

Physics Guided AI for Urban Heat Mitigation Planning

Abdul Ghafoor Nizamani

`abdul.g.nizamani@hiof.no`

Østfold University of Applied Sciences

Dung Thi Vu

Phenikaa University

Rafeeqe Ahmed Nizamani

Wuhan University of Technology

Mostafa Aliyari

Østfold University of Applied Sciences

Cunrong Li

Wuhan University of Technology

Article

Keywords:

Posted Date: October 8th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10949253/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Physics Guided AI for Urban Heat Mitigation Planning

Abdul Ghafoor Nizamani^{a,b,*}, Dung Thi Vu^{c,d}, Rafeeqe Ahmed Nizamani^e,
Mostafa Aliyari^a, Cunrong Li^e

^a*Faculty of Computer Science, Engineering and Economics, Østfold University of Applied Sciences, BRA Veien 4, Halden, 1757, NO, Norway*

^b*Department of Civil Engineering, Delft University of Technology, Delft, 2628 CD, Netherlands*

^c*Department of Mathematics, Faculty of Fundamental Sciences, Phenikaa University, Nguyen Trac Street, Duong Noi Ward, Hanoi, 12116, Vietnam*

^d*Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, Viale Carlo Berti Pichat, 6/2, Bologna, 40127 BO, Italy*

^e*School of Mechanical and Electronic Engineering, Wuhan University of Technology, Wuhan, Hubei, P.R. China*

Abstract

Urban heat-mitigation planning requires ways to compare intervention locations when only part of the candidate set can be treated. We developed a time-dependent physics-guided model of land surface temperature for Oslo using Landsat, ERA5, and OpenStreetMap data, and used the fitted model to screen green-roof surface-property scenarios. Across the three fitted Landsat scenes, the Physics informed neural network achieved RMSE values of 0.2549–0.3959 °C, MAE values of 0.1961–0.3043 °C, and R^2 values of 0.9638–0.9676, although some data-driven models achieved lower fitted errors. Prescribing higher NDVI and emissivity at hotspot building pixels produced maximum local modelled LST reductions of 3.49–4.28 °C, while most affected pixels reported smaller responses. Under partial deployment, mean response depended on acquisition date, spatial group, and prioritisation strategy; no single proximity-based rule produced the more negative mean response across all conditions. The analysis supports comparative thermal screening rather than prediction of complete green-roof performance.

Keywords: green roofs, land surface temperature, physics-informed neural networks, urban heat, urban planning

*Corresponding author

Email address: abdul.g.nizamani@hiof.no (Abdul Ghafoor Nizamani)

1. Introduction

Urban heat is shaped by interactions among regional climate, urban form, surface materials, vegetation, water, and anthropogenic heat release [1–4]. These interactions produce substantial spatial heterogeneity within cities, making heat mitigation not only a question of reducing temperature but also of deciding where interventions should be placed. Green infrastructure has long been considered part of urban climate adaptation [5], while subsequent work has shown that the thermal effects of green infrastructure depend on intervention type, spatial configuration, surrounding urban form, and climatic setting [6, 7]. Evidence from Oslo also demonstrates spatial associations between green infrastructure, urban thermal conditions, and heat-related human exposure [8]. For urban planning, these findings make the spatial allocation of heat-mitigation measures an important part of the intervention problem.

Green roofs provide one possible means of introducing vegetation into densely built areas where ground-level space is limited. Their thermal effects arise through changes in surface radiative properties, evapotranspiration, heat storage, and sensible and latent heat exchange [9–11]. Modelling and experimental studies have nevertheless illustrated that their thermal response is not fixed: vegetation, substrate properties, substrate depth, moisture availability, irrigation, climate, and surrounding urban conditions can affect performance [12–14]. Recent studies continue to examine the potential of roof greening at building and larger urban scales [15, 16]. Consequently, treating every candidate roof as providing the same thermal response would overlook both spatial context and differences in the thermal state of the locations being considered.

This becomes particularly important when intervention coverage is limited. Urban greening studies increasingly distinguish between the total amount of vegetation and its spatial allocation. At the micro-scale, the configuration and location of green infrastructure affect the resulting heat-mitigation response [6]. Across 200 cities, Massaro *et al.* [17] described that targeting areas with high exposure to LST extremes required less additional vegetation to achieve the same exposure reduction than uniform treatment. Broader evidence on green and blue infrastructure likewise indicates that cooling effectiveness depends on characteristics of the intervention and its environmental and climatic context [7]. These studies provide a planning basis for examining coverage and prioritisation together: if only part of an eligible candidate set can be treated, the order in which locations are selected may influence the thermal response obtained at a given deployment level.

Satellite-derived land surface temperature (LST) provides one means of resolving the spatial thermal differences relevant to such screening. Thermal remote sensing has been used extensively to characterise urban surface-temperature patterns and

surface urban heat islands [18, 19]. Its spatial coverage is particularly useful where the aim is to distinguish thermally different parts of the urban surface. LST, however, is a surface variable and should not be treated as interchangeable with near-surface air temperature, mean radiant temperature, thermal comfort, or human heat exposure [20, 21]. The relationship between LST and air temperature can also vary with season and local atmospheric conditions [21]. In the present study, Landsat-derived LST is therefore used as a surface thermal screening variable for identifying hotspot building pixels and comparing modelled responses among intervention configurations, rather than as a direct measure of human thermal risk.

Moving from hotspot identification to intervention comparison requires a model that can represent changes in the thermal field after selected surface inputs are modified. Urban thermal modelling encompasses approaches with different levels of physical representation and data dependence [22, 23]. Physically based urban models explicitly represent surface-atmosphere and urban energy-balance processes [24, 25], while statistical and machine-learning methods can learn spatial and temporal thermal relationships from observational data [26–28]. Comparisons of urban energy-balance models have also displayed that model formulations differ in their representation of urban thermal processes [29]. More generally, combining data-driven learning with process understanding has been identified as an important direction for Earth-system modelling where neither observational learning nor physical representation alone resolves all modelling requirements [30].

Physics-informed neural networks (PINNs) provide one such hybrid approach by incorporating residuals from prescribed differential equations into the neural-network training objective alongside observational information [31, 32]. The physical residual acts as an additional model constraint rather than guaranteeing physical correctness, and the balance between data and physical regularisation can materially affect model behaviour [32, 33]. PINN-based methods have recently been applied to building and urban thermal modelling [34–36]. Their relevance to the present study is not predicated on achieving lower fitted error than every data-driven model. Instead, the physics-guided formulation is used as a common thermal representation through which prescribed surface-property changes and alternative spatial allocations can be compared.

The green-roof scenarios considered here remain deliberately narrower than a complete green-roof energy-balance simulation. Process-based studies represent interactions among vegetation, substrate, moisture, radiation, and heat exchange [9, 12], and experimental evidence illustrates that substrate depth, vegetation, and irrigation can alter green-roof thermal performance [13]. In the present analysis, green-roof-associated changes are represented by prescribed NDVI and emissivity values at

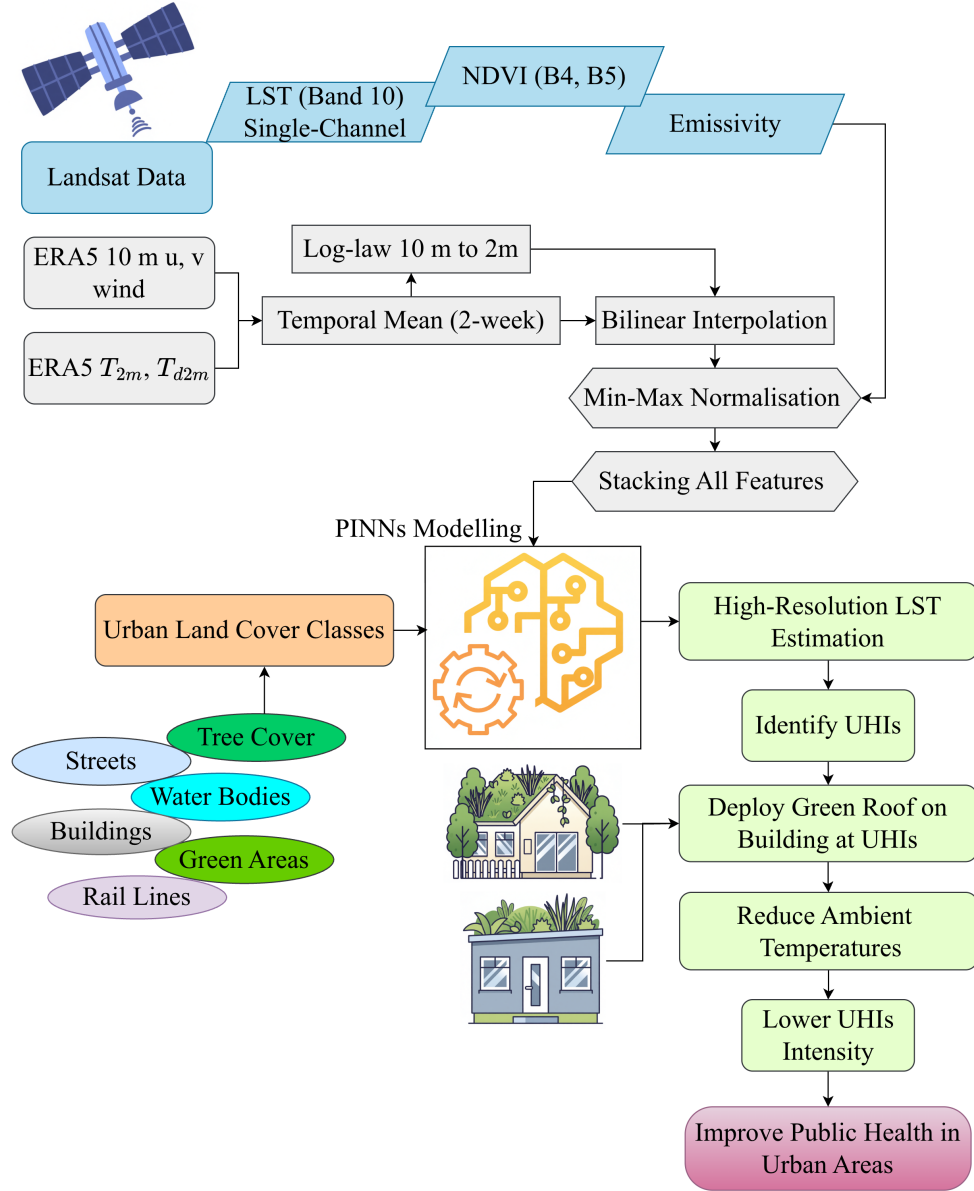


Figure 1: Overall methodological framework, from satellite, meteorological, and urban spatial data processing to PINN fitting, model comparison, and counterfactual green roof scenario analysis.

selected building pixels. Substrate moisture, irrigation, explicit evapotranspiration, roof thermal mass, and meteorological feedbacks are not modified. The resulting temperature differences therefore represent responses of the fitted thermal model to

prescribed surface-property changes rather than predictions of complete green-roof performance.

The analysis uses three clear-sky Landsat acquisitions from May and June 2025 for Oslo, together with ERA5 meteorological fields and OpenStreetMap urban features. A time-dependent PINN constrained by an advection–diffusion-type residual is fitted to the three Landsat-derived LST fields. Hotspot building pixels are identified from the observed Landsat LST distributions, and the fitted model is then used to compare prescribed surface-property changes across different candidate groups, deployment fractions, and spatial ranking rules. The overall methodological workflow is summarised in Figure 1.

