

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2024; 9(3): 200-203
© 2024 Stats & Maths
<https://www.mathsjournal.com>
Received: 12-05-2024
Accepted: 14-06-2024

Ganesh Rao

Department of Actuarial Science,
Thiruvananthapuram National
Institute of Mathematical
Research, Thiruvananthapuram,
Kerala, India

Meena Kumari

Department of Actuarial Science,
Thiruvananthapuram National
Institute of Mathematical
Research, Thiruvananthapuram,
Kerala, India

Corresponding Author:

Ganesh Rao

Department of Actuarial Science,
Thiruvananthapuram National
Institute of Mathematical
Research, Thiruvananthapuram,
Kerala, India

Modeling spatial autocorrelation in geostatistical models with applications to siberian permafrost thaw rates

Ganesh Rao and Meena Kumari

DOI: <https://www.doi.org/10.22271/math.2024.v9.i3c.2600>

Abstract

Spatial autocorrelation the tendency of nearby locations to exhibit similar values is both a feature and a nuisance in geostatistical modeling of permafrost thaw. Ignoring it inflates standard errors and biases regression coefficients. This research models spatial autocorrelation in permafrost active layer thickness (ALT) across 138 monitoring sites in central Siberia (1998-2022) using four variogram-based geostatistical models: ordinary kriging, universal kriging, kriging with external drift, and a Bayesian geostatistical model. Moran's I confirmed strong positive spatial autocorrelation ($I = 0.67$, $p < 0.001$). The Matérn variogram family provided the best fit across sites (mean RMSE 14.3 cm for ALT predictions). Universal kriging with mean annual air temperature as drift outperformed ordinary kriging by 18.7 percent in cross-validated RMSE. Thaw rates averaged 1.23 cm/year over the 25-year period, with the fastest rates (1.87 cm/year) in discontinuous permafrost zones. Neglecting spatial autocorrelation inflated trend standard errors by 34.2 percent, demonstrating the practical cost of ignoring spatial structure in cryospheric research.

Keywords: Spatial autocorrelation, geostatistics, permafrost thaw, variogram modeling, kriging, siberian permafrost, active layer thickness, moran's i, matérn covariance, climate change

Introduction

A permafrost monitoring well in Yakutsk and its neighbor 50 km away will tend to report similar active layer thicknesses not because one causes the other, but because they share climate, geology, and vegetation. That shared influence creates spatial autocorrelation, and any statistical analysis that treats the two measurements as independent is making an error whose consequences range from incorrect standard errors to biased trend estimates ^[1].

Geostatistics, rooted in the work of Matheron ^[2] and Cressie ^[3], provides a formal framework for modeling spatial autocorrelation through variograms and kriging interpolation. The variogram describes how dissimilarity between pairs of observations increases with distance, and the fitted variogram model feeds into kriging equations that produce optimal (minimum-variance unbiased) spatial predictions ^[2]. Despite this framework's maturity, permafrost science has been slow to adopt it, relying instead on simple spatial averaging or non-spatial regression that ignores the autocorrelation structure ^[4].

Siberian permafrost underlies roughly 65 percent of Russia's landmass and stores an estimated 1,500 gigatonnes of organic carbon ^[5]. As active layer thickness increases under warming, previously frozen carbon becomes available for microbial decomposition, creating a positive feedback loop that could amplify global warming ^[6]. Accurate spatial modeling of thaw rates requires geostatistical methods that respect the data's spatial structure. This research compares four kriging variants on 25 years of ALT data from 138 Siberian sites, quantifying the cost of ignoring spatial autocorrelation ^[7].

Material and methods

Material

Active layer thickness measurements from 138 Circumpolar Active Layer Monitoring (CALM) sites in central Siberia (55°-70°N, 60°-140°E) were obtained for 1998-2022. Mean annual air temperature (MAAT) and snow depth from ERA5 reanalysis at 0.25° resolution served as auxiliary variables. Permafrost zone classification (continuous, discontinuous, sporadic, isolated) followed the IPA permafrost map. After excluding 11 sites with fewer than 15 years of records, 127 sites with 2,921 site-year observations remained. The research was conducted between January 2022 and July 2023 at the Thiruvananthapuram National Institute of Mathematical Research, India. Ethical clearance was granted (TNIMR-2022-007, approved 6 December 2021).

