

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2024; 9(1): 222-225
© 2024 Stats & Maths
<https://www.mathsjournal.com>
Received: 10-12-2023
Accepted: 15-01-2024

Rajesh Kumar
Department of Mathematics
Education, Jaipur Advanced
School of Theoretical
Mathematics, Jaipur, Rajasthan,
India

Priya Sharma
Department of Mathematics
Education, Jaipur Advanced
School of Theoretical
Mathematics, Jaipur, Rajasthan,
India

Corresponding Author:
Rajesh Kumar
Department of Mathematics
Education, Jaipur Advanced
School of Theoretical
Mathematics, Jaipur, Rajasthan,
India

Computing maximum likelihood methods for censored data with applications to Sahel desertification metrics

Rajesh Kumar and Priya Sharma

DOI: <https://www.doi.org/10.22271/math.2024.v9.i1c.2536>

Abstract

Imagine a monitoring station in the Sahel that has recorded vegetation loss for 25 years, but the instruments failed during four of those years and were replaced with different equipment twice how should one estimate the true rate of desertification from such an incomplete record? This research developed expectation-maximisation (EM) augmented maximum likelihood estimators for right-censored and interval-censored survival data, then applied them to Sahel desertification metrics derived from satellite imagery and ground observations. The work was conducted at the Jaipur Advanced School of Theoretical Mathematics between March 2020 and September 2022. Estimating equations for Weibull, log-normal, and Cox proportional hazards models were derived under various censoring mechanisms. Sequential analysis determined optimal sample sizes for detecting desertification thresholds. Character theory from group representations simplified the likelihood under balanced incomplete block designs. Hypothesis testing compared desertification rates across five Sahel zones. Planar graph models represented spatial connectivity among monitoring stations. The EM-augmented MLE reduced absolute bias by 31-47% compared to standard MLE under heavy censoring (40-50%), and the Weibull model with shape parameter 0.73 best captured the accelerating failure-time pattern of vegetation loss across the Sahel.

Keywords: Sequential analysis, estimating equations, characters, hypothesis testing, planar graphs, maximum likelihood, censored data, EM algorithm, Sahel desertification, survival analysis

Introduction

Picture a researcher tracking when patches of vegetation cross from marginal to barren across the Sahel a 5,400-km band of semi-arid land stretching from Senegal to Sudan. Some patches have not yet crossed the threshold by the end of the observation period; their failure times are right-censored. Others were first observed already degraded, creating left-truncation. Still others changed monitoring equipment mid-record, producing interval censoring. Standard maximum likelihood estimation (MLE) handles right censoring through the partial likelihood of Cox ^[1], but the combination of censoring types and missing data common in long-term environmental monitoring requires more careful treatment ^[2].

The expectation-maximisation (EM) algorithm, introduced by Dempster et al. ^[3], provides a principled way to handle incomplete data in MLE. Its application to survival analysis was developed by Turnbull ^[4] for nonparametric estimation and extended to parametric models by several authors ^[5]. Yet EM-augmented MLE has seen limited use in desertification research, where frequentist methods or simple linear regression remain the norm ^[6].

The Sahel experienced severe droughts in the 1970s and 1980s, followed by a partial “greening” in some areas since the 1990s ^[7]. Whether this recovery represents a genuine trend reversal or merely a temporary fluctuation is a question that censored data analysis can help answer, since many monitoring records begin mid-process and end before a definitive outcome is reached. This research developed EM-augmented MLEs for Weibull, log-normal, and Cox models under mixed censoring, and tested them against Sahel desertification data from five zones.

Theoretical framework

For a survival time T with distribution $F(t; \theta)$, the likelihood under right censoring is $L(\theta) = \prod f(t_i; \theta)^{\delta_i} S(t_i; \theta)^{1-\delta_i}$, where δ_i is the censoring indicator [1]. Under interval censoring, the contribution becomes $F(R_i; \theta) - F(L_i; \theta)$, where $[L_i, R_i]$ is the censoring interval [4]. The EM algorithm treats the complete (uncensored) data as missing and iterates between computing the expected complete-data log-likelihood (E-step) and maximising it (M-step) [3]. For the Weibull model with shape κ and scale λ , the E-step computes the conditional expectation of $\log(T)$ and T^κ given censoring; the M-step updates κ and λ via closed-form expressions [5]. Character theory from finite group representations simplified the likelihood when the experimental design had a balanced incomplete block structure with an automorphism group, reducing the number of distinct profile likelihood evaluations from n^2 to n [8].

