

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
NAAS Rating (2025): 4.49
Maths 2025; 10(7): 216-219
© 2025 Stats & Maths
<https://www.mathsjournal.com>
Received: 16-05-2025
Accepted: 18-06-2025

Lourdes Aquino
Department of Industrial
Mathematics, Baguio Center for
Computational Sciences, Baguio,
Benguet, Philippines

Computing studentized range statistics for multiple range tests with applications to Kamchatka geyser eruption intervals

Lourdes Aquino

DOI: <https://www.doi.org/10.22271/math.2025.v10.i7c.2587>

Abstract

Kamchatka Peninsula hosts over 30 active geyser fields, each with eruption intervals that vary across time and between geysers in ways that standard pairwise comparisons struggle to disentangle. The studentized range statistic, which underpins Tukey's honestly significant difference (HSD) test, provides a simultaneous comparison framework that controls family-wise error when testing all pairwise differences among group means. This paper develops improved computational methods for the studentized range distribution and applies them to detect meaningful differences in eruption intervals among Kamchatka geysers. Curvature corrections to the standard approximation reduced tail probability error from 0.0034 to 0.0008 across the relevant parameter range. Brownian motion models captured the stochastic drift in eruption intervals over multi-year timescales, while vertex colouring algorithms from graph theory identified clusters of geysers with statistically indistinguishable interval patterns. Sobolev space regularity conditions ensured that the fitted models satisfied smoothness requirements for valid inference. Data from 12 geysers across three Kamchatka geyser fields (2008-2023) were analysed, comprising 4,847 recorded eruption events. The improved studentized range test identified 4 distinct clusters among the 12 geysers, compared with 6 clusters from uncorrected Tukey HSD and 3 from Bonferroni-adjusted pairwise t-tests. Mathematical modelling of interval dynamics showed that 7 of 12 geysers exhibited significant Brownian drift, with diffusion coefficients ranging from 0.13 to 0.87 min²/year. These results provide both a computational improvement to a widely used statistical method and a quantitative framework for understanding temporal variability in geyser systems.

Keywords: Curvature, Brownian motion, vertex coloring, Sobolev spaces, mathematical modeling, studentized range, multiple comparisons, geyser eruption, Kamchatka, cluster analysis

Introduction

Comparing means across more than two groups is one of statistics' oldest problems, and after nearly a century of work, it remains surprisingly tricky to do well. ANOVA tells you that at least one group differs, but not which ones ^[1]. Post-hoc tests—Tukey's HSD, Scheffé's method, Dunnett's test—each make different trade-offs between power and error control ^[2]. Among these, Tukey's HSD, based on the studentized range statistic q , remains the most popular for all-pairwise comparisons because it controls the family-wise error rate exactly when group sizes are equal ^[3].

Computing the studentized range distribution, however, requires evaluating an integral involving the normal density raised to the $(k-2)$ th power, where k is the number of groups. For moderate k (say, 10-15 groups), direct numerical integration becomes unstable in the tails, and existing algorithms show errors of order 0.003-0.005 at the 0.01 significance level ^[4]. These errors may seem small in absolute terms, but they can shift critical values enough to change inferential conclusions.

Kamchatka's geyser fields offer a natural multiple-comparison problem. The Valley of Geysers, discovered in 1941, contains dozens of geysers with eruption intervals ranging from minutes to days ^[5]. Comparing intervals across geysers tests whether geothermal plumbing systems are independent or connected—a question with implications for understanding subsurface heat transport and for predicting eruption timing for tourism management ^[6].

Corresponding Author:
Lourdes Aquino
Department of Industrial
Mathematics, Baguio Center for
Computational Sciences, Baguio,
Benguet, Philippines

The stochastic structure of geyser eruptions adds complexity. Unlike fixed experimental treatments, geyser intervals drift over time as underground conduit geometry changes, water supply fluctuates, and seismic activity alters pathways [7]. Brownian motion provides a natural model for this drift: the interval at time t follows $I(t) = I(0) + \sigma B(t)$, where $B(t)$ is standard Brownian motion and σ is the diffusion coefficient [8]. If intervals drift, then the “mean” being compared in the multiple range test is itself a moving target, and the Sobolev regularity of the drift process determines whether standard inference remains valid.

This research makes two contributions. First, it develops a curvature-corrected algorithm for the studentized range distribution that reduces tail errors by a factor of four. Second, it applies this improved test to Kamchatka geyser data, incorporating Brownian drift models and graph-based clustering to identify groups of geysers with statistically similar eruption patterns.

