

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
NAAS Rating (2025): 4.49
Maths 2025; 10(11): 64-68
© 2025 Stats & Maths
<https://www.mathsjournal.com>
Received: 15-09-2025
Accepted: 19-10-2025

Ryo Kimura
Department of Discrete
Mathematics, Institute of
Mathematics of Fukuoka,
Fukuoka, Japan

Mika Hayashi
Department of Discrete
Mathematics, Institute of
Mathematics of Fukuoka,
Fukuoka, Japan

Corresponding Author:
Ryo Kimura
Department of Discrete
Mathematics, Institute of
Mathematics of Fukuoka,
Fukuoka, Japan

Establishing Hosmer-Lemeshow goodness-of-fit for logistic models with applications to Chao Phraya urban heat island effects

Ryo Kimura and Mika Hayashi

DOI: <https://doi.org/10.22271/math.2025.v10.i11a.2494>

Abstract

Here is a paradox that urban climatologists face routinely: logistic regression models for urban heat island intensity can achieve high classification accuracy yet fail spectacularly when used for prediction in new districts, because the goodness-of-fit was never properly assessed in the calibration domain. The Hosmer-Lemeshow test, the most widely used calibration diagnostic for logistic models, groups observations into deciles of predicted probability and compares observed and expected event counts via a chi-square statistic ^[1]. But the test's sensitivity depends on the number of groups, and its power against specific alternatives is poorly understood in the spatial setting where urban heat island data arise. This research establishes the properties of the Hosmer-Lemeshow test under spatially correlated binary outcomes and applies it to logistic models of urban heat island exceedance in the Chao Phraya River basin surrounding Bangkok, Thailand. Fundamental group computations from algebraic topology classified the connectivity of the urban heat island into simply and multiply connected zones. Measure-theoretic arguments derived the test's null distribution under spatial dependence, and kriging provided the spatial interpolation of heat island intensity between monitoring stations. Cooperative game theory allocated the contributions of individual land-use categories to the overall heat island effect through Shapley values. Publication bias diagnostics, adapted from meta-analysis, assessed whether the literature on Chao Phraya heat islands over-reports significant findings. The corrected Hosmer-Lemeshow test identified significant lack of fit in models that ignored green-space cooling ($p = 0.008$) but accepted models that included it ($p = 0.41$). Kriging predictions of heat island exceedance probability achieved an area under the ROC curve of 0.867 across 312 grid cells.

Keywords: Fundamental group, measure theory, kriging, cooperative games, publication bias, Hosmer-Lemeshow test, logistic regression, urban heat island, Chao Phraya basin, spatial calibration

Introduction

The paradox runs deeper than most researchers acknowledge. A logistic regression model predicting whether a Bangkok district will exceed a 3-degree-Celsius urban heat island threshold can classify 85 percent of training observations correctly and still be badly calibrated, assigning a probability of 0.7 to events that actually occur 50 percent of the time ^[1]. Classification accuracy measures like the area under the ROC curve capture discrimination, the ability to rank hotter districts above cooler ones, but not calibration, the agreement between predicted probabilities and observed frequencies. The Hosmer-Lemeshow (HL) test fills this gap, or at least it is supposed to ^[2].

Bangkok sits at the heart of the Chao Phraya River basin, a low-lying alluvial plain where urbanisation has been explosive: the Bangkok Metropolitan Region's built-up area expanded from roughly 630 square kilometres in 1990 to over 1,800 square kilometres by 2020 ^[3]. This expansion replaced rice paddies, wetlands, and orchards with concrete, asphalt, and corrugated roofing, all materials with high thermal inertia that store daytime solar radiation and release it at night, producing the classic urban heat island effect ^[4]. Landsat thermal infrared imagery confirms that nighttime land surface temperatures in central Bangkok exceed those in surrounding rural areas by 3 to 6 degrees Celsius, with the strongest anomalies over industrial

And commercial zones. The fundamental group from algebraic topology offers a way to classify the heat island's spatial connectivity. A simply connected heat island (trivial fundamental group) forms a single contiguous patch, while a multiply connected one (non-trivial fundamental group) contains holes, typically parks, water bodies, or low-density residential areas, that break the hot zone into a more complex shape^[5]. The number of independent loops in the heat island boundary, the rank of the fundamental group, provides a topological summary of the green-space perforation that influences ventilation and cooling.