The central planning question is: *when intervention can cover only part of the eligible hotspot-building set, how do deployment level and prioritisation strategy affect the modelled LST response, and how consistent are those responses across the observed thermal states?* The study addresses this question through three objectives: (i) to fit and evaluate the physics-guided LST model across the three Landsat scenes, with the resulting statistics interpreted as fitted-scene agreement rather than independent predictive validation; (ii) to evaluate prescribed green-roof-associated surface-property changes at hotspot building pixels; and (iii) to compare modelled thermal responses across deployment coverage levels and spatial prioritisation strategies.

2. Results

2.1. Model Fit and Comparative behaviour

Model behaviour was examined across the three Landsat acquisition dates before the counterfactual analysis. The Full PINN was assessed alongside the data-driven comparison models and reduced PINN variants using RMSE, MAE, R^2 , and the predicted LST range. Across the three fitted scenes, the Full PINN had RMSE values between 0.2549 and 0.3959 °C and R^2 values between 0.9638 and 0.9676, although several comparison models produced lower fitted errors. Because the same Landsat scenes were used during model fitting, the statistics reported here describe fitted-scene agreement rather than independent predictive performance. Table 1 gives the complete numerical comparison.

Table 1: Fitted-scene comparison of the Full PINN, data-driven models, and reduced PINN variants for the three Landsat acquisition dates. RMSE and MAE are reported in degrees Celsius. A dash indicates that MAE was not retained for that specific model configuration in the archived results.

Date	Model	RMSE	MAE	R^2	Predicted LST ($^{\circ}\text{C}$)
19 May	<i>Observed LST ($^{\circ}\text{C}$)</i>	–	–	–	<i>23.3–38.3</i>
	Full PINN	0.3758	0.2890	0.9638	23.8–37.7
	MLP	0.5600	0.2718	0.9196	23.4–35.7
	LSTM	0.9244	–	0.7810	24.8–35.5
	ConvLSTM	0.1666	–	0.9929	23.3–38.5
	U-Net	0.1006	0.0549	0.9974	23.3–38.4
	GCN	0.3830	0.2674	0.9624	23.5–38.2
	GLSTM	0.6698	0.5155	0.8850	23.7–35.2
	RF	0.7298	0.5466	0.8635	23.7–38.1
	XGBoost	0.0162	0.0125	0.9999	23.3–38.3
	No Advection	0.5595	0.4280	0.9198	24.0–36.9
	No Q	0.3086	0.2316	0.9756	23.5–37.0
	No T_{2m}/T_{d2m}	0.5758	0.4411	0.9150	23.9–36.5
4 June	<i>Observed LST ($^{\circ}\text{C}$)</i>	–	–	–	<i>16.2–25.6</i>
	Full PINN	0.2549	0.1961	0.9663	16.2–25.6
	MLP	0.4385	0.1987	0.9002	16.1–24.1
	LSTM	0.4614	–	0.8896	16.5–24.3
	ConvLSTM	0.0897	–	0.9958	16.3–25.8
	U-Net	0.5902	0.3472	0.8192	16.2–26.0
	GCN	0.2064	0.1502	0.9779	16.1–25.8
	GLSTM	0.3217	0.2380	0.9463	16.1–25.1
	RF	0.5550	0.4106	0.8402	16.7–25.5
	XGBoost	0.0120	0.0091	0.9999	16.2–25.6
	No Advection	0.4273	0.3373	0.9053	16.3–24.9
	No Q	0.2125	0.1597	0.9766	15.9–25.6
	No T_{2m}/T_{d2m}	0.3412	0.2665	0.9396	16.2–25.4
20 June	<i>Observed LST ($^{\circ}\text{C}$)</i>	–	–	–	<i>21.2–37.7</i>
	Full PINN	0.3959	0.3043	0.9676	21.4–37.5
	MLP	0.6720	0.3056	0.9066	21.0–34.6
	LSTM	0.5859	–	0.9290	21.4–36.8
	ConvLSTM	0.1220	–	0.9969	21.3–37.9
	U-Net	1.3673	0.8872	0.6134	21.9–38.1
	GCN	0.3303	0.2354	0.9774	21.2–38.2
	GLSTM	0.4305	0.3178	0.9617	21.1–37.2
	RF	0.7780	0.5784	0.8748	21.6–37.5
	XGBoost	0.0181	0.0137	0.9999	21.2–37.7
	No Advection	0.5912	0.4488	0.9277	21.7–36.9
	No Q	0.3477	0.2563	0.9750	20.9–37.3
	No T_{2m}/T_{d2m}	0.5154	0.3951	0.9451	21.7–37.7

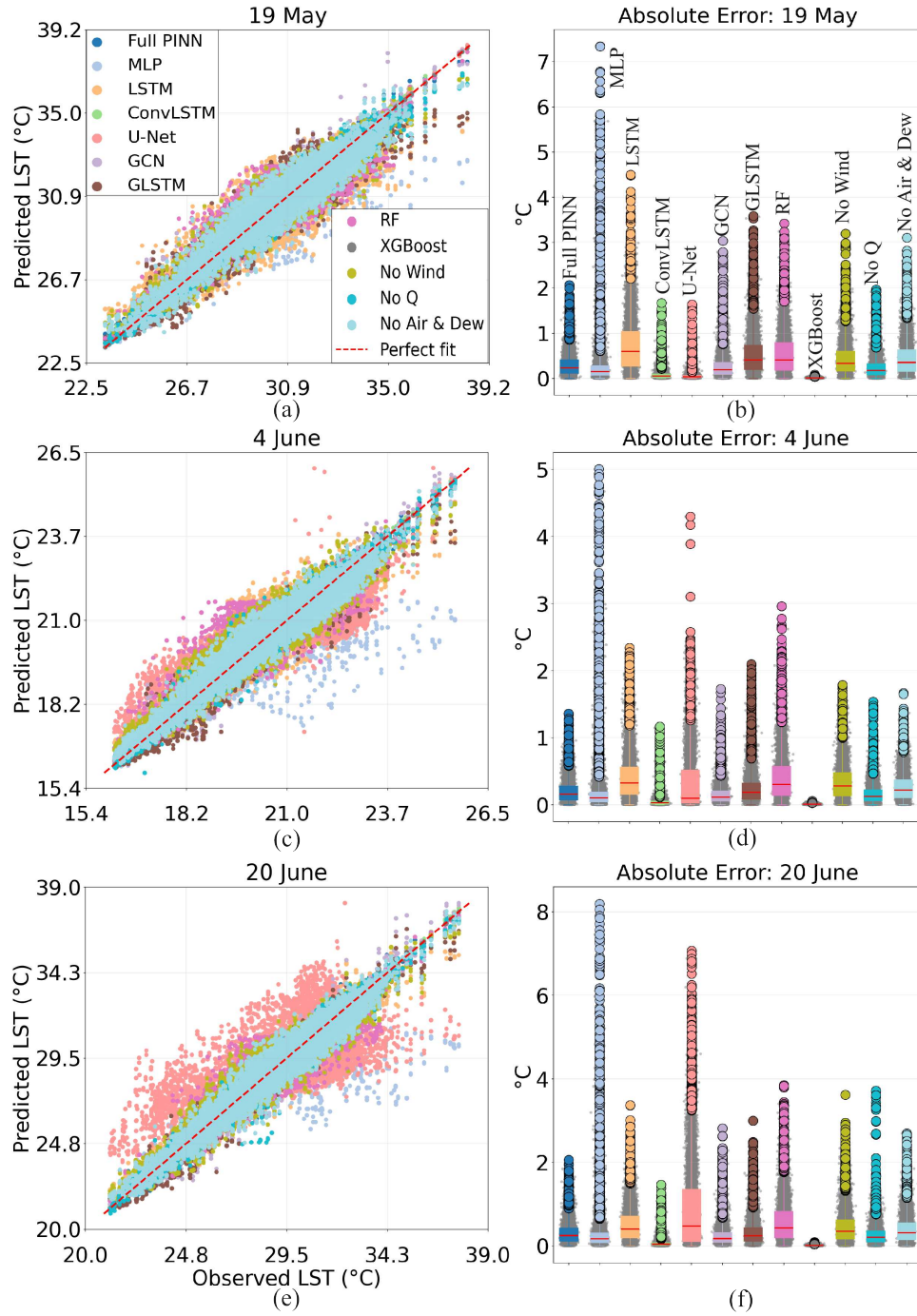


Figure 2: Predicted versus Landsat-derived LST and absolute-error distributions for 19 May and 4 June. Panels (a) and (c) display predicted versus observed LST with the 1:1 reference line; panels (b) and (d) illustrate the corresponding absolute-error distributions.

Performance varied considerably by architecture and acquisition date. ConvLSTM had lower RMSE than the Full PINN on all three fitted scenes, with values of 0.1666, 0.0897, and 0.1220 °C. GCN was close to the Full PINN on 19 May (0.3830 versus 0.3758 °C) and had lower RMSE on 4 June and 20 June. U-Net behaved differently across the three dates: RMSE increased from 0.1006 °C on 19 May to 0.5902 °C on 4 June and 1.3673 °C on 20 June, while R^2 decreased from 0.9974 to 0.6134 between the first and third scenes. The four-layer LSTM produced narrower fitted temperature ranges than observed, particularly on 19 May, when the modelled range was 24.8–35.5 °C compared with 23.3–38.3 °C in the Landsat-derived LST.

XGBoost produced the smallest numerical errors in the fitted-scene comparison: RMSE was 0.0162, 0.0120, and 0.0181 °C, with $R^2 = 0.9999$ on each date. Those values describe agreement with the fitted observations and are not evidence of comparable performance at unobserved locations or dates. Figure 2 describes the corresponding prediction–observation relationships and absolute-error distributions.

The reduced PINN variants separate the effects of the retained model components. Removing wind advection increased RMSE from 0.3758 to 0.5595 °C on 19 May, from 0.2549 to 0.4273 °C on 4 June, and from 0.3959 to 0.5912 °C on 20 June. Removing T_{2m} and T_{d2m} also increased RMSE on every fitted scene, to 0.5758, 0.3412, and 0.5154 °C, respectively.

The Q ablation moved in the opposite direction. Without the explicit source–sink term, RMSE decreased to 0.3086, 0.2125, and 0.3477 °C, while R^2 increased to 0.9756, 0.9766, and 0.9750. The fitted-scene comparison therefore provides no accuracy-based justification for retaining Q ; in the Full PINN it remains part of the adopted physics-guided residual rather than a component displayed here to improve agreement with Landsat LST. A spatial comparison for 4 June is given in Figure 3.

For 4 June, the Full PINN reproduced the observed 16.2–25.6 °C range and had a maximum absolute error of 1.36 °C. U-Net extended to 26.0 °C and reached a maximum absolute error of 4.30 °C. The reported XGBoost run reproduced the 16.2–25.6 °C range with a maximum absolute error of 0.06 °C.

The Full PINN was retained for the counterfactual analysis despite not having the lowest fitted-scene error. The following scenarios use its physics-guided formulation as a common model within which prescribed changes in surface properties are compared. Additional spatial maps and temperature-change distributions for the NDVI = 0.55 scenario, separated by railway-proximity group, are provided in Supplementary Appendix F.