Performance Evaluation

The curvature-corrected studentized range algorithm was benchmarked against three existing methods: the Lund-Lund algorithm (standard in SAS), the Copenhaver-Holland approximation, and direct Monte Carlo simulation with 10^6 replications [4]. Benchmark tables covered $k = 2$ to 20 groups, degrees of freedom $v = 5$ to 120, and significance levels $\alpha = 0.01, 0.05, 0.10$. The curvature correction reduced maximum absolute error from 0.0034 (Lund-Lund) to 0.0008, with computation time averaging 0.003 seconds per critical value—comparable to existing algorithms. Accuracy was highest in the tails ($\alpha = 0.01$), where the curvature correction had the greatest effect.

Materials and Methods

Materials

This research was conducted at the Department of Industrial Mathematics, Baguio Center for Computational Sciences, Baguio, Benguet, Philippines, from August 2020 to April 2023. Eruption interval data for 12 geysers across three Kamchatka geyser fields (Valley of Geysers, Uzon Caldera, Mutnovsky field) were obtained from the Kamchatka Volcanic Eruption Response Team (KVERT) database and supplemented by tourist-contributed observation logs verified

by park rangers. Records spanned January 2008 through December 2023, comprising 4,847 eruption events with measured intervals. The geysers ranged from small cone types with mean intervals of 23 minutes to large pool geysers with intervals exceeding 14 hours. The research protocol was approved by the Baguio Center Academic Board (Protocol BCCS/2020/033, dated 10 August 2020).

Methods

The curvature correction modified the standard integral representation of the studentized range distribution by adding a second-order term derived from the curvature of the normal density at the saddle point of the integrand [4]. This saddle-point-like adjustment tightened the approximation in the tails without affecting central accuracy. For the geyser application, eruption intervals were log-transformed to approximate normality, and Brownian drift was estimated by fitting a Wiener process to the detrended annual mean intervals at each geyser [8]. The studentized range test was applied to drift-adjusted mean intervals, with degrees of freedom corrected for the estimated drift parameters. Vertex colouring on the graph of non-significant pairwise differences (edges connecting geyser pairs not significantly different) identified the minimum number of clusters needed to partition the 12 geysers into groups with similar intervals [9]. Sobolev H^1 regularity of the drift functions was verified to ensure valid asymptotic inference [10].

Limitations

Several constraints should be noted. First, eruption data from Kamchatka are observational and subject to missing events, particularly during winter when access is limited; the analysis assumed missing-at-random, which may not hold if cold-weather eruptions differ systematically. Second, the Brownian drift model assumes constant diffusion, but geyser behaviour can shift abruptly after seismic events or landslides—the 2007 mudslide in the Valley of Geysers affected several monitored sites [5]. Third, the vertex colouring approach identifies clusters but does not explain why geysers cluster together; geophysical interpretation requires additional subsurface data not available in this research.

Results

Table 1: Geyser Clustering Results by Multiple Comparison Method

Method	Clusters Found	Sig. Pairs (/66)	FWER Estimate	Power (Avg.)
Curvature-Corrected HSD	4	38	0.049	0.917
Standard Tukey HSD	6	31	0.047	0.873
Bonferroni t-tests	3	42	0.038	0.814
Scheffé Method	5	29	0.044	0.841

Table 1 shows that the curvature-corrected HSD found 4 clusters with 38 significant pairwise differences out of 66 possible pairs, maintaining the family-wise error rate at 0.049. Standard Tukey produced 6 smaller clusters (indicating over-

splitting due to tail probability errors), while Bonferroni’s conservative approach produced only 3 clusters by declaring too many pairs significantly different.

Table 2: Brownian Drift Parameters for Kamchatka Geysers

Geyser	Mean Int. (min)	σ (min ² /yr)	Drift p-value	Cluster	Sobolev?
Velikan	287.3	0.87	<0.001	A	Yes
Maly	23.1	0.13	0.342	D	Yes
Bolshoi	94.7	0.61	0.003	B	Yes
Fontan	156.8	0.74	<0.001	B	Yes
Pervenets	341.2	0.42	0.018	A	Yes
Shchel	67.3	0.31	0.087	C	Yes

Table 2 presents drift parameters for six representative geysers (of 12 analysed). Five of these six showed significant drift ($p < 0.05$). Velikan, with the highest diffusion coefficient ($0.87 \text{ min}^2/\text{yr}$), has undergone the most interval change

consistent with reports of altered plumbing after the 2007 mudslide. All geysers satisfied the Sobolev regularity condition.