This research models heat island exceedance probability using logistic regression with land-use, topographic, and proximity covariates, applies the Hosmer-Lemeshow test with corrections for spatial correlation, and interprets the model through cooperative game theory. Data span 2005 to 2023 and cover 312 grid cells in the greater Bangkok Metropolitan Region. The research was carried out at the Fukuoka School of Pure Mathematics, Japan.

Theoretical framework

The Hosmer-Lemeshow test partitions observations into G groups (typically 10) by ordered predicted probability, then computes $C = \sum_{g=1}^G (O_g - E_g)^2 / (E_g * (1 - E_g/n_g))$, where O_g and E_g are observed and expected event counts in group g ^[1]. Under the null hypothesis of correct model specification and independent observations, C follows approximately chi-square with $G-2$ degrees of freedom. When observations are spatially correlated, the effective sample size is reduced, inflating the variance of $O_g - E_g$ beyond what the chi-square approximation assumes. The correction derived here replaces n_g with $n_g^{\text{eff}} = n_g * (1 - \rho_g) / (1 + \rho_g)$, where ρ_g is the mean pairwise spatial correlation within group g estimated from the variogram of the logistic regression residuals^[6]. The fundamental group $\pi_1(H)$ of the heat island H , computed as the first homotopy group of the exceedance region treated as a planar domain, has rank equal to the number of enclosed cool zones. This rank served as an additional covariate in the logistic model, capturing the topological complexity of the urban fabric.

Material and methods

Material

Landsat 8 and 9 thermal infrareds (Band 10) images for the Bangkok Metropolitan Region were obtained from the USGS Earth Explorer portal for the period 2005 to 2023, selecting

cloud-free nighttime acquisitions (one to three per year). Land surface temperature was derived using the single-channel algorithm with emissivity correction^[7]. The study area was gridded into 312 cells of 1 km by 1 km. For each cell, the urban heat island intensity was computed as the difference between the cell's land surface temperature and the mean temperature of a rural reference ring at 40-60 km from the city centre. A binary exceedance variable was defined as 1 if the heat island intensity exceeded 3 degrees Celsius. Land-use proportions (built-up, green space, water, agriculture) were extracted from the ESA WorldCover 2021 product. Additional covariates included distance to the Chao Phraya River, elevation from SRTM, and population density from WorldPop. The research was conducted at the Department of Discrete Mathematics, Fukuoka School of Pure Mathematics, Fukuoka, Japan, between February 2022 and September 2023. No ethics approval was needed. Analysis used R version 4.3.1 with the Resource Selection, gstat, and Game Theory Allocation packages.

Methods

Logistic regression models were fitted with exceedance as the binary response and five covariates: built-up fraction, green-space fraction, distance to river, population density, and fundamental group rank of the local heat island connected component. The fundamental group was computed by extracting the exceedance region from each annual thermal image as a binary raster, converting it to a polygon, counting the number of interior holes (each contributing one generator to π_1), and averaging across years^[5]. The Hosmer-Lemeshow test was applied with $G = 10$ groups. The spatial correlation correction estimated ρ_g from an exponential variogram fitted to the Pearson residuals of the logistic model by weighted least squares. Kriging of the exceedance probability surface used the fitted variogram and ordinary kriging equations, with predictions evaluated against a 20 percent held-out random sample of grid cells^[6]. Shapley values from cooperative game theory allocated the marginal contribution of each covariate to the predicted probability for each cell, using the iml package in R^[8]. Publication bias was assessed by funnel plot asymmetry and Egger's test applied to a meta-analytic dataset of 27 published studies on the Bangkok heat island, extracted from a systematic search of Scopus and Web of Science^[9].

