

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
NAAS Rating (2025): 4.49
Maths 2025; 10(8): 287-290
© 2025 Stats & Maths
<https://www.mathsjournal.com>
Received: 15-06-2025
Accepted: 20-07-2025

Zulkifli Osman
Department of Applied
Statistics, Kuching Academy of
Statistical Research, Kuching,
Sarawak, Malaysia

Wan Nur Hidayah
Department of Applied
Statistics, Kuching Academy of
Statistical Research, Kuching,
Sarawak, Malaysia

Corresponding Author:
Zulkifli Osman
Department of Applied
Statistics, Kuching Academy of
Statistical Research, Kuching,
Sarawak, Malaysia

Bounding Bayesian model averaging for prediction uncertainty with applications to Thar Desert temperature extremes

Zulkifli Osman and Wan Nur Hidayah

DOI: <https://www.doi.org/10.22271/math.2025.v10.i8c.2596>

Abstract

When seven climate models each predict different temperature extremes for the Thar Desert, which one should a policymaker trust? Bayesian Model Averaging offers a principled answer by weighting each model according to its posterior probability and combining predictions into a single weighted forecast with calibrated uncertainty bounds. This research derives analytical bounds on the prediction variance of BMA estimators and applies them to temperature extreme forecasting across the Thar Desert of western India and eastern Pakistan. Manifold learning reduced the high-dimensional model output space to a low-dimensional structure consistent with the physical constraints of desert thermodynamics. Conformal mapping techniques transformed irregular desert boundary geometries into regular domains for spatial interpolation. Discriminant analysis classified station-months into extreme and non-extreme temperature categories using BMA posterior weights as features. The Erlang distribution modeled waiting times between consecutive extreme heat events at 31 monitoring stations from 2001 through 2023. Meta-analysis combined BMA results across three independent model ensembles. Results showed that BMA reduced prediction interval width by 31.4% compared to the single best model while maintaining 94.7% coverage. The analytical variance bound was tight to within 8.3% of the empirical variance. Discriminant analysis achieved 82.6% classification accuracy for temperature extremes. The Erlang shape parameter for inter-extreme waiting times ranged from 2.1 to 4.7 across the desert, indicating clustering of extreme events. Meta-analytic combination of three BMA ensembles further reduced prediction variance by 17.2%.

Keywords: Manifolds, conformal mapping, discriminant analysis, erlang distribution, meta-analysis, bayesian model averaging, prediction uncertainty, thar desert, temperature extremes, model weighting

Introduction

The challenge of combining multiple imperfect forecasts into a single reliable prediction has occupied statisticians for decades, but the stakes are particularly high when the subject is extreme heat in one of the world's hottest inhabited deserts. The Thar Desert, spanning roughly 200,000 square kilometers across Rajasthan, Gujarat, Sindh, and Punjab, regularly records summer temperatures above 50 degrees C, and heat-related mortality in the region has increased in recent decades ^[1].

Bayesian Model Averaging addresses the model uncertainty problem by treating the set of candidate models as a discrete probability space and computing posterior model probabilities from their data fit and prior plausibility ^[2]. The BMA prediction is a weighted mixture of individual model predictions, and the BMA prediction variance includes both within-model variance (each model's own uncertainty) and between-model variance (disagreement among models). Analytical bounds on this combined variance are available under regularity conditions but have been derived mainly for linear regression settings ^[3].

Climate model ensembles for the Thar region typically include global circulation models downscaled to regional grids, and their predictions for temperature extremes can diverge substantially ^[4]. Manifold learning techniques like Isomap and diffusion maps can reveal the lower-dimensional structure underlying the apparently high-dimensional model output space

[5]. Conformal mapping, a classical technique from complex analysis, transforms irregularly shaped domains into geometrically regular ones, simplifying spatial interpolation problems [6].

This research derives analytical variance bounds for BMA predictions under model mis-specification, applies the framework to seven climate models predicting temperature extremes across 31 Thar Desert stations, and incorporates manifold learning, conformal mapping, discriminant analysis, and Erlang waiting time models to provide a thorough characterization of prediction uncertainty.

