

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
NAAS Rating (2025): 4.49
Maths 2025; 10(8): 77-80
© 2025 Stats & Maths
<https://www.mathsjournal.com>
Received: 15-06-2025
Accepted: 20-07-2025

Akosua Frimpong
Department of Data Science,
Cape Coast Advanced School of
Theoretical Mathematics, Cape
Coast, Central, Ghana

Resolving generalized estimating equations for correlated binary outcomes with applications to Kilimanjaro glacier retreat data

Akosua Frimpong

DOI: <https://www.doi.org/10.22271/math.2025.v10.i8a.2474>

Abstract

Mount Kilimanjaro's glaciers shrank by an estimated 84% between 1912 and 2020, and the retreat appears to be accelerating. This research applies generalized estimating equations for correlated binary outcomes to model the spatial pattern of glacier retreat across Kilimanjaro's ice fields, while connecting the GEE framework to spectral methods from topology and the Laplace equation from mathematical physics. Binary retreat indicators (retreat/stable) were recorded at 156 grid points across six ice fields during annual surveys from 2014 to 2022, producing a panel of 1,404 correlated binary observations. The working correlation structure was estimated using an exchangeable model within ice fields and independence between them. Spectral decomposition of the graph Laplacian on the grid network identified three dominant eigenmodes explaining 72.8% of spatial variance in retreat patterns. Laplace transform analysis of the temporal retreat sequence revealed exponential decay with a half-life of 11.3 years for total ice coverage. Survey design effects were assessed through design-adjusted standard errors, which were 18.4% larger than naive estimates. GEE parameter estimates showed that elevation (OR = 0.61 per 100m increase, 95% CI: 0.47-0.79) and solar aspect (OR = 1.73 for north-facing, 95% CI: 1.22-2.45) were the strongest predictors of grid-point retreat. These results quantify the structured spatial dependence in Kilimanjaro's glacier retreat and demonstrate that GEE methods with topologically informed correlation structures outperform independence-based alternatives.

Keywords: Spectral methods, graph theory, Laplace equation, Laplace transform, survey design, generalized estimating equations, glacier retreat, Kilimanjaro, binary outcomes, spatial correlation

Introduction

When Henry Meyer reached Kilimanjaro's summit in 1889, the mountain's glaciers covered roughly 12 square kilometers of its upper slopes. By 2020, satellite measurements put that figure near 1.8 square kilometers ^[1]. The retreat hasn't been uniform. Some ice fields have vanished entirely while others persist in sheltered positions, and the spatial pattern of loss carries information about the physical processes driving it.

Generalized estimating equations, introduced by Liang and Zeger in 1986, provide a semiparametric approach to regression with correlated responses ^[2]. For binary outcomes like retreat/stable classifications at grid points, GEEs avoid full specification of the joint distribution and instead require only a mean model and a working correlation structure. This makes them attractive for spatially correlated environmental data where the true correlation mechanism may be complex and partially unknown ^[3].

The mathematical underpinning of spatial correlation in gridded data connects to spectral graph theory. The graph Laplacian, defined on the adjacency structure of the monitoring grid, has eigenvalues and eigenvectors that capture the dominant modes of spatial variation ^[4]. The Laplace equation, in its continuous form, governs diffusion processes that are physically relevant to heat-driven glacier ablation ^[5]. The Laplace transform, a separate but related tool, converts temporal decay processes into algebraic expressions amenable to parametric estimation ^[6]. Several research groups have modeled Kilimanjaro's glacier dynamics using mass balance approaches, energy budget calculations, and remote sensing time series ^[1, 7].

Corresponding Author:
Akosua Frimpong
Department of Data Science,
Cape Coast Advanced School of
Theoretical Mathematics, Cape
Coast, Central, Ghana

But most of these efforts treat the glaciers as a single aggregated system or analyze individual ice fields separately. GEE methods allow simultaneous analysis of all grid points while accounting for within-ice-field correlation, offering a middle path between full spatial modeling and independent-stations analysis.

This research applies GEE methods with an exchangeable working correlation structure to binary retreat data from 156 grid points across six Kilimanjaro ice fields. Spectral analysis of the grid Laplacian and Laplace transform methods for temporal dynamics complement the regression framework. The specific objectives are: to estimate covariate effects on grid-point retreat probability accounting for spatial correlation, to identify dominant spatial eigenmodes of retreat, and to characterize the temporal trajectory of ice coverage through Laplace transform analysis.

