

Developing an urban temperature emulator using physics-based climate models and machine learning

Minn Lin Wong^{1*}, Manmeet Singh²

¹Future Cities Laboratory, Singapore-ETH Centre, Singapore (minnlin.wong@sec.ethz.ch); ²Jackson School of Geosciences, The University of Texas at Austin, Austin, USA

1 Context and Motivation

Urban heat affects outdoor thermal comfort, public health, and energy demand. The distribution of heat in a city is shaped by complex interactions between urban form and prevailing weather conditions (summarised in figure 1).

Challenges:
Current methods for modelling city-scale climate, such as when testing the effects of heat mitigation strategies or land use change, often rely on computationally intensive physics-based climate models which can be slow and resource-heavy for city-scale, high-resolution applications.

Objective:
Develop an urban temperature emulator for city-scale heat analysis that can efficiently predict temperature response based on changing urban land-use and surface characteristics.

Proposed approach:
Build a climate emulator using a hybrid approach that combines the strengths of physics-based climate model simulations—which provide high-resolution, physically consistent data—with computationally efficient machine learning models that require training on high-quality datasets.

2 Physical drivers of urban heat

Urban temperatures are shaped by urban morphology like building density and heights and the presence of green spaces, as well as surface characteristics such as surface albedo and permeability, which collectively influence heat storage and ventilation. Analysing how the modification of these features impacts local climate is critical for designing effective heat-resilient cities and adaptive urban policies².

