26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

Multiscale Analysis of Terrain–Lightning Relationships Using Machine Learning

Sima Rahmani[a,c], **Jason Evans**[a,c], **Jason Sharples**[b,c]

[a] *School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, 2052, Australia*
[b] *School of Science, University of New South Wales, Canberra, ACT, 2600, Australia*
[c] *NSW Bushfire and Natural Hazards Research Centre, University of New South Wales, Sydney, Australia*
Email: sima.rahmani@unsw.edu.au

Abstract: Understanding how terrain influences lightning occurrence is critical to improving spatial models of lightning distribution, particularly in the absence of atmospheric data. Topographic variables, such as elevation, slope, aspect, and local relief can influence atmospheric processes, particularly convective initiation, making them potentially useful proxies for lightning occurrence. Among these, local relief is inherently scale-dependent, capturing the vertical variability of terrain over a defined spatial window. Despite its relevance, the influence of multi-scale local relief on lightning occurrence remains under-investigated. This study focuses on local relief because, unlike point-based variables such as elevation, slope, and aspect, it encapsulates terrain variability over an area, potentially offering a more informative representation of land–atmosphere interactions.

Lightning occurrence is modelled using a machine learning framework based on topographic variables, including elevation, slope, aspect, and local relief—at spatial scales ranging from 1 to 165 km—derived from the 250 m-resolution GEODATA 9 Second digital elevation model (Version 3). Lightning strike records (2016–2021), sourced via a lightning detection network operated by Weatherzone/DTN with reported accuracy <200 m, were used to label each grid cell as lightning-affected or lightning-free. A Random Forest model with an 80:20 training/test split, used these variables to classify lightning occurrence within the study area of New South Wales, Australia. To isolate the terrain signal, no atmospheric predictors were included. Thirty-four models were developed using elevation, slope, and aspect, differing only in the spatial scale of local relief.

Permutation feature importance analysis reveals that among topographic variables, local relief and elevation are the most influential, whereas slope and aspect consistently contribute less to model performance across all scales. Due to this contrast, the analysis primarily focused on local relief and elevation. A comparison of permutation feature importance plots shows that local relief is a major contributor to the model's performance, with its contribution varying systematically across scales. While the values for elevation remain constant at all scales, local relief is recalculated at each scale. From 1 to 25 km, local relief consistently ranks as the second most important topographic variable; beyond 25 km, it matches or surpasses elevation in its contribution, indicating that larger-scale local relief encodes terrain structure not fully represented by static elevation values. To account for the correlation between elevation and local relief, which may affect feature importance ranking, the combined effect of these variables was measured. A grouped (elevation and local relief) feature importance measure was computed by permuting both simultaneously. The contribution of these grouped features increased gradually from 1 to 165 km. Results confirm that, despite correlation between elevation and local relief, the shift in rank order of elevation and local relief at larger scales is due to more helpful information provided by local relief at larger scales, whereas values for elevation remain constant. This shift reflects the added value from local relief rather than a mere reordering of feature importance caused by correlation.

The model performance was evaluated by different metrics including accuracy, recall, precision, F1-score, ROC-AUC. Across all evaluation metrics, performance generally improves steadily before gradually converging at the largest scales, with the most pronounced variations observed from 1 km to ~25 km, followed by a gradual, continuing upward trend at larger scales that eventually tends to converge at the final scales. These findings underscore the significance of topographic variables in spatial modelling of lightning occurrence, capturing 77.36% of actual events at the most informative scale of local relief, determined by the terrain characteristics of the study area. This represents a 15% improvement over the corresponding value in the model using local relief at the 1 km scale. Furthermore, it demonstrates how machine learning can elucidate complex land–atmosphere interactions. Local relief emerges as a valuable contextual predictor, capturing terrain variability that enhances model interpretability and performance. However, the optimal scale varied by terrain type, suggesting context-dependent effectiveness of this factor.

Keywords: *Lightning, Topography, local Relief, Data-Driven Multi-scale Analysis, Machine Learning*