Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Towards Rapid and Data-Efficient Urban Climate Assessment using Machine Learning Frameworks

Nisha Patel

nisha.patel@uni-kassel.de

University of Kassel

Lester Mackey

Microsoft Research New England

Jian Zhong

University of Greenwich

Yanzhi Lu

University of Greenwich

René Burghardt

Burghardt und Partner Ingenieure (BPI)

Koulis Pericleous

University of Greenwich

Britta Jaenicke

University of Kassel

Method Article

Keywords: Machine Learning Frameworks for Urban Cities, Climate Adaptation, Data-efficient urban planning, Urban Climate Assessment

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11241479/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Towards Rapid and Data-Efficient Urban Climate Assessment using Machine Learning Frameworks

Abstract

High-resolution physics-based urban climate models provide detailed representations of intra-urban thermal variability, yet their computational and data demands limit their use for rapid city-scale climate assessment and operational planning. This study develops and evaluates two complementary Machine Learning Frameworks (MLFs), FS-MLF (Fine-scale) and CS-MLF (Coarse Scale), that emulate near-surface air temperature from Atmospheric Dispersion Modelling System for urban environments (ADMS-Urban) simulations in Birmingham, UK. The study addresses two questions: (i) to what extent can machine learning support rapid prediction and spatial gap filling of high-resolution urban climate outputs? and (ii) how sensitive are fine- and coarse-scale representations to training-data availability and spatial configuration? In this study, FS-MLF predicts temperature at the Fine-scale using local neighbourhood information, while CS-MLF (U-MOSAIKS) learns compact coarse representations when only coarse-scale predictions are available. Both frameworks reproduced the dominant spatial thermal patterns with high accuracy, with FS-MLF achieving $R^2 = 0.85–0.93$ and CS-MLF $R^2 = 0.83–0.95$ on held-out test data. Performance remained strong using only 10% of training data, while geographically distributed random sampling (test $R^2 = 0.91–0.93$) matched or exceeded concentrated sampling spatially, highlighting the importance of spatial representativeness over sample density alone. Training-data experiments further showed convergence towards test $R^2 \approx 0.94–0.96$ with 50–70% of available samples. The findings demonstrate that machine learning can substantially reduce the computational burden of high-resolution urban climate emulation while retaining spatial fidelity, providing a scalable pathway for citywide climate adaptation assessment, urban digital twins, and climate-resilient planning.

Keywords

Machine Learning Frameworks for Urban Cities, Climate Adaptation, Data-efficient urban planning, Urban Climate Assessment

1. Introduction

Urban areas are on the front line of climate change due to the concentration of populations, infrastructure, and economic activity in urban environments. Moreover, cities are bearing the compounded burden of rising global temperatures and intensifying local heat extremes. This combination makes urban climate adaptation a critical priority for sustainable development (Pörtner et al., 2022; Oke et al., 2017; Pinheiro et al., 2026). Urban climate research shows that extreme heat is one of the most significant and rapidly growing urban health risks, with urban heat island (UHI) effects amplifying ambient temperatures and increasing heat-related morbidity and mortality (Morabito et al., 2019; Briegel et al., 2025; Unger et al., 2018; Stewart & Oke, 2012; Zhong et al., 2025). Despite growing importance of Urban Climate Adaptation Indicators (UCAIs), recent work (Patel et al., 2025; Mäkinen et al., 2020) highlights that their operational use remains limited. The major concerns listed are a lack of high-resolution datasets, as well as methodological challenges in translating heterogeneous urban data into reliable spatial adaptation maps. This constitutes a critical barrier to formulating effective climate adaptation and health resilience strategies at the urban scale (Tyler et al., 2016; Araos et al., 2016; Leiter, 2021).

Urban climate models provide a powerful means of addressing these challenges, offering the capability to generate detailed thermal maps that capture fine-coarse scale (10 m – 100 km) spatial heterogeneity across cities (Wang et al., 2025; Carmeliet et al., 2026; Lean et al., 2024). The high-resolution urban climate models such as, ADMS-Urban (Zhong et al., 2025; Biggart et al., 2021), PALM-4U (Maronga et al., 2019), ENVI-met (Bruse et al., 2023), and U-WRF (Barlage et al., 2016) are capable of representing thermal heterogeneity across a range of spatial scales, from microscale processes to meso- and local-scale atmospheric patterns. Moreover, these models incorporate urban morphology, land-cover composition, terrain characteristics, anthropogenic heat emissions, and meteorological forcing (Wang et al., 2025; Masson et al., 2002; Michau et al., 2023; Hertwig et al., 2021; B. Yang et al., 2012) to generate efficient urban climatic maps. However, despite their capability and physical realism, these urban climate models remain computationally expensive, data-intensive, and operationally demanding (Briegel et al., 2025).

Recent review studies (Heo et al., 2026; Masson et al., 2020; Villefranque et al., 2022; Briegel et al., 2023; Samad et al., 2026; Briegel et al., 2025) emphasise that conventional urban climate simulations frequently require extensive pre-processing workflows, repeated parameter calibration, and high-performance computing resources, limiting their operational applicability for rapid scenario exploration for climate-resilient planning. Moreover, even minor modifications to urban geometry, meteorological forcing, or land-cover characteristics typically necessitate re-running the entire simulation, incurring substantial computational time and operational cost (Briegel et al., 2023; Wang et al., 2025). For example, Wu et al. (2021) demonstrated that urban climate simulations at 250 m resolution require more than two orders of magnitude greater computational resources than an equivalent domain at 2.5 km resolution. Moreover, this work (Wu et al., 2021) underscores the non-linear increase in computational demand associated with higher-resolution urban climate modelling. These limitations reduce the practical feasibility of physically based urban climate models for real-time applications and rapid assessment of urban climate adaptation strategies (Heo et al., 2026).

These computational and operational constraints have motivated a growing number of researchers to investigate Machine Learning (ML) frameworks as efficient alternatives for emulating complex urban climate simulations (Amnuaylojaroen, 2025; Briegel et al., 2023, 2025; Gu et al., 2025; Miniandi et al., 2025; Samad et al., 2026; Syeda et al., 2025; Wu et al., 2021; Yang et al., 2012; Zhang et al., 2024). These data-driven ML methods are now widely recognised as efficient alternatives, or complements, to conventional physics-based modelling frameworks (Samad et al., 2026; Luo & Chen, 2025; Vahid & Aly, 2025). Among these, image-to-image deep learning architectures, particularly U-Net-based models, have become widely used due to their strong capability for spatial prediction and feature preservation (Ronneberger et al., 2015; Briegel et al., 2025). For instance, Lu et al. (2023) and Briegel et al. (2025) employed U-Net frameworks which are used to emulate large-eddy simulation (LES). They used SOLWEIG outputs, predicting wind speed (U), mean radiant temperature (Tmrt), and air temperature (Ta), respectively. However, Delgado Enales et al. (2024) noticed in their work at Bilbao, Spain that a defining characteristic across these studies have heavy reliance on large, well-sampled training datasets, typically following conventional 80/20 train–test splits (Briegel et al., 2025; Amnuaylojaroen, 2025; Hsu et al., 2022; Zumwald et al., 2021; Samad et al., 2026b).

This dependence on data-rich training conditions means that the robustness of existing ML frameworks under limited or sparsely distributed observational and simulation data remains largely less explored (Amnuaylojaroen, 2025; Samad et al., 2026b). Consequently, there is a need to develop ML-based approaches that maintain reliable predictive performance when training data are scarce. In this context, the Multi-task Observation using Satellite

Imagery and Kitchen Sinks (MOSAIKS) framework offers a promising alternative. By combining simple image encoding with linear models, MOSAIKS has demonstrated competitive predictive performance while substantially reducing computational cost, enabling scalable and globally transferable applications in data-limited settings (Rolf et al., 2021). Despite these advantages, its potential has been less investigated for emulating high-resolution micro- and meso-scale simulation outputs (Briegel et al., 2023). In particular, the ability to generate city-wide urban maps from only a limited number of micro-scale simulations could provide a valuable solution for cities where comprehensive simulation datasets are unavailable (Li et al., 2026). Recent work by Li et al. (2026) demonstrates the potential of machine learning to emulate high-fidelity urban microclimate simulations, showing that a deep neural network trained on only 36 high-fidelity simulations can accurately reconstruct three-dimensional temperature fields, with a mean absolute percentage error of less than 1.2%. While Li et al. (2026) primarily demonstrate the accuracy and computational efficiency of DNN-based reconstruction of high-fidelity microclimate fields, the present study extends this approach towards scalable city-scale urban climate emulation. Specifically, we develop and evaluate complementary fine- and coarse-scale ML frameworks and systematically investigate the effects of training-data availability and spatial sampling configuration on prediction accuracy and spatial gap filling. This provides new insights into how ML-based emulation can balance spatial fidelity, data requirements, and computational efficiency for rapid citywide urban climate assessment. This wok demonstrate the potential of ML for predicting urban microclimate conditions from limited simulation data. Although recent advances in numerical approaches such as the Hu et al. (2025) have used Lattice Boltzmann Method (LBM) to improve the feasibility of high-resolution urban airflow simulations through their efficient parallelisation and suitability for large-scale urban domains, generating detailed simulations over extensive areas, these can still involve substantial computational resources, particularly when repeated simulations are required across different conditions and scenarios. Developing a framework capable of learning from such simulation data and transferring the learned relationships across a wider urban area therefore provides a complementary strategy for reducing the amount of high-fidelity simulation data required, while supporting efficient and data-robust urban temperature prediction.

To address this gap, we introduce FS-MLF and CS-MLF, as an adaptation of the MOSAIKs framework for urban climate applications (U-MOSAIKS) emulating high-resolution urban climate model predictions. This study systematically evaluates the effects of training sample size and spatial training tile size on ML predictions using identical predictor datasets derived from ADMS-Urban outputs (Zhong et al., 2025), at a neighbourhood-scale spatial resolution. Moreover, the ADMS-Urban simulations are treated as the truth, with the objective of developing a computationally efficient surrogate emulator that reproduces their outputs at substantially lower computational cost. Weather-station observations are also used, but only as an independent reference for evaluating the realism of the simulated fields, rather than as the direct prediction target.

Specifically, the study addresses two key questions: (1) To what extent can ML Frameworks support rapid prediction, spatial gap filling, and scalable urban climate model outputs? (2) How sensitive are the proposed fine (pixel-based) and coarse (U-MOSAIKS) frameworks to variations in training sample size? Both these proposed frameworks are aim to substantially reduce computational cost while maintaining high spatial fidelity. Moreover, they provide an efficient pathway for generating citywide, high-resolution urban climate information to support as an assessment tools for climate adaptation, digital twins, and climate-resilient urban planning.

