Journal of Climate Change Research 2025, Vol. 16, No. 5-1, pp. 739~763
DOI: https://doi.org/10.15531/KSCCR.2025.16.5.739

Quantifying undetected urban heat: Identifying monitoring gaps using public urban IoT sensor network

Aida, Ayano[*], Park, Chan[**†], Jeon, Jong-June[***] and Choi, Jaeyeon[****]

[*]*Doctoral Candidate, Dept. of Urban Planning and Design, University of Seoul, Seoul, Korea*
[**]*Professor, Dept. of Landscape Architecture, College of Urban Sciences, University of Seoul, Seoul, Korea*
[***]*Professor, Dept. of Statistics, University of Seoul, Seoul, Korea*
[****]*Researcher, Dept. of Urban Planning and Design, University of Seoul, Seoul, Korea*

ABSTRACT

With rising summer temperatures causing significant health hazards, understanding fine-scale temperatures patterns is becoming increasingly important. Public urban IoT sensor networks have the potential to provide such information; however, the high cost of these networks represents a considerable constraint, particularly for smaller municipalities. Therefore, this study aimed to quantitatively evaluate to quantitatively evaluate their effectiveness by using the high-density public urban IoT sensor network (S-DoT) installed by the Seoul Metropolitan Government as a case study. The difference in daily maximum temperature (ΔTX) between conventional reference stations and S-DoT was quantified, and its relationships with variables representing urban landscape and structure were examined. Key findings include: (1) ΔTX ranged from -0.93°C to +2.36°C. (2) Positive ΔTX (S-DoT > reference) trend was more prominent in compact low- to mid-rise urban areas near green spaces. (3) Negative ΔTX (S-DoT < reference) trend was stronger in areas with forest land cover on steep-slope, likely due to the cooling effect of tree shade. (4) Large ΔTX values were also derived by the prediction uncertainty of the reference station network. This study demonstrates that S-DoT sensors can identify critical thermal risk areas in compact low- to mid-rise urban zones near green spaces that cannot be captured by conventional reference stations. In addition, it provides insights into the strategic placement of IoT sensors to bridge undetected temperature gaps between vertical levels, thereby offering valuable information for decision-makers to strengthen urban heat risk assessment and adaptation strategies.

Key words: Internet of Things (IoT), S-DoT, Urban Heat Risk, Monitoring, Random Forest

1. Introduction

High summer air temperatures contribute to an increase in morbidity and mortality in urban areas worldwide. Urban environments retain and generate heat more effectively than rural areas due to various physical and structural factors, including high-rise, densely building fabric, impervious surfaces, and limited vegetation (Oke et al., 2017). Additionally, anthropogenic heat sources, such as outdoor air conditioning units and vehicle exhaust exacerbate urban heat retention. It is assessed with high confidence that the adverse impacts of hot extremes on human health, livelihoods, and key infrastructure will continue to intensify in the future as a result of human-caused climate change (IPCC, 2023).

To mitigate the adverse effects of extreme heat, policymakers and urban planners have developed high-temperature adaptation strategies and public health management plans. Identifying and quantifying high-temperature zones is a critical step toward targeted

†Corresponding author : chaneparkmomo7@uos.ac.kr (No.204 Baebong-hall, 163, Seoulsiripdae-ro, Dongdaemun-gu, Seoul, Korea. Tel. +82-2-6490-2489)

ORCID Aida, Ayano 0000-0003-3554-8992 Jeon, Jong-June 0000-0002-1423-4292
Park, Chan 0000-0002-4994-6855 Choi, Jaeyeon 0000-0001-5837-6279

Received: August 28, 2025 / Revised: September 19, 2025 1st, October 10, 2025 2nd / Accepted: October 20, 2025

and efficient adaptation. A deeper understanding of urban air temperatures can provide scientific evidence to support policymaking strategies (G. Chen et al., 2022; Romero Rodríguez et al., 2024).

Over the years, many studies have been conducted across various cities to monitor and analyze spatial distributions of urban temperature. The most fundamental and traditional approach to temperature monitoring is fixed-point measurement, where official governmental meteorological stations, operated by meteorological agencies in each country, systematically and reliably collect meteorological data at a synoptic scale as reference stations (Oke et al., 2017). However, a key limitation of fixed-point measurement is the low spatial density of stations. Statistical methods have been used to overcome this limitation by spatially interpolating temperatures in areas without measurements, enabling city-wide temperature mapping (Shen et al., 2020). Recently, high-resolution temperature estimation has been increasingly developed through the integration of nonlinear models, such as machine learning and deep learning algorithms, with temperature data from reference stations and remote sensing datasets like MODIS and Landsat 8 imagery (dos Santos 2020; Ho et al., 2014; Shen et al. 2020; Wang et al., 2023).

To address the spatial limitations of reference stations, the past decade has witnessed the emergence of urban meteorological networks as a standard approach for high-resolution urban climate monitoring, facilitated by advances in IoT technology (Chapman et al., 2023; World Meteorological Organization, 2023). Urban meteorological networks can provide abundant low-cost data potentially with high density compared to traditional meteorological networks (World Meteorological Organization, 2023). These networks have been evaluated to improve intra-urban temperature monitoring practices when combined with existing traditional sensors (Honjo et al., 2015; Lam et al., 2021; Sakthivel and Sengupta, 2025; Shi et al., 2021).

Fixed-point observation urban meteorological networks include citizen weather stations (CWS) networks, which are installed by individuals for private purposes and public urban IoT sensor networks installed and operated by municipal IoT projects. Some CWSs, such as the Netatmo sensor network, have their own data platforms and have attracted considerable attention in the scientific literature (Chapman et al., 2023). Several studies have also examined how to perform quality control (QC) for the network, which is the most important concern when dealing with data from non-traditional sensor networks (Fenner et al., 2021; Meier et al., 2017; Napoly et al., 2018). However, since CWSs are often located outside of private buildings, it has been pointed out that the distribution of population and the income level of residents affect station placement and density, making it difficult to collect data in areas with low population density or low income (Brousse et al., 2024).

In contrast, public urban IoT sensor networks are deployed by municipalities and have the potential to provide spatially uniform and continuous data, complementing conventional meteorological observation networks. Despite this potential, a major challenge of these sensors is their high cost, which represents a considerable constraint, particularly for smaller municipalities.

We therefore attempted to quantitatively evaluate the effectiveness of public urban IoT sensor networks as a tool for monitoring high summer temperatures and to demonstrate their practical value of these sensors in urban climate adaptation strategies. As a case study, we utilized the Smart Seoul Data of Things (S-DoT) sensor network, which has been installed and operated by the Seoul Metropolitan Government since April 2020 as part of its IoT policy (Seoul Metropolitan Government, 2020a, 2020b). The S-DoT sensors measure air temperature within the urban canyon, and in 2024, the network maintains an exceptionally high density of 1.8 stations per km^2. As one of the largest urban meteorological networks worldwide, the network provides a valuable data foundation for urban climate research.