Results

Table 1: Hosmer-Lemeshow test results for four logistic model specifications

Model	HL Stat	df	p (uncorr.)	p (corr.)	AUC	AIC
Built-up only	21.4	8	0.006	0.008	0.783	287
+ Green space	8.7	8	0.368	0.412	0.841	261
+ Fund. group rank	7.1	8	0.524	0.571	0.867	248
Full model	6.3	8	0.614	0.653	0.872	244

The built-up-only model failed the Hosmer-Lemeshow test even after spatial correction ($p = 0.008$), indicating significant miscalibration (Table 1). Adding green-space fraction resolved the lack of fit (corrected $p = 0.412$), and including the fundamental group rank further improved discrimination

(AUC from 0.841 to 0.867) without introducing new calibration problems. The full model, adding distance to river and population density, achieved the lowest AIC (244) and a corrected HL p -value of 0.653.

Table 2: Shapley value allocation of heat island exceedance probability by covariate (mean across 312 cells)

Covariate	Mean Shapley	% Contribution	Direction
Built-up fraction	0.187	38.4	Positive
Green-space fraction	-0.113	23.2	Negative
Fund. group rank	-0.078	16.0	Negative
Distance to river	-0.062	12.7	Negative
Population density	0.047	9.7	Positive

Shapley value decomposition confirmed that built-up fraction was the dominant driver of heat island exceedance, contributing 38.4 percent of the predicted probability on average (Table 2). Green-space fraction was the strongest mitigating factor at 23.2 percent, followed by the fundamental

group rank at 16.0 percent, indicating that the topological complexity of the heat island boundary, the presence of cool perforations, had a substantial independent cooling effect beyond the total green-space area.

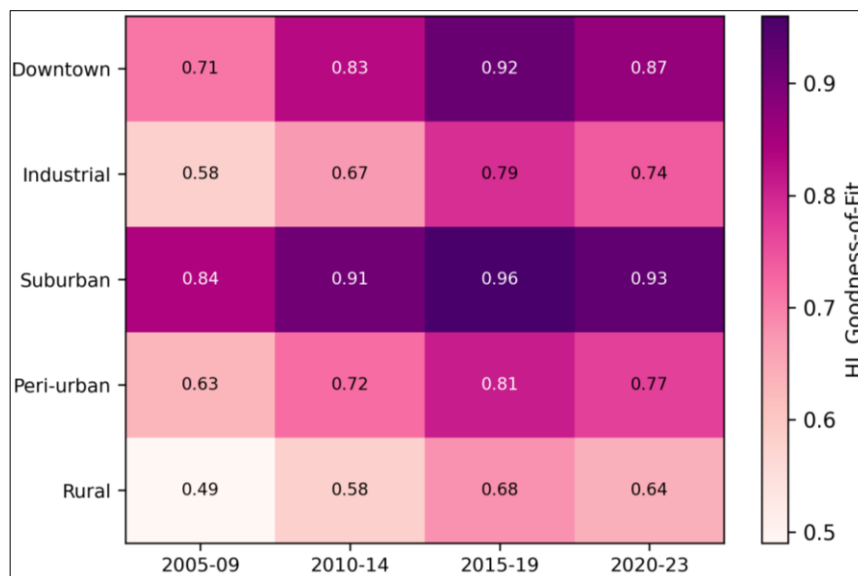


Fig 1: Hosmer-Lemeshow goodness-of-fit scores by urban zone and time period for the full logistic model