2. Study design and Machine Learning Frameworks (MLFs)

2.1 Study area

Birmingham is the second-largest city in the United Kingdom, with a population of approximately 1.15 million (Office for National Statistics, 2023), and exhibits a heterogeneous urban structure spanning dense city-centre development, low-rise residential suburbs, industrial and commercial zones, and extensive green space. This is reflected in a diversity of Local Climate Zones (Stewart and Oke, 2012), several of which have been identified across the city under the WUDAPT Europe mapping initiative (Demuzere et al., 2019), corresponding to marked variation in building height, density, and surface cover. The city also spans a notable elevation gradient, sitting on the Birmingham Plateau at roughly 100–300 m above sea level, rising to 315 m in the adjoining Lickey and Clent Hills to the SouthWest, terrain shown to influence local temperature patterns, with the comparatively higher ground of southwest Birmingham associated with cooler conditions. Birmingham has a temperate maritime climate (Köppen–Geiger classification), with urban temperature patterns further shaped by the distribution of built surfaces, vegetation, and water bodies, and by a well-documented urban heat island effect (Chapman et al., 2015; Bassett et al., 2016). This combination of heterogeneous urban morphology, terrain, and surface characteristics makes Birmingham a suitable testbed for evaluating whether ML frameworks trained on spatially limited

simulation data can reproduce temperature patterns across a wider urban area and for assessing the transferability of the proposed approach to other cities with comparable or contrasting urban and topographic characteristics.

2.2 ADMS-urban model output for Birmingham, UK

The present study uses an existing high-resolution ADMS-Urban Temperature and Humidity (ADMS-Urban T&H) simulation dataset for Birmingham, UK, originally developed and evaluated by Zhong et al. (2025). ADMS-Urban T&H, developed by Cambridge Environmental Research Consultants (CERC), is a physics-based urban climate model that represents spatial variations in air temperature by accounting for ground and latent heat fluxes associated with surface thermal properties and moisture availability. Previous studies have demonstrated its ability to reproduce neighbourhood-scale air temperature patterns (Hamilton et al., 2014; Maggiotto et al., 2014; Wang et al., 2019; Tiwari et al., 2021). Zhong et al. (2025) generated the Birmingham simulations at a 50-m spatial resolution and 1-hour temporal resolution using thermal admittance, surface resistance to evaporation, and albedo to represent key controls on the urban surface energy balance. Their simulations incorporated land cover, terrain, normalised building volume, surface roughness, and hourly upwind meteorological conditions. The model domain covered the minimum bounding square enclosing Birmingham with an approximately 15% outward buffer, and the simulations were performed on the University of Birmingham's BlueBEAR supercomputer. The present study uses the resulting ADMS-Urban T&H output for 11:00 on 26 August 2019 as the reference dataset for developing and evaluating the proposed machine-learning frameworks. Zhong et al. (2025) evaluated this model output against Landsat 8-derived land surface temperature, providing an independently assessed physics-based representation of the spatial thermal pattern. Thus, rather than conducting new ADMS-Urban simulations, this study builds on the computationally intensive simulations already produced by Zhong et al. (2025) and investigates their emulation using machine-learning approaches.

2.3 Study design

Surface characteristic layers for Birmingham at 50 m × 50 m spatial resolution, including (i) thermal admittance (thc), (ii) surface resistance to evaporation (evp), (iii) albedo (alb), (iv) surface roughness length (ruf), (v) normalised building volume (nbv), and (vi) terrain elevation (ter), were used as input variables to predict target near-surface air temperature for year 2019. Using QGIS 3.40, all raster layers were partitioned into a regular grid of 50 × 50 pixel (where 50 x 50 pixel spatial resolution = 1 tile = 2.5 km x 2.5 km), with alternative spatial resolution tile configurations evaluated to assess sensitivity to spatial partitioning (Supplementary Figure S1). Each tile contained a co-registered stack of the six-predictor layers and the corresponding near-surface air temperature values over an identical spatial extent. This approach provides a consistent training and prediction unit for both modelling frameworks (Figure 1). This common tiling strategy ensured spatial alignment between predictor and target datasets, minimised the risk of pixel-level misregistration, and facilitated systematic quality control through the identification and removal of incomplete or corrupted spatial patches before ML framework development.

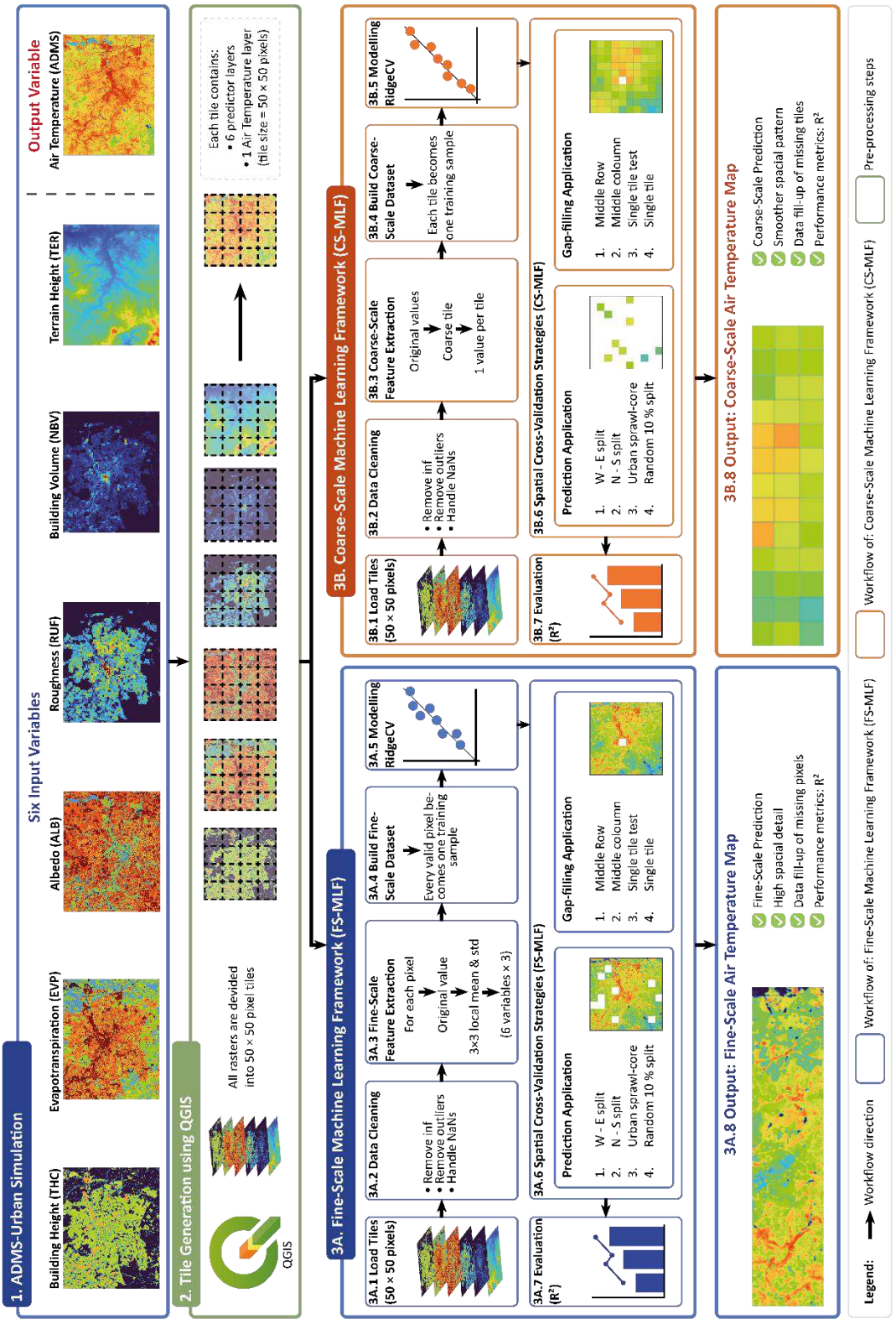


Figure 1: Workflow two Machine Learning (ML) frameworks for predicting air temperature from ADMS-Urban simulation outputs: (3A) a fine resolution, spatially explicit framework, and (3B) a coarse Urban-MOSAIKS-style framework Solid arrows indicate workflow direction; dashed arrows indicate data/information flow. Blue boxes denote fine-resolution workflow, orange boxes denote coarse-resolution workflow, and green boxes denote shared pre-processing steps.

2.4 Machine Learning Frameworks (MLFs)

This study proposes two ML frameworks for predicting near-surface air temperature from ADMS-Urban derived outputs. These two frameworks are the i) Fine-scale Machine Learning Framework (FS-MLF) and ii) Coarse Scale MOSAIKS based MLF (CS-MLF). These two methods are complementary to each other and operate at different spatial scales.

2.4.1 Fine-scale Machine Learning Framework (MLF-FC)

This framework treats every individual pixel within every tile as a training observation. The workflow of this framework is mentioned in Figure 1 (3A) and explained below.

i) **Tile loading and data cleaning** (Figure 1, Section 3A.1-3A.2): Each 50 x 50 pixel tile is loaded for all six input variable layers and the target layer. Then data cleaning performed to remove infinite values and outliers before feature construction.

ii) **Pixel-wise feature extraction** (Figure 1, Section 3A.3)**:** For each pixel, 18 features are derived from six input-layer variables. For each variable, three features are calculated: the original pixel value, the mean value calculated from the target pixel and its eight immediately adjacent pixels (3 × 3 neighbourhood), and the corresponding standard deviation within the same neighbourhood (6 variables x 3 statistics = 18). This neighbourhood-based representation embeds local spatial context directly into each pixel's feature vector.

A spatial context sensitivity analysis was conducted for the pixel-based reconstruction model to quantify the influence of neighbouring information. The neighbourhood size k was systematically varied over

$$k \in \{1, 3, 5, 7, 9\}$$

For each configuration, adjacent grid cells surrounding each target pixel were included as input features. Each model variant was independently trained and evaluated to assess the effect of spatial context on predictive accuracy, spatial coherence, smoothness of reconstructed fields, and preservation of local thermal gradients. This procedure enables identification of an optimal neighbourhood configuration that balances local detail preservation with spatial stability. Where $k \in \{1, 3, 5, 7, 9\}$ represents a spatial neighbourhood window of adjacent grid cells. This formulation explicitly incorporates local spatial context to improve continuity and reduce discontinuities in reconstructed temperature fields.