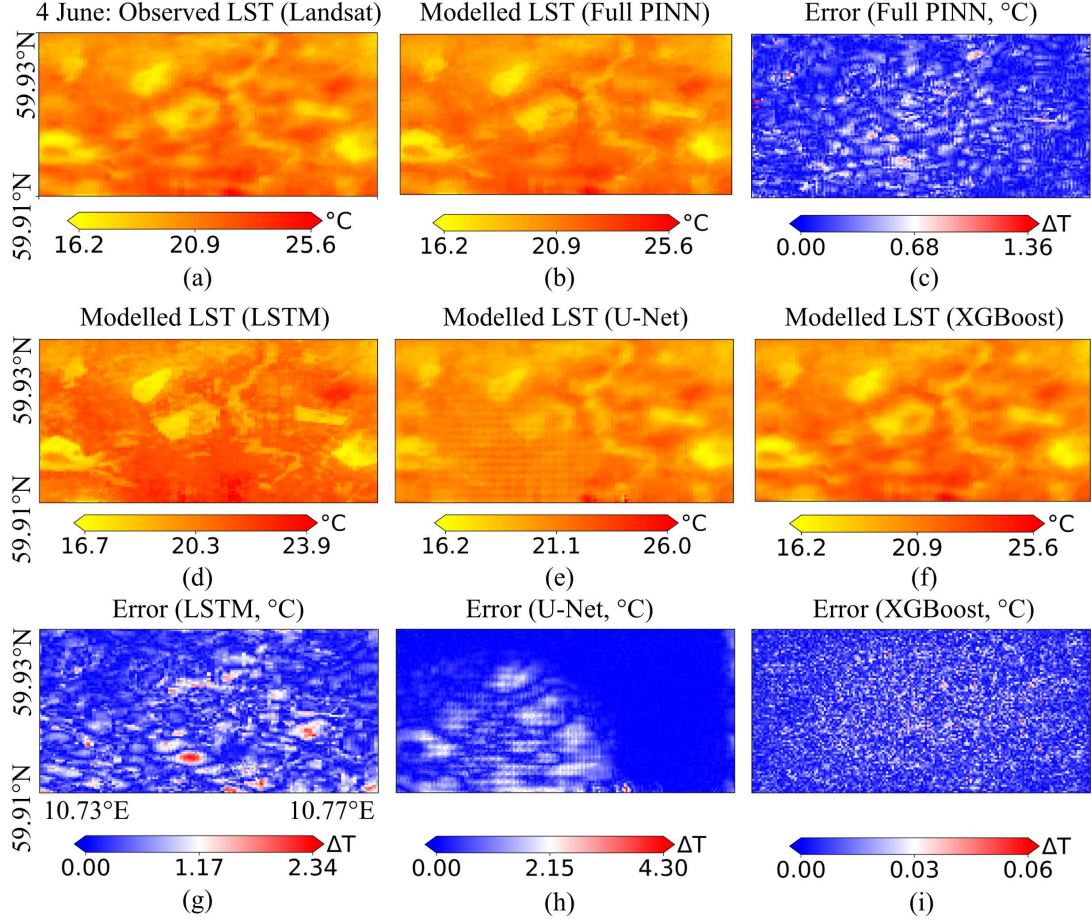


Figure 3: Landsat-derived and modelled LST and absolute-error maps for Oslo on 4 June. (a) Landsat-derived LST; (b) Full PINN prediction; (c) Full PINN absolute error; (d) LSTM prediction; (e) U-Net prediction; (f) XGBoost prediction; (g) LSTM absolute error; (h) U-Net absolute error; and (i) XGBoost absolute error.

2.2. Thermal Hotspots and Candidate Buildings

Hotspot locations differed among the three Landsat acquisitions because the thermal classes were defined separately for each scene. Pixels between the 85th and 90th percentiles formed the first hotspot class, those between the 90th and 95th percentiles the second, and pixels at or above the 95th percentile the third (Figure 4). All three classes were retained, so the candidate screening started from pixels within the upper 15% of the Landsat-derived LST distribution for each date. The resulting thresholds were therefore date-specific rather than fixed temperatures applied across the three scenes.

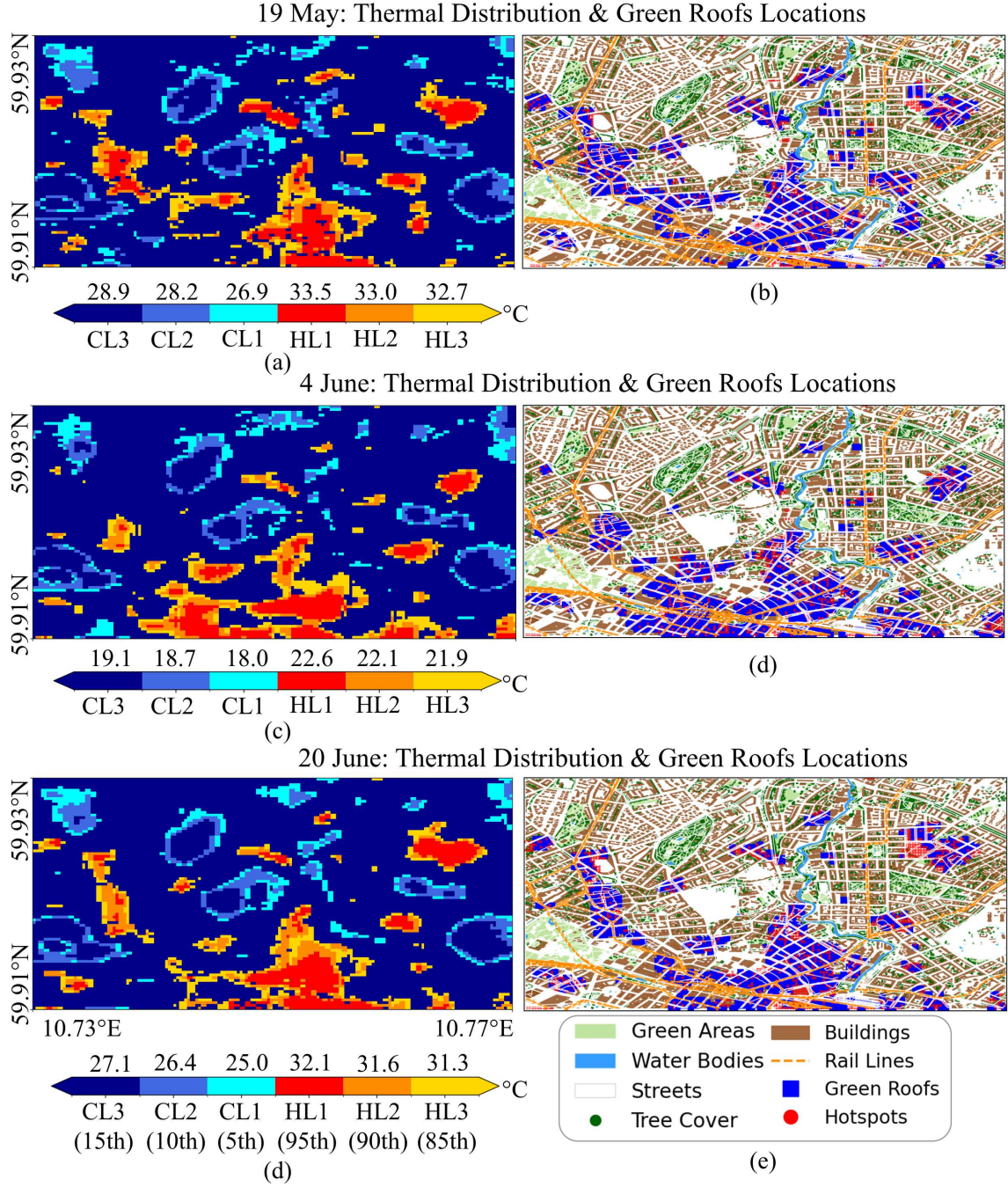


Figure 4: Landsat-derived thermal classes and hotspot building locations used in the counterfactual analysis. Left panels describe cool classes (CL1–CL3) and hotspot classes (HL1–HL3) for (a) 19 May, (c) 4 June, and (e) 20 June. Right panels display building locations associated with the identified hotspots for (b) 19 May, (d) 4 June, and (f) 20 June for the green roof scenario (NDVI = 0.55).

Intersecting the hotspot masks with the OpenStreetMap building mask restricted the analysis to building pixels located within these date-specific thermal classes. Their spatial distribution consequently changed between 19 May, 4 June, and 20 June rather than forming a single fixed candidate set.

2.3. Thermal Response and Prioritisation

At $\text{NDVI} = 0.75$, the response under partial deployment varied with coverage level, acquisition date, candidate group, and ranking strategy. Coverage refers to the proportion of eligible candidate pixels within each date and spatial group rather than to total building or roof area. Figure 5 describes the mean ΔT across the analysed coverage levels for the three ranking strategies.

Table 2: Mean temperature change ΔT ($^{\circ}\text{C}$) over selected hotspot building pixels at 15% and 100% coverage for park/water-proximity and railway-proximity prioritisation under the $\text{NDVI} = 0.75$ counterfactual. Values are reported as 15% to 100% coverage; negative values denote lower predicted LST.

Candidate group and date	Park/water proximity	Railway proximity
Outside near-rail group		
19 May	−2.334 to −2.256	−2.072 to −2.256
4 June	−1.618 to −1.369	−1.348 to −1.369
20 June	−1.723 to −1.702	−1.646 to −1.702
Near-rail group		
19 May	−1.994 to −2.002	−2.228 to −2.002
4 June	−1.750 to −1.700	−1.961 to −1.700
20 June	−1.830 to −2.084	−2.576 to −2.084
All candidates		
19 May	−2.134 to −2.060	−2.033 to −2.060
4 June	−1.707 to −1.626	−1.967 to −1.626
20 June	−1.797 to −1.901	−2.277 to −1.901

The coverage-response curves were not uniform across dates or spatial groups. Under baseline-LST prioritisation, the near-rail group on 20 June reached a mean ΔT of approximately -3.45°C at low coverage. The two wind-based rankings showed different coverage-response trajectories, with their relative position changing among dates and candidate groups rather than following a single ordering throughout the analysis. Table 2 provides a separate comparison of the park/water-proximity and railway-proximity rankings at 15% and 100% coverage. At 15%, park/water proximity gave the more negative mean ΔT outside the near-rail group on all three dates:

-2.334 versus -2.072 °C on 19 May, -1.618 versus -1.348 °C on 4 June, and -1.723 versus -1.646 °C on 20 June. Within the near-rail group, railway proximity gave the more negative value on each date, with mean ΔT values of -2.228 , -1.961 , and -2.576 °C compared with -1.994 , -1.750 , and -1.830 °C under park/water proximity.

For all candidates combined, the ordering at 15% coverage changed with date. Park/water proximity produced the more negative mean ΔT on 19 May (-2.134 versus -2.033 °C), whereas railway proximity was more negative on 4 June (-1.967 versus -1.707 °C) and 20 June (-2.277 versus -1.797 °C). At 100% coverage, both proximity rankings contain the full eligible candidate set and therefore give the same mean ΔT within each date and spatial group. No single proximity-based ranking produced the more negative mean response across every date and candidate group. Additional coverage-response results for the NDVI = 0.75 prioritisation analysis are provided in Supplementary Appendix G.