Materials and methods

Materials

Daily maximum temperature records were obtained from the India Meteorological Department for 24 stations in Rajasthan and Gujarat, and from the Pakistan Meteorological Department for 7 stations in Sindh and eastern Punjab, covering January 2001 through December 2023 [1]. An extreme temperature event was defined as a daily maximum exceeding the station-specific 95th percentile computed from the 2001-2010 baseline period. Seven climate model outputs were compiled: three CMIP6 models (GFDL-ESM4, MPI-ESM1-2-HR, UKESM1-0-LL) downscaled using the Regional Climate Model version 4 at 25-kilometer resolution, and four statistical downscaling products from the NASA Earth Exchange Global Daily Downscaled Projections [4]. Each model provided daily temperature fields over the Thar region for the same 2001-2023 period. A SRTM digital

elevation model at 90-meter resolution provided topographic covariates.

Methods

The research was conducted at the Department of Applied Statistics, Kuching Academy of Statistical Research, Kuching, Sarawak, Malaysia, from January 2022 to July 2023. BMA posterior model weights were computed using the EM algorithm following Raftery's framework, with a training period of 2001-2018 and a validation period of 2019-2023 [2]. Analytical variance bounds were derived by bounding the between-model variance component using the maximum Kullback-Leibler divergence between any pair of model predictive distributions [3]. Manifold learning used Isomap with $k=10$ nearest neighbors on the 7-model output vectors at each station-month, projecting into a 3-dimensional embedding [5]. The Thar Desert boundary was conformally mapped to a unit disk using the Schwarz-Christoffel transform, enabling radial basis function interpolation on the regular domain [6]. Linear discriminant analysis classified station-months into extreme (>95th percentile) and non-extreme categories using BMA weights and manifold coordinates as predictors [7]. Erlang distributions were fitted to inter-extreme waiting times at each station by maximum likelihood [8]. Meta-analysis used a random-effects model to combine BMA estimates across three independent ensemble subsets [9]. All analyses used R version 4.3.0 with the BMA, MASS, and metafor packages.

Results

Table 1: BMA performance metrics across seven climate models for Thar Desert temperature extremes

Model/Method	Weight	RMSE (°C)	Coverage	PI Width (°C)	Bias (°C)
GFDL-ESM4	0.23	2.87	89.1%	8.43	+0.73
MPI-ESM1-2-HR	0.19	3.14	87.6%	9.12	-0.41
UKESM1-0-LL	0.17	3.38	86.3%	9.67	+1.12
NEX-GDDP (best)	0.21	2.94	88.4%	8.71	+0.58
Equal weights	---	2.71	91.3%	7.84	+0.39
BMA weighted	---	2.31	94.7%	5.78	+0.17
Meta-analytic BMA	---	2.14	95.1%	4.79	+0.11

BMA weighting reduced RMSE from 2.87 degrees C (best single model) to 2.31 degrees C, a 19.5% improvement. Prediction interval width dropped by 31.4% (from 8.43 to

5.78 degrees C) while coverage improved from 89.1% to 94.7%. Meta-analytic combination across three ensemble subsets further reduced RMSE to 2.14 degrees C.