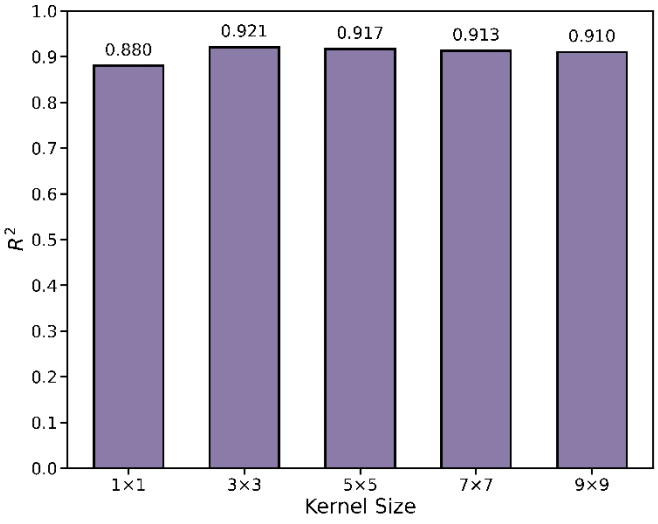


Figure 2: Relationship between neighbourhood kernel size and model performance using the coefficient of determination (R^2) on held-out test data for the pixel-based spatially explicit machine learning framework.

iii) **Dataset construction (**Figure 1, Section 3A.4): Every valid pixel becomes one training sample, so the resulting dataset has pixel-level granularity.

iv) **Model training (**Figure 1, Section 3A.5): A ridge regression model with regularisation parameter selected via cross-validation (RidgeCV) is fitted to the standardised pixel-level feature set using scikit-learn.

v) **Spatial cross-validation and Output** (Figure 1, Section 3A.6-10)**:** Spatial generalisation is tested eight application scenarios: i) four involving prediction and ii) four involving gap filling (details in results). For example, if we are using a 10% random train-test split scheme, the rest of the datasets will be used as training tiles (rest 90%). The various applications have been explored in the results section. Then, the results are evaluated against the original ADMS-Urban data using test R^2, that is, the coefficient of determination on held-out test data. The result is a fine-resolution air temperature map at native pixel-level prediction, preserving spatial details.

2.4.2 Coarse Scale (U-MOSAIKS) Machine Learning Framework (CS-MLF)

This framework treats each 50 x 50 pixel tile as a single training observation. This approach summarises its spatial content into one descriptive vector in the style of the MOSAIKS (Multi-task observation using satellite Imagery Sinks) approach introduced by (Rolf et al., 2021). The CS-MLF provides a practical solution for applications where meteorological datasets are spatially sparse or unavailable at fine spatial scales. By aggregating information at the coarse level, it allows CS-MLF to apply in data-scarce regions while still capturing the broader spatial patterns relevant to the analysis. This is particularly useful during preprocessing urban climatic maps, where fine-scale spatial detail may not be necessary and a coarser representation can provide sufficient information with lower data and computational requirements.

i) **Tile loading and data cleaning** (Figure 1, section 3B.1-3B.2): As in framework 1, each tile's six input variable layers and target layer are loaded and infinite values, outliers and NaNs are removed.

ii) **Urban-MOSAIKS tile representation** (Figure 1, section 3B.3): In this framework, each tile is reduced to a single feature vector as an input variable layer. This produces six coarse, tile-level representations of the input layer and predicts the air temperature.

iii) **Dataset construction (**Figure 1, section 3B.4): Each tile becomes one training sample, so the resulting dataset has coarse (tile-level) granularity. This framework mimics the common scenario in which only coarse-grained predictor data is available for simulation and emulation.

iv) **Model training (**Figure 1, section 3B.5): As in FS-MLF, a RidgeCV regression model is fitted to the tile-level datasets.

v) **Spatial cross-validation and Output** (Figure 1, section 3B.6-10)**:** Spatial generalisation is tested in eight application scenarios: i) four involving prediction and ii) four involving gap filling (details in results) for coarse resolution. The tiles are generated tile-level test sets and reassembled into coarse spatial maps, filling in missing gaps. Performance is quantified using test R^2. The result is a coarse-resolution air temperature map at the coarse resolution.

Overall, the two frameworks represent complementary approaches to spatial temperature prediction. FS-MLF preserves fine-scale spatial variability by operating at the pixel level and explicitly incorporating neighbourhood information, enabling the reconstruction of continuous high-resolution temperature fields. Moreover, CS-MLF adopts a coarse, tile-based representation, where random fourier features provide a nonlinear representation of the spatial information within each tile, enabling efficient prediction of a single representative temperature value. While both frameworks employ RidgeCV regression and identical train–test partitioning, they differ fundamentally in how spatial information is represented, the granularity of prediction, and the required granularity of the input data (Table 1).

Table 1: ***Description of the Fine-scale resolution (FS-MLF) and Coarse Scale resolution U-MOSAIKS (CS-MLF) Machine learning frameworks.*** *This table summarises the prediction strategy, target formulation, feature representation, spatial*

information integration, training configuration, regression approach, pre-processing steps, validation strategy, and output characteristics used in each framework.

	Fine-scale MLF (FS-MLF)	**Coarse Scale MLF (CS-MLF)**
Prediction Unit	Individual pixels	Entire tile
Target Variable	Pixel temperature	Mean temperature per tile
Features	Raw variables + spatial statistics	Deep neural network embeddings
Spatial Information	Explicitly calculated (3×3 mean/std)	Learned implicitly by CNN
Training Samples	Thousands of pixels per tile	One sample per tile
Regression Model	RidgeCV	RidgeCV
Standardisation	Standard Scaler applied	No explicit scaling shown
Output	Full temperature map	One value per tile
Validation	All pixels in random test tiles	Random test tiles
Detailed spatial Prediction	Yes	No

2.5 Spatial holdout experiments

To assess the spatial generalisability of Machine Learning Frameworks (MLF), spatial holdout experiments were conducted using ADMS-Urban near-surface air temperature field for Birmingham, UK as reference data (Figure 3). In each experiment, the model was trained on a spatial subset and evaluated on a geographically distinct, unseen subset. For the experiments 50 x 50 pixel tiles (2.5 km x 2.5 km) have been used. In addition to that, the sensitivity of the framework to spatial tile size was evaluated using additional configurations of 10 × 10 (500 m x 500 m), 25 × 25 (1.25 km x 1.25 km), and 100 × 100 pixels (5 km x 5 km) (Supplementary Figure 1).

We conducted eight different experiments to test different skills and potential future applications:

Prediction-based applications

1) **Spatial generalisation from sparse training data (Random 10%) (Figure 3,5 (a)):** The objective here is to assess whether the proposed approach can produce accurate, city-wide, fine-resolution temperature predictions when expensive simulator is run only on a limited subset of locations to reduce total simulation time.
2) **Urban Core - Sprawl extrapolation (Figure 3,5 (b)):** This scenario focuses on extending the learned spatial relationships beyond the urban core, with the aim of predicting temperatures across surrounding urban-sprawl areas using training data obtained exclusively from densely developed regions.
3) **North - South spatial transfer (Figure 3,5 (c)):** To examine the transferability of the approach across geographically separated regions, temperature predictions are generated for one part of the city using simulation data from the Northern region of Birmingham.
4) **West - East spatial transfer (Figure 3,5 (d)):** A complementary spatial-transfer scenario is considered by training the approach on Western Birmingham and applying it to the Eastern region. This evaluates its ability to generate reliable temperature estimates when simulation data are available only in a geographically distinct area.

Spatial gap-filling experiments

5) **Row-wise spatial gap filling (Figure 4,6 (a)):** Here, a continuous strip of temperature data is intentionally removed from the city domain, and the proposed approach is used to reconstruct the missing region based on simulations available on either side of the gap.
6) **Column-wise spatial gap filling (Figure 4,6 (b)):** The orientation of the missing region is changed to provide a complementary gap-filling scenario. Temperature values within a continuous column-wise gap are reconstructed using information from the neighbouring simulated areas.
7) **Localised single-tile gap filling (Figure 4,6 (c)):** A more localised form of missing-data reconstruction is examined by removing an isolated tile from the simulation domain. The approach must infer the temperature within this area from the surrounding spatial context.
8) **Extreme sparse-data reconstruction (Figure 4,6 (d)):** The final experiment represents an extreme data-scarcity scenario, in which only a single tile is retained for training. The resulting model is then used to reconstruct temperature predictions across the wider city, providing a stringent assessment of its ability to generalise from extremely limited simulation data.

3. Results

The following sections 3.1, 3.2 and 3.3 evaluate both MLFs for 8 experimental scenarios (4 prediction application-based + 4 gap-filling based) and their performance with varying training sample sizes.

3.1 Fine-scale prediction based and gap-filling scenarios (FS-MLF)

The performance of the Fine-scale spatially explicit MLF was evaluated for pixel-level prediction and spatial gap filling experiments. The framework was first assessed by varying the spatial neighbourhood size (k = 1,3,5,7 and 9) to examine the local spatial context accuracy (Figure 2). The neighbourhood kernel size 3x3 shows high performance with a test set coefficient of determination (test R^2) value of 0.92. Therefore, for all the experiments we have used a 3 x 3 kernel size. For each configuration, the model was trained using the pixel-wise feature set and evaluated through spatial cross-validation using independent tile-based prediction schemes. Model performance was quantified using test R^2.