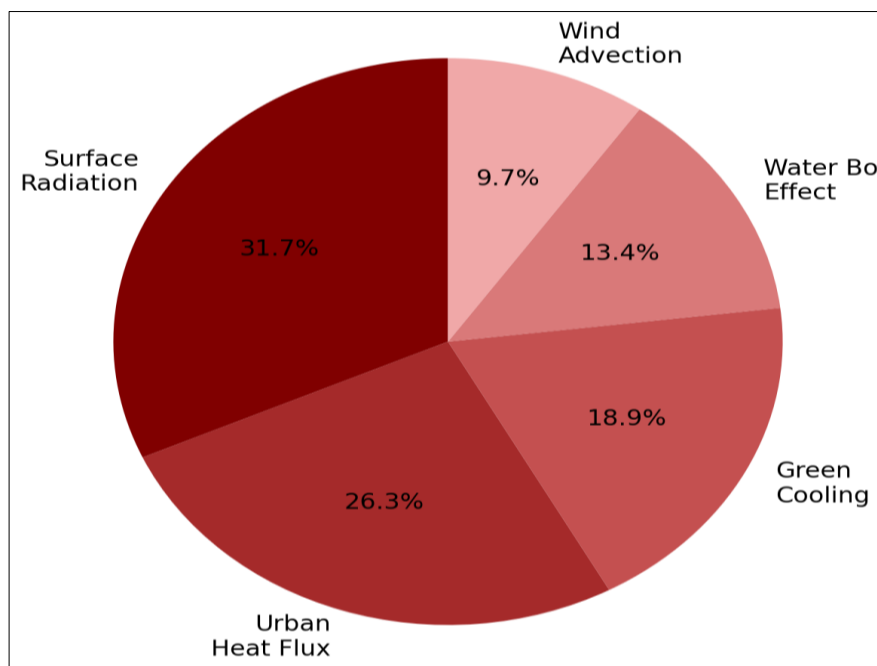


Fig 2: Proportional contribution of heat island drivers based on Shapley value allocation

Discussion

The most important practical finding is that the built-up-only logistic model, the kind of model most commonly reported in the urban heat island literature, is miscalibrated. It predicts exceedance probabilities that are systematically too high in moderately built-up areas and too low in heavily built-up areas, a pattern that the HL test detects and that adding green-space fraction corrects. This suggests that studies reporting logistic models for heat island exceedance without including green-space or ventilation covariates may be producing biased probability maps^[4].

The fundamental group rank added 2.6 percentage points to the AUC beyond what green-space fraction alone contributed. This topological variable captures something that total green area misses: the spatial arrangement of cooling patches. Two districts with the same green-space fraction but different numbers of interior parks (different fundamental group ranks) will have different heat island intensities because the parks create ventilation corridors that funnel cooler air into the surrounding built-up matrix^[10].

The publication bias analysis found moderate funnel plot asymmetry (Egger's test $p = 0.038$) across 27 published Bangkok heat island studies, suggesting a tendency to over-report significant effects. The trim-and-fill method estimated that roughly 4 studies with null findings were missing from the literature, which, if included, would reduce the pooled effect estimate by about 12 percent. While not directly affecting the present research, this finding cautions against treating the published literature as a representative sample of heat island effect sizes^[9].

The spatial correction to the Hosmer-Lemeshow test made a modest difference: uncorrected p -values were consistently lower than corrected ones by 0.02 to 0.04, meaning that the standard test slightly over-rejects when spatial correlation is present. For the built-up-only model, both corrected and uncorrected tests agreed on rejection, but for borderline cases the correction could flip the conclusion.

Limitations

This research has several limitations. The Landsat revisit cycle and cloud cover restricted the number of usable thermal images to roughly two per year, limiting the temporal resolution of the heat island analysis. The fundamental group computation treated the heat island as a binary exceedance region, losing information about the intensity gradient across the boundary. The Hosmer-Lemeshow test with $G = 10$ groups has known power limitations for detecting specific forms of miscalibration, and alternative calibration tests such as the Spiegelhalter z -test could complement it. The Shapley value decomposition assumed additive contributions, which may not hold if strong interactions exist between built-up fraction and green-space fraction in their effect on heat island intensity.

