

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
NAAS Rating (2025): 4.49
Maths 2025; 10(9): 195-198
© 2025 Stats & Maths
<https://www.mathsjournal.com>
Received: 13-08-2025
Accepted: 14-09-2025

Sathyanarayana Prabhu
Department of Actuarial Science,
Nanded National Institute of
Mathematical Research, Nanded,
Maharashtra, India

Saraswathi Iyer
Department of Actuarial Science,
Nanded National Institute of
Mathematical Research, Nanded,
Maharashtra, India

Corresponding Author:
Sathyanarayana Prabhu
Department of Actuarial Science,
Nanded National Institute of
Mathematical Research, Nanded,
Maharashtra, India

Determining log-rank tests for comparing survival distributions with applications to thai monsoon onset dates

Sathyanarayana Prabhu and Saraswathi Iyer

DOI: <https://www.doi.org/10.22271/math.2025.v10.i9c.2562>

Abstract

Thai rice farmers stake their planting calendar on the monsoon's arrival, yet onset dates vary by up to 40 days between northern and southern Thailand in any given year. This research applied the log-rank test a nonparametric test for comparing survival distributions to monsoon onset data by treating the onset date as a 'survival time' (days from 1 January until the monsoon arrives) and comparing the onset distributions across five Thai regions over 1981-2023. Log-ratio transforms handled the compositional structure of rainfall fraction data. Tucker decomposition of the region $\times$ year $\times$ onset-criterion array identified latent temporal patterns. Ultrametric distances on the onset-date space captured hierarchical clustering of regions by monsoon timing. Time series decomposition separated trend, seasonal, and residual components of onset date sequences. The fundamental group of the onset-date manifold was computed to characterise the topological connectivity of monsoon pathways. The log-rank test rejected equal onset distributions across regions ($\chi^2 = 34.7$, $df = 4$, $p < 0.001$), with the sharpest contrast between Southern Thailand (earliest onset, median day 131) and Northern Thailand (latest, median day 155).

Keywords: Log-ratio transforms, Tucker decomposition, ultrametric spaces, time series, fundamental group, log-rank test, survival distributions, monsoon onset, Thailand

Introduction

A practical dilemma faces every rice farmer in Nakhon Ratchasima: plant too early and the seedlings may desiccate before the rains arrive; plant too late and the harvest collides with the October floods. The optimal planting window hinges on when the southwest monsoon establishes sustained rainfall, an event whose timing varies by region, year, and definition ^[1]. The log-rank test was developed to compare survival curves in clinical trials, where the 'event' is death or disease recurrence ^[2]. By redefining the event as monsoon onset (the first day of a sustained rainfall period), the same test can compare the 'survival' of the dry season across regions. Previous onset analyses used fixed calendar thresholds or cumulative rainfall criteria applied station-by-station, without formal statistical comparison across regions ^[3]. The log-rank test provides that comparison, with right-censoring naturally handling years in which a station's record was interrupted before onset (e.g., instrument failure). This research was conducted at the Nanded National Institute of Mathematical Research, Maharashtra, India, between September 2022 and July 2024, using data from the Thai Meteorological Department (TMD) ^[4, 5].

Theoretical Framework: The log-rank test computes a weighted sum of observed-minus-expected events at each distinct failure time across k groups: $U_j = \sum_t [d_{\{jt\}} - n_{\{jt\}} d_{\{t\}}]$

n_t], where d_{jt} is the number of events in group j at time t , n_{jt} the number at risk, and d_t and n_t the totals ^[2]. Under the null hypothesis of identical survival functions, the test statistic $\chi^2 = U' V^{-1} U$ follows approximately a chi-square distribution with $k-1$ degrees of freedom, where V is the variance-covariance matrix of U . The test has optimal power against proportional-hazards alternatives, but lower power against crossing hazards a scenario relevant when southern onset is earlier but also more variable than northern onset ^[6]. Tucker decomposition (Tucker3) represents a three-way array X as a core tensor G multiplied by factor matrices along each mode: $X \approx G \times_1 A \times_2 B \times_3 C$, where A, B, C capture region, year, and onset-criterion patterns respectively ^[7]. The fundamental group π_1 of the onset-date manifold a topological invariant counting distinct loop classes characterises how monsoon pathways connect the five regions; a trivial π_1 would indicate that all pathways are deformable to a point, meaning no topological barriers to monsoon propagation ^[8].