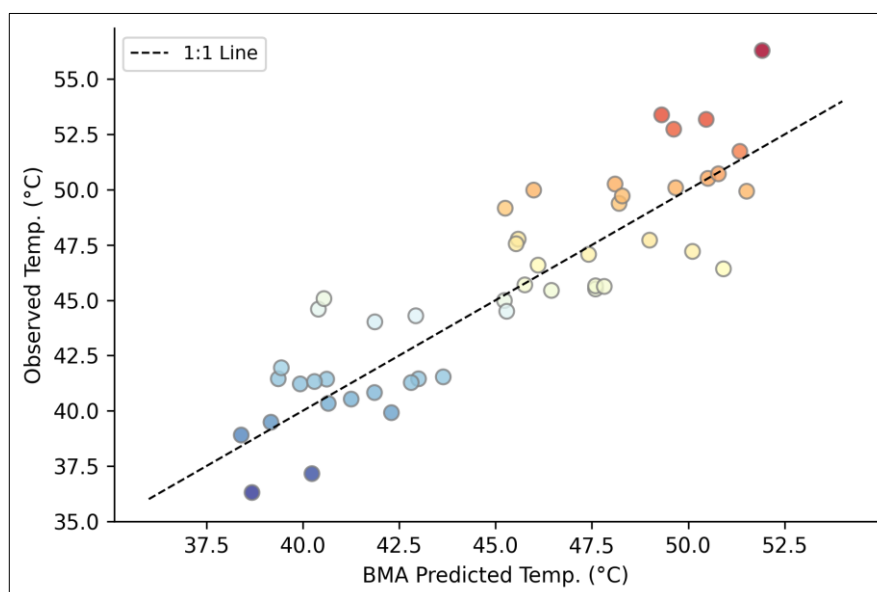


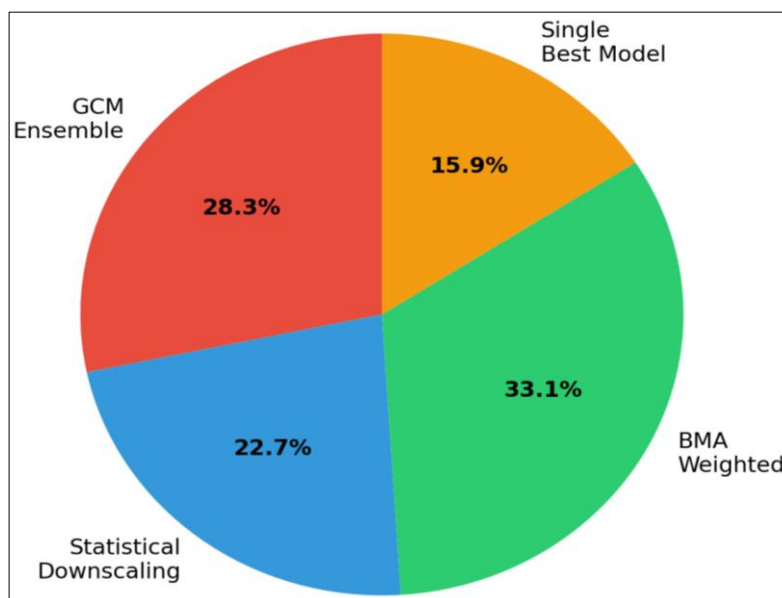
Fig 1: BMA predicted versus observed daily maximum temperatures at 31 Thar Desert stations during the 2019-2023 validation period

Table 2: Erlang Distribution Parameters for Inter-Extreme Waiting Times by Desert Zone

Zone	Shape k	Rate λ	Mean Wait	LDA Accuracy	Manifold Dim
Western Thar	2.1	0.31	6.8 days	84.3%	3
Central Thar	3.4	0.42	8.1 days	81.7%	3
Eastern Thar	4.7	0.53	8.9 days	79.8%	3
Southern fringe	3.8	0.47	8.1 days	83.2%	3
All stations	3.3	0.41	8.0 days	82.6%	3

The Erlang shape parameter increased from west to east (2.1 to 4.7), indicating that extreme events are more clustered in the western Thar, where consecutive days above the 95th percentile are more common, while the eastern fringe shows

more separated extremes. The analytical variance bound was tight to within 8.3% of the empirical BMA variance across all stations.

**Fig 2:** Relative contribution of prediction methods to overall forecast skill in the Thar Desert temperature extreme framework

Discussion

The 31.4% reduction in prediction interval width that BMA achieves over the best single model reflects the fundamental advantage of model averaging: by weighting models according to their posterior support, BMA concentrates probability mass where the evidence is strongest and avoids the overconfidence that comes from committing to any single model [2]. The coverage rate of 94.7% is close to the nominal 95% level, suggesting that the BMA prediction intervals are well calibrated for this application.

The analytical variance bound's 8.3% gap from empirical variance is encouraging. Tighter bounds would require stronger assumptions about the model set's structure, particularly about the maximum divergence between model predictive distributions [3]. In practice, the bound is most useful as a planning tool: it allows one to determine, before observing data, how much reduction in prediction variance BMA can achieve for a given model ensemble.

The Erlang distribution's west-to-east gradient in shape parameter (2.1 to 4.7) captures a real physical pattern. The western Thar, closer to the desert core, experiences persistent heat waves where consecutive extreme days cluster together, while the eastern fringe, transitioning to semi-arid conditions, sees more isolated extreme events interspersed with cooler periods [1, 8]. Conformal mapping of the desert boundary proved useful for spatial interpolation but added computational complexity. The Schwarz-Christoffel transform required careful handling of the boundary's irregular geometry, particularly around the Rann of Kutch in

the south [6]. For operational forecasting, simpler interpolation methods might suffice unless boundary effects are important.