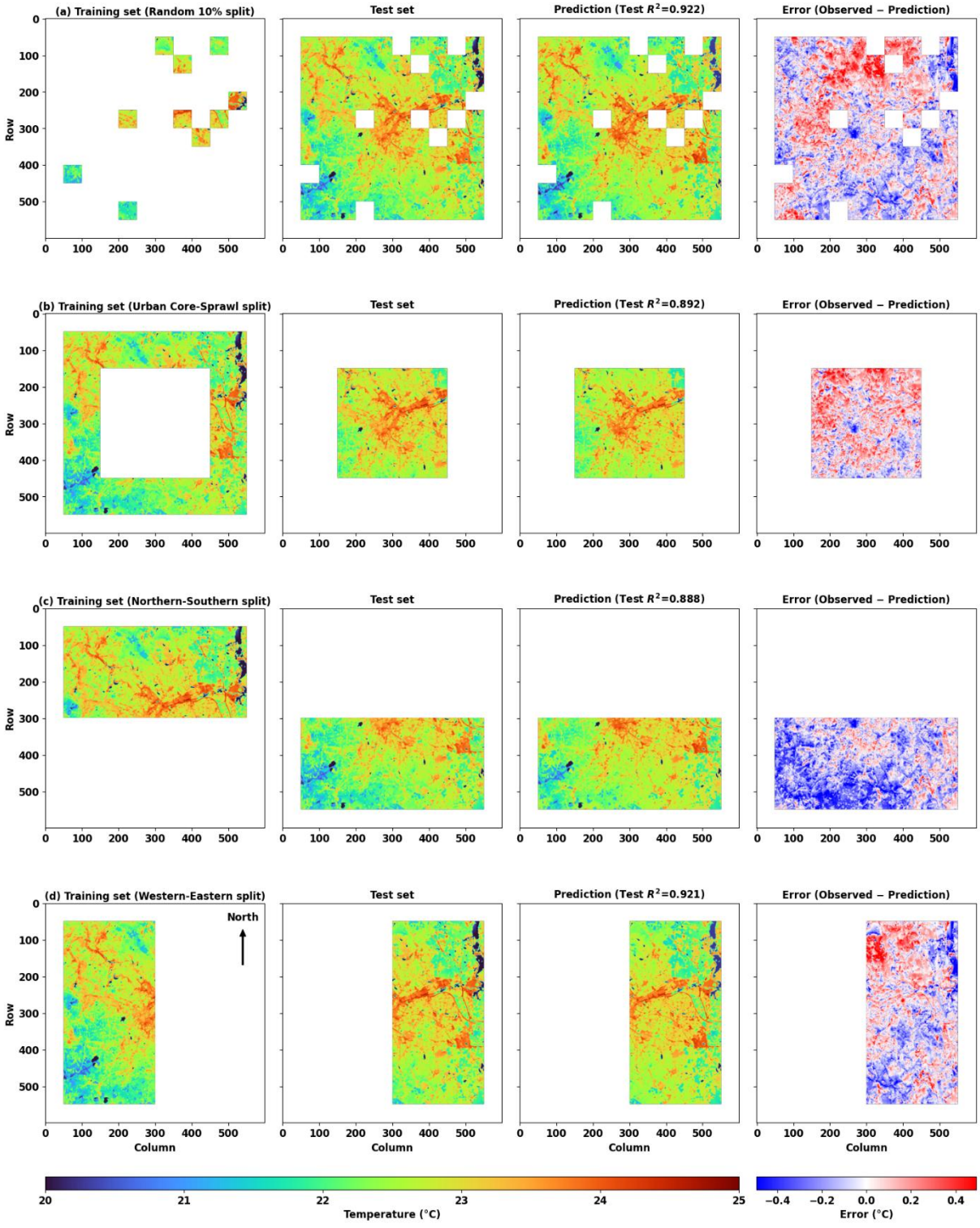


Figure 3: ***Prediction-based applications of the Machine Learning Framework (FS-MLF) for Fine-scale resolution, pixel-based near-surface air temperature estimation.*** *For each scenario, column 1 presents the spatial distribution of the training data, column 2 shows the reference ADMS-Urban simulation, column 3 shows the FS-MLF predicted near-surface air temperature, and column 4 presents the corresponding prediction error (reference − prediction).* ***(a)*** *FS-MLF is trained using a random 10% sample of the available data and applied to predict temperature across the remaining 90% of the city.* ***(b)*** *The model is trained within the urban core and extrapolated to the surrounding urban-sprawl areas.* ***(c)*** *The model is trained using data from the northern part of Birmingham and applied to predict temperatures in the southern part.* ***(d)*** *The model is trained using the western part of Birmingham and applied to predict temperatures across the eastern region. The scenarios progressively assess the ability of FS-MLF to generalise from sparse, spatially concentrated, and geographically separated training data to previously unseen areas.*

Fine-scale prediction-based scenarios: Figure 3 presents the performance of the pixel-based fine resolution machine-learning framework (FS-MLF) across two complementary applications: spatial prediction and gap filling. The spatial prediction experiments (Figure 3 (a-d)) assess the framework's ability to evaluate different training configurations by varying the spatial distribution of the training data. These include prediction within a citywide reconstruction using randomly selected 10% of available data (Figure 3a), urban core – sprawl (Figure 3b), Northern-Southern split (Figure 3c), and Western-Eastern split (Figure 3d). Moreover, the gap-filling experiments (Figure 4 a-c) assess the capability of the frameworks to reconstruct missing regions by withholding an entire row of tiles (Figure 4a), an entire column of tiles (Figure 4b), and a single tile prediction using all the datasets (Figure 4c) and prediction of the entire map using single tile test data (Figure 4d). For each experiment, the Figure presents the spatial distribution of the training data, the reference ADMS-Urban modelled data as test sets, FS-MLF predicted test results and corresponding prediction error (Observed-Predicted). Together, these experiments provide a comprehensive assessment of the framework's ability to reproduce Fine-scale temperature patterns, preserve local thermal variability and accurately reconstruct missing spatial information across urban environments in this case, Birmingham, UK, presented domain.

Training on a random 10% sample and predicting the remaining 90% yielded a high R^2 of 0.92 (Figure 4a), demonstrating that FS-MLF can achieve reliable fine-scale predictions from relatively sparse training data. The other proportions of the training data (10-70%) were also explored, as shown in Figure 8. Similarly, training on the urban core and predicting the surrounding urban sprawl resulted in R^2=0.89 (Figure 4b), indicating that the model can effectively transfer information from data-rich, thermally heterogeneous urban centres to less-observed peripheral areas. Directional extrapolation further showed strong performance, with West-to-East prediction achieving R^2=0.92 (Figure 4c), whereas North-to-South prediction yielded a comparatively lower R^2=0.88 (Figure 4d). This difference suggests greater spatial heterogeneity along the North–South direction, potentially influenced by variations in environmental conditions such as wind patterns. However, this mechanism cannot be directly established from the present experiments. Overall, the consistently high test R^2 values (0.88–0.92) demonstrate the robustness of FS-MLF for spatial extrapolation and its potential to fill fine-scale informative predictions where complete city-wide high-resolution mapping is constrained by data availability or computational requirements.

Fine-scale gap-filling scenarios: The gap-filling experiments demonstrated consistently strong reconstruction performance across all tested spatial configurations, confirming the ability of FS-MLF to infer missing Fine-scale information from surrounding observations. Withholding the middle row of tiles (Figure 4a) yielded the highest performance (test R^2=0.93), followed by the middle-column gap (Figure 4b; test R^2=0.91), indicating that FS-MLF can effectively reconstruct relatively large, spatially contiguous missing regions. Reconstruction of a single missing tile produced a slightly lower test R^2 of 0.89 (Figure 4c), while the most challenging configuration – using only a single tile for testing and predicting the remainder of the domain (Figure 4d) yielded test R^2=0.85. Nevertheless, under this most restrictive scenario, the model retained substantial predictive skill, highlighting its potential for fine-scale gap filling in cities or domains where observations are sparse and computationally intensive high-resolution microclimate simulations cannot be performed comprehensively.

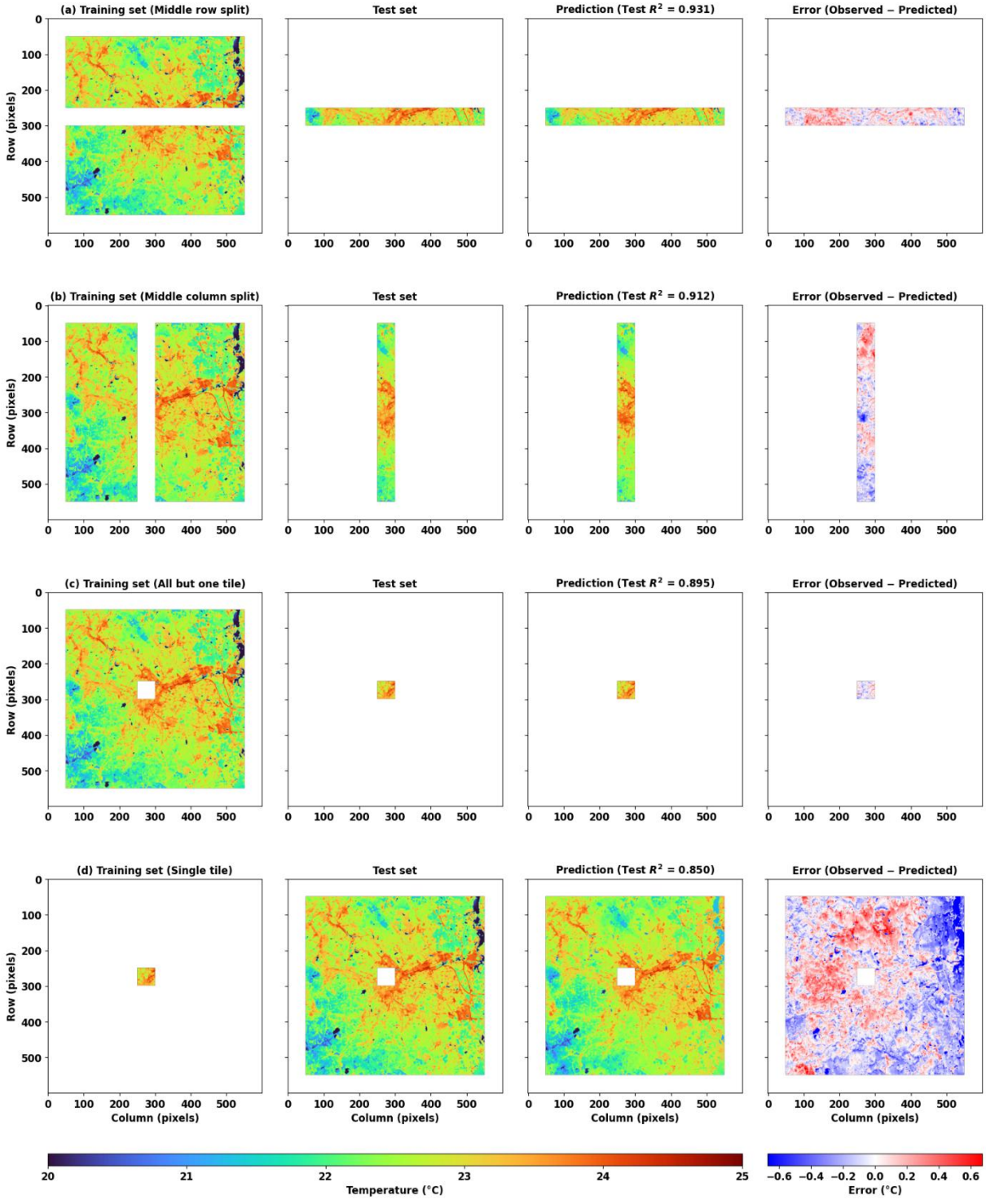


Figure 4: Spatial gap-filling capability of the fine-scale, pixel-based FS-MLF framework under progressively challenging data-availability scenarios. *(a) All tiles except row 6 are used for model training, with the withheld row predicted as a spatial gap. (b) All tiles except column 6 are used for training, with the withheld column reconstructed by the model. (c) A single tile is withheld from the complete domain and predicted using the remaining tiles as training data, representing localised gap filling. (d) Only a single tile is used for training, while the remaining domain is predicted, representing an extreme sparse-data scenario for large-scale spatial reconstruction.*

3.2 Coarse-scale Machine Learning Framework (CS-MLF)

To evaluate model performance at coarser spatial resolution and assess its suitability for city-scale applications, CS-MLF was evaluated using the same four spatial holdout configurations adopted for FS-MLF. CS-MLF and FS-MLF and therefore provide complementary capabilities. While FS-MLF is designed for fine-resolution, pixel-based prediction where fine-scale input and model output data is available for training, CS-MLF is designed for applications where only coarse-scale model outputs or coarse-scale input data are available for training. Together, the two frameworks provide flexibility for addressing a range of urban environmental prediction and data-gap-filling requirements.