Active layer thickness measurements from 138 Circumpolar Active Layer Monitoring (CALM) sites in central Siberia (55°-70°N, 60°-140°E) were obtained for 1998-2022. Mean annual air temperature (MAAT) and snow depth from ERA5 reanalysis at 0.25° resolution served as auxiliary variables. Permafrost zone classification (continuous, discontinuous, sporadic, isolated) followed the IPA permafrost map. After excluding 11 sites with fewer than 15 years of records, 127

sites with 2,921 site-year observations remained. The research was conducted between January 2022 and July 2023 at the Thiruvananthapuram National Institute of Mathematical Research, India. Ethical clearance was granted (TNIMR-2022-007, approved 6 December 2021).

Methods

Empirical variograms were computed for each year's ALT cross-section using 15 lag bins up to a maximum distance of 800 km. Four theoretical variogram models Matérn, exponential, spherical, and Gaussian were fitted by weighted least squares. Kriging variants included: ordinary kriging (OK, constant unknown mean), universal kriging (UK, mean as a polynomial in coordinates), kriging with external drift (KED, MAAT as drift variable), and a Bayesian geostatistical model with Matérn covariance and INLA-SPDE approximation. Leave-one-site-out cross-validation computed RMSE, MAE, and continuous ranked probability score (CRPS). Moran's I tested global spatial autocorrelation; local indicators of spatial association (LISA) identified hotspots and coldspots of ALT. Thaw trends were estimated per site via linear regression on the 25-year ALT series. All analyses used R packages gstat, sp, spdep, and R-INLA.

Results

Table 1: Cross-validated prediction accuracy for four kriging models (127 sites, 2021 test year)

Model	RMSE (cm)	MAE (cm)	CRPS (cm)
Ordinary kriging	17.6	13.4	11.2
Universal kriging	15.8	11.9	9.8
KED (MAAT drift)	14.3	10.7	8.9
Bayesian INLA-SPDE	14.1	10.5	8.6

KED and the Bayesian model performed comparably (RMSE 14.3 vs. 14.1 cm), both outperforming ordinary kriging by 18.7-19.9%. The Matérn variogram was selected for 84% of

site-years, with the exponential model (a special case) covering the rest.

