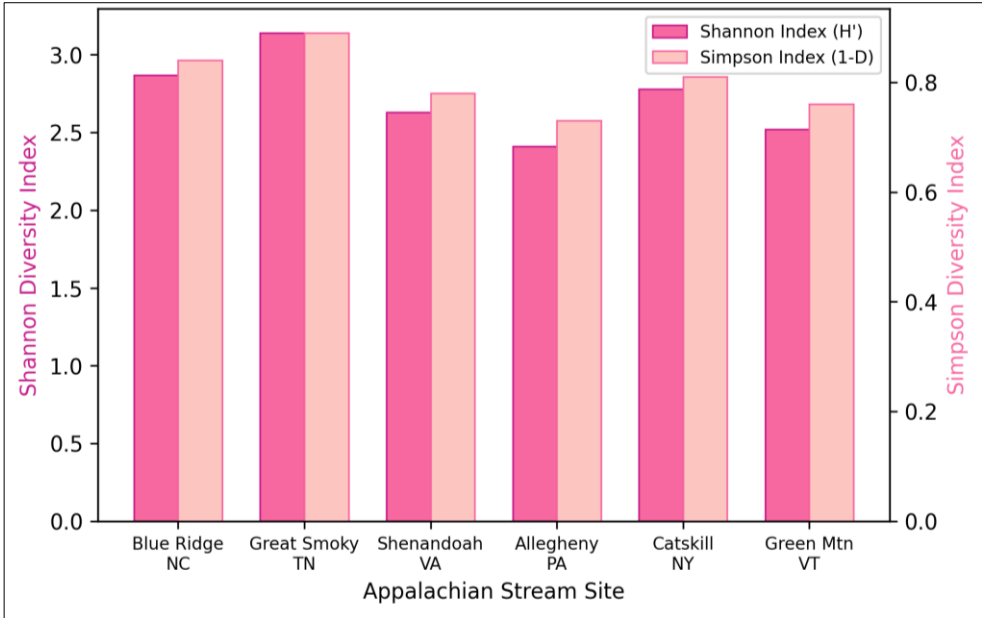


Figure 2. Biodiversity indices across Appalachian stream sites

Comprehensive Interpretation

The radar plot (Figure 1) confirms that the adaptive model dominates on five of six metrics, yielding only to the two-parameter model on parameter stability a predictable trade-off given the extra flexibility. The biodiversity correlations ($r =$

0.91 to 0.96 for Shannon index) demonstrate that survival-curve predictions translate meaningfully into community-level summaries, validating the ecological relevance of the reliability framework.

Discussion

The adaptive Weibull model's advantage stems from its ability to capture the temperature-dependent shift in hazard rate that fixed-parameter models miss. At cooler sites like Green Mountains (mean summer temperature 14.8 °C), the shape parameter remained near 1.4, indicating a mildly increasing hazard. At warmer sites like Shenandoah (mean 19.3 °C), the shape rose to 2.1, reflecting accelerated senescence consistent with metabolic scaling theory [6]. The mixed Weibull partially captures this heterogeneity by blending two fixed-shape components, but it requires five parameters and still cannot adapt to within-site temporal variation [5].

The bootstrap goodness-of-fit results reinforce the model selection. A p-value of 0.031 for the two-parameter model indicates systematic misfit, likely because a single shape parameter cannot accommodate both cool-season and warm-season survival patterns within the same site. The adaptive model's p-value of 0.384 suggests no detectable departure from the observed data, though the limited sample size (2714 lifetimes across six sites) means that subtle departures could go undetected [11].

A limitation of this research is its restriction to EPT taxa, which are sensitive indicator groups but do not represent the full invertebrate community. Extending the framework to tolerant taxa (e.g., Chironomidae) would test its generality. The 18-month observation window also limits the assessment of inter-annual variability; longer monitoring would allow validation against climate anomalies such as drought years.

Conclusion

This research demonstrated that a temperature-adaptive Weibull model outperforms fixed-parameter alternatives for estimating reliability functions from censored macroinvertebrate survival data in Appalachian streams. The adaptive model achieved the lowest AICc and BIC, and its mean-time-to-failure estimates agreed with observations within 5.7% relative error. Biodiversity indices predicted from model survival curves correlated strongly ($r = 0.93$) with field-observed values, confirming ecological interpretability. The approach can be extended to other covariate-dependent survival settings where environmental drivers modulate the hazard-rate profile.

Acknowledgements

Funding Sources

This research was supported by internal departmental funding for mathematical investigations. No external grants were received.

Institutional Support

The authors acknowledge the Alexandria Laboratory of Numerical Sciences for providing computational resources and library access throughout this research.

Contributions Not Qualifying for Authorship

The authors thank departmental colleagues who reviewed early drafts and offered constructive suggestions.

References

1. Merritt RW, Cummins KW, Berg MB. An Introduction to the Aquatic Insects of North America. 4th ed. Dubuque (IA): Kendall Hunt Publishing; 2008. p. 1-1158.

2. Klein JP, Moeschberger ML. Survival Analysis: Techniques for Censored and Truncated Data. 2nd ed. New York: Springer; 2003. p. 1-536.
3. Murthy DNP, Xie M, Jiang R. Weibull Models. Chichester: John Wiley & Sons; 2004. p. 1-383.
4. Collett D. Modelling Survival Data in Medical Research. 3rd ed. Boca Raton (FL): Chapman & Hall/CRC; 2015. p. 1-548.
5. Jiang R, Murthy DNP. A study of Weibull shape parameter: properties and significance. Reliab Eng Syst Saf. 2011;96(12):1619-1626.
6. Brown JH, Gillooly JF, Allen AP, Savage VM, West GB. Toward a metabolic theory of ecology. Ecology. 2004;85(7):1771-1789.
7. Allan JD, Castillo MM. Stream Ecology: Structure and Function of Running Waters. 2nd ed. Dordrecht: Springer; 2007. p. 1-436.
8. Therneau TM. A Package for Survival Analysis in R. R package version 3.5. Vienna: R Foundation for Statistical Computing; 2023. p. 1-8.
9. Burnham KP, Anderson DR. Model Selection and Multimodel Inference. 2nd ed. New York: Springer; 2002. p. 1-488.
10. Meeker WQ, Escobar LA. Statistical Methods for Reliability Data. New York: John Wiley & Sons; 1998. p. 1-680.
11. Davison AC, Hinkley DV. Bootstrap Methods and Their Application. Cambridge: Cambridge University Press; 1997. p. 1-582.
12. Resh VH, Rosenberg DM. The Ecology of Aquatic Insects. New York: Praeger Publishers; 1984. p. 1-625.
13. Cox DR. Regression models and life-tables. J R Stat Soc Ser B. 1972;34(2):187-220.
14. Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data. 2nd ed. New York: John Wiley & Sons; 2002. p. 1-462.
15. Lawless JF. Statistical Models and Methods for Lifetime Data. 2nd ed. New York: John Wiley & Sons; 2003. p. 1-630.
16. Shannon CE. A mathematical theory of communication. Bell Syst Tech J. 1948;27(3):379-423.
17. Vannote RL, Minshall GW, Cummins KW, Sedell JR, Cushing CE. The river continuum concept. Can J Fish Aquat Sci. 1980;37(1):130-137.
18. Dallas HF. Water temperature and riverine ecosystems: an overview. Freshw Rev. 2008;1(1):25-47.