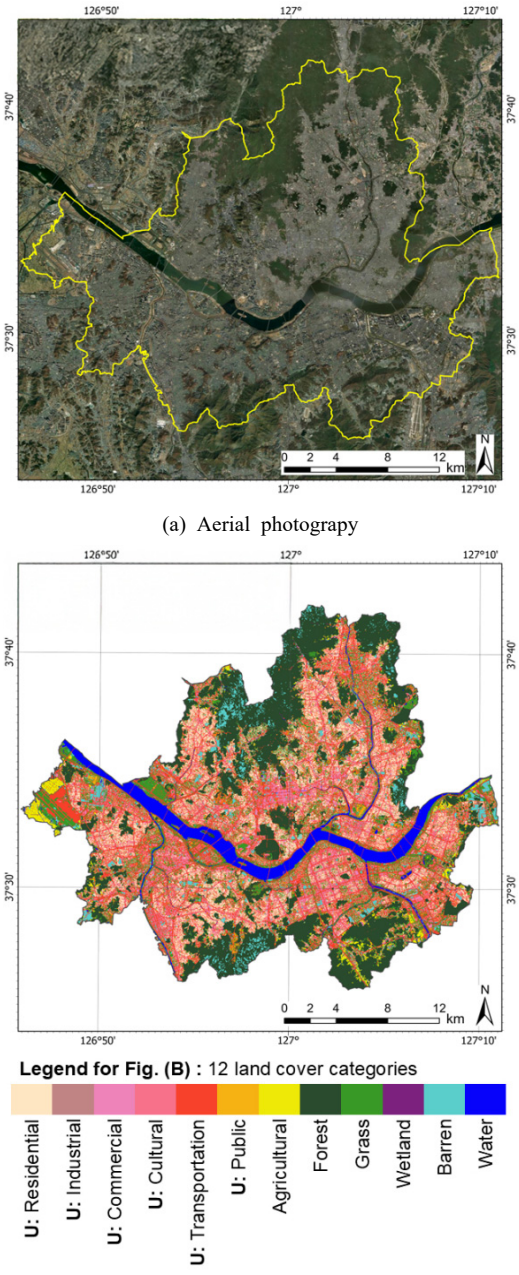
Specifically, this study aimed to quantify high summer temperatures that cannot be detected by reference stations but can be identified using the S-DoT sensor network. To analyze urban heat distribution, we compared summer

temperature maps estimated from S-DoT sensors with those derived from Korea's official government meteorological sensors. This comparison allowed us to quantify temperature differences between the two datasets. Furthermore, we examined where these differences occur and explored their relationship with spatial characteristics to better understand the underlying patterns. This study addresses the following two research questions: 1) How are high temperatures—undetected by the conventional reference station network located on rooftops but captured by street-level S-DoT observations—spatially distributed? 2) What spatial factors are associated with these previously undetected high-temperature zones? The findings provide insights for urban planners and decision-makers into priority areas of elevated temperatures, ultimately contributing to the improvement of urban temperature monitoring systems that support adaptation strategies in vulnerable areas.

2. Material and Methods

Seoul (latitude: 37.41°N-37.72°N, longitude: 126.74°E -127.21°E, Fig. 1), the study site, is the capital of the Republic of Korea, with a population of approximately 9.6 million, and 61.4% of its total land area (605 km^2) is devoted to urban use (Seoul Metropolitan Government, 2025). According to the Köppen climate classification, Seoul is Dwa, and summers are humid and hot, with the average annual temperature of 12.8℃, with the hottest month being August, when the average temperature is 26.1℃ (KMA, 2020). Seoul is geographically surrounded by several mountains, with the Han River flowing through its center. Seoul is characterized by a diverse landscape, where highly developed urban areas coexist with surrounding natural spaces, reflecting its status as a major metropolis in Asia.

This study was designed following the flow shown in Fig. 2. In this study, we defined summer high temperatures not detected by the reference station network as the calculated temperature difference between the two sensor datasets. To achieve this, we compared gridded temperature maps estimated from each sensor model. In



(a) Aerial photograpy

(b) Land cover category

Fig. 1. Seoul, the study area: (a) shows an aerial image of Seoul, and (b) presents a land cover classification based on categories defined by the Korean Ministry of Environment. The urban cover areas are subdivided into six categories (indicated by U in the legend) to show more detail, and the other categories refer to the criteria of the major classification

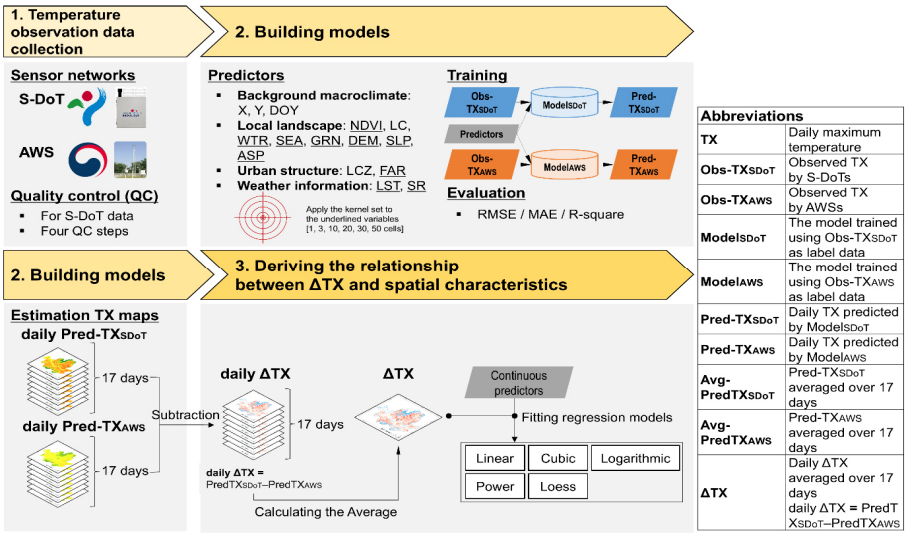


Fig. 2. Research flow. Abbreviations are re-listed here to aid the reader's understanding of the following content

Section 2.1, QC was applied to the raw data from all S-DoT sensors to ensure the reliability of micro-scale temperature measurements. In Section 2.2, daily maximum temperatures (Obs-TX_{SDoT}, Obs-TX_{AWS}) for each station were extracted from quality-controlled S-DoT data and observations from reference stations, and a random forest (RF) algorithm and auxiliary variables were used to construct two temperature estimation models ($Model_{SDOT}$, $Model_{AWS}$). These models were then used to perform spatial interpolation of areas without stations to generate 30m spatial resolution gridded TX maps (Pred-TX_{SDoT}, Pred-TX_{AWS}) for the entire Seoul area. In Section 2.3, we compared the gridded TX maps derived from S-DoT and reference stations, analyzing the relationship between ⊿TX and spatial characteristics.