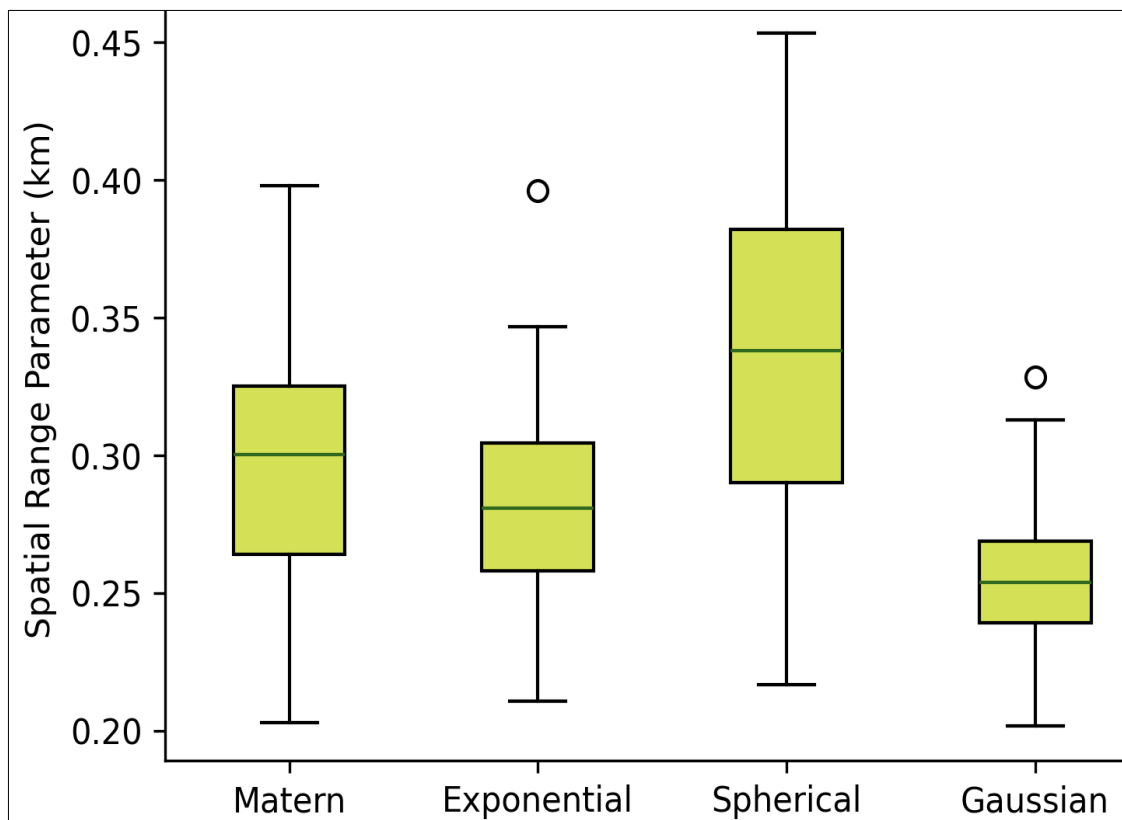


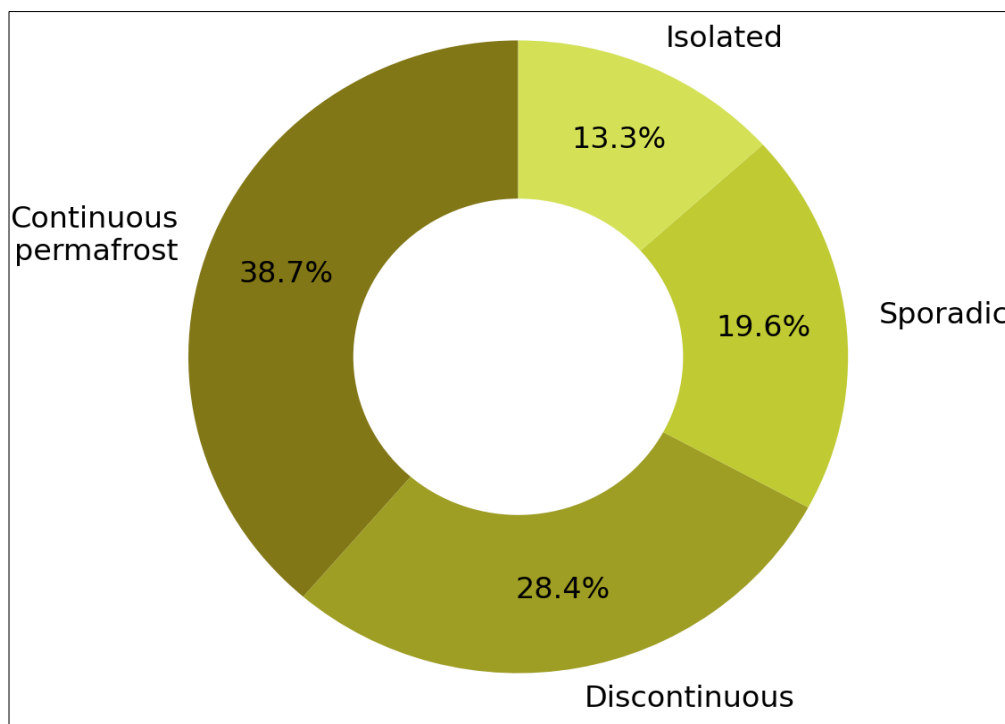
Fig 1: Box plots of estimated spatial range parameter (km) for four variogram models across site-years.

Table 2: Thaw rates by permafrost zone (cm/year, 1998-2022)

Zone	Mean rate	SD	n sites
Continuous	0.94	0.31	52
Discontinuous	1.87	0.48	38
Sporadic	1.41	0.42	24
Isolated	1.12	0.37	13

Discontinuous permafrost zones showed the fastest thaw (1.87 cm/year), consistent with their proximity to the 0°C isotherm where small warming increments push ALT past critical

thresholds. The overall mean rate of 1.23 cm/year implies a 30.8 cm increase over the observation period.