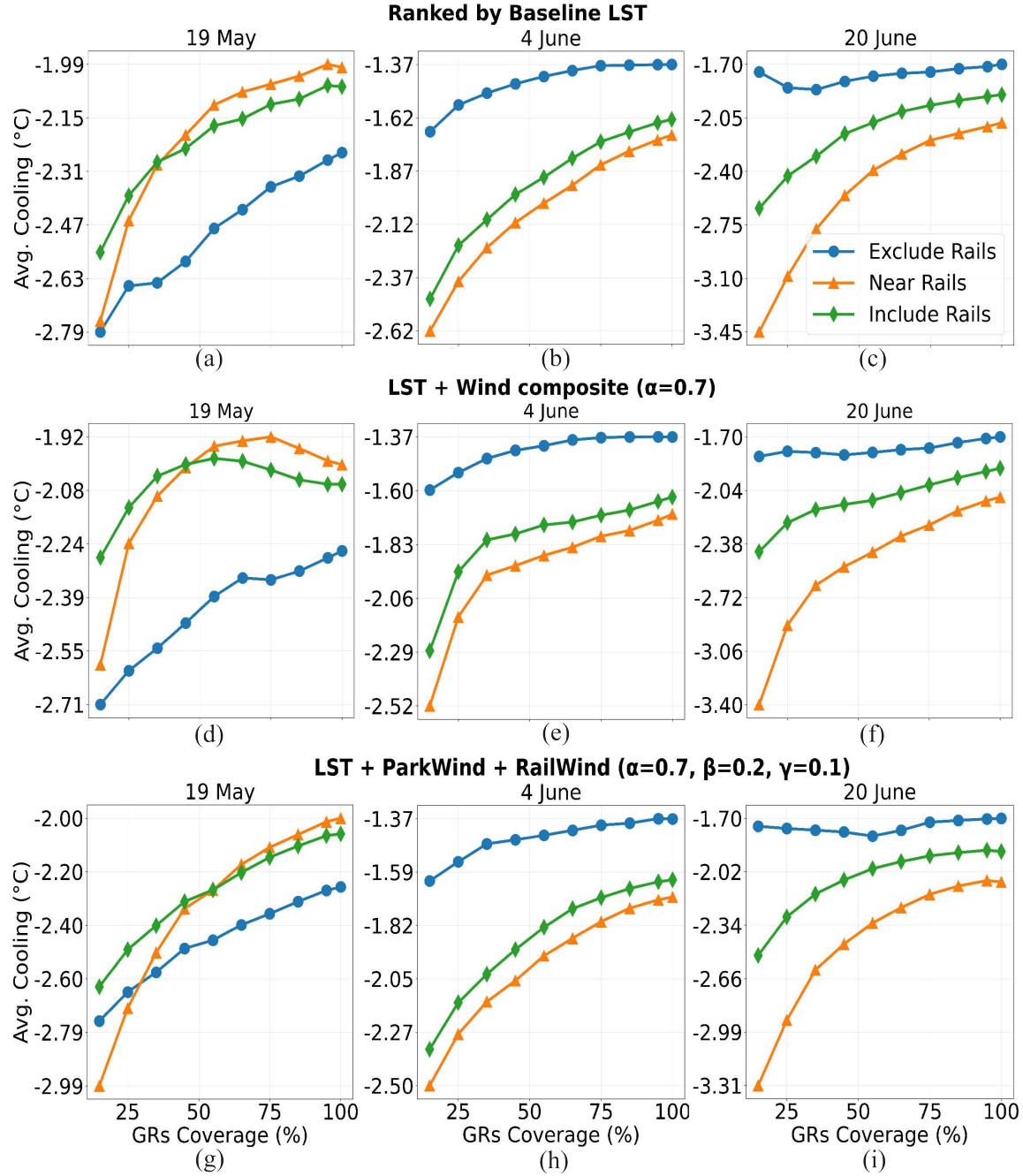


Figure 5: Mean ΔT over selected hot-spot building pixels as a function of deployment coverage for three ranking strategies under the NDVI = 0.75 counterfactual. Top, middle, and bottom rows present baseline-LST prioritisation, park/water wind-exposure ranking, and combined park/water and railway wind-exposure ranking. Columns correspond to 19 May, 4 June, and 20 June. Negative ΔT indicates lower predicted LST under the counterfactual.

3. Discussion

Urban heat-mitigation planning rarely involves unconstrained deployment. Decisions instead concern where limited interventions should be placed and whether the same spatial rule remains useful as the candidate set or thermal conditions change. This distinction is increasingly recognised in urban greening research: spatial targeting can reduce the amount of greening required to achieve a given reduction in population exposure to LST extremes, while the effectiveness of green and blue infrastructure varies with intervention type, urban context, and climatic conditions [7, 17]. The present results add a building-level example of this allocation problem. At 15% coverage, park/water proximity produced the more negative mean ΔT outside the near-rail group on all three acquisition dates, whereas railway proximity produced the more negative response within the near-rail group. For all candidates combined, the ordering changed with date: park/water proximity was favoured on 19 May, while railway proximity was favoured on 4 June and 20 June. The analysis therefore does not identify a spatial criterion that dominates across the conditions examined.

The largest local temperature changes need to be separated from this planning comparison. Under the NDVI = 0.55 counterfactual, minimum ΔT reached -4.09 , -3.49 , and -4.28 °C on the three dates, but more than 98% of affected pixels in both spatial groups belonged to the lower-magnitude response class. A planning argument based on the extrema alone would consequently misrepresent the distribution of modelled responses. There is an additional distinction between surface temperature and human heat exposure. Satellite-derived LST is useful for resolving spatial heterogeneity in the urban surface thermal environment, but it is not interchangeable with near-surface air temperature, mean radiant temperature, thermal comfort, or physiological heat exposure [20, 21]. The ΔT values reported here should therefore remain explicitly interpreted as changes in modelled LST.

The coverage experiments address a related question: *what happens to the mean modelled response as progressively larger portions of the eligible candidate set are included?* Under baseline-LST prioritisation, the response changed with coverage, and the three ranking strategies produced different trajectories across dates and spatial groups. This variation matters because spatially targeted greening has been visualised elsewhere to outperform uniform allocation for specific heat-exposure objectives, but the preferred configuration depends on the objective and spatial setting [7, 17]. Our analysis does not establish an optimal coverage percentage. It instead reports that conclusions drawn at low coverage need not describe the behaviour of the same ranking as the candidate set expands. At 100% coverage, the two proximity rankings necessarily converge because every eligible candidate has been selected.

The model comparison also changes how the physics-guided component should be discussed. The Full PINN reproduced the three fitted Landsat scenes with R^2 values of 0.9638–0.9676, but it was not the lowest-error model. ConvLSTM produced lower RMSE on all three scenes, and XGBoost gave the smallest fitted residuals. Physics-informed neural networks combine agreement with observations and satisfaction of a prescribed physical residual within the training objective; those objectives need not produce the same model ranking as prediction error alone, and optimisation and loss balancing remain recognised issues in PINN applications [32]. In this study, the Full PINN was retained as the common physics-guided model for the surface-property scenarios rather than because the fitted-scene comparison demonstrated predictive superiority. Since the evaluation uses observations already represented during fitting, the reported metrics cannot establish out-of-sample generalisation.

The ablations give a narrower basis for discussing individual terms. Removing wind advection increased fitted RMSE on all three dates, as did removing T_{2m} and T_{d2m} . The explicit source–sink term did not illustrate the same behaviour: removing Q reduced RMSE and increased R^2 on every fitted scene. Its inclusion therefore cannot be justified from fitted-scene accuracy. This is particularly important because Q is an uncalibrated spatial index constructed from mapped urban features rather than an estimate of energy flux. It provides a spatial source–sink constraint within the adopted formulation, but it should not be interpreted as a measured or dimensionally calibrated urban surface-energy term.

The green-roof interpretation requires a similar boundary. Actual green-roof thermal behaviour depends on interacting properties of the vegetation, substrate, water availability, irrigation regime, radiative characteristics, and roof construction. Process-based green-roof models explicitly represent several of these mechanisms, while experimental work provides that substrate depth, moisture content, vegetation properties, and irrigation can materially alter thermal performance [9–11, 13]. The present scenarios modify only NDVI and emissivity at selected building pixels; substrate moisture, irrigation, plant water stress, albedo, latent heat exchange, and roof thermal mass are not explicitly changed. Near-surface meteorological inputs are also held fixed. The scenarios are therefore green-roof-associated surface-property counterfactuals, not simulations of complete green-roof energy and water balances.

There is a further constraint within the prescribed surface properties themselves. The scenario emissivity formulation reaches 0.985 once NDVI exceeds 0.50, meaning that the NDVI = 0.55 and NDVI = 0.75 scenarios use the same prescribed emissivity. Their difference is therefore driven primarily by the NDVI input rather than by two distinct emissivity states. Describing the two values as validated representations of particular green-roof construction types would go beyond what was modelled.

Temporal sampling restricts the planning interpretation more sharply than the fitted R^2 values might suggest. Three clear-sky Landsat acquisitions represent only three late-spring and early-summer surface states in 2025. Satellite thermal observations also capture conditions around the overpass rather than the diurnal thermal cycle, and relationships between LST and near-surface air temperature can vary substantially across seasons and climatic settings [18, 20, 21]. With three observation times, the temporal component of the PINN is correspondingly weakly constrained. Neither the model comparison nor the prioritisation results demonstrate that the same ordering would persist through the rest of the warm season, during prolonged heat events, or in another year.

These limitations do not remove the planning value of the analysis, but they define what kind of planning information it provides. Spatial heat-mitigation studies increasingly distinguish between identifying hot surfaces and deciding where interventions should be allocated, particularly when exposure, implementation constraints, and the configuration of green infrastructure are considered [7, 17]. Here, the model provides a thermal screening layer: it allows candidate building pixels, deployment fractions, and alternative prioritisation rules to be compared under the same fitted thermal representation. A municipal deployment decision would require additional criteria not represented in the model, such as structural roof suitability, ownership, implementation and maintenance costs, and, where reduction of human heat risk is the objective, information on population exposure and vulnerability. LST screening alone should not substitute for those dimensions [20, 21].

The contribution is therefore narrower than prediction of green-roof performance but more directly connected to the allocation problem faced in urban heat-mitigation planning. The analysis combines a physics-guided representation of the observed surface thermal field with explicit tests of partial deployment and spatial prioritisation. Within the three observed thermal states, modelled response changed with acquisition date, spatial grouping, coverage, and ranking rule. For planning applications, this cautions against treating intervention priority as a fixed property of location: the ranking depends on the thermal state, the candidate set, and the criterion used to select among candidates.

4. Methods

4.1. Study Area, Data and Preprocessing

The study domain is located in Oslo, Norway, centred at approximately 59.92°N and 10.75°E. All analyses were carried out on a 30 m grid in UTM zone 32N containing 84×174 pixels.

Urban spatial features were obtained from OpenStreetMap [37]. Six classes were used: buildings, roads, railways, parks, water, and trees. The vector layers were projected to the coordinate reference system of the Landsat grid and converted to binary masks at the model grid locations. Figure 6 depicts the study domain and these urban feature layers.

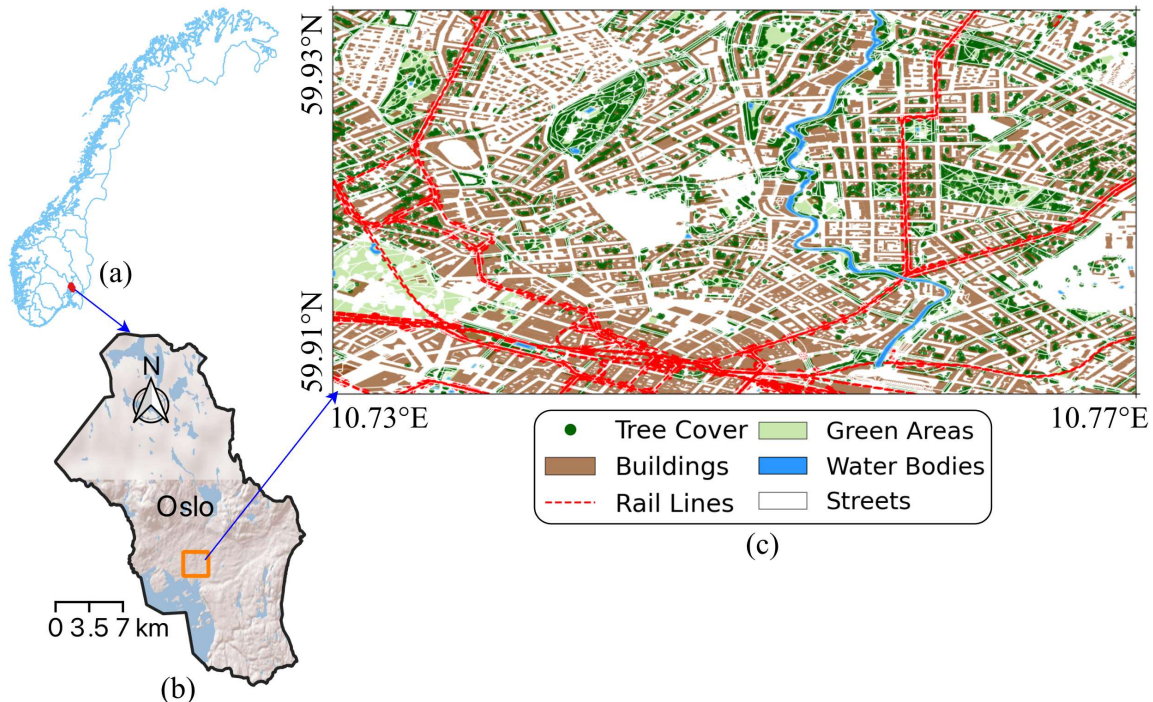


Figure 6: Oslo study domain and urban spatial features used in the analysis. The mapped features are buildings, roads, railways, parks, water, and trees derived from OpenStreetMap.