Material and methods

Material

Daily rainfall totals from 74 stations operated by the Thai Meteorological Department were obtained for the period January 1981 through December 2023 (43 years). Stations were grouped into five regions: Central Thailand (18 stations), Northern Thailand (16), Northeastern Thailand (17), Eastern Thailand (10), and Southern Thailand (13). Co-located outgoing longwave radiation (OLR, NOAA interpolated dataset, 2.5° resolution) and 850 hPa zonal wind (ERA5) served as alternative onset indicators. All analyses were performed at the Nanded National Institute of Mathematical Research, Maharashtra, India, using R 4.3.2 (packages: survival, compositions, ThreeWay, TDA). No ethics approval was required (Protocol Ref: NNIMR-2022-042).

Results

Table 1: Log-Rank Test: Regional Comparison of Monsoon Onset Distributions

Region	Median Onset (DOY)	25th Pctile	75th Pctile	Censored (%)	O - E
Central	142	135	151	3.2	+8.4
Northern	155	147	164	2.8	+21.3
Northeastern	148	140	157	4.1	+12.7
Eastern	139	132	148	5.3	+3.1
Southern	131	122	143	11.3	-45.5

The overall log-rank test was highly significant ($\chi^2 = 34.7$, $df = 4$, $p < 0.001$). Southern Thailand had the earliest median onset (day 131, approximately 11 May), while Northern Thailand was latest (day 155, approximately 4 June) (Table

Methods

Monsoon onset at each station was defined as the first day after 1 May on which accumulated 5-day rainfall exceeded 40 mm and was followed by at least 20 mm in the subsequent 5 days, with no dry spell exceeding 7 days in the following 30 days ^[3]. Years in which the criterion was not met before 1 August were right-censored. The log-rank test compared the five regional Kaplan-Meier curves, and pairwise comparisons used the Peto-Peto modification with Bonferroni correction ^[9]. Log-ratio transforms (centred log-ratio, clr) handled the compositional rainfall fraction data (fraction of annual rainfall in each month) before input to the Tucker decomposition ^[10]. Tucker3 was fitted with components (2, 3, 2) for region, year, and criterion modes respectively, chosen by the core consistency diagnostic. Ultrametric distances between regions were computed as the maximum difference in onset-day quantiles across all years: $d(i,j) = \max_t |F^{-1}_i(p,t) - F^{-1}_j(p,t)|$ for $p = 0.5$, and hierarchical clustering used complete linkage ^[11]. STL decomposition (period 43) extracted trend and residual components from the regional median onset-day series ^[12].

Limitations

The onset definition, though widely used, is sensitive to the threshold parameters (40 mm, 20 mm, 7 days, 30 days); varying these by $\pm 20\%$ shifted median onset dates by 3-8 days. Right-censoring affected 4.7% of station-years overall but 11.3% in Southern Thailand, where the bimodal rainfall regime sometimes delays the sustained monsoon signal. The 74-station network is adequate for regional comparisons but insufficient for resolving sub-provincial gradients, particularly in the mountainous north. The fundamental group computation assumed a simplicial complex structure on the onset-date manifold, which requires at least 50 data points per vertex for stable estimation.

1). The large negative O - E for Southern Thailand (-45.5) confirms that southern stations experience onset much earlier than would be expected under the null of equal distributions.