Coarse-scale prediction-based scenarios: The coarse-resolution experiments demonstrated consistently strong spatial generalisation across all four configurations. Training on a random 10% of tiles and predicting the remaining 90% produced the best overall performance, with a test R^2 of 0.93 (Figure 5a), indicating that CS-MLF can effectively learn spatial relationships from a relatively small subset of the available data. Training on the urban sprawl and predicting the urban core (Figure 5b) yielded a test R^2 of 0.89, demonstrating robust transferability between areas with contrasting urban characteristics. For directional extrapolation, training on the northern half and predicting the southern half (Figure 5c) resulted in the lowest performance, with test R^2=0.87 and, whereas training on the western half and predicting the eastern half (Figure 5d) achieved higher performance (test R^2=0.91). The comparatively lower performance for the North-South configuration suggests greater directional variability in the underlying temperature field, consistent with the spatial heterogeneity observed in the fine-resolution experiments.

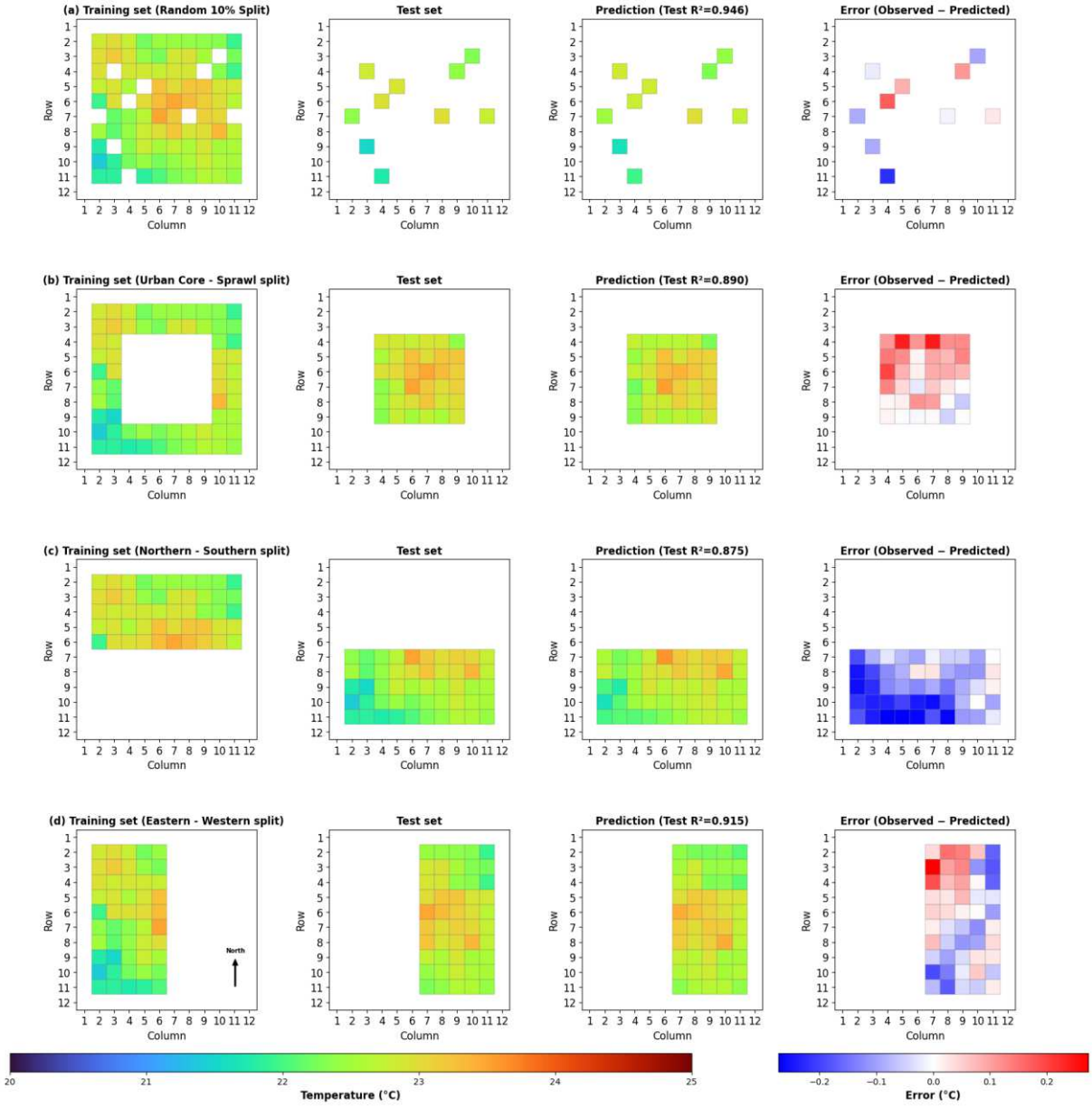


Figure 5: ***Coarse-resolution prediction performance of the Machine Learning Framework (CS-MLF) under four spatial training–prediction configurations.*** *For each configuration, column 1 presents the training data, column 2 the corresponding ADMS-Urban reference field, column 3 the CS-MLF predicted field (Urban-MOSAIKs), and column 4 the prediction error (Observation − Prediction).* ***(a)*** *CS-MLF is trained using a random 10% subsample and applied to predict the remaining 90% of the city.* ***(b)*** *The model is trained in the urban core and applied to predict the urban sprawl.* ***(c)*** *The model is trained in Northern Birmingham and applied to predict the Southern half.* ***(d)*** *The model is trained in Western Birmingham and applied to predict the Eastern half. These configurations assess the ability of CS-MLF to generalise across sparse, contrasting, and spatially separated training domains at coarse resolution.*

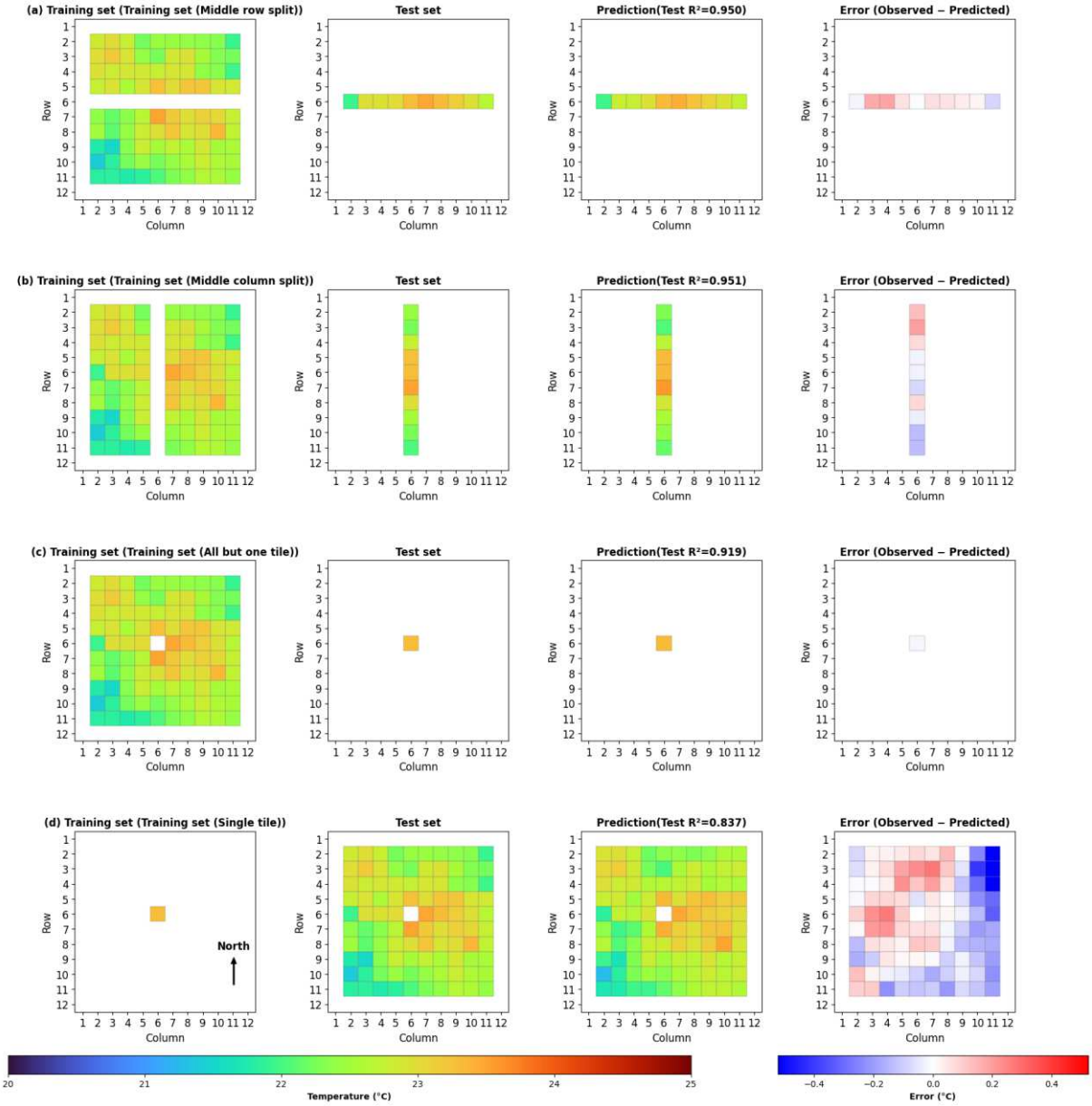


Figure 6: ***Spatial gap-filling capability of the coarse-scale (2.5 km × 2.5 km) CS-MLF framework under progressively challenging data-availability scenarios. (a)*** *All tiles except row 6 are used for training, with row 6 withheld for prediction.* ***(b)*** *All tiles except column 6 are used for training, with column 6 withheld for prediction.* ***(c)*** *A single tile is withheld for prediction, with all remaining tiles used for training, and (d) only a single tile is used for training, while the remaining domain is predicted.*

Coarse scale gap-filling based scenarios: To understand the performance of gap-filling scenarios, four gap-filling configurations were evaluated using CS-MLF, with an additional extreme sparse-data scenario in which only a single tile was used for training and the remainder of the domain was predicted. Withholding the middle row (Figure 6a) yielded test R^2=0.95, withholding the middle column (Figure 6b) produced test R^2=0.95, demonstrating strong reconstruction performance. Predicting the single withheld tile (tile 6, 6) provided a representative case of localised gap filling with $R^2 = 0.92$ (Figure 6c) . Consequently, the result is interpreted as an illustrative measure of single-tile reconstruction accuracy rather than a generalisable performance metric. In the most restrictive scenario, where only one tile was used for training and the remaining domain was predicted, CS-MLF achieved test R^2=0.84 (Figure 6d), demonstrating that the framework retains substantial predictive capability even under extremely limited training-data availability. Overall, these results highlight the strong potential of CS-MLF for coarse-scale spatial gap filling, particularly in data-sparse applications where complete citywide observations are unavailable.