2.1. Temperature Observation Data Collection

2.1.1. Overview of Temperature Observation Sensor Networks

This study utilized two temperature observation sensor networks: the Seoul Metropolitan Government's S-DoT and the Automatic Weather System (AWS) operated by the Korea Meteorological Administration (KMA). S-DoT, operating since 2020, consists of about 1,100 street-level

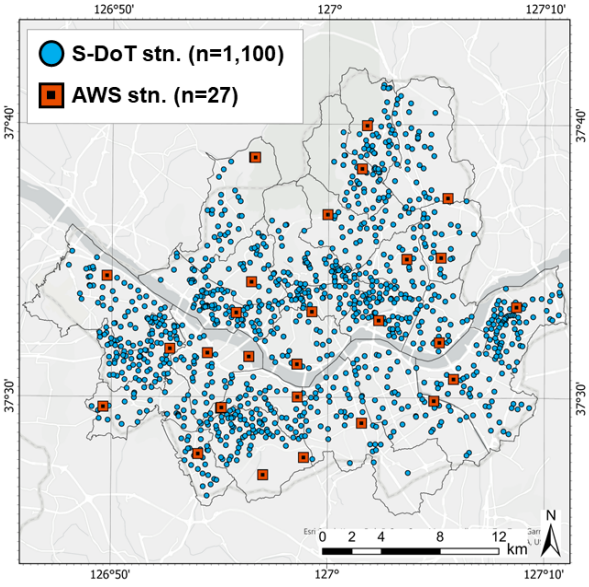


Fig. 3. Spatial distribution of two types of observation stations. The borders in the background represent Seoul's 25 autonomous districts

sensors. In contrast, AWS comprises 27 reference stations established under strict national standards, providing highly reliable baseline data. Fig. 3 shows the spatial distribution of the two networks, and the key characteristics of both are summarized in Table 1.

All S-DoT sensor devices are installed outdoors and

Table 1. Key characteristics of the two types of observation stations

Feature	S-DoT	AWS
Operator	Seoul Metropolitan Government	Korea Meteorological Administration (KMA)
Operation since	April 2020–	1980s–
Num. of stations in Seoul	about 1,100	27
Observation interval	2 min (processed to hourly averages)	1 min (published hourly/daily)
Measured variables	Particulate matter (PM2.5/PM10), Temperature, Humidity, Illumination, Noise, Vibration, Ultraviolet (UV), Wind speed and direction, and Floating population	Temperature, Humidity, Precipitation, Wind speed and direction, and Air pressure
Installation height	Most sensors are installed on CCTV poles at a height of 3–4 m, with some installed on walls or rooftops	Sensors are positioned 1.0–1.2 m above ground level or rooftop surfaces.
Atmospheric layer of installation	Urban canopy layer	Between the top of the urban canopy layer and the top of the roughness sublayer
Data source	Seoul Open Data Plaza (https://data.seoul.go.kr/dataList/OA-15969/S/1/datasetView.do.)	KMA Data Portal (https://data.kma.go.kr/cmmn/main.do)

measure temperature with an accuracy of ±0.3℃ using sensors housed inside a shell measuring 35 cm (W)×40 cm (H). Most S-DoT sensor devices are mounted on CCTV poles at a height of approximately 3–4 meters, as shown in Fig. 4, while some are installed on building walls and rooftops (Seoul Metropolitan Government, 2020c). The majority of these stations are located within the urban canopy layer, such as along roads or near buildings, where they are influenced by heat sources such as cooling units, vehicle exhaust, and radiant heat. As a result, S-DoT captures street-level temperature variations that closely reflect the thermal environment experienced by pedestrians. The Seoul Metropolitan Government processes the raw data collected every two minutes into hourly average temperature values and makes them publicly available.

AWS serves as the standard reference for administrative districts in urban areas, as well as in mountainous, coastal, and island regions. This reference network is established and maintained under strict regulations on installation and data processing (KMA, 2016, 2019), ensuring high reliability of the published data. Most AWS stations in Seoul are located on rooftops, with temperature sensors positioned at a height of 1.0 to 1.2 m above the installation surface. These stations are located between the top of the urban canopy layer and the top of the roughness sublayer, where complex flows are likely being measured, as contributions from individual facets are well mixed due to turbulence (Oke et al., 2017).

2.1.2. Quality Control of S-DoT Observation Data

To ensure the quality of observations used for predictions, S-DoT data must undergo QC for the following reasons: (1) metadata may contain incomplete or incorrect station location information; (2) some stations may have missing observation data or timestamps; and (3) errors may occur during data transmission or when converting 2-minute observations into hourly data.

Various preprocessing processes have been applied to urban meteorological networks in previous studies, with a focus on systematic QC processes developed by several studies (Fenner et al., 2021; Meier et al., 2017; Napoly et al., 2018; Romero Rodríguez et al., 2024; Yang et al., 2022). In particular, the QC process for temperature data from Netatmo, a CWS, has been continuously refined since Meier et al. (2017) and is now available as Crowd QC plus v.1.1, a standardized QC framework (Fenner et al., 2021). This process systematically removes errors, including missing data, inaccurate location information, outliers, and persistent errors, which are common in CWS networks.

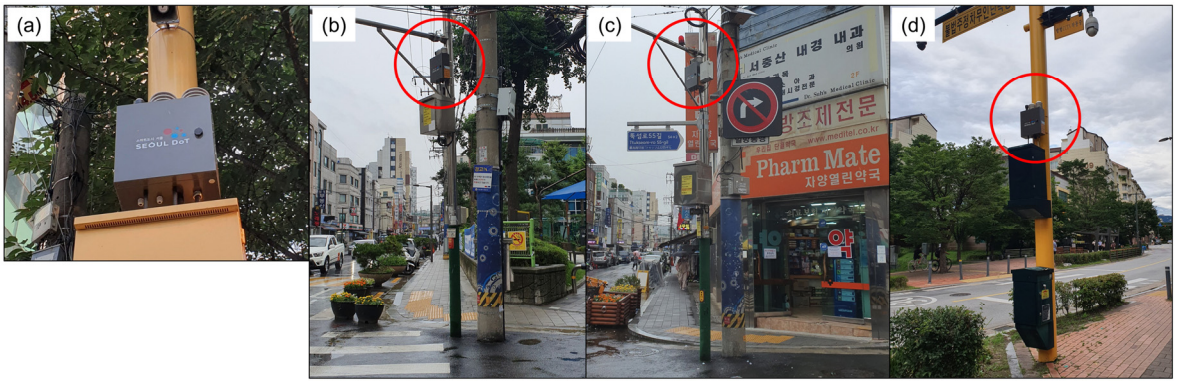


Fig. 4. S-DoT sensor devices taken by the author. (a) shows an S-DoT sensor device installed on a CCTV pole, while (b) ~ (d) depict S-DoT sensor devices mounted on CCTV poles along with their surrounding installation environments. The red circles in images (b) ~ (d) indicate the locations of the S-DoT sensor devices