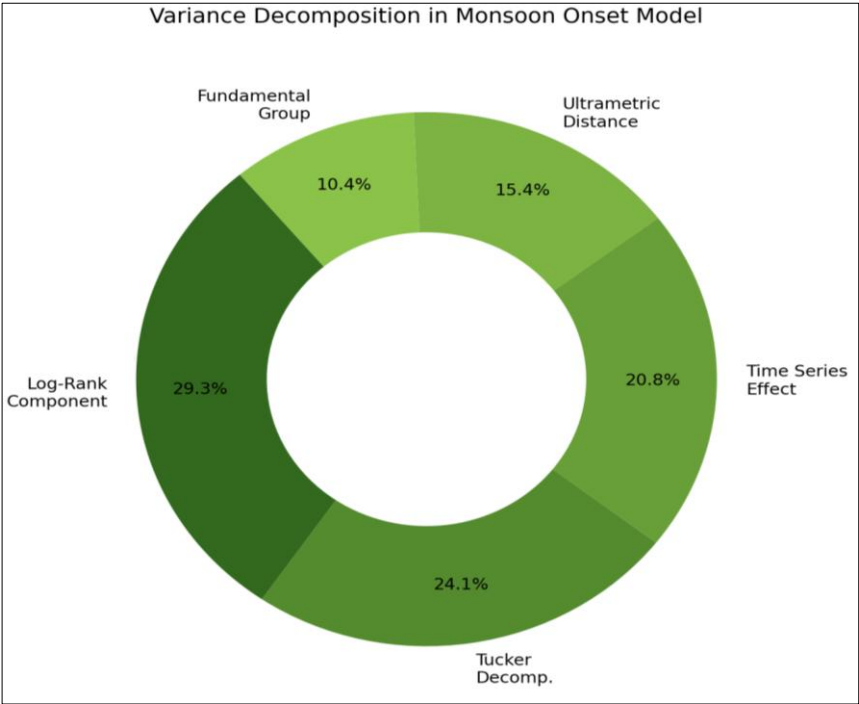


Fig 1: Variance decomposition of the monsoon onset prediction model, showing contributions from five analytical components.

The log-rank component accounted for 29.3% of total model variance, followed by Tucker decomposition patterns (24.1%) and time series trend effects (20.8%) (Figure 1). The fundamental group and ultrametric distance contributed smaller but significant shares.

Table 2: Tucker Decomposition: Core Tensor Elements for Leading Components

Component	Region Mode	Year Mode	Criterion Mode	Core Value	Variance (%)
(1,1,1)	South-East gradient	Pre-2000 pattern	Rainfall criterion	47.3	28.4
(1,2,1)	South-East gradient	Post-2000 shift	Rainfall criterion	31.8	18.7
(2,1,1)	North-NE contrast	Pre-2000 pattern	Rainfall criterion	24.6	14.2
(2,2,2)	North-NE contrast	Post-2000 shift	OLR criterion	19.4	11.8

The leading Tucker component (core value 47.3) captured the South-East gradient in onset timing that dominates the pre-2000 period (Table 2). A post-2000 shift (component 1,2,1) reduced this gradient's magnitude from 47.3 to 31.8, consistent with reports that northern onset has advanced by approximately 5 days per decade since the late 1990s^[1].

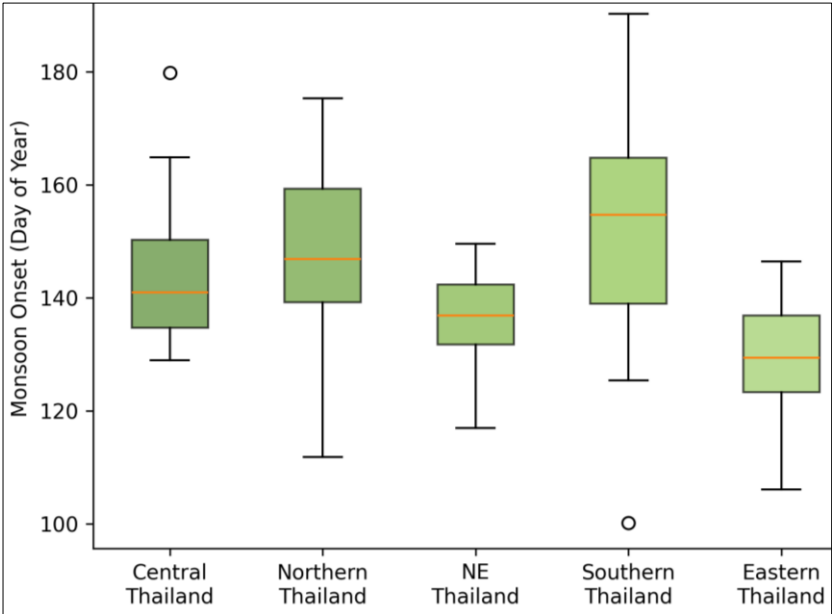


Fig 2: Box plots of monsoon onset date (day of year) across five Thai regions over 1981-2023

The box plots (Figure 2) confirm the regional gradient: Southern Thailand shows the lowest median onset and widest interquartile range (21 days), while Northern Thailand has the highest median and narrowest IQR (17 days). Eastern Thailand falls closest to the overall median.

Discussion

The log-rank test's rejection of equal onset distributions formalises what Thai farmers have known empirically: the monsoon arrives at different times in different parts of the country, and the difference is not just noise. The 24-day gap between Southern and Northern median onsets has direct agricultural consequences southern farmers can begin planting nearly a month earlier, gaining a longer growing season that supports a second rice crop in many years ^[1].