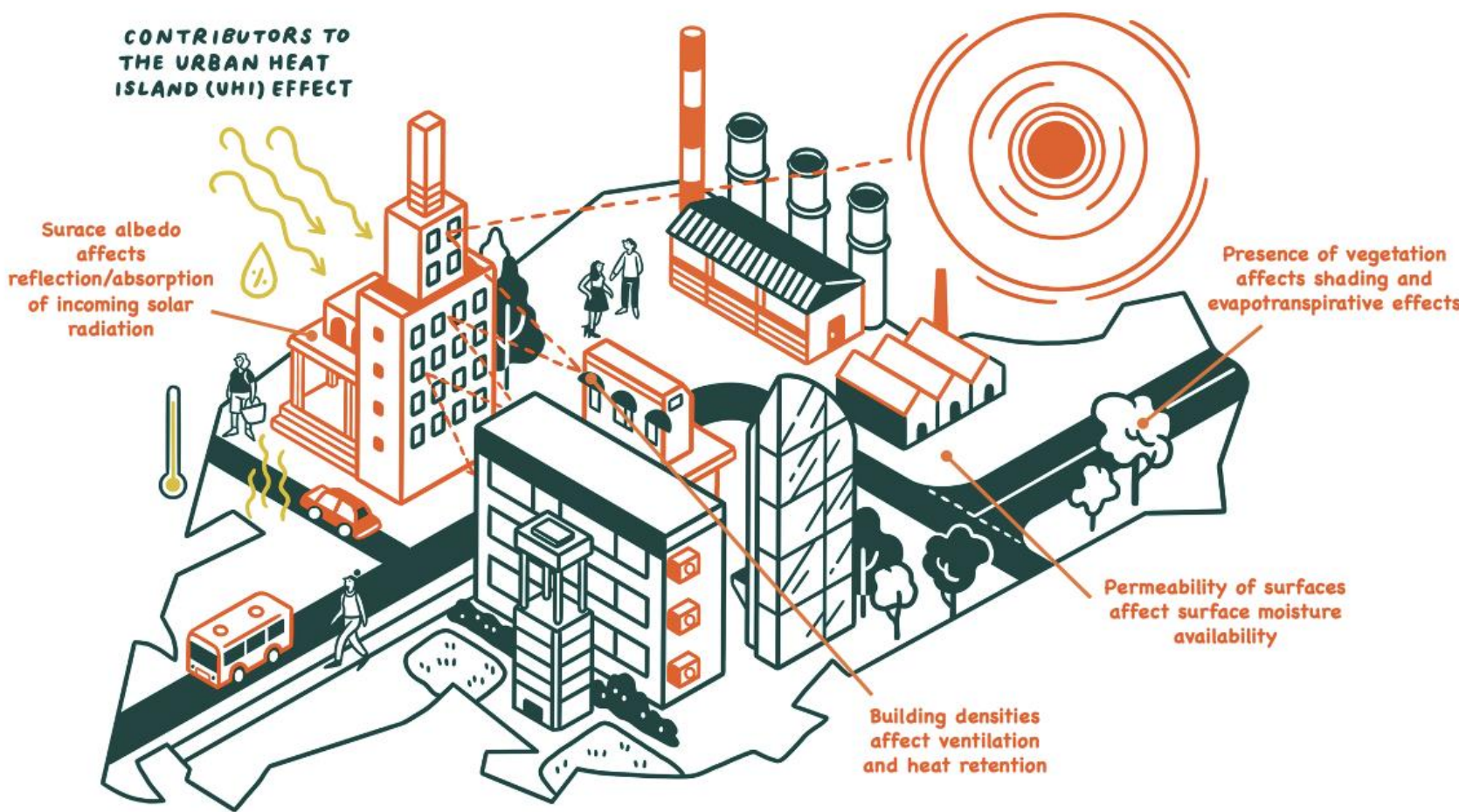


Figure 1: Schematic representation of factors that contribute to urban heat (Image modified from Cooling Singapore).

3 Proposed pipeline for building of city-scale temperature emulator

Table 1. Proposed pipeline outlining the steps from data generation to building the emulator model, illustrated using the city of Singapore as an example.

Identification of characteristic weather patterns	Input: ERA5 meteorological variables over study region - Specific humidity, Air temperature, U component of wind and V component of wind. Method: Weather Patterns Clustering¹ Output: A set of specific dates that represent characteristic weather conditions over the study region.
Generating high-resolution climatic data	Model: Weather Research Forecasting (WRF) model V4.2 at 300×300m resolution (resolution shown in Figure 2b). Input : Global Data Assimilation System (GDAS) 6-hourly final analysis (GDAS/FNL) for initialization. Updated local climate zone representation, updated building parameters, and anthropogenic heat profile³. Output: A variety of climatic variables across the 210×129 grid cells and at 37 levels vertical levels of atmosphere at hourly timesteps.
Land surface characteristics considerations	Data source : Sentinel-2, Open Street Maps, Municipal data sets Data considerations : Normalised Difference Vegetation Index Normalised Difference Building Index Building Heights and footprint Land use category
Model development for predicting 2m Temperature	Climatic predictive variables: Prioritize atmospheric fields known to influence urban heat dynamics (listed in Figure 2c). Variables are considered at the hourly timestep (t), and at various lead times (t-1hr, t-6hr, t-12hr), for each variable. P-value correlation tests will be used to reduce the input variables for meteorological data with the aim of retaining the variables with the strongest predictive influence on 2m temperatures. Data Split: Weather Patterns 1–8: Training set Weather Patterns 9–10: Testing set Baseline Models (initial emulator development): Random Forest, XGBoost Advanced ML Models: Diffusion Models (for higher resolution generation and downscaling) Flow Matching (to capture pattern evolution, such as heatwave development) Knowledge Distillation (compressing emulators into lightweight models for deployment)
Model evaluation	Evaluation: Evaluated the ML predicted 2m temperature with the WRF simulated 2m Temperature. Metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE)