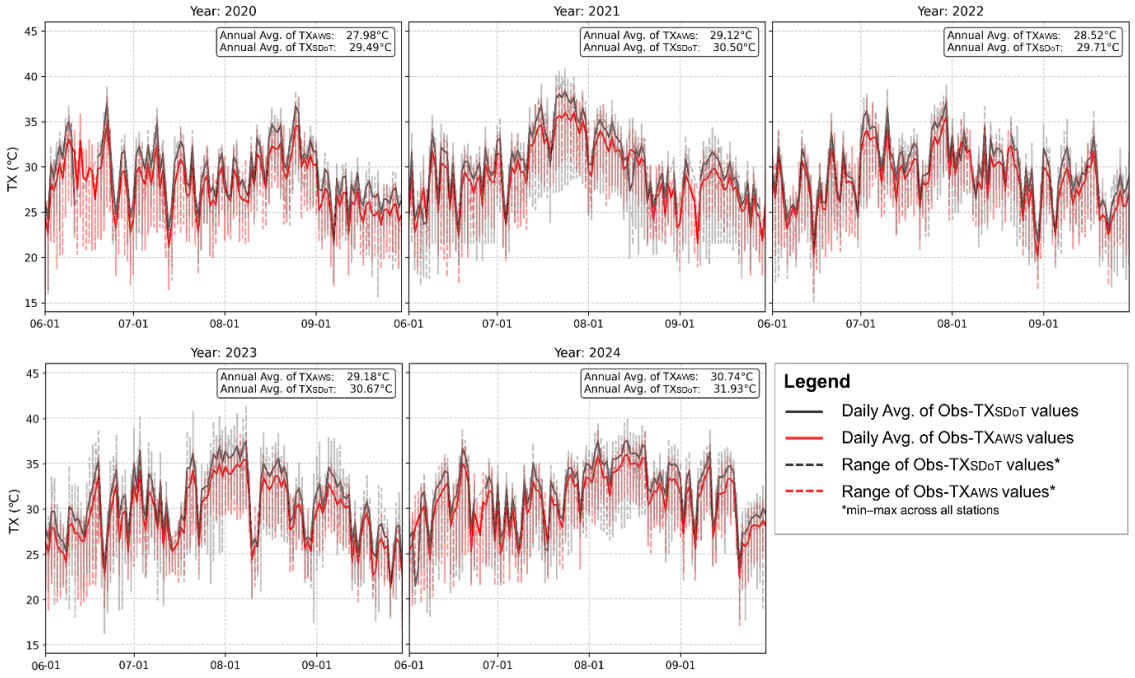


Fig. 5. Temporal variations of Obs-TXs by S-DoT and AWS. The period covers June 1 to September 30 from 2020 to 2024. Gaps in the temporal variation lines indicate periods when temperature data were missing for all stations

The S-DoT QC process implemented in this study was adapted from Crowd QC plus v.1.1 (Fenner et al., 2021) and applied to hourly temperature data recorded at all S-DoT stations from June to September during 2020–2024, corresponding to the analysis period. For further details of the QC process, see Table A1 in Appendix A. As a result of the QC process, 86.75% (11,431,942 rows) of the raw data were retained as usable. The results retained at each step of the QC process are shown in Table A2 in Appendix A.

Fig. 5 shows the temporal variation of the observed S-DoT TX values (Obs-TX_{SDOT}), retained after the QC

process, and the observed AWS TX values (Obs-TX_{AWS}) from June to September during 2020–2024, as well as the mean across all stations. Overall, the Obs-TX_{SDOT} were higher than that by AWS, and the variances per station were larger.

2.2. Building Models for Estimating Temperature

2.2.1. Model and Response Variables

We built two different RF models that estimate the TX for each dataset. The first model ($Model_{SDoT}$) used Obs-TX_{SDOT} as the response variable. These TXs were calculated from the valid hourly observation data after QC as response variable for the model predicting street-level TX. The other ($Model_{AWS}$) used Obs-TX_{AWS} of KMA as the response variable to estimate reference TXs. Both $Model_{SDoT}$ and $Model_{AWS}$ used the RF algorithm. RF is a widely used machine learning algorithm that aggregates the outputs of multiple decision trees to produce a single prediction (IBM, 2025). It has demonstrated strong performance in temperature estimation studies (Wang et al., 2023), and was therefore considered suitable for use in this study.

2.2.2. Predictors and Data Preprocessing

We selected 15 types of temporal and spatial predictors that causing the spatial urban temperature variability are known from theory (Lowry, 1977; Oke, 1982; Oke et al., 2017; Zumwald et al., 2021) and widely used in existing machine learning-based temperature estimation (Wang et al., 2023). The following 15 predictors were adopted: longitude (X), latitude (Y), and Day of the year (DOY) as variables representing background macroclimate, Normalized Difference Vegetation Index (NDVI), Land Cover (LC), Distance to fresh water (WTR), Distance to sea (SEA), Distance to green space (GRN), Digital Elevation Model (DEM), Slope (SLP), Aspect (ASP) as variables representing local landscape, Local Climate Zone (LCZ), Floor area ratio (FAR) as variables representing urban structure, Land Surface Temperature (LST), and Solar radiation (SR) as weather information. All predictors were constructed with a spatial resolution of 30 m×30 m. LC and LCZ are categorical variables, while all other predictors are continuous. Detailed explanations of each predictor, including construction methods and data sources, are provided in Table 2.

Temperature variations are influenced not only by conditions at the station's location but also by the surrounding environment. Previous studies have used buffer-averaged predictor variables to account for the effects of nearby land surface and land cover on local temperature measurements (Ho et al., 2014). Based on this approach, we applied a kernel centered on each station and tested the use of average values within the kernel as predictor variables.

Among the two-dimensionally constructed variables, spatial aggregation was applied to 10 continuous variables, while the categorical variables LC and LCZ were excluded. The kernel size was set at six levels: [1 cell, 3 cells, 10 cells, 20 cells, 30 cells, 50 cells].

2.2.3. Model Training and Evaluation

Both models were selected for optimal tuning parameters as well as for the kernel of predictors. The key tuning parameters of a RF include the number of trees, maximum depth, splitting criterion, minimum samples required for splitting, minimum samples required for a leaf, and the maximum number of features. In this study, the following candidates were given for each tuning parameter, and the best parameter was selected using grid search: the number of trees was [500, 1000, 1500], maximum depth was [10, 20, None], minimum samples required for splitting was [2, 5, 10], minimum samples required for a leaf was [1, 3, 5], the maximum number of features was [sqrt, log2, None].

Model performance was evaluated using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2), averaged over five results of 5-fold cross-validation. Due to the small size of $Model_{AWS}$ dataset, a process of averaging the five test results obtained through 5-fold cross-validation prevented overfitting to a particular set of samples.

Table 2. Fifteen types of predictors used to build the TX estimation models. The ● in the kernel set column indicates the ten continuous variables that were tested by changing the kernel size when inputting the models. The parentheses in the data source column indicate the publication year of the source

Category	Predictors	Kernel set	Construction methods	Data source	Reference
BMC	X	-	Obtained from META data	KMA and Seoul Metropolitan Government, Metadata (2024)	Shen et al. (2020); S. Chen et al. (2022)
	Y	-			
	DOY	-	Obtained from observation data	KMA and Seoul Metropolitan Government, observation data (2020~2024)	Benali et al. (2012); Shen et al. (2020)
LL	NDVI	●	Calculated from bands 4 and 5 of Landsat 8	USGS, Landsat 8 (2020-2024)	Benali et al. (2012); Ho et al. (2014); dos Santos (2020); Shen et al. (2020); Yu et al. (2024)
	LC	-	Obtained the land cover map sorted into 22 categories	ME, 41 categories land cover map (2020)	Shen et al. (2020); S. Chen et al. (2022)
	WTR	●	Calculated Euclidean distance using land cover map		Benali et al. (2012); Yu et al. (2024)
	SEA	●			Benali et al. (2012)
	GRN	●			Chajaei and Bagheri (2024)
	DEM	●	Obtained SRTM	NASA, SRTM 1 Arc-Second Global (2014)	Ho et al. (2014); Shen et al. (2020)
	SLP	●	Calculated using DEM		S. Chen et al. (2022)
	ASP	●			S. Chen et al. (2022)
US	LCZ	-	Obtained from the Global LCZ Map	Demuzere et al. (2023)	S. Chen et al. (2022)
	FAR	●	Calculated using information on the number of floors in a building. It differs from the actual floor-area ratio in city planning.	MOLIT, Building integrated information (2023)	Lau and Lin (2024)
WI	LST	●	Calculated from bands 4, 5 and 10 of Landsat 8	USGS, Landsat 8 (2020-2024)	Ho et al. (2014); dos Santos (2020); Shen et al. (2020); Yu et al. (2024)
	SR	●	Using the DSM created by overlaying building data on the DEM, the ArcGIS Points Solar Radiation tool calculates insulation for the whole day.	NASA, SRTM 1 Arc-Second Global (2014) & MOLIT, Building integrated information (2023)	Ho et al. (2014); Yu et al. (2024)