Three cloud-free Landsat 8 Collection 2 Tier 1 acquisitions were analysed: 19 May, 4 June, and 20 June 2025. Land surface temperature (LST), normalised difference vegetation index (NDVI), and land surface emissivity were derived for each acquisition on the common 30 m grid.

LST was retrieved from Landsat thermal Band 10 using the single-channel formulation of Jiménez-Muñoz *et al.* [38]:

$$\text{LST} = \gamma \left[\frac{1}{\varepsilon} (\psi_1 L_{\text{TOA}} + \psi_2) + \psi_3 \right] + \delta \quad (1)$$

where L_{TOA} is the top-of-atmosphere thermal radiance, ε is surface emissivity, γ and δ are Planck-function parameters, and ψ_1 , ψ_2 , and ψ_3 are atmospheric func-

tions. Total precipitable water used in the atmospheric correction was obtained from NCEP/NCAR reanalysis data [39, 40].

NDVI was calculated from Landsat surface-reflectance near-infrared and red bands:

$$\text{NDVI} = \frac{R_{\text{NIR}} - R_{\text{red}}}{R_{\text{NIR}} + R_{\text{red}}} \quad (2)$$

where R_{NIR} and R_{red} are surface reflectance from Bands 5 and 4, respectively.

Surface emissivity (EM) was estimated using the NDVI-based procedure adopted for the Landsat data [41]. Emissivity was set to 0.990 for water and 0.989 for snow/ice. For other surfaces,

$$\varepsilon = \varepsilon_{\text{veg}} P_v + \varepsilon_{\text{soil}} (1 - P_v) \quad (3)$$

where $\varepsilon_{\text{veg}} = 0.99$, $\varepsilon_{\text{soil}}$ was obtained from the ASTER Global Emissivity Dataset [42], and P_v is fractional vegetation cover. Fractional vegetation cover was calculated as

$$P_v = \left(\frac{\text{NDVI} - \text{NDVI}_{\min}}{\text{NDVI}_{\max} - \text{NDVI}_{\min}} \right)^2 \quad (4)$$

with $\text{NDVI}_{\min} = 0.20$ and $\text{NDVI}_{\max} = 0.86$.

The spatial fields and distributions of LST, NDVI, and emissivity are shown in Figure 7. Pairwise distributions and correlations among LST, NDVI, and emissivity for the three acquisition dates are provided in Supplementary Appendix A. NDVI enters directly into the emissivity parameterisation, so their statistical association is partly imposed by the retrieval procedure and is not treated as an independent relationship.

Meteorological inputs were derived from ERA5 [43]. The model used zonal and meridional wind components, 2 m air temperature (T_{2m}), and 2 m dewpoint temperature (T_{d2m}). Period-mean meteorological fields were associated with the three Landsat acquisitions using 15–31 May, 1–15 June, and 16–30 June 2025, respectively (Table 3).

Wind components were converted from 10 m to nominal 2 m values using a logarithmic wind-profile relation [44]:

$$u_2 = u_{10} \frac{\ln(2/z_0)}{\ln(10/z_0)}, \quad v_2 = v_{10} \frac{\ln(2/z_0)}{\ln(10/z_0)} \quad (5)$$

with $z_0 = 0.5$ m. This scaling does not resolve airflow around individual buildings. Figure 8 illustrates the wind conditions and period-mean T_{2m} and T_{d2m} used with the three acquisitions.

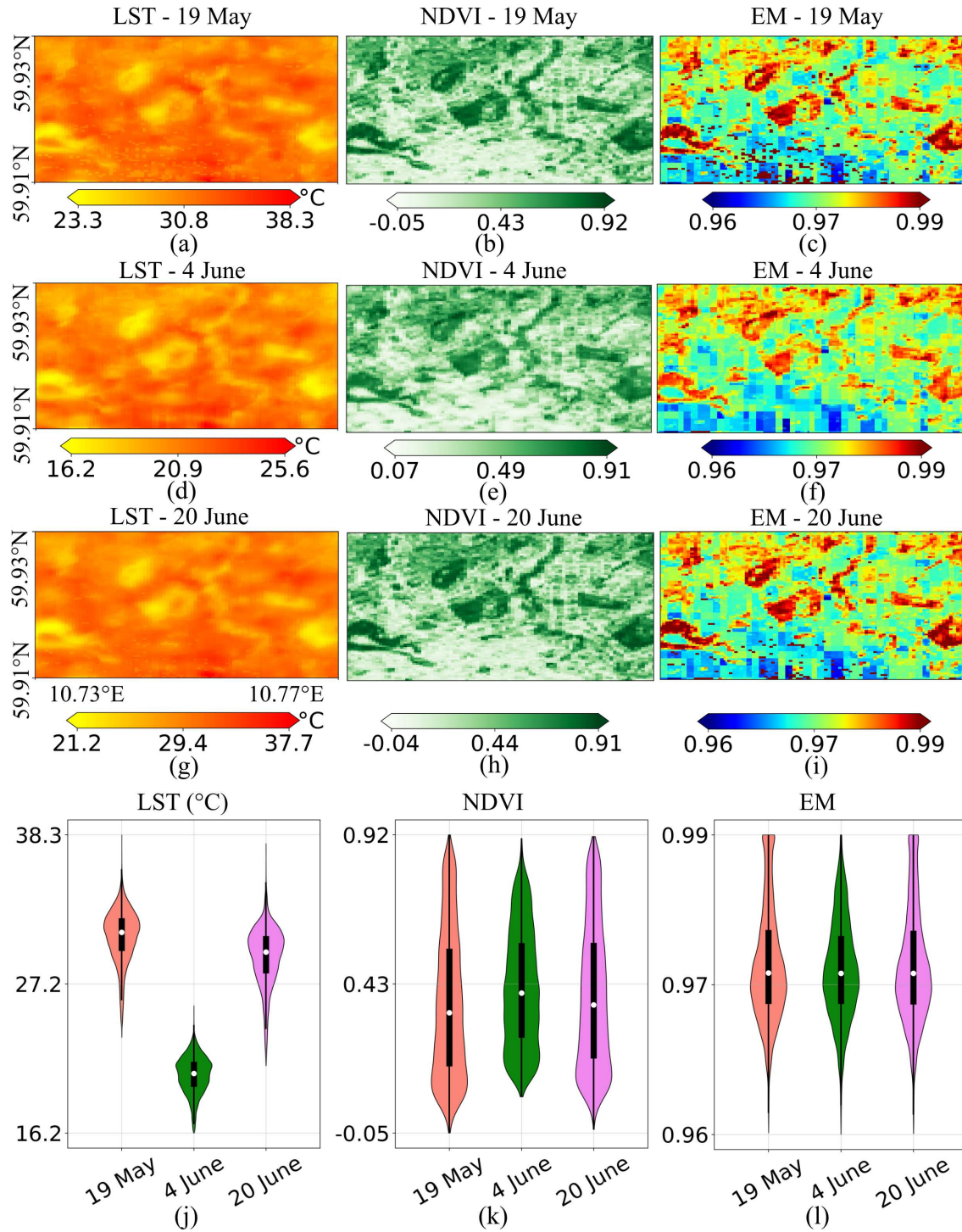


Figure 7: Landsat-derived LST, NDVI, and EM for the Oslo study domain on three periods. Panels (a-i) visualise spatial fields; while (j-l) display corresponding variable distributions.

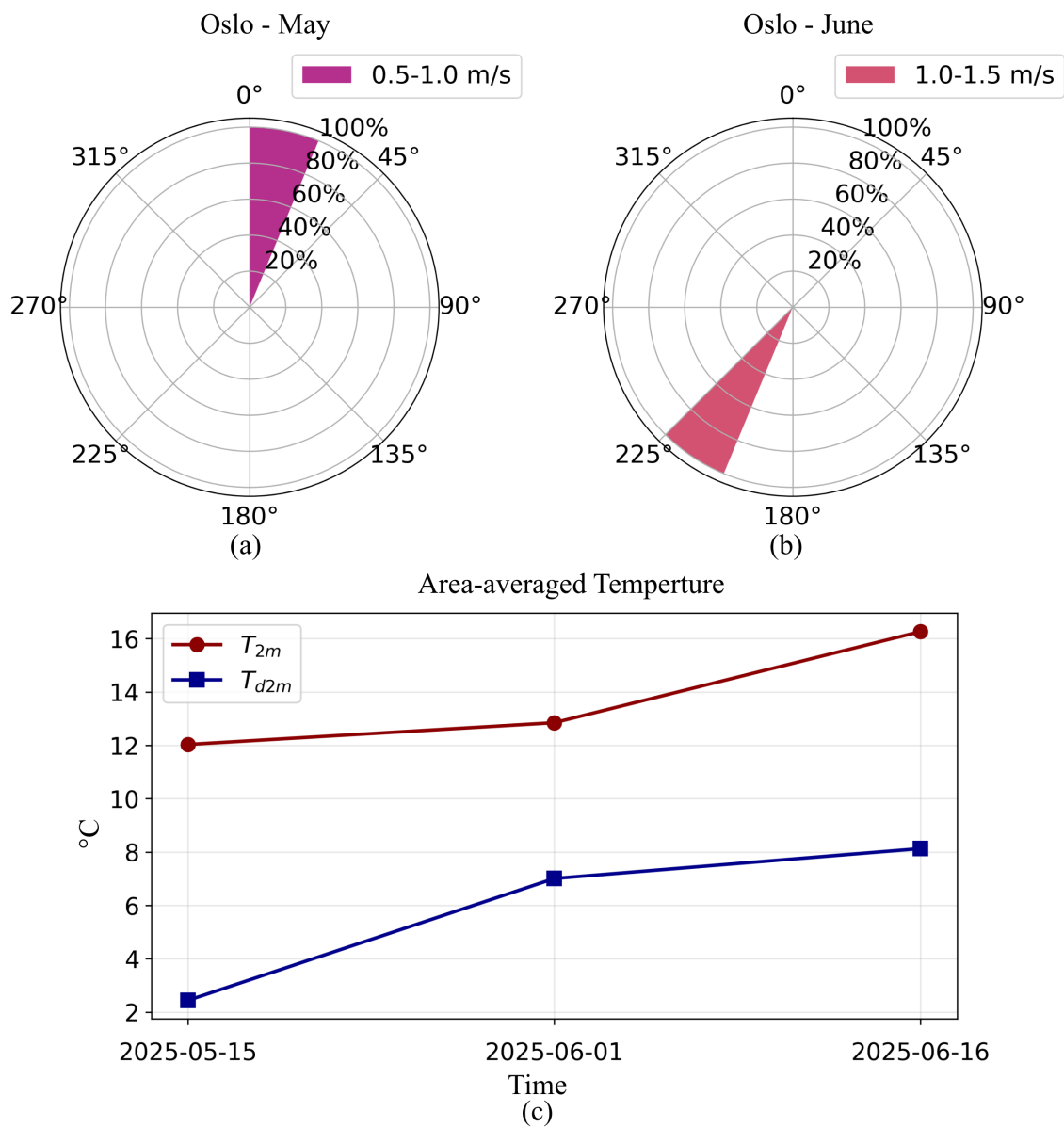


Figure 8: Meteorological conditions associated with the Landsat acquisitions. Panels (a) and (b) present wind direction and speed distributions for the May and June analysis periods, and panel (c) describe period-mean 2m air temperature (T_{2m}) and 2m dewpoint temperature (T_{d2m}).