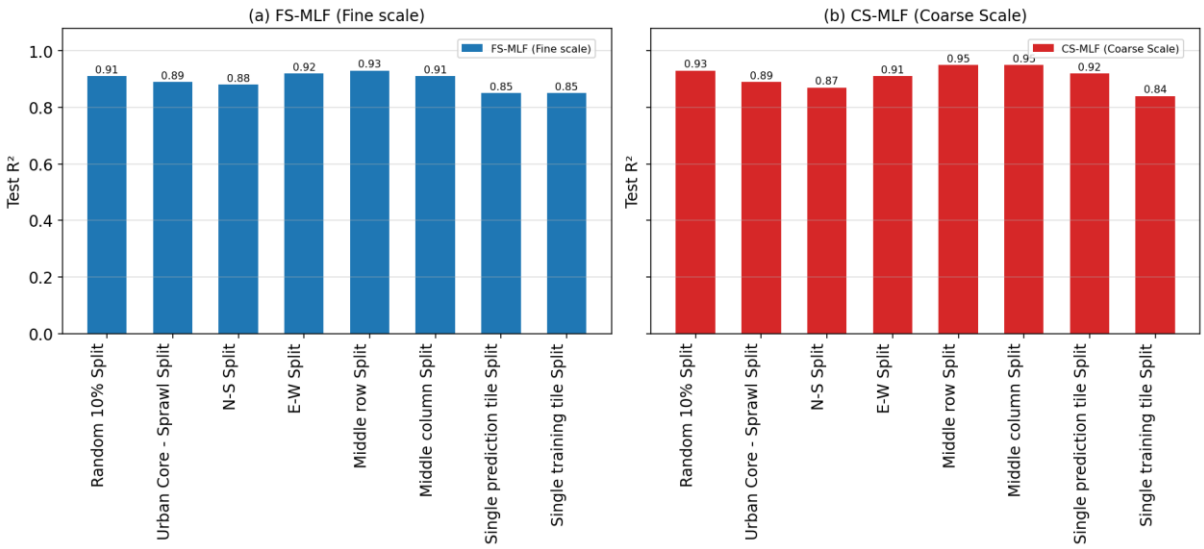


Figure 7: ***Summary of test R^2 performance of FS-MLF (Fine-scale, blue) and CS-MLF (coarse scale, red) across eight spatial validation scenarios:*** *Random 10% sampling, Urban Core–Sprawl, North–South (N–S), West–East (W–E), Middle Row, Middle Column, Single Prediction Tile, and Single Training Tile. Bar height represents the test R^2, with values annotated above each bar.*

Figure 7 summarises the test R^2 values obtained for FS-MLF and CS-MLF across eight spatial training and prediction configurations. Both frameworks achieved high predictive performance across the tested scenarios, with R^2 values generally ranging from 0.84 to 0.95. For FS-MLF, R^2 values ranged from 0.85 to 0.93, with the highest value observed for the Middle row Split ($R^2 = 0.93$) and the lowest values for the Single prediction tile and Single training tile scenarios ($R^2 = 0.85$). Moreover, CS-MLF produced R^2 values between 0.84 and 0.95, with the highest values obtained for the Middle row split and Middle column split scenarios ($R^2 = 0.95$), while the lowest value was observed for the Single training tile split ($R^2 = 0.84$). The Random 10% split also resulted in high R^2 values for both frameworks (0.91 for FS-MLF and 0.93 for CS-MLF). Overall, the results show that both frameworks maintained high predictive performance across different spatial configurations, while variations in R^2 indicate that the spatial arrangement of the training and prediction data influences model performance.

3.3 Assessment of model robustness under varying training data availability

The influence of training-data availability on predictive performance was evaluated by incrementally increasing the training fraction from 10% to 70%, with each configuration assessed over 100 independent iterations to characterise variability arising from data sampling (Figure 8). Both FS-MLF and CS-MLF exhibit pronounced data efficiency, attaining median test R^2 values of approximately 0.85 and 0.86, respectively, using only 10% of the available training data. At this stage, the 25th–75th percentile bands are comparatively wide (approximately 0.79–0.90 for FS-MLF and 0.81–0.90 for CS-MLF), reflecting heightened sensitivity to sample composition when training examples are scarce. Median R^2 increases steadily with training fraction for both frameworks, reaching approximately 0.91–0.92 at 30% and 0.93–0.95 at 50%, before plateauing at 0.94 (FS-MLF) and 0.955 (CS-MLF) at 70% training data. The percentile bands narrow progressively and consistently across this range, collapsing to a near-negligible width by 60–70%, indicating that predictive performance becomes increasingly stable and insensitive to the specific data subset sampled as training data availability increases.

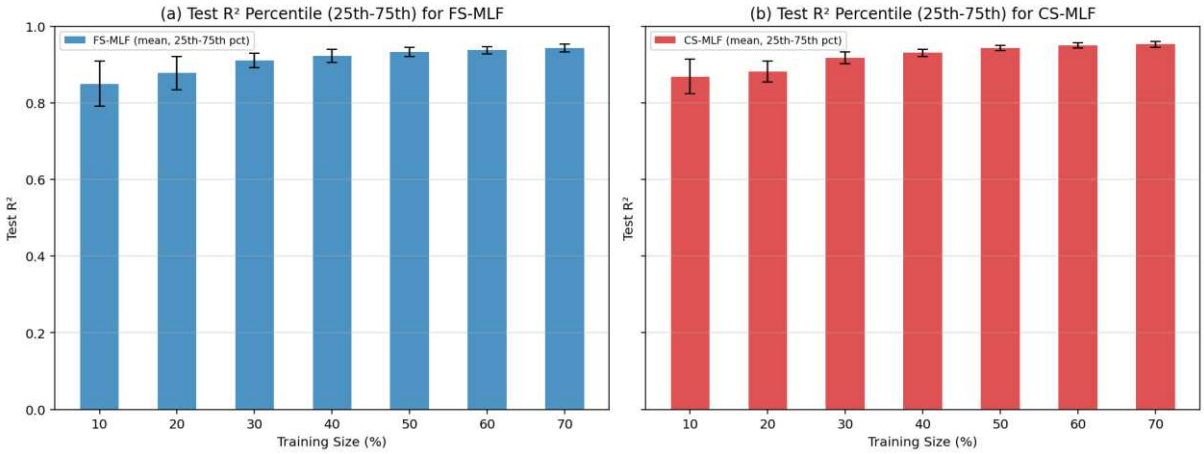


Figure 8: ***Performance of FS-MLF and CS-MLF predictive performance across increasing training data sizes (10–70%) with 100 repetitions****, showing* ***(a)*** *Test R^2 for FS-MLF (blue) and* ***(b)*** *Test R^2 for CS-MLF (red). Solid lines represent mean values, and shaded bands indicate the 25th–75th error percentile range.*

The two frameworks demonstrate broadly consistent predictive performance for 100 repetitions across the range of training data, while serving complementary modelling objectives. CS-MLF generally achieves a marginally higher median (R^2) than FS-MLF across most training-data proportions; however, these differences should not be interpreted as a direct measure of superiority, given that the two frameworks address different prediction tasks. Rather, the results indicate that both approaches maintain stable predictive performance for their respective tasks as the amount of training data increases. Collectively, these results indicate that both frameworks approach near-optimal predictive skill well before the full training dataset is required. Overall, across the evaluated range, median R^2 increases from approximately 0.85 to 0.94 for FS-MLF and from approximately 0.86 to 0.955 for CS-MLF between 10% and 70% training data, corresponding to absolute gains of approximately 0.09 R^2 units (9 percentage points) for both frameworks, or a relative improvement of approximately 10–11% over the 10% training-data baseline.

4. Discussion

In this work, we developed and evaluated two ML frameworks, FS-MLF and CS-MLF, for urban near-surface air temperature modelling. As part of this effort, we introduce U-MOSAIKS, to our knowledge the first adaptation of the MOSAIKS approach (Rolf et al., 2021) to an urban climate context, providing a data-efficient means of encoding spatial patterns of near-surface air temperature. The following sections discuss the performance, transferability, and limitations of both frameworks in light of these results. Rather than representing competing approaches, the two frameworks were designed to address different requirements within urban climate modelling. FS-MLF (Figures 3 and 4) explicitly resolves spatial heterogeneity by operating at the Fine-scale (pixel-level) and incorporating local neighbourhood statistics into the feature space. This enables the reconstruction of continuous near-surface air temperature fields and the identification of intra-urban thermal contrasts, including hotspots and cooling corridors. In contrast, CS-MLF (Figures 5 and 6) compresses each spatial unit into a learned CNN-based embedding and predicts a single temperature value per tile. This provides an emulation solution when only coarse-scale training data is available, at the expense of resolving sub-tile variability (Table 1). Overall, the findings suggest that FS-MLF and CS-MLF provide complementary solutions, with FS-MLF being more suitable for applications requiring detailed characterisation of intra-urban thermal variability, while CS-MLF offers advantages where computational efficiency and broader spatial prediction are prioritised.

The consistent sensitivity of both frameworks to the geographical configuration of training data, most pronounced under directional (Figure 3 and 5) (West–East, North–South) and urban core-sprawl partitioning, suggests that the observed performance variability is not an artefact of model architecture but rather reflects the intrinsic spatial organisation of Birmingham's thermal landscape. This interpretation is consistent with a growing body of evidence that spatial heterogeneity and urban morphology exert a stronger control on model transferability than algorithmic choice alone (Zhong et al., 2025; Briegel et al., 2025; Amnuaylojaroen, 2025; Samad et al., 2026b). Taken together with the gap-filling (Figure 4 and 6) and training-density experiments, these results point to two at least partially independent sources of predictive skill: local spatial proximity, which improves performance as neighbouring training-tile density increases, and spatial representativeness, which allows sparse but geographically distributed samples such as Random 10% configuration to match or exceed spatially concentrated strategies despite minimal

local adjacency. This finding echoes prior work showing that training data composition and sampling design can influence predictive error more strongly than data volume alone (Li & Sharma, 2024) and reframes an apparently technical modelling choice as a substantive methodological consideration for future urban climate applications, where training data are rarely both abundant and evenly distributed across the domain of interest. These results also motivate a shift away from framing FS-MLF and CS-MLF as competing solutions and toward viewing them as complementary components of a broader modelling strategy. FS-MLF offers superior spatial resolution but is relies on fine-scale training data, whereas CS-MLF provides a robust coarse-scale alternative when only coarse-scale training data is avaialbe.

The relationship between training sample size and predictive uncertainty (Figure 8) further underscores the practical viability of both frameworks for data-constrained urban climate applications. Recent studies have demonstrated the potential of machine learning to achieve high predictive accuracy in urban climate applications. However, many approaches rely on relatively large and well-sampled training datasets and predominantly adopt conventional 80/20 train–test splits (Briegel et al., 2025; Amnuaylojaroen, 2025; Hsu et al., 2022; Zumwald et al., 2021; Delgado-Enales et al., 2024; Samad et al., 2026b). Consequently, the extent to which high predictive performance can be maintained with substantially reduced training data remains less explored. In contrast, the present study systematically evaluates training-data availability across 10–70% of the available simulations, providing a more explicit assessment of data requirements and predictive performance under data-constrained conditions.