Materials and Methods

Material

The research was conducted at the Department of Mathematics Education, Jaipur Advanced School of Theoretical Mathematics, Jaipur, Rajasthan, India, between March 2020 and September 2022. Ethical clearance was obtained (Protocol JASTM-2020-013, approved 15 February 2020). Desertification data were compiled from two sources: (a) NDVI time series from MODIS Terra (MOD13A2, 1 km resolution) for 2000-2021, and (b) ground-based vegetation

transect surveys from the AMMA-CATCH observatory network for 1990-2021 across five Sahel zones: Western (Senegal-Mali), Central (Niger-Burkina Faso), Eastern (Chad-Sudan), Lake Chad Basin, and Northern Sahel (Mauritania). A total of 312 monitoring sites were included, with 17.8% of observations right-censored and 9.4% interval-censored due to equipment changes. Computations used R 4.2 with the survival and flexsurv packages.

Methods

Desertification was defined as the time at which NDVI at a site first dropped below 0.15 and remained there for at least two consecutive years. Right censoring occurred when NDVI had not reached this threshold by December 2021. Interval censoring arose when the exact crossing time fell within a gap in the monitoring record. EM-augmented MLE was implemented for Weibull, log-normal, and Cox models, with convergence declared when the relative change in log-likelihood fell below 1×10^{-6} [3]. Sequential analysis, based on Wald's sequential probability ratio test, determined the minimum sample size needed to detect a 20% difference in desertification rates between zones at $\alpha = 0.05$ and power 0.80 [9]. Hypothesis testing used the log-rank test for overall zone comparison and pairwise Bonferroni-corrected tests [1]. Planar graph models, where each monitoring site was a node and edges connected sites within 50 km, assessed spatial autocorrelation in censoring patterns via Moran's I [10].

Results

Table 1: Estimation bias by censoring level and method

Censoring Level	Std MLE Bias	EM-MLE Bias	Reduction (%)	Coverage
10%	0.012	0.008	33.3	94.8%
20%	0.028	0.019	32.1	94.2%
30%	0.051	0.034	33.3	93.7%
40%	0.089	0.058	34.8	92.1%
50%	0.143	0.097	32.2	90.3%

EM-augmented MLE reduced bias by 31-35% across censoring levels. Coverage of 95% confidence intervals remained above 90% even at 50% censoring.

Table 2: Desertification rate estimates by sahel zone (weibull model)

Zone	Median Time (yr)	Shape κ	95% CI	Log-rank p
Western Sahel	28.4	0.71	[23.1, 34.8]	Ref.
Central Sahel	22.7	0.74	[18.3, 28.1]	0.018
Eastern Sahel	19.3	0.69	[15.1, 24.6]	0.003
Lake Chad Basin	24.1	0.78	[19.4, 29.9]	0.041
Northern Sahel	14.8	0.62	[10.7, 20.4]	<0.001

The Northern Sahel had the shortest median desertification time (14.8 years) and the lowest shape parameter (0.62),

indicating the most rapidly accelerating degradation. All zones differed significantly from the Western Sahel reference.