Acquisition time was normalised between 1 May and 30 June 2025:

$$t = \frac{t_{\text{acq}} - t_{\text{base}}}{t_{\text{end}} - t_{\text{base}}} \quad (6)$$

The values were 0.300 for 19 May, 0.567 for 4 June, and 0.833 for 20 June. The summer solstice on 21 June corresponds to $t = 0.850$ under the same normalisation.

Table 3: Landsat acquisition dates, associated meteorological averaging periods, and normalised acquisition times.

Meteorological period	Landsat acquisition	Normalised time t
15–31 May 2025	19 May 2025	0.300
1–15 June 2025	4 June 2025	0.567
16–30 June 2025	20 June 2025	0.833

Spatial coordinates were generated from the Landsat raster transform and normalised over the study grid. For each acquisition date, NDVI, emissivity, the two wind components, T_{2m} , and T_{d2m} were independently min–max scaled to $[0, 1]$; the six urban-feature masks remained binary. LST was also normalised separately for each date and converted back using the corresponding date-specific minimum and maximum when model outputs were reported.

Each pixel-date sample contained 15 model inputs: normalised x and y coordinates, normalised acquisition time, NDVI, emissivity, six urban-feature masks, u_2 , v_2 , T_{2m} , and T_{d2m} . There were 14,616 pixels per acquisition and 43,848 pixel-date samples across the three dates. A 5-pixel perimeter was defined around the grid for use in the training losses.

4.2. Physics Guided Urban Thermal Model

LST was represented with a time-dependent physics-informed neural network (PINN). PINNs combine agreement with observational data and residuals derived from prescribed differential equations within the training objective [31]. For an input vector $\mathbf{Z}$ containing the 15 variables defined in Section 4.1, the network predicts normalised LST as

$$\tilde{T} = \mathcal{N}(\mathbf{Z}; \boldsymbol{\theta}) \quad (7)$$

where $\boldsymbol{\theta}$ contains the trainable parameters.

The network has seven hidden linear layers. The first two contain 256 neurons with hyperbolic tangent (tanh) activation. They are followed by layers of 256 and 128

neurons using LeakyReLU activation with a negative slope of 0.01. The remaining hidden layers contain 128, 64, and 32 neurons and use tanh activation. A final linear layer returns one LST value:

$$15 \rightarrow 256 \rightarrow 256 \rightarrow 256 \rightarrow 128 \rightarrow 128 \rightarrow 64 \rightarrow 32 \rightarrow 1.$$

No dropout or batch-normalisation layers are used. The architecture contains 195,457 trainable parameters. Figure 9 displays the network and the three loss components used during training.

Training minimised:

$$\mathcal{L}_{\text{comp}} = \mathcal{L}_{\text{sup}} + \xi_p \mathcal{L}_{\text{phy}} + \xi_b \mathcal{L}_{\text{bnd}} \quad (8)$$

with $\xi_p = 1.0$ and $\xi_b = 0.3$. The supervised component was calculated over pixels outside the 5-pixel perimeter:

$$\mathcal{L}_{\text{sup}} = \frac{1}{N_{\text{int}}} \sum_{i \in \mathcal{I}} \left(\tilde{T}_i - T_{\text{obs},i} \right)^2 \quad (9)$$

where $\mathcal{I}$ is the interior-pixel set and T_{obs} is the Landsat-derived normalised LST. The physics term was based on an advection–diffusion-type residual:

$$r = \frac{\partial \tilde{T}}{\partial t} + u \frac{\partial \tilde{T}}{\partial \tilde{x}} + v \frac{\partial \tilde{T}}{\partial \tilde{y}} - \left[\alpha^* \left(\frac{\partial^2 \tilde{T}}{\partial \tilde{x}^2} + \frac{\partial^2 \tilde{T}}{\partial \tilde{y}^2} \right) + \frac{Q^*}{\rho c_p^*} \right] \quad (10)$$

and

$$\mathcal{L}_{\text{phy}} = \frac{1}{N_{\text{int}}} \sum_{i \in \mathcal{I}} r_i^2 \quad (11)$$

First- and second-order derivatives were obtained by automatic differentiation [45] with respect to the normalised x , y , and acquisition-time inputs. The wind components were supplied to the network in normalised form but were converted back using the minimum and maximum of the corresponding meteorological period before calculation of the advection term. Because the derivatives are taken with respect to scaled coordinates and time, Equation 10 is used as a training constraint in the model-variable space rather than as a dimensionally calibrated surface-energy-balance equation.

Relative thermal coefficients were assigned using the six binary urban-feature masks. Buildings, roads, and railways used $\alpha^* = 1.0$ and $\rho c_p^* = 1.0$; parks and trees used $\alpha^* = 0.333$ and $\rho c_p^* = 1.5$; and water used $\alpha^* = 0.187$ and $\rho c_p^* = 4.18$.

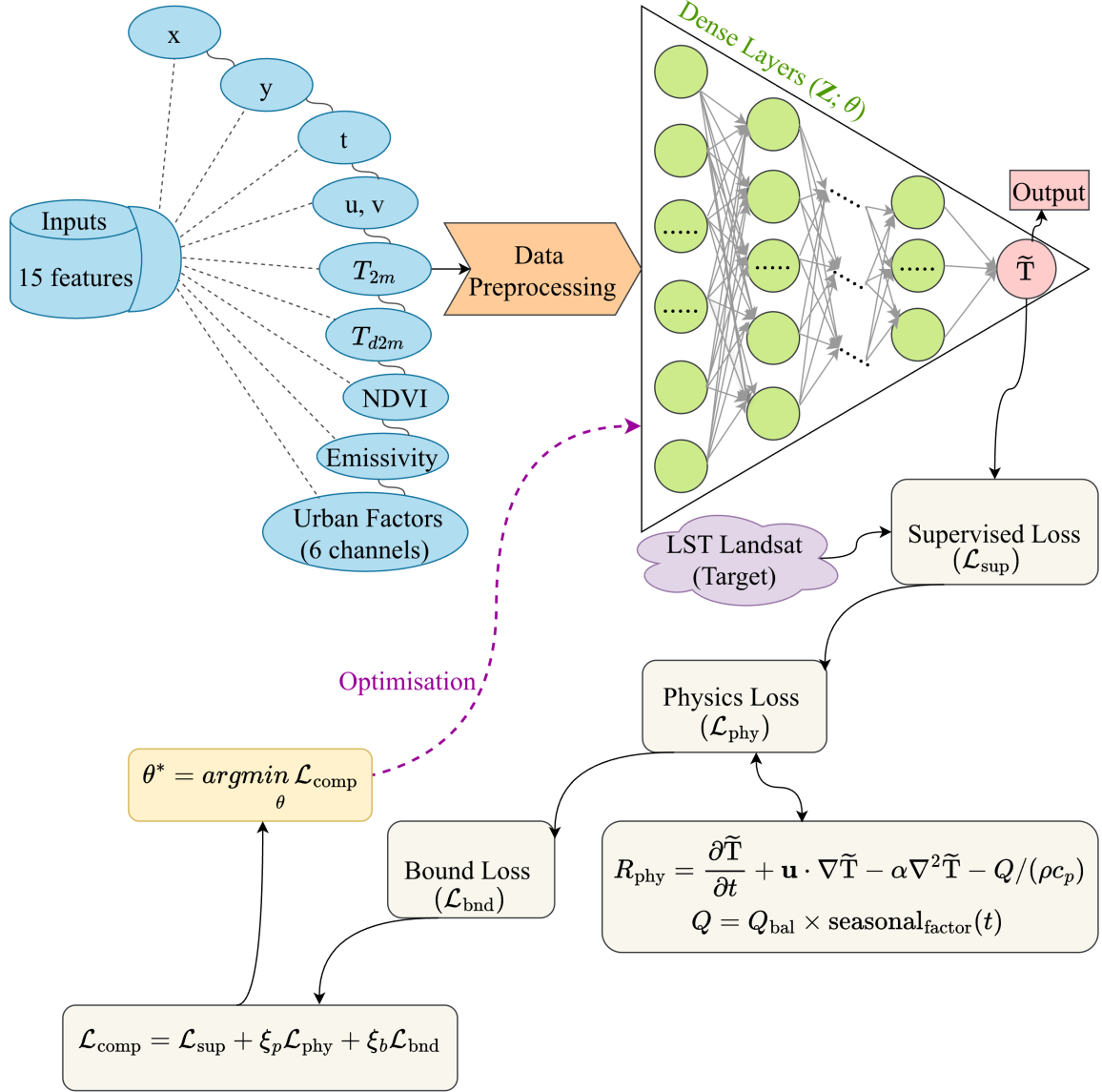


Figure 9: Time-dependent PINN architecture and training losses. The network receives 15 spatial, surface, urban-feature, and meteorological inputs and predicts normalised LST. Training combines an LST fitting term, an advection–diffusion residual term, and a perimeter data-anchoring term.

For cells in which these masks overlapped, coefficients were calculated as weighted averages. The urban weight was the sum of the building, road, and railway masks, the vegetation weight was the sum of the park and tree masks, and water entered as a separate weight. These coefficients are relative model parameters and are not interpreted as calibrated material properties.

The training implementation also contains a spatial source–sink term, Q^* . Its OSM-based formulation combines four positive and four negative spatial components. Building density, road density, inverse distance to major roads, and a building perimeter-to-area proxy enter with weights 0.30, 0.25, 0.20, and 0.15, respectively. Park density, water density, tree density, and proximity to parks, water, and trees enter with weights -0.15 , -0.10 , -0.10 , and -0.10 . Each component is scaled to $[0, 1]$ before combination:

$$Q_{\text{bal}}^* = 0.30B + 0.25R + 0.20D_r + 0.15C_b - 0.15P - 0.10W - 0.10T_r - 0.10D_c \quad (12)$$

Here B and R are smoothed building and road density, D_r is inverse distance to major roads, C_b is the building perimeter-to-area proxy after spatial smoothing, P , W , and T_r are the smoothed park, water, and tree terms, and D_c is proximity to parks, water, and trees. Supplementary Appendix B reports the tested source–sink weightings. The resulting quantity is an uncalibrated spatial source–sink index rather than an estimate of physical heat flux.

The archived training implementation attempts to construct this OSM-based field and contains a fallback formulation if that construction fails. The saved model checkpoint does not record which branch generated the Q^* field during the reported training run. Equation 12 therefore documents the OSM-based formulation contained in the archived implementation but is not used to infer the branch executed from the checkpoint alone.

For the OSM-based formulation, temporal modulation is defined as:

$$Q^*(t) = Q_{\text{bal}}^* \left[1 + 0.25 \sin \left(2\pi(t - t_{\text{solstice}}) + \frac{\pi}{2} \right) \right] \quad (13)$$

with $t_{\text{solstice}} = 0.85$. The multiplier therefore varies between 0.75 and 1.25.