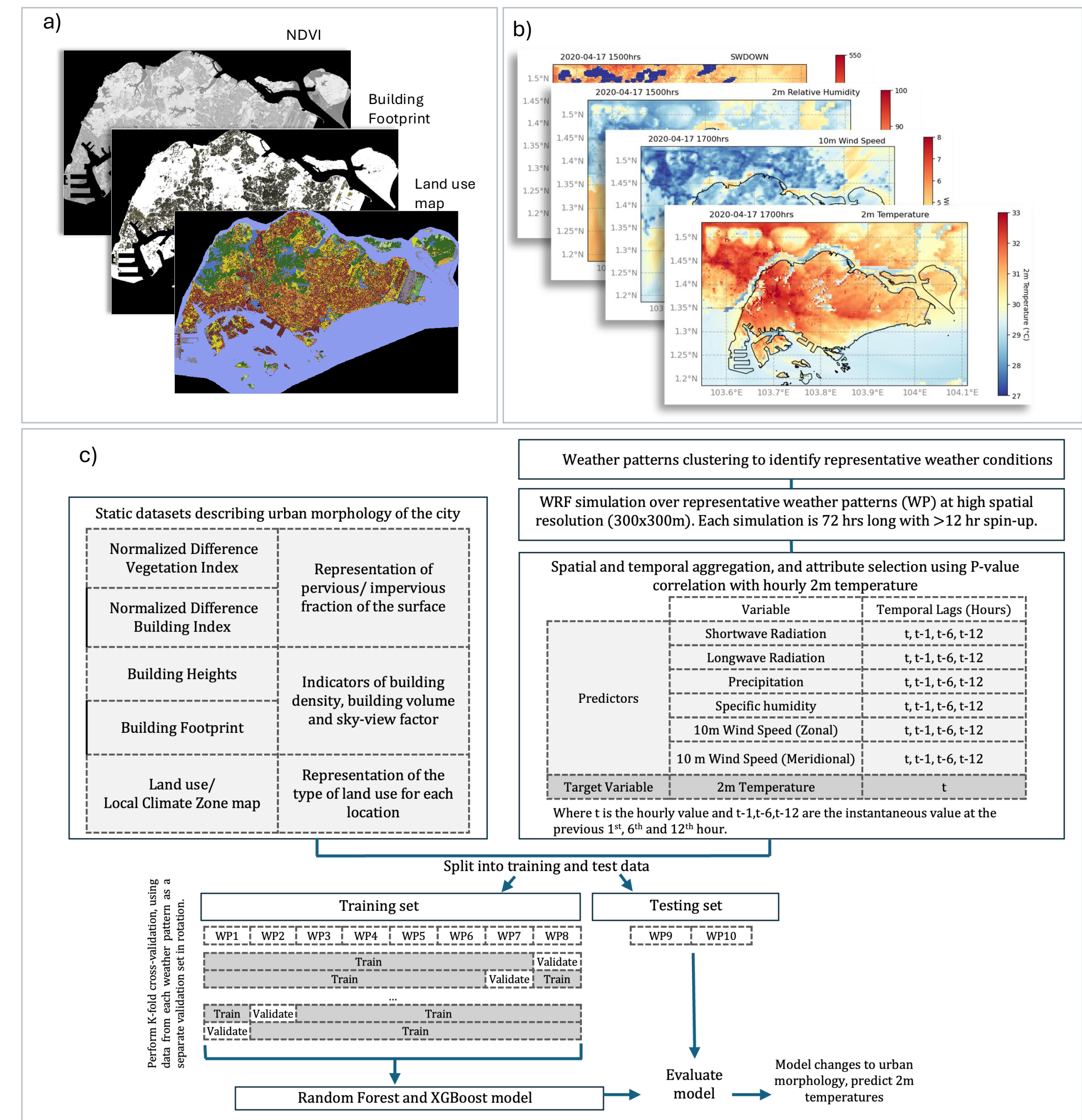


Figure 2: Proposed initial model workflow. Examples of the static and climatic datasets are shown in panels (a) and (b) respectively, for the city of Singapore, and the schematic of proposed methodology is shown in panel (c).

4. Pathway to Impact

Benefits to stakeholders, insights for urban design :
Trained emulators can provide fast temperature predictions in response to changes in urban form (e.g., greening, densification), enabling urban planners and policy makers to evaluate heat mitigation strategies without relying on computationally expensive climate models. Identifying key drivers of urban heat through feature importance rankings can further provide interpretable insights into the relative influence of different urban characteristics on temperature that can guide policy and design.

Potential drawbacks and other considerations :

- Training on weather patterns enables to model to generalize over different climatic conditions but may fail to capture extreme conditions.
- Emulator's performance may not be easily transferable to other cities without retraining, due to differences in urban morphology, climate, and data availability.

Acknowledgements and References

¹Ido Nevat and Juan A. Acero. Methodology for Identifying Mesoscale Weather Patterns from High-Dimensional Climate Datasets. Environmental Modeling & Assessment. (2024). ISSN 1573-2967. doi: 10.1007/s10666-024-09995-5.
²Ruefenacht, L., & Acero, J. A. (Eds.). (2017). Strategies for Cooling Singapore. A catalogue of 80+ measures to mitigate urban heat island and improve outdoor thermal comfort. Cooling Singapore (CS). Doi: 10.3929/ethz-b-000258216.
³Wong, M. & Nevat, I. (in prep.). A validation of the climate model embedded within DUCT (WRF). In Deliverable Technical Report: Vol. D 2.10.A [Report]. Singapore-ETH Centre (SEC).