BMC: Background MacroClimate; LL: Local landscape; US: Urban structure; WI: Weather information; MOLIT: Ministry of Land,Infrastructure and Transport, South Korea; ME: Ministry of Environment, South Korea; DSM: Digital Surface Model

2.2.4. Estimation of Gridded TX Maps

Using the models, TX maps were predicted with 30m×30m spatial resolution for the entire Seoul for 17 days during the period from June to September 2020 to 2024. Five predicted TX_{SDOT} and TX_{AWS} were estimated for each of the five models built by 5-fold cross-validation. These five predicted TX values were averaged to obtain daily predicted TX_{SDOT} (Pred-TX_{SDoT}) and daily Pred-TX_{AWS} (Pred-TX_{AWS}). We further obtained Avg-PredTX_{SDoT} and Avg-PredTX_{AWS} by calculating the average of these 17-day Pred-TXs, respectively.

2.3. Deriving the Relationship between ⊿TX and Spatial Characteristics

To compare the spatial distribution of Pred-TX_{SDoT} and Pred-TX_{AWS}, we calculated the difference between Pred-TX values for each day (daily ⊿TX) during the 17 days for which TX maps were generated. The calculation follows the formula (1):

$$daily\ \Delta TX = \mathrm{PredTX}_{SDoT} - \mathrm{PredTX}_{AWS} \quad (1)$$

A positive daily ⊿TX indicates that Pred-TX_{SDoT} is higher than Pred-TX_{AWS}, while a negative daily ⊿TX indicates that Pred-TX_{SDoT} is lower than Pred-TX_{AWS}.

We obtained ⊿TX map by calculating the average of these 17-day daily ⊿TX maps. In addition, to identify spatial areas where statistically significant differences in hot and cold spots occurred, we derived ⊿Gi Bin value maps. The Gi Bin values identify statistically significant hot and cold spots, adjusted for multiple testing and spatial dependence using the False Discovery Rate correction (Esri, 2025). Features in ±3 bins are significant at the 99% confidence level, ±2 bins at 95%, and ±1 bins at 90%, while 0 bins indicate no significant clustering. The ⊿Gi Bin values were obtained as integers ranging from −6 to +6.

To further investigate the spatial patterns of ⊿TX, we analyzed its relationship with variables representing the spatial characteristics of Seoul. First, we calculated the Pearson correlation coefficient between ⊿TX and 10 predictor variables used in the model. These variables included NDVI, LC, WTR, SEA, GRN, DEM, SLP, and ASP as local landscape variables, and LCZ and FAR as urban structure variables. We then explored these relationships by fitting the data with four regression models: linear, cubic, logarithmic, and power. In addition, local polynomial regression was performed using the Locally Weighted Scatterplot Smoothing (LOWESS) regression from the Python package *statsmodels*.

3. Results

3.1. Deriving ⊿TX between Two Sensor Networks

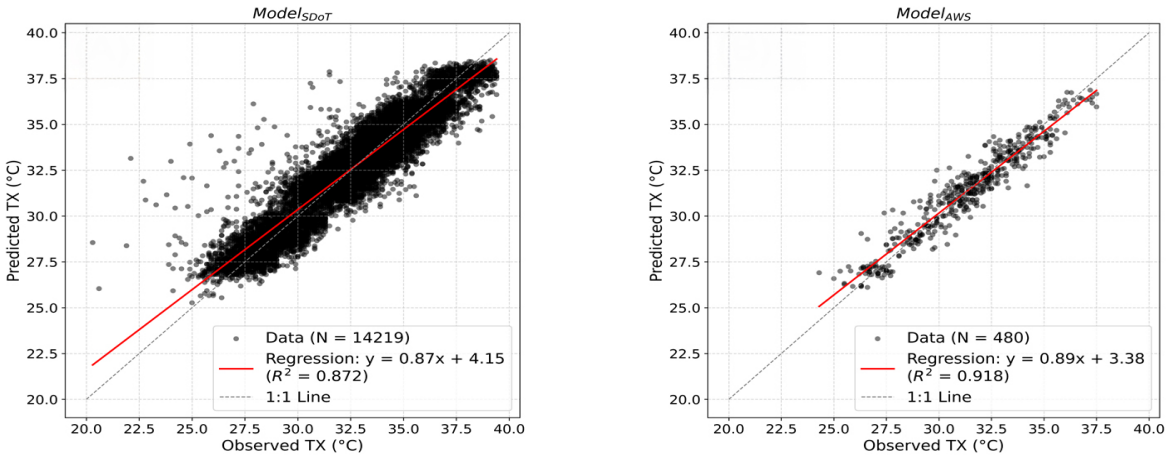
3.1.1. Performance of TX Estimation Models

Table 3 summarizes the evaluation results of the best-performing models ($Model_{SDoT}$ and $Model_{AWS}$) for each label dataset. $Model_{SDoT}$ achieved the best performance with a kernel size of 3 cells, yielding an RMSE of 1.019℃, MAE of 0.772℃, and R^2 of 0.872. $Model_{AWS}$ performed best with a kernel size of 30 cells, achieving an RMSE of 0.762℃, MAE of 0.582℃, and R^2 of 0.913. Overall, $Model_{AWS}$ tended to outperform $Model_{SDoT}$, but $Model_{SDoT}$'s performance was still comparable to that reported in previous AI-based urban temperature estimation studies (Ho et al., 2014; Shen et al., 2020; Venter et al., 2020; Wang et al., 2023; Zumwald et al., 2021). Additional test results for other kernel sizes, as well as the evaluation outcomes from the five tests conducted with five-fold cross-validation, are presented in Table B1 in Appendix B.

Fig. 6 shows the distribution of predicted TX and observed TX values for the test samples of each fold used to train the best models. Both $Model_{SDoT}$ and $Model_{AWS}$ exhibited a tendency to overestimate temperatures in the lower range and underestimate temperatures in the higher range. Since this trend has also been observed in previous studies (Zumwald et al., 2021), it was considered to be within an acceptable range of errors.

Table 3. Performance of the best models. The values represent the average evaluation metrics (RMSE, MAE, R^2) across five tests with five-fold cross-validation

Label data	Kernel size	RMSE	MAE	R^2	Tuning parameters
S-DoT	3	1.019	0.772	0.872	Number of trees = 2,000; maximum depth = None; minimum samples required for splitting = 10; minimum samples required for leaf = 5; maximum number of features = None
AWS	30	0.762	0.582	0.913	Number of trees = 1,500; maximum depth = None; minimum samples required for splitting = 2; minimum samples required for leaf = 3; maximum number of features = None
Other tuning parameters of the Random Forest regressor were kept at their default values as implemented in scikit-learn (version 1.6.1).					