Notably, both FS-MLF and CS-MLF achieve strong median performance ($R^2 \approx 0.85–0.86$) using as little as 10% of the available training data, demonstrating that neither framework requires extensive labelled datasets to capture the dominant spatial patterns governing near-surface air temperature variability across the domain. While the comparatively wide 25th–75th percentile bands at low training fractions (approximately 0.79–0.90 for FS-MLF and 0.81–0.90 for CS-MLF) indicate greater sensitivity to the specific subset of tiles sampled, this uncertainty diminishes rapidly and consistently as training data availability increases, with both frameworks converging toward stable, narrow-band performance ($R^2 \approx 0.94–0.96$) by 50–70% training data. Importantly, this progressive stabilisation, rather than a delayed onset of skill, suggests that additional training data primarily serves to constrain variance rather than to unlock fundamentally new predictive capability. This shows a distinction with direct practical relevance, given that reliable performance under limited data conditions is often more operationally valuable than marginal gains in peak accuracy. This robustness under sparse-sampling conditions is particularly encouraging for real-world deployment, where extensive, spatially exhaustive training data are frequently unavailable due to limited sensor coverage, computational constraints, or the cost of high-resolution urban morphology datasets. This supports the broader transferability of both FS-MLF and CS-MLF to cities or regions where only partial training coverage can be obtained. Running time and computational resources are important considerations when comparing ML-based emulation with conventional urban climate models. Although extending the simulations to 24 hours would increase the amount of input data and prediction steps, this would not result in a similarly large increase in computational time for the trained ML frameworks. Once trained, the models can generate predictions efficiently without rerunning the underlying physics-based simulations. Thus, even for longer simulation periods, the computational demand is expected to remain substantially lower than that of repeated high-resolution urban climate model simulations.

Taken together, these results demonstrate that FS-MLF and CS-MLF achieve robust and spatially transferable performance when supplied with an equivalently rich set of morphological and land-cover predictors, with predictive skill remaining consistently high ($R^2 > 0.83$) even under the most spatially demanding single-tile train-to-rest configuration. This confirms that the frameworks are well suited to spatial emulation of ADMS-Urban output, provided the full complement of informative input datasets is available, a condition readily met in most urban climate modelling contexts since the required predictors (albedo, normalised building volume, terrain elevation, thermal admittance, surface resistance to evaporation, and surface roughness length) are routinely derivable from standard land-use, terrain, and building datasets. It should be noted that both frameworks are currently conFigured for spatial rather than temporal prediction, having been trained and evaluated on a single representative snapshot (11:00, 26 August 2019) whose underlying ADMS-Urban output was independently validated against Landsat 8-derived land surface temperature, UK Met Office and Weather Underground observations, and complementary satellite datasets (Zhong et al., 2025). However, the workflow is, in principle, readily extensible to other hours of the day given equivalent inputs and reference data, a promising direction for future work. Where the input feature set is more limited, for instance, when only albedo, normalised building volume, and terrain elevation are available, performance declines accordingly (Supplementary Figures S2 and S3). This effect is most pronounced under the more spatially demanding splits, for the Fine-scale configuration, R^2 falls from 0.931 to 0.436 under the row split and from 0.850 to 0.073 under the single-tile-to-rest split, while the Coarse Scale configuration shows an analogous decline from 0.950 to 0.649 (middle row) and 0.837 to 0.190 (single tile to rest). The comparatively modest reduction observed for the middle-column splits (0.951 to 0.892 and 0.912 to 0.429) suggests that residual spatial overlap and land-cover heterogeneity between training and test regions can

partially offset the loss of thermophysical predictors. This reflects that the omitted parameters (thermal admittance, surface resistance to evaporation, surface roughness length) carry thermophysical information on heat storage, evaporative cooling, and aerodynamic mixing that is not recovered from morphological and terrain variables alone. Rather than pointing to a weakness in the modelling approach, this sensitivity highlights the value of these land-cover-related parameters and reinforces that, wherever such datasets are available, FS-MLF and CS-MLF provide more complete information. Moreover, Meyer et al. (2019, 2022) confirm that predictor informativeness, rather than algorithmic choice, is often the dominant factor governing spatial prediction accuracy, reinforcing that, wherever informative morphological and land-cover predictors are retained, FS-MLF and CS-MLF offer an efficient and reliable route to high-resolution spatial gap-filling and predictive characterisation of urban temperature fields.

5. Summary and future scope

This study makes three principal contributions to urban climate emulation. First, it presents, to our knowledge, the first translation of the MOSAIKS approach into an urban climate context, introducing U-MOSAIKS as a data-efficient means of encoding spatial patterns of near-surface air temperature from satellite and morphological imagery, extending a framework previously applied mainly in socioeconomic and land-cover mapping. Building on this, two complementary frameworks, FS-MLF and CS-MLF, were developed and evaluated for spatial emulation of near-surface air temperature in Birmingham: FS-MLF, operating at the pixel level with local neighbourhood information, captured fine-scale thermal heterogeneity, while CS-MLF used tile-level CNN embeddings to provide a more computationally efficient and comparatively robust representation at the expense of sub-tile variability. Second, both frameworks proved valuable and efficient under limited data conditions. This retains strong predictive skill (median test $R^2 \approx 0.85–0.86$) with as little as 10% of the available training data, with performance stabilising at $R^2 \approx 0.94–0.96$ once 50–70% of the data were used. Third, the frameworks demonstrated that a single modelling approach can serve a dual purpose, delivering both high-accuracy spatial prediction and reliable gap-filling of missing urban data, sustaining R^2 values from 0.83 up to 0.95 across most spatial cross-validation configurations when informative morphological and land-cover predictors were retained.

Taken together, these findings show that ML-based emulation can provide useful, high-resolution urban climate information at a fraction of the computational and data cost of physics-based modelling alone, and that the range of spatial training–predicting configurations tested here supports the frameworks' applicability for rapid, spatially explicit urban climate assessment. The present results, however, represent a spatial case study based on a single validated snapshot (11:00, 26 August 2019). Future work should therefore extend the frameworks from spatial emulation towards spatiotemporal prediction, evaluating performance across different times of day, weather conditions, and seasons, and testing transferability across multiple cities with contrasting urban forms and climatic contexts. Further investigation of adaptive or optimised sampling strategies could help identify the minimum number and spatial arrangement of training observations required for reliable predictions. Collectively, such developments would strengthen the potential of ML frameworks to generate operational urban heat mapping, climate adaptation assessment, and planning decisions in data-constrained urban environments.

Reference

Amnuaylojaroen, T. (2025). Advancements and challenges of artificial intelligence in climate modeling for sustainable urban planning. *Frontiers in Artificial Intelligence*, *8*. https://doi.org/10.3389/frai.2025.1517986

Araos, M., Berrang-Ford, L., Ford, J. D., Austin, S. E., Biesbroek, R., & Lesnikowski, A. (2016). Climate change adaptation planning in large cities: A systematic global assessment. *Environmental Science & Policy*, *66*, 375–382. https://doi.org/10.1016/j.envsci.2016.06.009

Barlage, M., Miao, S., & Chen, F. (2016). Impact of physics parameterizations on high-resolution weather prediction over two Chinese megacities. *Journal of Geophysical Research: Atmospheres*, *121*(9), 4487–4498. https://doi.org/10.1002/2015JD024450

Biggart, M., Stocker, J., Doherty, R. M., Wild, O., Carruthers, D., Grimmond, S., Han, Y., Fu, P., & Kotthaus, S. (2021). Modelling spatiotemporal variations of the canopy layer urban heat island in Beijing at the neighbourhood scale. *Atmospheric Chemistry and Physics*, *21*(17), 13687–13711. https://doi.org/10.5194/acp-21-13687-2021

Briegel, F., Makansi, O., Brox, T., Matzarakis, A., & Christen, A. (2023). Modelling long-term thermal comfort conditions in urban environments using a deep convolutional encoder-decoder as a computational shortcut. *Urban Climate*, *47*, 101359. https://doi.org/10.1016/j.uclim.2022.101359

Briegel, F., Schrodi, S., Sulzer, M., Brox, T., Pinto, J. G., & Christen, A. (2025). Deep learning enables city-wide climate projections of street-level heat stress. *Urban Climate*, *62*, 102564. https://doi.org/10.1016/j.uclim.2025.102564

Bruse, M., Simon, H., & Sinsel, T. (n.d.). *Development and implementation of a high-resolution dynamical wall and roof model for ENVI-met Part 2: Vegetated walls and roofs*.

Bassett, R., Cai, X., Chapman, L., Heaviside, C., Thornes, J.E., Muller, C.L., Young, D.T. and Warren, E.L. (2016), Observations of urban heat island advection from a high-density monitoring network. Q.J.R. Meteorol. Soc., 142: 2434-2441. https://doi.org/10.1002/qj.2836

Carmeliet, J., Nevers, C., Strebel, D., Rubin, A., Kubilay, A., & Derome, D. (2026). Changing Urban Climate and Its Mitigation: An Explicit Urban Climate Simulation Approach. In B. Nagy & Z. Szalay (Eds.), *Proceedings of CESBP 2025—6th Central European Symposium on Building Physics* (pp. 414–422). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-14019-7_30

Chapman, L., Bell, C., & Bell, S. (2016). Can the crowdsourcing data paradigm take atmospheric science to a new level?: A case study of the Urban Heat Island of London quantified using Netatmo weather stations. International Journal of Climatology. https://doi.org/10.1002/joc.4940

Delgado Enales, I. (2024). Advanced optimization and data modeling techniques to improve accesibility and thermal comfort in urban plannning. https://doi.org/10/73521

Demuzere, M., Bechtel, B., Middel, A., & Mills, G. (2019). Mapping Europe into local climate zones. PloS One, 14(4), e0214474.

Gu, J., Kim, D., Jun, C., & Son, S. (2025). Quantitative assessment of factors that influence heat vulnerability in residential areas using machine learning and unmanned aerial vehicle. *City and Environment Interactions*. https://doi.org/10.1016/j.cacint.2025.100214

Hamilton, I., Stocker, J., Evans, S., Davies, M., & Carruthers, D. (2014). The impact of the London Olympic Parkland on the urban heat island. Journal of Building Performance Simulation, 7(2), 119–132. https://doi.org/10.1080/19401493.2013.791343

Hashad, K., Gu, J., Yang, B., Rong, M., Chen, E., Ma, X., & Zhang, K. M. (2021). Designing roadside green infrastructure to mitigate traffic-related air pollution using machine learning. *The Science of the Total Environment*, *773*, 144760. https://doi.org/10.1016/j.scitotenv.2020.144760

Heo, Y., Setyantho, G. R., & Hong, T. (2026). Toward a new paradigm for urban climate modelling: Challenges and opportunities. *Journal of Building Performance Simulation*, *19*(3), 395–402. https://doi.org/10.1080/19401493.2025.2540925

Hertwig, D., Ng, M., Grimmond, S., Vidale, P. L., & McGuire, P. C. (2021). High-resolution global climate simulations: Representation of cities. *International Journal of Climatology*, *41*(5), 3266–3285.