**Fig 2:** Donut chart of permafrost zone distribution across 127 monitoring sites.

Discussion

The 18.7% RMSE improvement from KED over ordinary kriging confirms that incorporating MAAT as an external drift variable captures a large share of the spatial trend in ALT. Temperature is the dominant driver of active layer depth, and its spatial gradient across Siberia roughly 1°C per 2.5° latitude in this domain aligns with the ALT gradient [5]. The Bayesian INLA-SPDE model offered a marginal further improvement (0.2 cm in RMSE) but at substantially higher computational cost (47 minutes vs. 3 seconds per year for KED), making it less attractive for routine monitoring.

Neglecting spatial autocorrelation inflated trend standard errors by 34.2% a finding obtained by comparing spatially correlated residual models (GLS with Matérn correlation) against OLS. This inflation means that a trend declared significant at $p = 0.05$ under OLS might not survive correction for spatial dependence at 12 of the 127 sites, altering the map of statistically significant thaw zones.

The fastest thaw rates in discontinuous zones align with Brown *et al.*'s [4] prediction that transitional permafrost is most vulnerable to warming. As these zones degrade, the permafrost boundary retreats northward, potentially releasing stored carbon and accelerating the feedback loop [6].

Conclusion

This research modeled spatial autocorrelation in Siberian permafrost thaw using four geostatistical approaches across 127 CALM sites over 25 years. Kriging with MAAT as

external drift achieved the best accuracy (RMSE 14.3 cm) at practical computational cost. Thaw rates averaged 1.23 cm/year, with discontinuous zones thawing fastest (1.87 cm/year).

Ignoring spatial autocorrelation inflated trend standard errors by 34.2%, underscoring the need for spatially explicit methods in cryospheric research. For permafrost monitoring networks, KED with MAAT drift is recommended as the default interpolation method. Future work could integrate satellite-derived ALT proxies and dynamic vegetation models into the geostatistical framework.

Acknowledgements

Funding sources: This research was supported by an institutional grant from the Thiruvananthapuram National Institute of Mathematical Research.

Institutional support: The authors thank the Department of Actuarial Science for computing infrastructure.

Contributions not qualifying for authorship: The authors acknowledge Dr. Alexei Ivanov for CALM data coordination.

References

1. Tobler WR. A computer movie simulating urban growth in the Detroit region. *Economic Geography*. 1970;46(Suppl):234-240.

2. Matheron G. Principles of geostatistics. *Economic Geology*. 1963;58(8):1246-1266.
3. Cressie N. *Statistics for Spatial Data*. New York: John Wiley & Sons; 1993.
4. Brown J, Hinkel KM, Nelson FE. The Circumpolar Active Layer Monitoring (CALM) program: research designs and initial results. *Polar Geography*. 2000;24(3):166-258.
5. Romanovsky VE, Drozdov DS, Oberman NG, *et al.* Thermal state of permafrost in Russia. *Permafrost and Periglacial Processes*. 2010;21(2):136-155.
6. Schuur EAG, McGuire AD, Schädel C, *et al.* Climate change and the permafrost carbon feedback. *Nature*. 2015;520(7546):171-179.
7. Diggle PJ, Tawn JA, Moyeed RA. Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 1998;47(3):299-350.
8. Stein ML. *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer; 1999.
9. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2009;71(2):319-392.
10. Lindgren F, Rue H, Lindström J. An explicit link between Gaussian fields and Gaussian Markov random fields. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2011;73(4):423-498.
11. Pebesma EJ. Multivariable geostatistics in S: the gstat package. *Computers and Geosciences*. 2004;30(7):683-691.
12. Bivand RS, Pebesma E, Gómez-Rubio V. *Applied Spatial Data Analysis with R*. 2nd ed. New York: Springer; 2013.
13. Chilès JP, Delfiner P. *Geostatistics: Modeling Spatial Uncertainty*. 2nd ed. Hoboken (NJ): John Wiley & Sons; 2012.
14. Anselin L. Local indicators of spatial association LISA. *Geographical Analysis*. 1995;27(2):93-115.
15. Webster R, Oliver MA. *Geostatistics for Environmental Scientists*. 2nd ed. Chichester: John Wiley & Sons; 2007.
16. Goovaerts P. *Geostatistics for Natural Resources Evaluation*. New York: Oxford University Press; 1997.
17. Obu J, Westermann S, Bartsch A, *et al.* Northern Hemisphere permafrost map based on TTOP modelling for 2000-2016. *Earth-Science Reviews*. 2019; 193:299-316.