The boundary term acts on the 5-pixel perimeter. For perimeter pixels that are not classified as park or water, the implementation penalises the difference between predicted and Landsat-derived normalised LST:

$$\mathcal{L}_{\text{bnd}} = \frac{1}{N_{\text{bnd}}} \sum_{i \in \mathcal{B}} \left(\tilde{T}_i - T_{\text{obs},i} \right)^2 \quad (14)$$

where $\mathcal{B}$ denotes the non-park and non-water perimeter pixels. The archived boundary-loss function also contains a gradient penalty intended for park and water perimeter pixels. In that implementation, the gradient is requested with respect to an input slice that is not used directly to calculate the corresponding predictions; with unused gradients permitted, this term does not contribute effectively to the reported boundary loss. Equation 14 therefore describes the operative perimeter term.

4.3. Model Training Evaluation and Comparison

The Full PINN was implemented in PyTorch [46] and trained with the Adam optimiser [47] using an initial learning rate of 1×10^{-3} . Samples from the three acquisition dates were concatenated and shuffled during training. The batch size was 512. A ReduceLROnPlateau scheduler monitored the mean composite training loss with a patience of five epochs.

Training was run for 200 epochs with $\xi_p = 1.0$ and $\xi_b = 0.3$. The model state after the 200th epoch was saved and used in the subsequent analyses; no earlier checkpoint was selected on the basis of minimum loss.

Supplementary Appendix C reports tests of training duration, loss weighting, activation functions, and network size. The 200-epoch configuration reduced fitted-scene error substantially relative to the 10-epoch run, while extending training to 300 epochs produced smaller changes. Physics-loss weights of 0.01, 1.0, and 1.5 and boundary-loss weights of 0.3 and 0.5 were examined. A physics-loss weight of 0.01 gave lower fitted-scene error than 1.0, but the reported model retained $\xi_p = 1.0$ to give the residual a larger contribution to the composite loss. The selected weighting was therefore not the configuration with the lowest fitted-scene RMSE.

Model agreement with Landsat-derived LST was evaluated separately for each acquisition after converting predictions back to degrees Celsius. Root mean squared error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2) were used to quantify fitted-scene agreement [48]:

$$\text{RMSE} = \sqrt{\frac{1}{M} \sum_{i=1}^M (T_{\text{pred},i} - T_{\text{obs},i})^2} \quad (15)$$

$$\text{MAE} = \frac{1}{M} \sum_{i=1}^M |T_{\text{pred},i} - T_{\text{obs},i}| \quad (16)$$

and

$$R^2 = 1 - \frac{\sum_{i=1}^M (T_{\text{obs},i} - T_{\text{pred},i})^2}{\sum_{i=1}^M (T_{\text{obs},i} - \bar{T}_{\text{obs}})^2} \quad (17)$$

Here $M = 14,616$ is the number of pixels in one acquisition. The three scenes used for these calculations were also used during model fitting, so the reported statistics measure fitted-scene agreement rather than performance on an independent test set.

The PINN was compared with several data-driven architectures using the same Oslo input variables. Supplementary Appendix D records the tested configurations and fitted-scene results. The comparison included a multilayer perceptron (MLP), long short-term memory network (LSTM) [49], convolutional LSTM (ConvLSTM) [50], U-Net [51], graph-convolutional models based on GCN operations [52], Random Forest [53], and XGBoost [54]. The surviving code archive contains the LSTM, U-Net, GCN, GLSTM, Random Forest, and XGBoost implementations. The MLP, ConvLSTM, and separate PINN-ablation scripts are not present in the surviving archive; their configurations and fitted-scene results are documented in Supplementary Appendix D.

The archived LSTM treats the three acquisition dates as a sequence for each pixel. Following the LSTM recurrent formulation [49], the implementation contains four recurrent layers with 128 hidden units, dropout of 0.2, and a linear output layer. The model was trained for 200 epochs with Adam, an initial learning rate of 1×10^{-3} , a batch size of 512, and a ReduceLROnPlateau scheduler.

The U-Net follows an encoder–decoder architecture with skip connections [51] and operates here on 32×32 spatial patches. Its encoder uses feature dimensions of 64, 128, 256, and 512, with two 3×3 convolution–batch-normalisation–ReLU operations in each block. Max pooling is used in the encoder and transposed convolutions in the decoder. Training used a batch size of 8 for 200 epochs.

For the GCN, each 32×32 patch was represented as 1024 graph nodes. Graph convolution was implemented on a regular spatial graph following the general GCN formulation [52]. Edges connected the four cardinal neighbours, and self-loops were added before symmetric normalisation of the adjacency matrix. The archived model contains four graph-convolution layers with 128 hidden units and dropout of 0.2. It was trained for 200 epochs with a batch size of 8.

The GLSTM used the same four-neighbour grid graph with self-loops, but graph propagation was incorporated into recurrent gates. The recurrent component follows an LSTM-type formulation [49], while spatial propagation uses the same graph structure adopted for the GCN [52]. The archived configuration contains four recurrent graph layers with 128 hidden units and dropout of 0.2. Each sequence contains the three acquisition dates. Training used a batch size of 8 and 200 epochs.

The archived Random Forest implementation [53] contains 1000 trees with a maximum depth of 10, a minimum split size of 5, and a minimum leaf size of 2. Supplementary Appendix D also records tests with 200 and 500 trees and with a maximum depth of 30.

The surviving XGBoost script [54] contains an active configuration with 500 estimators, maximum depth 10, learning rate 0.1, row subsampling of 0.8, and column subsampling of 1.0. Supplementary Appendix D documents additional XGBoost runs with 200–1000 estimators and row and column subsampling values between 0.8 and 1.0. The XGBoost configurations documented in the supplementary analysis are not identical to the active configuration in the surviving script; the main-text comparison therefore reports only results associated with configurations that can be identified from the retained records rather than assuming that all archived XGBoost outputs originated from the active script configuration.

Supplementary Appendix D also records an MLP without the physics-guided loss. The documented network contains hidden layers of 256, 128, 64, and 32 neurons with tanh and LeakyReLU activations, and configurations trained for 10 and 200 epochs are reported. Because the original MLP script is not present in the surviving archive, this description is based on the supplementary record.

Three reduced PINN formulations were evaluated. The No Advection variant removed the wind-advection terms from the residual while retaining diffusion and the source–sink term. The No Q variant removed $Q^*/(\rho c_p^*)$. The No Air and Dewpoint Temperature variant omitted T_{2m} and T_{d2m} from the network input, reducing the input dimension from 15 to 13. The separate ablation scripts are not present in the surviving code archive; their configurations and fitted-scene results are retained in Supplementary Appendix D.

4.4. *Green Roof Scenarios*

Counterfactual scenarios were restricted to building pixels within the upper tail of the observed Landsat LST distribution. Hotspots were identified separately for each acquisition date. The 85th, 90th, and 95th percentiles defined three hot classes; pixels at or above the 85th percentile were retained.

The hotspot mask was intersected with the OpenStreetMap building mask. Three candidate groups were then considered: all hotspot building pixels, hotspot building pixels near railways, and hotspot building pixels outside the near-rail group.

The near-rail group was defined from the railway raster using a 15-pixel binary dilation with a 31×31 square structuring element. The dilated railway mask was intersected with the building mask. At 30 m resolution, this corresponds to 15 grid cells, or 450 m, in the horizontal and vertical directions; the square structuring el-

ement extends farther along the diagonals. It is therefore a raster neighbourhood rather than an isotropic 450 m Euclidean buffer.

Candidate pixels were screened using their original NDVI and emissivity values. The procedure was applied separately for each date and candidate group. For a variable x , values outside

$$Q_1 - 1.5 IQR \leq x \leq Q_3 + 1.5 IQR \quad (18)$$

were first removed, where $IQR = Q_3 - Q_1$. Quartiles were then recalculated from the retained values and a second set of $1.5 \times IQR$ limits was obtained. The final candidate mask retained only pixels whose original NDVI and emissivity both satisfied their respective second-stage limits. The original and IQR-screened NDVI and emissivity distributions for candidate pixels outside the near-rail group and within the near-rail group are shown in Supplementary Appendix E.

The intervention itself consisted only of prescribed changes to NDVI and emissivity at selected candidate pixels. The scenario scripts calculate emissivity from the prescribed NDVI using

$$P_{v,sc} = \text{clip} \left(\frac{\text{NDVI}_{sc} - 0.20}{0.50 - 0.20}, 0, 1 \right) \quad (19)$$

followed by

$$\varepsilon_{sc} = 0.985 P_{v,sc} + 0.960 (1 - P_{v,sc}) \quad (20)$$

This is not the emissivity formulation used to prepare the original Landsat input in Equation 3. In the scenario formulation, $P_{v,sc}$ is clipped at 1 for NDVI values of 0.50 or greater. The documented NDVI targets of 0.55 and 0.75 therefore both give $\varepsilon_{sc} = 0.985$.

For each acquisition date, the trained PINN was first run with the original inputs to obtain a baseline LST field. A counterfactual input was then constructed by replacing NDVI and emissivity at the selected pixels with the prescribed scenario values. No other input variable was changed. Spatial coordinates, acquisition time, urban-feature masks, wind components, T_{2m} , and T_{d2m} remained fixed, and the PINN was not retrained.

Temperature change was calculated pixel by pixel as

$$\Delta T = T_{\text{counterfactual}} - T_{\text{baseline}} \quad (21)$$

Negative values therefore denote lower PINN-predicted LST after the prescribed NDVI–emissivity change.

The counterfactual does not represent a complete green-roof process model. Substrate depth, substrate moisture, irrigation, albedo, roof thermal mass, evapotranspiration, and feedbacks from the modified surface to the meteorological inputs were not altered explicitly. The reported ΔT values should therefore be read as responses of the fitted PINN to the prescribed surface-property changes.

4.5. Spatial prioritisation and deployment coverage

Spatial targeting provides a way to compare intervention locations when only part of the available candidate set can be treated [17]. In the present analysis, prioritisation was carried out separately for each acquisition date and for each of the three candidate groups defined in Section 4.4. Ranking was applied after hotspot identification and NDVI–emissivity screening.

Coverage was defined relative to the eligible candidate set for the corresponding date and spatial group. For a nominal coverage fraction c , the number of selected pixels was:

$$N_{\text{select}} = \text{round}(c \times N_{\text{candidate}}) \quad (22)$$

where $N_{\text{candidate}}$ is the number of eligible building-hotspot pixels remaining after the IQR screening. If candidates were available and rounding produced zero selected pixels, one pixel was retained. The tested fractions were 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95, and 1.00. These values refer to the eligible candidate pixels, not to total roof area or to the proportion of all buildings in the study domain.

Three ranking rules were used for the NDVI = 0.75 counterfactual.

The first ranked candidate pixels by baseline PINN-predicted LST. Candidates were ordered from highest to lowest predicted LST, and the first N_{select} pixels were assigned the counterfactual surface properties at each coverage level.

The second ranking used proximity to parks and water. The two feature masks were combined, and Euclidean raster distance to the nearest park or water pixel was calculated with a distance transform. Candidate pixels were ordered from the smallest to the largest distance.

The third ranking used distance to the railway mask. Candidate pixels were again ordered from the smallest to the largest Euclidean raster distance. This ranking should not be confused with the near-rail candidate grouping: the grouping is based on the 15-pixel square dilation described in Section 4.4, whereas the ranking uses continuous raster distance to order pixels within the eligible set.

At each coverage level, NDVI and emissivity were changed only for the selected pixels using Equations 19 and 20. The remaining inputs and trained network parameters were unchanged.

The response for a ranking–coverage combination was summarised by the mean temperature change across the selected pixels:

$$\overline{\Delta T} = \frac{1}{N_{\text{select}}} \sum_{i=1}^{N_{\text{select}}} \Delta T_i \quad (23)$$

This quantity represents the mean modelled response across the selected candidate pixels. It is not an area-integrated temperature change for the study domain.

Funding:

This research received no external funding.

Conflicts of Interest:

The authors declare no conflicts of interest.