(a) $Model_{SDoT}$ (b) $Model_{AWS}$

Fig. 6. Relationship between predicted and observed TX values in test samples from 5-fold cross-validation. A red line indicates the linear regression fit

Table 4. Descriptive statistics of the Avg-PredTX (17 days) estimated by $Model_{SDoT}$ and $Model_{AWS}$

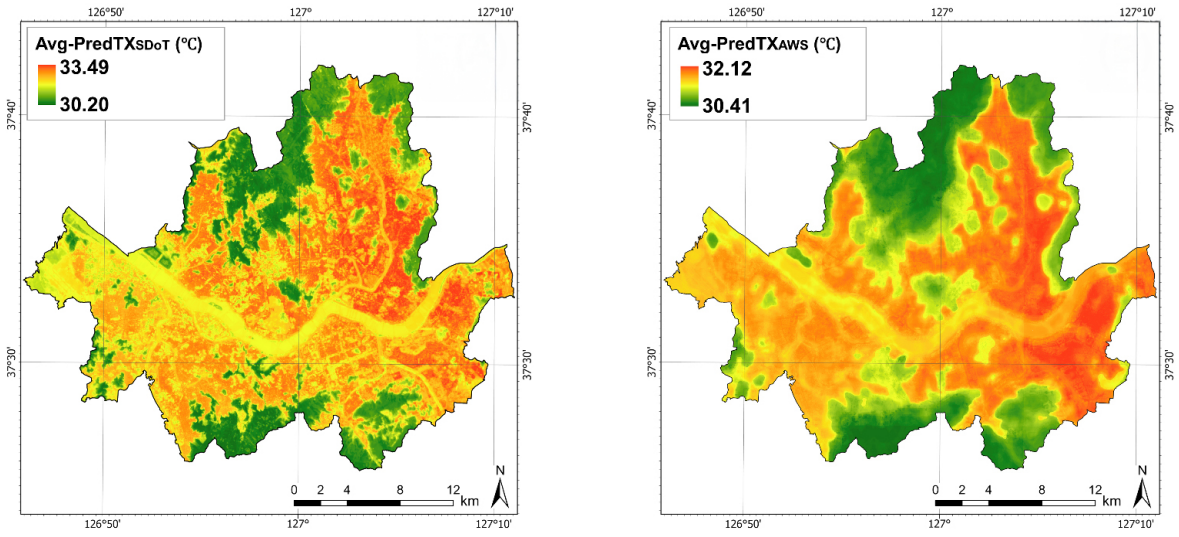
Model	$Model_{SDoT}$				$Model_{AWS}$			
Statistic	Min.	Max.	Ave.	Std.	Min.	Max.	Ave.	Std.
Pred-TX(℃)	30.2	33.49	32.04	0.78	30.41	32.12	31.35	0.44

3.1.2. Estimation of Gridded TX Maps

Table 4 presents the descriptive statistics of the Avg-PredTX over 17 days estimated by $Model_{SDoT}$ and $Model_{AWS}$, and Fig. 7 shows the corresponding maps. The descriptive statistics of the daily Pred-TX are provided in Table B2 in Appendix B. The Avg-$PredTX_{SDoT}$ exhibited a wider temperature distribution than Avg-$PredTX_{AWS}$, with values ranging from 30.2℃ to 33.49℃, whereas Avg-$PredTX_{AWS}$ ranged from 30.41℃ to 32.12℃. Compared to Avg-$PredTX_{AWS}$, Avg-$PredTX_{SDoT}$ predicted hot areas to be hotter, and cool areas to be cooler, indicating a stronger contrast in temperature variations. These results were consistent with the relationship between TX variations observed by S-DoT and AWS shown in Fig. 5.

3.1.3. Calculation of ⊿TX

The descriptive statistics of the 17-day average ⊿TX ranged from -0.93℃ to 2.36℃, with a mean value of 0.69℃, as shown in Table 5. The descriptive statistics of the daily ⊿TX calculated between $Model_{SDoT}$ and



(a) (b)

Fig. 7. Avg-PredTX maps averaged over 17 days by each model. (a) represents Avg-PredTX$_{SDoT}$, predicted by Model$_{SDoT}$, while (b) represents Avg-PredTX$_{AWS}$, predicted by Model$_{AWS}$.

Table 5. Descriptive statistics of ⊿TX (17-day average)

Statistic	Min.	Max.	Ave.
⊿TX (℃)	-0.93	2.36	0.69

Model$_{AWS}$ are provided in Appendix B (Table B3).

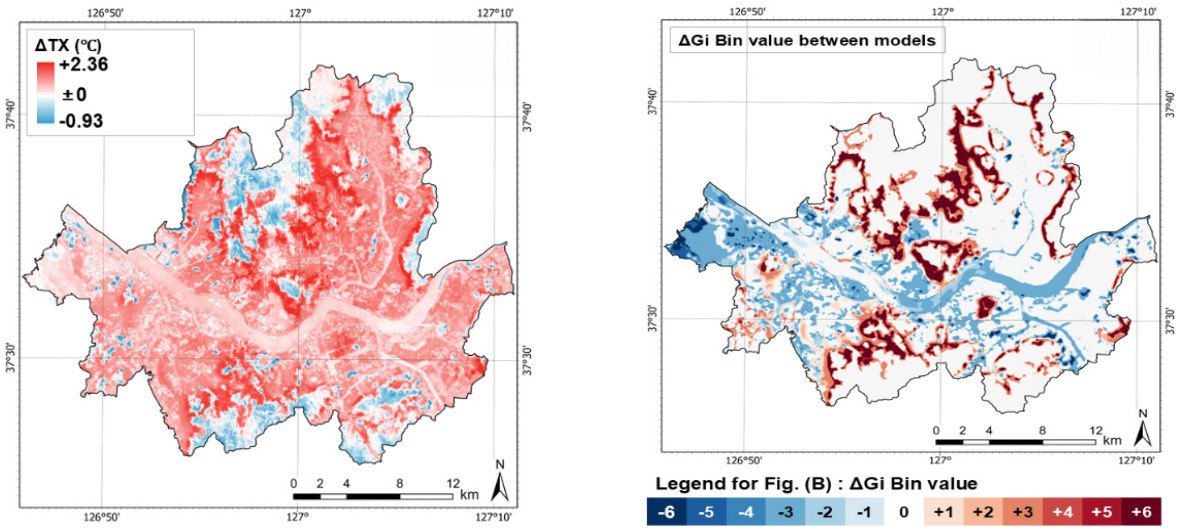
Fig. 8(a) illustrates the gridded ⊿TX map derived from the daily ⊿TX maps. Fig. 8(b) presents the gridded map of difference in Gi Bin values (⊿ Gi Bin), calculated from Avg-PredTX$_{SDoT}$ and Avg-PredTX$_{AWS}$.

3.2. The Relationship between ⊿TX and Spatial Characteristics

Table 6 presents the Pearson correlation coefficients between ⊿TX and continuous predictor variables representing T_L and T_U used in the TX estimation model. It also includes the fitting results for Linear, Log, Cubic, Power and LOWESS regressions.