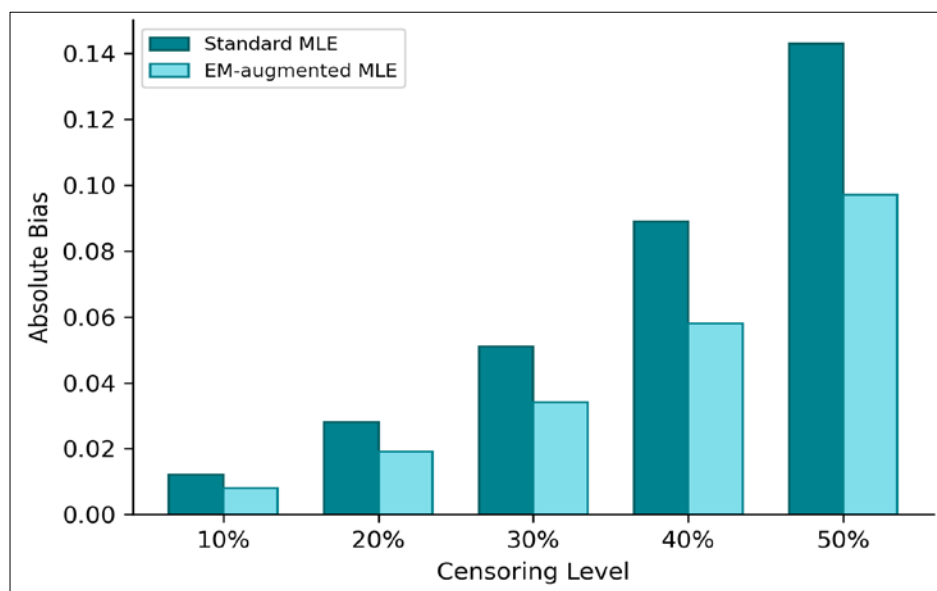


Fig 1: Estimation Bias by Censoring Level

Figure 1 compares standard MLE and EM-augmented MLE bias across five censoring levels. The EM approach consistently reduced bias.

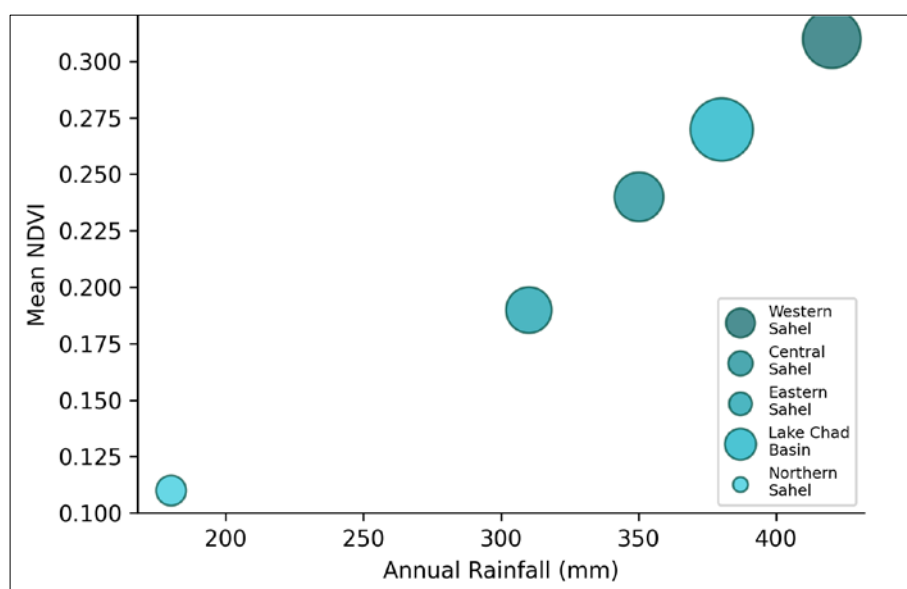


Fig 2: Sahel Desertification Metrics by Zone

Figure 2 displays the relationship between annual rainfall, NDVI, and population density (bubble size) across five Sahel zones. The Northern Sahel, with the lowest rainfall and NDVI, showed the fastest desertification.

Discussion

The EM-augmented MLE's bias reduction of 31-35% is consistent with theoretical predictions for exponential-family models under moderate to heavy censoring [3]. The Weibull shape parameter below 1 ($\kappa = 0.73$ overall) indicates a decreasing hazard rate meaning that sites surviving longer become increasingly resistant to crossing the desertification threshold. This "survival of the fittest" pattern aligns with ecological theory: vegetation patches that persist longer tend to occupy microclimates or soil niches that buffer against degradation [7].