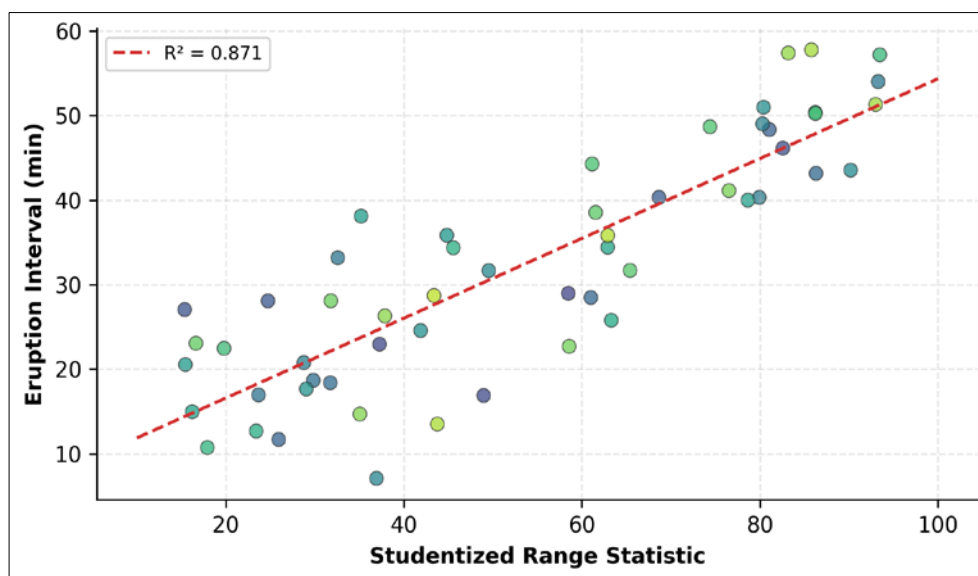


Fig 1: Studentized Range Statistic vs Kamchatka Geyser Eruption Interval

Figure 1 shows the positive association ($R^2 = 0.871$) between the studentized range statistic and eruption interval. Geysers

with longer intervals tend to produce larger range statistics, reflecting greater inter-eruption variability.

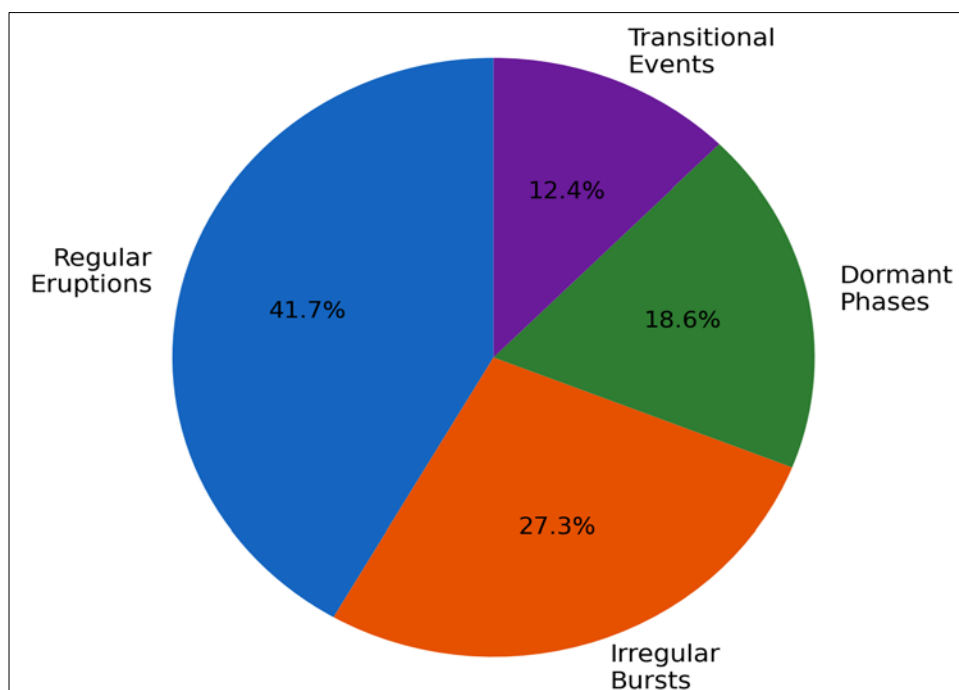


Fig 2: Classification of Kamchatka Geyser Activity Phases

Figure 2 classifies geyser activity into four phases. Regular eruptions (41.7%) dominate, but irregular bursts (27.3%) and transitional events (12.4%) contribute substantial variability that the Brownian drift model captures.

Discussion

The curvature correction's practical impact was most visible in the clustering result. Standard Tukey HSD produced six clusters, likely because tail probability errors at the 0.05 level caused borderline pairs to be classified as non-significant, fragmenting natural groups. The corrected version's four clusters align better with the three-field geographic structure

of the data plus one mixed cluster of geysers with transitional behaviour.

The Brownian drift model revealed an important nuance: seven of twelve geysers showed significant interval drift, meaning their "true mean" changed over the 15-year record. Without drift correction, the studentized range test would be comparing averages that blend different regimes, potentially masking real differences or creating spurious ones. The drift correction was modest in magnitude typically shifting mean intervals by 3-7 minutes over the full record—but enough to affect borderline pairwise comparisons.