Among the eight variables, NDVI, GRN, DEM, SLP and FAR showed the most significant correlations with ⊿TX, based on both the Pearson correlation coefficients and the model fitting results. These variables exhibited a statistically significant strong correlation with ⊿TX, with the correlation coefficients ranked in the following order: NDVI, SLP, DEM, GRN, and FAR. Additionally, the fitting analysis revealed that the LOWESS regression provided the best explanation for the variation in ⊿TX. The R^2 values were highest for FAR, NDVI, DEM, GRN, SLP, in that order.

Fig. 9 presents the results of fitting the relationship between ⊿TX and four variables (NDVI, GRN, DEM, SLP and FAR) to the LOWESS regression. The relationship between NDVI and ⊿TX in Fig. 9(a) shows that the increase in ⊿TX was the largest when NDVI was 0.09, and the decrease in ⊿TX was the largest when NDVI was 0.46. And at NDVI = 1.07 the trend of ⊿TX > 0 was maximum. The relationship between GRN and ⊿TX in Fig. 9(b) shows that ⊿TX increased as one moved away from the green space, reaching a maximum at GRN = 90 m. Thereafter, ⊿TX decreased slowly up to 216.33 m, and when GRN was greater than that, ⊿TX remained positive and its value decreased. It is also characteristic that samples with negative ⊿TX are concentrated at GRN = 0. The relationship between DEM and ⊿TX in Fig. 9(c) shows a strong trend of positive ⊿TX in the range of 12.40 m to 59.20 m DEM. On the other hand, the value of ⊿TX becomes smaller as DEM increases, especially in the range of 150 m to 400 m



(a) (b)

Fig. 8. Map of the difference between two Avg-PredTX values. (a) shows the gridded ⊿TX map, while (b) shows the gridded ⊿Gi Bin map. Both ⊿TX and ⊿Gi Bin were calculated by subtracting Avg-PredTX$_{AWS}$ from Avg-PredTX$_{SDoT}$

Table 6. Results of Pearson correlation coefficients and four types of regression fits between ⊿TX and continuous variables representing spatial characteristics

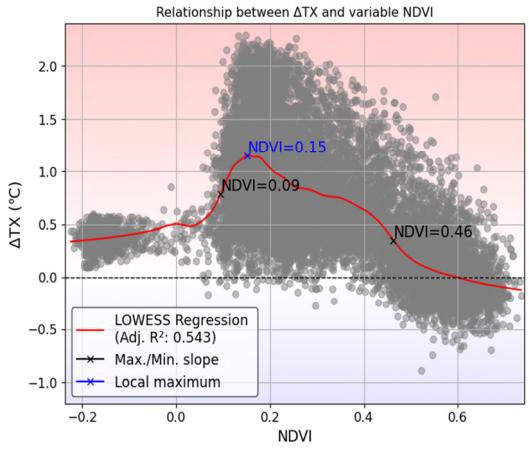
Category	Predictors	Corr.	Adjusted R^2				
			Linear	Cubic	Log	Power	LOWESS
Local landscape	NDVI	-0.544**	0.296	0.516	0.242	0.112	0.543 (0.1)
	WTR	-0.017**	-0.001	0.017	0.013	-0.03	0.037 (0.1)
	SEA	0.067**	0.001	0.006	0.002	-0.036	0.036 (0.1)
	GRN	0.285**	0.089	0.318	0.3	0.269	0.341 (0.4)
	DEM	-0.415**	0.183	0.212	0.088	0.022	0.394 (0.2)
	SLP	-0.455**	0.223	0.254	0.14	0.089	0.275 (0.2)
	ASP	0.055**	0.003	0.01	0.003	-0.036	0.014 (0.1)
Urban structure	FAR	0.160**	0.027	-0.228	0.165	0.088	0.546 (0.5)

**$p < 0.001$ indicates statistical significance.

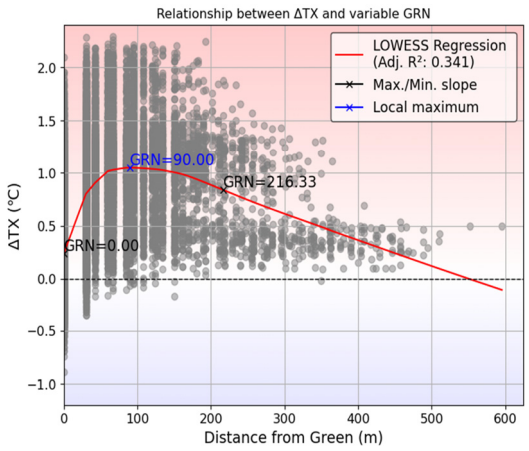
Values in parentheses for LOWESS regressions represent optimal smoothing parameters selected within the range of 0.1 to 0.7.

DEM, where the trend of negative ⊿TX is stronger. Fig. 9(d) shows the relationship between SLP and ⊿TX. ⊿TX reaches its maximum at a slope of 5.58°, but the value of ⊿TX is smaller at slopes of 11.58° or higher. Finally, Fig. 9(e) shows the relationship between FAR and ⊿TX, where ⊿TX reached a maximum when FAR = 1.07%, and then decreases significantly as FAR increases. As with GRN in Fig. 9(b), there was a concentration of samples with negative ⊿TX when FAR = 0.

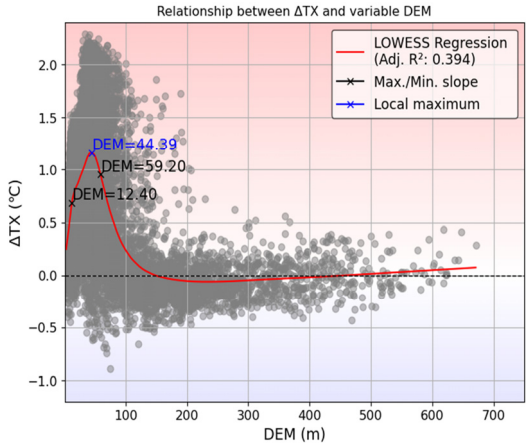
Fig. 10 presents the ⊿TX values for each category of LC and LCZ, which are categorical variables in the TX estimation model. While the average ⊿TX for Seoul was 0.69℃, Fig. 10(a) shows that ⊿TX was notably high in urban land cover categories, including Residential, Industrial, Commercial, Cultural, Transportation, and Public, with the highest values observed in Residential areas. In contrast, the median ⊿TX for forest land cover was 0℃, indicating that half of all samples in this