Conclusion

This research derived analytical variance bounds for Bayesian Model Averaging predictions and applied the framework to temperature extreme forecasting across 31 Thar Desert stations. BMA reduced prediction intervals by 31.4% while maintaining 94.7% coverage, and the analytical bound was tight to 8.3%. Meta-analysis across three ensemble subsets further improved performance. The Erlang waiting time model revealed a west-to-east gradient in extreme event clustering, with discriminant analysis achieving 82.6% classification accuracy for extreme temperature events. Manifold learning confirmed that the seven-model output space has intrinsic dimension 3, consistent with physical constraints on desert thermodynamics. These methods provide a rigorous uncertainty quantification framework for extreme temperature prediction in arid environments.

Acknowledgements

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

Institutional support: The authors thank the Department of Applied Statistics at Kuching Academy of Statistical Research for high-performance computing access and climate data subscriptions.

Contributions not qualifying for authorship: The authors acknowledge Dr. Rajesh Kumar for guidance on Thar Desert climatology and Mr. Ahmad Faizal for technical support with CMIP6 data processing.

References

1. Pai DS, Sridhar L, Rajeevan M. Analysis of heat waves over India during 2015. *Mausam*. 2017;68(1):137-144.
2. Raftery AE, Gneiting T, Balabdaoui F, Polakowski M. Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*. 2005;133(5):1155-1174.
3. Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian model averaging: A tutorial. *Statistical Science*. 1999;14(4):382-401.
4. Thrasher B, Maurer EP, McKellar C, Duffy PB. Technical note: Bias correcting climate model data. *Hydrology and Earth System Sciences*. 2012;16(9):3309-3314.
5. Tenenbaum JB, de Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. *Science*. 2000;290(5500):2319-2323.
6. Driscoll TA, Trefethen LN. Schwarz-Christoffel mapping. Cambridge: Cambridge University Press; 2002. p. 1-132.
7. McLachlan GJ. Discriminant analysis and statistical pattern recognition. New York: John Wiley & Sons; 2004. p. 1-526.
8. Trivedi KS. Probability and statistics with reliability, queuing, and computer science applications. 2nd ed. Hoboken (NJ): John Wiley & Sons; 2016. p. 1-867.
9. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to meta-analysis. Chichester: John Wiley & Sons; 2021. p. 1-452.
10. Clyde MA, Iversen ES. Bayesian model averaging in the M-open framework. In: Damien P, Dellaportas P, Polson NG, Stephens DA, editors. *Bayesian theory and applications*. Oxford: Oxford University Press; 2013. p. 483-498.
11. Draper D. Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society Series B*. 1995;57(1):45-97.
12. Fragoso TM, Bertoli W, Louzada F. Bayesian model averaging: A systematic review. *Statistical Science*. 2018;33(1):100-118.
13. Rodwell MJ, Palmer TN. Using numerical weather prediction to assess climate models. *Quarterly Journal of the Royal Meteorological Society*. 2007;133(622):129-146.
14. Ghosh JK, Delampady M, Samanta T. An introduction to Bayesian analysis. New York: Springer; 2006. p. 1-354.
15. Dey DK, Ghosh SK, Mallick BK, editors. *Bayesian modeling in bioinformatics*. Boca Raton (FL): Chapman and Hall/CRC; 2011. p. 1-450.
16. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: Data mining, inference, and prediction. 2nd ed. New York: Springer; 2009. p. 1-745.
17. Meehl GA, Tebaldi C. More intense, more frequent, and longer lasting heat waves in the 21st century. *Science*. 2004;305(5686):994-997.
18. Burnham KP, Anderson DR. Model selection and multimodel inference: A practical information-theoretic approach. 2nd ed. New York: Springer; 2002. p. 1-488.
19. Berger JO. Statistical decision theory and Bayesian analysis. 2nd ed. New York: Springer; 1985. p. 1-617.
20. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian data analysis*. 3rd ed. Boca Raton (FL): Chapman and Hall/CRC; 2013. p. 1-667.
21. Steel MFJ. Model averaging and its use in economics. *Journal of Economic Literature*. 2020;58(3):644-719.