Hsu, A., Wang, X., Tan, J., Toh, W., & Goyal, N. (2022). Predicting European cities' climate mitigation performance using machine learning. Nature Communications, 13(1), 7487. https://doi.org/10.1038/s41467-022-35108-5

Lean, H. W., Theeuwes, N. E., Baldauf, M., Barkmeijer, J., Bessardon, G., Blunn, L., Bojarova, J., Boutle, I. A., Clark, P. A., Demuzere, M., Dueben, P., Frogner, I.-L., de Haan, S., Harrison, D., Heerwaarden, C. van, Honnert, R., Lock, A., Marsigli, C., Masson, V., … Yang, X. (2024). The hectometric modelling challenge: Gaps in the current state of the art and ways forward towards the implementation of 100-m scale weather and climate models. *Quarterly Journal of the Royal Meteorological Society*, *150*(765), 4671–4708. https://doi.org/10.1002/qj.4858

Leiter, T. (2021). Do governments track the implementation of national climate change adaptation plans? An evidence-based global stocktake of monitoring and evaluation systems. *Environmental Science & Policy*, *125*, 179–188. https://doi.org/10.1016/j.envsci.2021.08.017

Li, J., Wang, Y., Zhai, Z., & Ding, Y. (2026). Reconstructed 3D temperature field of microclimate by deep neural network (DNN) based on high-fidelity numerical simulation. *Indoor and Built Environment*, *35*(5), 631–647. https://doi.org/10.1177/1420326X261418727

Lu, Y., Zhou, X.-H., Xiao, H., & Li, Q. (2023). Using Machine Learning to Predict Urban Canopy Flows for Land Surface Modeling. Geophysical Research Letters, 50(1), e2022GL102313. https://doi.org/10.1029/2022GL102313

Luo, T., & Chen, M. (2025). Advancements in supervised machine learning for outdoor thermal comfort: A comprehensive systematic review of scales, applications, and data types. *Energy and Buildings*, *329*, 115255. https://doi.org/10.1016/j.enbuild.2024.115255

Maronga, B., Gross, G., Raasch, S., Banzhaf, S., Forkel, R., Heldens, W., Kanani-Sühring, F., Matzarakis, A., Mauder, M., Pavlik, D., Pfaffenrott, J., Schubert, S., Seckmeyer, G., Sieker, H., & Winderlich, K. (2019). Development of a new urban climate model based on the model PALM - Project overview, planned work, and first achievements. *Meteorologische Zeitschrift*, 1–15.

Masson, V., Grimmond, C. S. B., & Oke, T. R. (2002). *Evaluation of the Town Energy Balance (TEB) Scheme with Direct Measurements from Dry Districts in Two Cities*. https://journals.ametsoc.org/view/journals/apme/41/10/1520-0450_2002_041_1011_eotteb_2.0.co_2.xml

Masson, V., Heldens, W., Bocher, E., Bonhomme, M., Bucher, B., Burmeister, C., De Munck, C., Esch, T., Hidalgo, J., Kanani-Sühring, F., Kwok, Y.-T., Lemonsu, A., Lévy, J.-P., Maronga, B., Pavlik, D., Petit, G., See, L., Schoetter, R., Tornay, N., … Zeidler, J. (2020). City-descriptive input data for urban climate models: Model requirements, data sources and challenges. *Urban Climate*, *31*, 100536. https://doi.org/10.1016/j.uclim.2019.100536

Mäkinen, K., Sorvali, J., Lipsanen, A., & Hildén, M. (2020). Implementation of Finland's National Climate Change Adaptation Plan 2022: A Mid-term Evaluation. fi=Maa- ja metsätalousministeriö|sv=Jord- och skogsbruksministeriet|en=Ministry of Agriculture and Forestry|. https://julkaisut.valtioneuvosto.fi/handle/11111/1483

Maggiotto, G., Miani, A., Rizzo, E., Castellone, M. D., & Piscitelli, P. (2021). Heat waves and adaptation strategies in a mediterranean urban context. Environmental Research, 197, 111066. https://doi.org/10.1016/j.envres.2021.111066

Michau, Y., Lemonsu, A., Lucas-Picher, P., & Caillaud, C. (2023). Evaluation of the Urban Heat Island of 12 cities of France in a high-resolution regional climate model simulation. *Urban Climate*, *47*, 101386. https://doi.org/10.1016/j.uclim.2022.101386

Miniandi, N. D., Jamal, M. H., Muhammad, M. K. I., Sharrar, L., & Shahid, S. (2025). Machine Learning in Modeling Urban Heat Islands: A Data-Driven Approach for Kuala Lumpur. *Earth Systems and Environment*, *9*(2), 1037–1060. https://doi.org/10.1007/s41748-025-00584-4

Morabito, M., Messeri, A., Noti, P., Casanueva, A., Crisci, A., Kotlarski, S., Orlandini, S., Schwierz, C., Spirig, C., Kingma, B. R. M., Flouris, A. D., & Nybo, L. (2019). An Occupational Heat–Health Warning System for Europe: The HEAT-SHIELD Platform. *International Journal of Environmental Research and Public Health*, *16*(16), Article 16. https://doi.org/10.3390/ijerph16162890

Oke, T. R., Mills, G., Christen, A., & Voogt, J. A. (2017). *Urban Climates*. Cambridge University Press.

Office for National Statistics (2023) Towns and cities, characteristics of built-up areas, England and Wales: Census 2021. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/housing/articles/townsandcitiescharacteristicsofbuiltupareasenglandandwales/census2021

Patel, N., Jänicke, B., Burghardt, R., Vulova, S., & Otto, F. (2025). Assessing Progress in Urban Climate Adaptation: A Review of Indicators for Heat- and Water-Sensitive Urban Development. *Climate Resilience and Sustainability*, *4*(1), e70009. https://doi.org/10.1002/cli2.70009

Pinheiro, M. J. de A., Coelho, C. M. T., Bevilaqua, D. V., Domingues, L. C. S. M., & Almeida, R. dos S. de. (2026). Climate Change and Sustainability: Challenges for Cultural Heritage and Local Communities in Rio de Janeiro. In *The Routledge Handbook on Cultural Heritage and Climate Justice*. Routledge.

Pörtner, H.-O., Roberts, D., Tignor, M., Poloczanska, E., Mintenbeck, K., Alegría, A., Craig, M., Langsdorf, S., Löschke, S., Möller, V., Okem, A., Rama, B., Belling, D., Dieck, W., Götze, S., Kersher, T., Mangele, P., Maus, B., Mühle, A., & Weyer, N. (2022). *Climate Change 2022: Impacts, Adaptation and Vulnerability Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. https://doi.org/10.1017/9781009325844

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 (pp. 234–241). Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28

Rolf, E., Proctor, J., Carleton, T., Bolliger, I., Shankar, V., Ishihara, M., Recht, B., & Hsiang, S. (2021). A generalizable and accessible approach to machine learning with global satellite imagery. *Nature Communications*, *12*(1), 4392. https://doi.org/10.1038/s41467-021-24638-z

Samad, N., Farnaghi, M., Ostermann, F. O., & Zurita-Milla, R. (2026). A systematic literature review of machine learning and deep learning for urban heat island modelling. *Discover Cities*, *3*(1), 43. https://doi.org/10.1007/s44327-026-00223-1

Stewart, I. D., & Oke, T. R. (2012). *Local Climate Zones for Urban Temperature Studies*. https://doi.org/10.1175/BAMS-D-11-00019.1

Syeda, A. Q., Castillo-Villar, K., & Alaeddini, A. (2025). Sustainable Urban Heat Island Mitigation Through Machine Learning: Integrating Physical and Social Determinants for Evidence-Based Urban Policy. *Sustainability*. https://doi.org/10.3390/su17157040

Tyler, C. J., Reeve, T., Hodges, G. J., & Cheung, S. S. (2016). The Effects of Heat Adaptation on Physiology, Perception and Exercise Performance in the Heat: A Meta-Analysis. *Sports Medicine*, *46*(11), 1699–1724. https://doi.org/10.1007/s40279-016-0538-5

Tiwari, A., Kumar, P., Kalaiarasan, G., & Ottosen, T.-B. (2021). The impacts of existing and hypothetical green infrastructure scenarios on urban heat island formation. Environmental Pollution, 274, 115898. https://doi.org/10.1016/j.envpol.2020.115898

Unger, J., Skarbit, N., & Gál, T. (2018). Evaluation of outdoor human thermal sensation of local climate zones based on long-term database. *International Journal of Biometeorology*, *62*(2), 183–193. https://doi.org/10.1007/s00484-017-1440-z

Vahid, R., & Aly, M. (2025). A Comprehensive Systematic Review of Machine Learning Applications in Assessing Land Use/Cover Dynamics and Their Impact on Land Surface Temperatures. *Urban Science*. https://doi.org/10.3390/urbansci9070234

Villefranque, N., Hourdin, F., d'Alençon, L., Blanco, S., Boucher, O., Caliot, C., Coustet, C., Dauchet, J., El Hafi, M., Eymet, V., Farges, O., Forest, V., Fournier, R., Gautrais, J., Masson, V., Piaud, B., & Schoetter, R. (2022). The "teapot in a city": A paradigm shift in urban climate modeling. *Science Advances*, *8*(27), eabp8934. https://doi.org/10.1126/sciadv.abp8934

Wang, C., Zhao, Y., Li, Q., Wang, Z.-H., & Fan, J. (2025). Ultrafine-Resolution Urban Climate Modeling: Resolving Processes Across Scales. *Journal of Advances in Modeling Earth Systems*, *17*(6), e2025MS005053. https://doi.org/10.1029/2025MS005053

Wu, Y., Teufel, B., Sushama, L., Belair, S., & Sun, L. (2021). Deep Learning-Based Super-Resolution Climate Simulator-Emulator Framework for Urban Heat Studies. *Geophysical Research Letters*, *48*(19), e2021GL094737. https://doi.org/10.1029/2021GL094737

Yang, B., Zhang, Y., & Qian, Y. (2012). Simulation of urban climate with high-resolution WRF model: A case study in Nanjing, China. *Asia-Pacific Journal of Atmospheric Sciences*, *48*(3), 227–241. https://doi.org/10.1007/s13143-012-0023-5

Zhang, M., Tan, S., Zhang, C., & Chen, E. (2024). Machine learning in modelling the urban thermal field variance index and assessing the impacts of urban land expansion on seasonal thermal environment. *Sustainable Cities and Society*. https://doi.org/10.1016/j.scs.2024.105345

Zhong, J., Lu, Y., Stocker, J., Hamilton, V., & Johnson, K. (2025). Modelling temperature variations in Birmingham, UK at the neighbourhood scale: Using ADMS-Urban climate model. *Sustainable Cities and Society*, *131*, 106666. https://doi.org/10.1016/j.scs.2025.106666

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.docx