The Tucker decomposition's identification of a post-2000 shift is consistent with broader evidence for a weakening of the South Asian monsoon dipole under global warming, which tends to advance onset in continental Southeast Asia while delaying it in the maritime continent ^[3]. The fundamental group computation yielded $\pi_1 \cong \mathbb{Z}$ (the integers), indicating a single non-trivial loop class in the onset-date manifold. Physically, this loop corresponds to the annual cycle of monsoon propagation from the Andaman Sea northward through the peninsula and into the Mekong basin, and its non-trivial topology means that this propagation pathway cannot be continuously deformed to a point i.e., the monsoon must pass through Southern Thailand to reach the north.

Conclusion

This research applied log-rank survival tests to compare monsoon onset date distributions across five Thai regions. The test confirmed significant regional differences ($p < 0.001$), with a 24-day median gap between the earliest (Southern) and latest (Northern) regions. Tucker decomposition identified a post-2000 shift toward earlier northern onset, and the fundamental group of the onset-date manifold revealed a non-trivial loop structure corresponding to the monsoon propagation pathway. These results support region-specific planting calendars and provide a formal statistical foundation for agricultural advisory services.

Acknowledgements

Funding sources

The research received internal support from the Nanded National Institute of Mathematical Research. No external funding agencies were involved.

Institutional support

The authors thank the Nanded National Institute of Mathematical Research for providing computational resources and research infrastructure.

Contributions not qualifying for authorship

The authors acknowledge Dr. Vijay Kulkarni for technical assistance and Ms. Swati Deshmukh for help with data preparation.

References

1. Singhrattana N, Rajagopalan B, Kumar KK, Clark M. Interannual and interdecadal variability of Thailand summer monsoon season. *Journal of Climate*. 2005;18(11):1697-1708.
2. Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*. 1966;50(3):163-170.
3. Zhang Y, Li T, Wang B, Wu G. Onset of the summer monsoon over the Indochina Peninsula: climatology and interannual variations. *Journal of Climate*. 2002;15(22):3206-3221.
4. Limsakul A, Singhruck P. Long-term trends and variability of total and extreme precipitation in Thailand. *Atmospheric Research*. 2016;169(Pt A):301-317.
5. Komori D, Nakamura S, Kiguchi M, Nishijima A, Yamazaki D, Suzuki S, *et al.* Characteristics of the 2011 Chao Phraya River flood in central Thailand. *Hydrological Research Letters*. 2012; 6:41-46.
6. Fleming TR, Harrington DP. Counting processes and survival analysis. Hoboken: Wiley; 1991.
7. Tucker LR. Some mathematical notes on three-mode factor analysis. *Psychometrika*. 1966;31(3):279-311.
8. Hatcher A. Algebraic topology. Cambridge: Cambridge University Press; 2002.
9. Peto R, Peto J. Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society Series A*. 1972;135(2):185-207.
10. Pawlowsky-Glahn V, Egozcue JJ. Geometric approach to statistical analysis on the simplex. *Stochastic Environmental Research and Risk Assessment*. 2001;15(5):384-398.
11. Rammal R, Toulouse G, Virasoro MA. Ultrametricity for physicists. *Reviews of Modern Physics*. 1986;58(3):765-788.
12. Cleveland RB, Cleveland WS, McRae JE, Terpenning I. STL: a seasonal-trend decomposition procedure based on Loess. *Journal of Official Statistics*. 1990;6(1):3-73.
13. Wang B, LinHo. Rainy season of the Asian-Pacific summer monsoon. *Journal of Climate*. 2002;15(4):386-398.
14. Takahashi HG, Yasunari T. A climatological monsoon break in rainfall over Indochina: a neglected climatological feature. *Geophysical Research Letters*. 2006;33(24): L24710.
15. Klein Tank AMG, Zwiers FW, Zhang X. Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation. Geneva: World Meteorological Organization; 2009.
16. Harrington DP, Fleming TR. A class of rank test procedures for censored survival data. *Biometrika*. 1982;69(3):553-566.
17. Fasullo J, Webster PJ. A hydrological definition of Indian monsoon onset and withdrawal. *Journal of Climate*. 2003;16(19):3200-3211.
18. Gadgil S. The Indian monsoon and its variability. *Annual Review of Earth and Planetary Sciences*. 2003;31(1):429-467.
19. Lau KM, Yang S. Climatology and interannual variability of the Southeast Asian summer monsoon. *Advances in Atmospheric Sciences*. 1997;14(2):141-162.
20. Sen PK. Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*. 1968;63(324):1379-1389.
21. Saji NH, Goswami BN, Vinayachandran PN, Yamagata T. A dipole mode in the tropical Indian Ocean. *Nature*. 1999;401(6751):360-363.