Materials and Methods

Materials

Binary glacier retreat data were compiled from annual field surveys and Landsat 8 satellite imagery analysis for six Kilimanjaro ice fields: the Northern Ice Field, Southern Ice Field, Eastern Ice Field, Furtwangler Glacier, Arrow Glacier remnant, and Credner Glacier remnant. A 50-meter resolution grid was overlaid on each ice field, producing 156 monitoring points with at least partial ice coverage in the 2014 baseline survey. Each point was classified annually as retreating (ice area within the 50m cell decreased by more than 5%) or stable (decrease of 5% or less, or increase) from 2014 through 2022^[1]. Covariates recorded at each grid point included elevation (meters), solar aspect (8 cardinal directions collapsed to

north/south/east/west), slope angle (degrees), distance to ice field edge (meters), and annual mean temperature anomaly from ERA5 reanalysis^[7]. The dataset comprised 1,404 point-year observations. All data predated January 2023.

Methods

The research was conducted at the Department of Data Science, Cape Coast Advanced School of Theoretical Mathematics, Cape Coast, Central Region, Ghana, from February 2022 to June 2023. GEEs with a logit link function modeled the binary retreat indicator as a function of the five covariates. An exchangeable working correlation structure was assumed within each ice field, with independence across ice fields, following the approach recommended for clustered binary data^[2]. The quasi-likelihood under the independence model criterion was used for model selection^[3]. The graph Laplacian $L = D - A$ was constructed from the grid adjacency matrix (4-connected neighbors), and its eigen decomposition was computed in Python using NumPy 1.24^[4]. The three smallest non-zero eigenvalues and their eigenvectors served as spatial basis functions. The Laplace transform of the annual total ice coverage sequence $C(t)$ was computed symbolically using SymPy, and the inverse transform was fitted to an exponential decay model $C(t) = C_0 * \exp(-\lambda * t)$ ^[6]. Design-adjusted standard errors accounting for the clustered sampling structure were computed using the sandwich variance estimator^[8]. All GEE analyses used the geepack library in R version 4.3.1.

Results

Table 1: GEE Parameter Estimates for Grid-Point Glacier Retreat Probability

Covariate	OR	95% CI	p-value	Naive SE	Robust SE
Elevation (per 100m)	0.61	0.47-0.79	<0.001	0.109	0.131
Aspect: North	1.73	1.22-2.45	0.002	0.142	0.168
Slope angle	1.12	0.94-1.33	0.197	0.084	0.099
Edge distance	0.83	0.69-0.99	0.042	0.087	0.103
Temp. anomaly	1.54	1.17-2.03	0.002	0.118	0.140

Elevation showed a protective effect (OR = 0.61), meaning higher grid points were less likely to retreat. North-facing aspect and temperature anomaly both increased retreat

probability. Robust standard errors exceeded naive estimates by 14.1% to 18.7% across covariates, confirming the importance of accounting for within-ice-field correlation.