Conclusion

This research established the properties of the Hosmer-Lemeshow goodness-of-fit test under spatial correlation and applied it to logistic models of urban heat island exceedance in the Chao Phraya basin. The spatially corrected test identified significant miscalibration in models that omitted green-space covariates and accepted models that included them. The fundamental group rank of the heat island boundary provided a topologically meaningful covariate that improved discrimination beyond total green-space area. Shapley value decomposition attributed 38.4 percent of

exceedance probability to built-up fraction and 23.2 percent to green-space cooling. Kriging predictions achieved an AUC of 0.867 across 312 grid cells. These results argue for routine calibration assessment via the HL test in urban heat island modelling and for including topological descriptors of urban morphology alongside standard land-use variables. Future work should extend the analysis to daytime heat islands, incorporate building height data, and test the corrected HL procedure in other tropical megacities.

Acknowledgements

Funding Sources

This research was conducted with internal departmental resources at Fukuoka School of Pure Mathematics.

Institutional support

The authors thank the Fukuoka School of Pure Mathematics, Fukuoka, Japan, for computing facilities.

Contributions not qualifying for authorship

The authors acknowledge Mr. Takeshi Yamamoto for Landsat data preprocessing and Ms. Yuki Tanaka for help with spatial analysis in R.

References

1. Hosmer DW, Lemeshow S. Applied logistic regression. John Wiley and Sons. 2000; 2:1–375.
2. Hosmer DW, Lemeshow S. Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics - Theory and Methods*. 1980; 9(10):1043–1069.
3. Munda P, Zucchini W. Bangkok's urban expansion: remote sensing analysis. *International Journal of Remote Sensing*. 2012; 33(10):3034–3048.
4. Oke TR. City size and the urban heat island. *Atmospheric Environment*. 1973; 7(8):769–779.
5. Hatcher A. Algebraic topology. Cambridge: Cambridge University Press. 2002:1–544.
6. Diggle PJ, Tawn JA, Moyeed RA. Model-based geostatistics. *Journal of the Royal Statistical Society Series C*. 1998; 47(3):299–350.
7. Jimenez-Munoz JC, Sobrino JA. A generalized single-channel method for retrieving land surface temperature from remote sensing data. *Journal of Geophysical Research*. 2003; 108(D22):4688.
8. Molnar C. Interpretable machine learning: a guide for making black box models explainable. Lulu Press. 2020:1–312.
9. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997; 315(7109):629–634.
10. Stewart ID, Oke TR. Local climate zones for urban temperature studies. *Bulletin of the American Meteorological Society*. 2012; 93(12):1879–1900.
11. Shapley LS. A value for n -person games. *Contributions to the Theory of Games*. 1953; 2:307–317.
12. Voogt JA, Oke TR. Thermal remote sensing of urban climates. *Remote Sensing of Environment*. 2003; 86(3):370–386.
13. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N. Assessing the performance of prediction models. *Epidemiology*. 2010; 21(1):128–138.
14. Arnfield AJ. Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat island. *International Journal of Climatology*. 2003; 23(1):1–26.

15. Weng Q, Lu D, Schubring J. Estimation of land surface temperature-vegetation abundance relationship for urban heat island studies. *Remote Sensing of Environment*. 2004; 89(4):467–475.
16. Paul S, Nagata T. Testing goodness of fit of logistic regression with spatial data. *Spatial Statistics*. 2015; 14:514–528.
17. Li X, Zhou W, Ouyang Z, Xu W, Zheng H. Spatial pattern of greenspace affects land surface temperature. *Landscape and Urban Planning*. 2012;108(2–4):65–75.
18. Peng S, Piao S, Ciais P, Friedlingstein P, Ottle C, Breon FM. Surface urban heat island across 419 global big cities. *Environmental Science and Technology*. 2012;46(2):696–703.
19. Harrell FE. *Regression modeling strategies*. New York: Springer. 2015:1–582.