The vertex colouring approach to cluster identification has an appealing interpretation: it finds the minimum number of “colours” (clusters) needed so that no two geysers sharing an edge (i.e., not significantly different) receive the same colour^[9]. This graph-theoretic framing is equivalent to finding connected components in the complement graph, but the colouring formulation generalises naturally to weighted graphs where edge weights reflect p-value magnitudes.

The Sobolev regularity check, while technically necessary for valid inference, was satisfied by all 12 geysers without difficulty. This likely reflects the physical smoothness of geothermal processes operating over month-to-year timescales, where abrupt discontinuities are rare except during catastrophic events like the 2007 mudslide.

Conclusion

This research contributed a curvature-corrected algorithm for the studentized range distribution that reduced tail probability errors by a factor of four and applied the improved test to identify eruption interval clusters among 12 Kamchatka geysers. The correction produced a more parsimonious and geophysically interpretable clustering (4 vs 6 clusters) while maintaining exact family-wise error control.

Brownian motion modelling of interval drift affected 7 of 12 geysers and was necessary for accurate mean comparison. The vertex colouring algorithm provided an efficient and interpretable method for translating pairwise test results into a global clustering structure.

For the computational statistics community, the curvature correction is simple to implement and could improve any application that relies on studentized range critical values. For geophysicists, the four-cluster structure raises questions about subsurface connectivity that future work could address with geochemical or seismic data. The Brownian drift findings suggest that long-term geyser monitoring programmes should routinely check for interval drift before applying standard comparison tests.

Acknowledgements

Funding Sources

This research received support from the Philippine Council for Industry, Energy and Emerging Technology Research through its computational sciences programme.

Institutional Support

The Baguio Center for Computational Sciences provided high-performance computing access and statistical software licences.

Contributions Not Qualifying for Authorship

Dr. Sergei Kamchatkin provided access to KVERT eruption logs and verified event timestamps.

References

1. Fisher RA. Statistical methods for research workers. Edinburgh: Oliver and Boyd; 1925:1-239.
2. Miller RG. Simultaneous statistical inference. New York: Springer; 1981:1-299.
3. Tukey JW. The problem of multiple comparisons. Princeton: Princeton University; 1953:1-396.
4. Lund RE, Lund JR. Algorithm AS 190: probabilities and upper quantiles for the studentized range. J R Stat Soc Ser C. 1983;32(2):204-210.
5. Belousov A, Belousova M, Nechayev A. The Valley of Geysers in Kamchatka before and after the catastrophe of

2007. In: Proceedings of the World Geothermal Congress. 2010:1-10.

6. Hurwitz S, Manga M. The fascinating and complex dynamics of geyser eruptions. Annu Rev Earth Planet Sci. 2017;45:31-59.
7. Ingebritsen SE, Rojstaczer SA. Geyser periodicity and the response of geysers to deformation. J Geophys Res. 1996;101(B10):21891-21905.
8. Karlin S, Taylor HM. A first course in stochastic processes. 2nd ed. New York: Academic Press; 1975:1-557.
9. West DB. Introduction to graph theory. 2nd ed. Upper Saddle River: Prentice Hall; 2001:1-512.
10. Adams RA, Fournier JJF. Sobolev spaces. 2nd ed. Amsterdam: Elsevier; 2003:1-305.
11. Hochberg Y, Tamhane AC. Multiple comparison procedures. New York: Wiley; 1987:1-450.
12. Copenhaver MD, Holland BS. Computation of the distribution of the maximum studentized range statistic. J Stat Comput Simul. 1988;30(1):1-15.
13. Manga M, Brodsky E. Seismic triggering of eruptions in the far field. Annu Rev Earth Planet Sci. 2006;34:263-291.
14. Steinberg GS, Merzhanov AG, Steinberg AS. Geyser process: its theory, modeling, and field experiment. Mod Geol. 1981;8:67-86.
15. Kemp CD. On a contagious distribution suggested for accident data. Biometrics. 1967;23(2):241-260.
16. Dunnett CW. A multiple comparison procedure for comparing several treatments with a control. J Am Stat Assoc. 1955;50(272):1096-1121.
17. Scheffé H. A method for judging all contrasts in the analysis of variance. Biometrika. 1953;40(1):87-104.
18. Rinehart JS. Geysers and geothermal energy. Berlin: Springer; 1980:1-223.
19. Nichols TE, Holmes AP. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp. 2002;15(1):1-25.
20. Brillinger DR. Time series: data analysis and theory. Philadelphia: SIAM; 2001:1-540.
21. Abramowitz M, Stegun IA. Handbook of mathematical functions with formulas, graphs, and mathematical tables. New York: Dover Publications; 1964:1-1046.