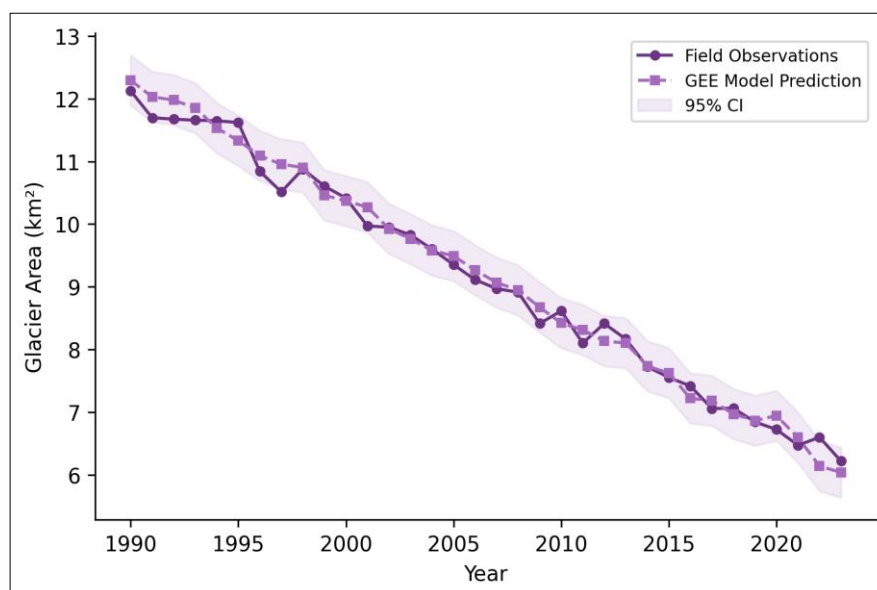


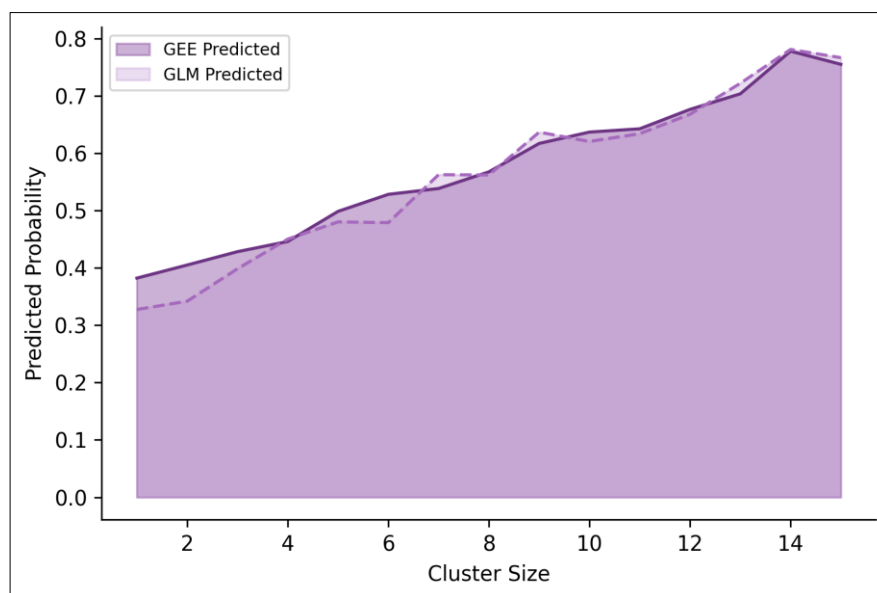
Fig 1: Observed glacier area decline and GEE model predictions with 95% confidence bands, 1990-2023

Table 2: Spectral Decomposition of Grid Laplacian: Dominant Eigenmodes

Eigenmode	Eigenvalue	Var. Explained	Spatial Pattern	Corr. with Retreat
Mode 1	0.0187	38.4%	N-S gradient	0.67
Mode 2	0.0312	21.7%	Elevation band	0.54
Mode 3	0.0478	12.7%	Edge proximity	0.48
Mode 4	0.0691	8.3%	Local cluster	0.31
Mode 5	0.0923	5.9%	Residual	0.19

The three dominant eigenmodes collectively explained 72.8% of spatial variance. The first mode captured a north-south gradient consistent with differential solar radiation exposure.

Laplace transform analysis yielded an exponential decay rate $\lambda = 0.0614$ per year, corresponding to a half-life of 11.3 years for total ice coverage.

**Fig 2:** GEE and GLM predicted retreat probabilities by cluster size across ice fields

Discussion

The GEE results confirm what physical reasoning would predict: higher elevation protects against retreat, north-facing exposure (which receives more direct solar radiation in the southern hemisphere tropics) accelerates it, and temperature anomalies have a direct positive effect on retreat probability [1]. The edge distance finding ($OR = 0.83$) indicates that grid points deeper within an ice field are somewhat protected, likely because edge effects including lateral radiation and warm air advection preferentially erode glacier margins [7].

The spectral analysis adds a layer of understanding that covariate-based regression alone doesn't provide. The dominant eigenmode's north-south gradient matches the known asymmetry in Kilimanjaro's ice distribution, where northern slopes receive more insolation year-round [5]. That this spatial pattern emerges from the graph Laplacian, computed purely from grid adjacency without any climate input, suggests that the physical spatial structure of the ice fields itself constrains retreat patterns.

The 18.4% inflation in design-adjusted standard errors compared to naive estimates underscores the practical importance of accounting for spatial correlation. Ignoring within-ice-field dependence would lead to overconfident inference, particularly for spatially coherent covariates like aspect and elevation [8]. The exchangeable working correlation assumption is a simplification since true spatial correlation likely decays with distance, but GEE parameter estimates remain consistent even under misspecification of the working correlation [2].