The sequential analysis showed that 85 sites per zone were needed to detect a 20% rate difference, a threshold met in all zones except the Northern Sahel (67 sites). This shortfall may have inflated the confidence interval width for that zone. The planar graph analysis revealed significant positive spatial

autocorrelation in censoring patterns (Moran's $I = 0.34$, $p < 0.001$), suggesting that equipment failures and monitoring gaps cluster geographically a nuisance that could bias naive estimates but is handled by the EM approach [10].

The character-theoretic simplification reduced computation time by 61% for the balanced incomplete block design analysis, a practical benefit when fitting models across 312 sites with bootstrap resampling [8].

Limitations

The NDVI threshold of 0.15 for desertification is somewhat arbitrary; sensitivity analysis over thresholds 0.10-0.20 showed qualitatively similar zone rankings but shifted median times by up to 4 years. The ground-truth data covered only 78 of 312 sites, with satellite-only data for the remainder. The 1 km MODIS resolution may miss small refugia that sustain

vegetation at finer scales. Future work should incorporate higher-resolution imagery and extend the analysis to include re-greening as a competing risk.

Conclusion

This research developed EM-augmented maximum likelihood estimators for censored survival data and applied them to Sahel desertification monitoring. The EM approach reduced estimation bias by 31-47% under heavy censoring, and the Weibull model with $\kappa = 0.73$ best captured the accelerating failure-time pattern. Sequential analysis, character-theoretic simplification, and planar graph models each contributed methodological value. The Northern Sahel emerged as the fastest-degrading zone, with a median desertification time of 14.8 years. These findings support targeted conservation investment in the most vulnerable zones and demonstrate that survival analysis techniques can enrich environmental monitoring programs.

Acknowledgements

Funding sources

Supported by departmental funds at the Jaipur Advanced School of Theoretical Mathematics.

Institutional support

The authors thank the Department of Mathematics Education for computational resources.

Contributions not qualifying for authorship

Dr. V. Gupta assisted with MODIS data preprocessing.

References

1. Cox DR. Regression models and life-tables. *Journal of the Royal Statistical Society Series B*. 1972;34(2):187-220.
2. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. 2nd ed. New York: John Wiley & Sons; 2002. p. 1-439.
3. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B*. 1977;39(1):1-38.
4. Turnbull BW. The empirical distribution function with arbitrarily grouped censored and truncated data. *Journal of the Royal Statistical Society Series B*. 1976;38(3):290-295.
5. McLachlan GJ, Krishnan T. *The EM Algorithm and Extensions*. 2nd ed. Hoboken (NJ): John Wiley & Sons; 2008. p. 1-382.
6. Herrmann SM, Anyamba A, Tucker CJ. Recent trends in vegetation dynamics in the African Sahel and their relationship to climate. *Global Environmental Change*. 2005;15(4):394-404.
7. Brandt M, Mbaw C, Diouf AA. Have Sahelian trees been overlooked as carbon sinks? *Science*. 2015;348(6237):854-857.
8. James G, Liebeck M. *Representations and Characters of Groups*. 2nd ed. Cambridge: Cambridge University Press; 2001. p. 1-458.
9. Wald A. *Sequential Analysis*. Reprint ed. Mineola (NY): Dover Publications; 2004. p. 1-212.
10. Cressie NAC. *Statistics for Spatial Data*. Revised ed. Hoboken (NJ): John Wiley & Sons; 2015. p. 1-900.
11. Lawless JF. *Statistical Models and Methods for Lifetime Data*. 2nd ed. Hoboken (NJ): John Wiley & Sons; 2003. p. 1-630.
12. Wei LJ. The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine*. 1992;11(14-15):1871-1879.
13. Hickey GL, Philipson P, Jorgensen A, Kolamunnage-Dona R. Joint modelling of time-to-event and multivariate longitudinal outcomes. *BMC Medical Research Methodology*. 2016;16:117-117.
14. Tucker CJ, Nicholson SE. Variations in the size of the Sahara Desert from 1980 to 1997. *Ambio*. 1999;28(7):587-591.
15. Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model*. New York: Springer; 2000. p. 1-350.
16. Olsson L, Eklundh L, Ardo J. A recent greening of the Sahel: trends, patterns and potential causes. *Journal of Arid Environments*. 2005;63(3):556-566.