Data Availability Statement:

Wind data were obtained from the ERA5 reanalysis (<https://cds.climate.copernicus.eu/>). Landsat-derived LST, NDVI, and emissivity data were accessed through Google Earth Engine (<https://earthengine.google.com/>), and urban morphology data were obtained from OpenStreetMap (<https://www.openstreetmap.org/>). The code developed for this study is available in the Zenodo repository at <https://doi.org/10.5281/zenodo.22540060>

References

- [1] T. R. Oke, “The energetic basis of the urban heat island,” *Quarterly journal of the royal meteorological society*, vol. 108, no. 455, pp. 1–24, 1982.
- [2] S. Grimmond, “Urbanization and global environmental change: local effects of urban warming,” *The geographical journal*, vol. 173, no. 1, pp. 83–88, 2007.
- [3] L. Zhao, X. Lee, R. B. Smith, and K. Oleson, “Strong contributions of local background climate to urban heat islands,” *Nature*, vol. 511, no. 7508, pp. 216–219, 2014.
- [4] D. J. Sailor, “A review of methods for estimating anthropogenic heat and moisture emissions in the urban environment,” *International journal of climatology*, vol. 31, no. 2, pp. 189–199, 2011.

- [5] S. E. Gill, J. F. Handley, A. R. Ennos, and S. Pauleit, “Adapting cities for climate change: the role of the green infrastructure,” *Built environment*, vol. 33, no. 1, pp. 115–133, 2007.
- [6] T. Zölch, J. Maderspacher, C. Wamsler, and S. Pauleit, “Using green infrastructure for urban climate-proofing: An evaluation of heat mitigation measures at the micro-scale,” *Urban Forestry & Urban Greening*, vol. 20, pp. 305–316, 2016.
- [7] P. Kumar *et al.*, “Urban heat mitigation by green and blue infrastructure: Drivers, effectiveness, and future needs,” *The Innovation*, vol. 5, no. 2, p. 100588, 2024.
- [8] Z. S. Venter, N. H. Krog, and D. N. Barton, “Linking green infrastructure to urban heat and human health risk mitigation in oslo, norway,” *Science of the total environment*, vol. 709, p. 136193, 2020.
- [9] D. J. Sailor, “A green roof model for building energy simulation programs,” *Energy and buildings*, vol. 40, no. 8, pp. 1466–1478, 2008.
- [10] M. Santamouris, “Cooling the cities—a review of reflective and green roof mitigation technologies to fight heat island and improve comfort in urban environments,” *Solar energy*, vol. 103, pp. 682–703, 2014.
- [11] U. Berardi, A. GhaffarianHoseini, and A. GhaffarianHoseini, “State-of-the-art analysis of the environmental benefits of green roofs,” *Applied energy*, vol. 115, pp. 411–428, 2014.
- [12] E. Alexandri and P. Jones, “Temperature decreases in an urban canyon due to green walls and green roofs in diverse climates,” *Building and environment*, vol. 43, no. 4, pp. 480–493, 2008.
- [13] A. Pianella, L. Aye, Z. Chen, and N. S. G. Williams, “Substrate depth, vegetation and irrigation affect green roof thermal performance in a mediterranean type climate,” *Sustainability*, vol. 9, no. 8, p. 1451, 2017.
- [14] J. Yan, F. Zhang, F. K. S. Chan, L. E. Yang, Y. He, and S. Zhang, “Adapting green roofs to climate change: Insights from case studies across four typical climates in china,” *Building and Environment*, p. 113477, 2025.
- [15] S. Jia, Q. Weng, C. Yoo, H. Xiao, and Q. Zhong, “Building energy savings by green roofs and cool roofs in current and future climates,” *Npj Urban Sustainability*, vol. 4, no. 1, p. 23, 2024.

- [16] H. Bai, Z. Yu, S. Nie, M. Luo, X. Hu, J. Xiao, B. Xue, C. Xu, X. Chen, L. Fan, *et al.*, “Cooling potential of global urban roof greening,” *npj Urban Sustainability*, vol. 6, no. 1, p. 49, 2026.
- [17] E. Massaro, R. Schifanella, M. Piccardo, L. Caporaso, H. Taubenböck, A. Cescatti, and G. Duveiller, “Spatially-optimized urban greening for reduction of population exposure to land surface temperature extremes,” *Nature Communications*, vol. 14, p. 2903, 2023.
- [18] Q. Weng, “Thermal infrared remote sensing for urban climate and environmental studies: Methods, applications, and trends,” *ISPRS Journal of photogrammetry and remote sensing*, vol. 64, no. 4, pp. 335–344, 2009.
- [19] D. Zhou, J. Xiao, S. Bonafoni, C. Berger, K. Deilami, Y. Zhou, S. Frohling, R. Yao, Z. Qiao, and J. A. Sobrino, “Satellite remote sensing of surface urban heat islands: Progress, challenges, and perspectives,” *Remote Sensing*, vol. 11, no. 1, p. 48, 2018.
- [20] V. K. Turner, M. L. Rogers, Y. Zhang, A. Middel, F. A. Schneider, J. P. Ocón, M. Seeley, and J. Dialesandro, “More than surface temperature: mitigating thermal exposure in hyper-local land system,” *Journal of Land Use Science*, vol. 17, no. 1, pp. 79–99, 2022.
- [21] N. Muse, A. Clement, and K. J. Mach, “Daytime land surface temperature and its limits as a proxy for surface air temperature in a subtropical, seasonally wet region,” *PLOS Climate*, vol. 3, no. 10, p. e0000278, 2024.
- [22] P. A. Mirzaei and F. Haghighat, “Approaches to study urban heat island–abilities and limitations,” *Building and environment*, vol. 45, no. 10, pp. 2192–2201, 2010.
- [23] J. Voelkel and V. Shandas, “Towards systematic prediction of urban heat islands: Grounding measurements, assessing modeling techniques,” *Climate*, vol. 5, no. 2, p. 41, 2017.
- [24] M. Bruse and H. Fleer, “Simulating surface–plant–air interactions inside urban environments with a three dimensional numerical model,” *Environmental modelling & software*, vol. 13, no. 3-4, pp. 373–384, 1998.
- [25] V. Masson, “A physically-based scheme for the urban energy budget in atmospheric models,” *Boundary-layer meteorology*, vol. 94, no. 3, pp. 357–397, 2000.

- [26] J. W. Oh, J. Ngarambe, P. N. Duhirwe, G. Y. Yun, and M. Santamouris, “Using deep-learning to forecast the magnitude and characteristics of urban heat island in seoul korea,” *Scientific reports*, vol. 10, no. 1, p. 3559, 2020.
- [27] S. Vulova, F. Meier, D. Fenner, H. Nouri, and B. Kleinschmit, “Summer nights in berlin, germany: modeling air temperature spatially with remote sensing, crowdsourced weather data, and machine learning,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5074–5087, 2020.
- [28] R. S. dos Santos, “Estimating spatio-temporal air temperature in london (uk) using machine learning and earth observation satellite data,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 88, p. 102066, 2020.
- [29] C. S. B. Grimmond, M. Blackett, M. J. Best, J. Barlow, J. Baik, S. Belcher, S. Bohnenstengel, I. Calmet, F. Chen, A. Dandou, *et al.*, “The international urban energy balance models comparison project: first results from phase 1,” *Journal of applied meteorology and climatology*, vol. 49, no. 6, pp. 1268–1292, 2010.
- [30] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and F. Prabhat, “Deep learning and process understanding for data-driven earth system science,” *Nature*, vol. 566, no. 7743, pp. 195–204, 2019.
- [31] M. Raissi, P. Perdikaris, and G. E. Karniadakis, “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations,” *Journal of Computational physics*, vol. 378, pp. 686–707, 2019.
- [32] S. Cuomo, V. Schiano Di Cola, F. Giampaolo, G. Rozza, M. Raissi, and F. Piccialli, “Scientific machine learning through physics-informed neural networks: Where we are and what’s next,” *Journal of Scientific Computing*, vol. 92, p. 88, 2022.
- [33] Y. Chen, H. Wang, and Z. Chen, “Sensitivity analysis of physical regularization in physics-informed neural networks (pinns) of building thermal modeling,” *Building and Environment*, vol. 273, p. 112693, 2025.
- [34] Y. Cui, X. Zhou, C. Fan, H. Lin, and S. Deng, “Application of physics-informed neural networks-based hybrid prediction models in campus thermal environment prediction,” *Building and Environment*, p. 113880, 2025.

- [35] A. G. Nizamani, D. T. Vu, R. A. Nizamani, and C. Li, “Physics-informed neural network calibration and application for urban cooling,” *Sustainable Cities and Society*, p. 107525, 2026.
- [36] D. T. Vu, R. A. Nizamani, and A. G. Nizamani, “Smart urban design with physics-informed neural networks: Quantifying temperature reductions from green infrastructure using satellite thermal data,” *EarthArXiv eprints*, p. X51T9D, 2025.
- [37] G. Boeing, “Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks,” *Computers, environment and urban systems*, vol. 65, pp. 126–139, 2017.
- [38] J. C. Jimenez-Munoz, J. A. Sobrino, D. Skoković, C. Mattar, and J. Cristobal, “Land surface temperature retrieval methods from landsat-8 thermal infrared sensor data,” *IEEE Geoscience and remote sensing letters*, vol. 11, no. 10, pp. 1840–1843, 2014.
- [39] K. E. Trenberth and C. J. Guillemot, “Evaluation of the atmospheric moisture and hydrological cycle in the ncep/ncar reanalyses,” *Climate Dynamics*, vol. 14, no. 3, pp. 213–231, 1998.
- [40] R. Kistler, E. Kalnay, W. Collins, S. Saha, G. White, J. Woollen, M. Chelliah, W. Ebisuzaki, M. Kanamitsu, V. Kousky, *et al.*, “The ncep–ncar 50-year reanalysis: monthly means cd-rom and documentation,” *Bulletin of the American Meteorological society*, vol. 82, no. 2, pp. 247–268, 2001.
- [41] S. L. Ermida, P. Soares, V. Mantas, F. M. Göttsche, and I. F. Trigo, “Google earth engine open-source code for land surface temperature estimation from the landsat series,” *Remote Sensing*, vol. 12, no. 9, p. 1471, 2020.
- [42] G. C. Hulley, S. J. Hook, E. Abbott, N. Malakar, T. Islam, and M. Abrams, “The aster global emissivity dataset (aster ged): Mapping earth’s emissivity at 100 meter spatial scale,” *Geophysical Research Letters*, vol. 42, no. 19, pp. 7966–7976, 2015.
- [43] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, *et al.*, “The era5 global reanalysis, quarterly journal of the royal meteorological society,” 2020.
- [44] T. R. Oke, *Boundary layer climates*. Routledge, 2002.

- [45] M. Bartholomew-Biggs, S. Brown, B. Christianson, and L. Dixon, “Automatic differentiation of algorithms,” *Journal of Computational and Applied Mathematics*, vol. 124, no. 1-2, pp. 171–190, 2000.
- [46] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, “Pytorch: An imperative style, high-performance deep learning library,” in *Advances in Neural Information Processing Systems*, vol. 32, pp. 8024–8035, 2019.
- [47] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [48] C. J. Willmott, “Some comments on the evaluation of model performance,” *Bulletin of the American meteorological society*, vol. 63, no. 11, pp. 1309–1313, 1982.
- [49] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [50] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, “Convolutional lstm network: A machine learning approach for precipitation nowcasting,” in *Advances in Neural Information Processing Systems*, vol. 28, pp. 802–810, 2015.
- [51] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.
- [52] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” in *5th International Conference on Learning Representations*, 2017.
- [53] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, pp. 5–32, 2001.
- [54] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, ACM, 2016.

Supplementary Files

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

- [SupplementaryAppendix.pdf](#)