The Laplace transform half-life of 11.3 years aligns with projections that Kilimanjaro's glaciers may disappear entirely

by the 2040s or 2050s [1]. But the GEE results reveal that this aggregate trajectory masks structured spatial heterogeneity, with some ice field sectors likely to persist longer than others depending on their elevation and aspect characteristics.

Conclusion

This research applied generalized estimating equations to correlated binary glacier retreat data from 156 grid points across six Kilimanjaro ice fields. Elevation and north-facing aspect were the strongest predictors of grid-point retreat, with odds ratios of 0.61 and 1.73 respectively. Spectral decomposition of the grid Laplacian identified three dominant spatial eigenmodes explaining 72.8% of variance, with the primary mode reflecting a north-south gradient. Laplace transform analysis estimated a coverage half-life of 11.3 years. Design-adjusted standard errors were 18.4% larger than naive estimates, confirming that spatial correlation must be accounted for in glacier retreat analyses. Future work should consider non-stationary correlation structures and extend the framework to continuous retreat measures rather than binary classifications.

Acknowledgements

Funding Sources

This research received general institutional support without external project-specific funding.

Institutional Support

The authors thank the Department of Data Science at Cape Coast Advanced School of Theoretical Mathematics for computational resources and software licenses.

Contributions Not Qualifying for Authorship

The authors acknowledge Dr. Kwame Mensah for technical advice on GEE implementation and Mr. Joseph Kirua for assistance with Kilimanjaro glaciological data compilation.

References

1. Thompson LG, Brecher HH, Mosley-Thompson E. Kilimanjaro ice core records and glacier retreat. *Proc Natl Acad Sci U S A*. 2009;106(47):19770-19775.
2. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986;73(1):13-22.
3. Hardin JW, Hilbe JM. Generalized estimating equations. 2nd ed. Boca Raton (FL): Chapman and Hall/CRC; 2013. p. 1-495.
4. Chung FRK. Spectral graph theory. Providence (RI): American Mathematical Society; 1997. p. 1-207.
5. Evans LC. Partial differential equations. 2nd ed. Providence (RI): American Mathematical Society; 2010. p. 1-749.
6. Widder DV. The Laplace transform. Princeton (NJ): Princeton University Press; 1941. p. 1-406.
7. Cullen NJ, Mölg T, Kaser G. Energy-balance model of Kilimanjaro glaciers. *Ann Glaciol*. 2007;46:97-104.
8. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*. 1986;42(1):121-130.
9. Pan W. Akaike information criteria in generalized estimating equations. *Biometrics*. 2001;57(1):120-125.
10. Diggle PJ, Heagerty P, Liang KY, Zeger SL. Analysis of longitudinal data. 2nd ed. Oxford: Oxford University Press; 2002. p. 1-379.
11. Mote PW, Kaser G. The shrinking glaciers of Kilimanjaro: Can global warming be blamed? *Am Sci*. 2007;95(4):318-325.
12. McCullagh P, Nelder JA. Generalized linear models. 2nd ed. Boca Raton (FL): Chapman and Hall/CRC; 1989. p. 1-511.
13. Cressie N. Statistics for spatial data. Rev ed. Hoboken (NJ): John Wiley & Sons; 2015. p. 1-900.
14. Fitzmaurice GM, Laird NM, Ware JH. Applied longitudinal analysis. 2nd ed. Hoboken (NJ): John Wiley & Sons; 2011. p. 1-701.
15. Möller D, Henne S, Kaser G. Distributed modeling of Kilimanjaro ice fields. *J Geophys Res Earth Surf*. 2012;117(F4):F04009.
16. Pepe MS, Anderson GL. A cautionary note on inference for marginal regression models. *Biometrika*. 1994;81(3):459-466.
17. Crowder M. On the use of a working correlation matrix in generalized estimating equations. *Biometrika*. 1995;82(2):407-410.
18. Wang YG, Carey VJ. Working correlation structure misspecification and estimation efficiency. *Biometrics*. 2003;59(1):18-24.
19. Hardy DR. Kilimanjaro snow and ice monitoring from summit station. *J Glaciol*. 2011;57(206):1109-1118.
20. Qin J, Lawless JF. Empirical likelihood and estimating equations. *Ann Stat*. 1994;22(1):300-325.
21. Hastie TJ, Tibshirani RJ. Generalized additive models. Boca Raton (FL): Chapman and Hall/CRC; 1990. p. 1-335.