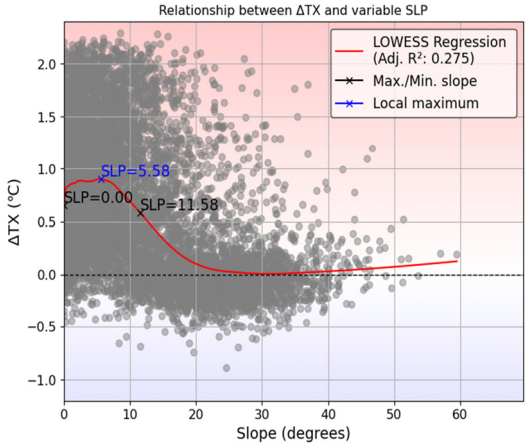
(a) NDVI vs. ⊿TX



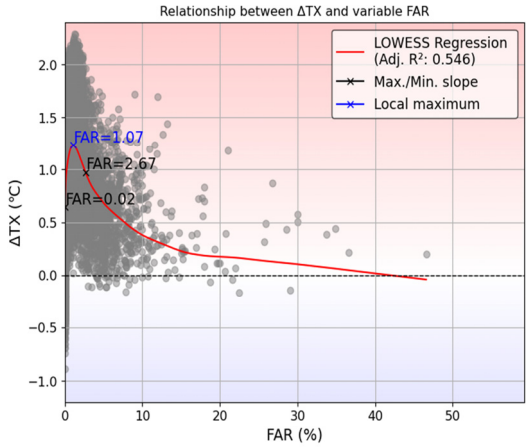
(b) GRN vs. ⊿TX



(c) DEM vs. ⊿TX

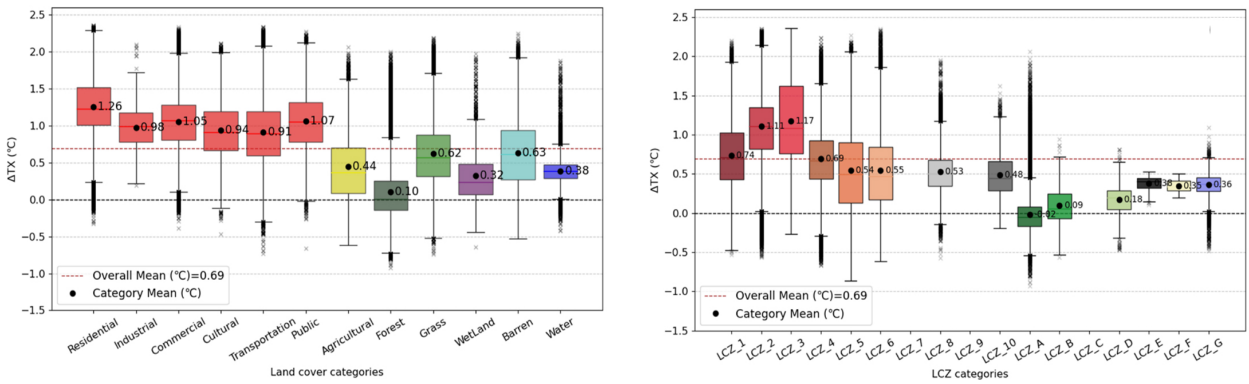


(d) SLP vs. ⊿TX



(e) FAR vs. ⊿TX

Fig. 9. Plots of the relationship between ⊿TX and five variables (NDVI, GRN, DEM, SLP, and FAR) with fitted LOWESS regressions. The LOWESS regressions illustrate the fitted relationships for each variable, allowing identification of the values at where ⊿TX shows stronger sensitivity



(a) (b)

Fig. 10. ⊿TX values for each LC and LCZ category. (a) shows ⊿TX by LC category, and (b) shows ⊿TX by LCZ category. In (a), the original land cover classification map contained 22 categories used for model training; however, for better visualization, all non-urban categories were grouped into five major categories, while the six urban land cover types were shown separately

category have negative ⊿TX. In Fig. 10(b), LCZ1, LCZ2, LCZ3, and LCZ4 demonstrated a strong tendency for positive ⊿TX, with LCZ2 and LCZ3 showing particularly high values. These categories correspond to Compact High-rise (LCZ1), Compact Mid-rise (LCZ2), Compact Low-rise (LCZ3), and Open High-rise (LCZ4). Conversely, LCZ-A through G, representing natural land cover types, had relatively small ⊿TX values, with LCZ-A (Dense Trees) in particular showing a significantly negative ⊿TX trend.

4. Discussions

4.1. Spatial Characteristics Affecting ⊿TX

By integrating the results, it was observed that the tendency for positive ⊿TX is most pronounced in urban land cover near the boundary between natural and urban areas, particularly in compact low- to mid-rise residential zones. In contrast, the tendency for negative ⊿TX is stronger in natural land cover, especially in mountainous regions characterized by relatively steep slopes and dense forest cover.

First, discussing the positive ⊿TX, the factors for which a relationship was identified—low DEM values, small slopes, low-positive NDVI values, urban land cover, and high-density LCZ categories—are characteristic features of urban canyons. Although measured by mobile measurements, previous studies have consistently demonstrated that temperatures within urban canyons tend to be higher than those recorded by reference stations on rooftops (Pigliautile and Pisello, 2020; Tsin et al., 2016; Vuckovic et al., 2017). Moreover, research on the relationship between LCZ classifications and temperature variations has confirmed that ⊿TX is significantly elevated in urban canyon areas characterized by compact mid- and low-rise urban morphology. For instance, Beck et al. (2018) reported that during summer, the daytime TX difference between micro- and local-scale levels reached approximately 2.0℃ in four urban LCZ categories. The highest positive temperature differences were observed in the following order: Compact Mid-rise, Open Mid-rise, Open Low-rise, and Large Low-rise (Beck et al., 2018).

In contrast, the relationship between positive ⊿TX and proximity to green spaces has not been clearly addressed in previous studies. While many studies have investigated temperature variability and the cooling effects of green spaces at the street level (Bowler et al., 2010; Galalizadeh et al., 2024; Jaganmohan et al., 2016; Shashua-Bar and Hoffman, 2000), very few have compared the cooling influence of green spaces or temperature variations near green areas across different vertical levels. However, in

the context of urban climate research, it has been demonstrated that thermal exchange processes differ substantially between the street-level and the rooftop-level (Oke et al., 2017; World Meteorological Organization, 2023). Consequently, the distance and intensity of the cooling influence of green spaces may vary depending on the vertical height at which measurements are taken, which could explain the observed relationship between positive ⊿TX and proximity to green spaces. However, this study alone is insufficient to fully elucidate the mechanism, and further detailed analyses are required. Nevertheless, our findings highlight an important implication: in urban areas near green spaces, pedestrians have experienced higher summer temperatures than those suggested by Pred-TX_{AWS} from conventional reference stations. This finding is particularly significant in Seoul, where residential areas are located immediately adjacent to major green spaces, as shown in Fig. 1(b).

The positive ⊿TX is a consequence of the S-DoT sensors ability to capture temperature variability within the urban canyon. The warming factors in the urban canopy layer, defined by Oke (1982) as increased absorption and delayed emission of solar radiation, reduced vegetation, and increased sensible heat from human activity (Stewart and Oke, 2012), are consistent with spatial characteristics that are significantly associated with a positive ⊿TX. These results for positive ⊿TX emphasize the important role of street-level sensors in dense urban areas, especially those bordering green spaces. Unlike reference stations, which primarily capture large-scale temperature trends, these street-level sensors can detect localized heat variations, providing valuable supplementary data. Their ability to capture fine-scale thermal dynamics highlights their importance in improving urban climate assessments and informing adaptation strategies.

We then examined the spatial characteristics that are negative ⊿TX and found that the installation environment between sensor networks affects the negative ⊿TX. Fig. 11 shows five S-DoT stations with negative ⊿TX and dense forest, and Fig. 12 shows the installation environment of AWS stations in dense forest. Although the land cover of AWS stations was forest, they were installed above the tree canopy or away from trees based on KMA's strict installation criteria. On the other hand, we found that these S-DoT stations were installed below

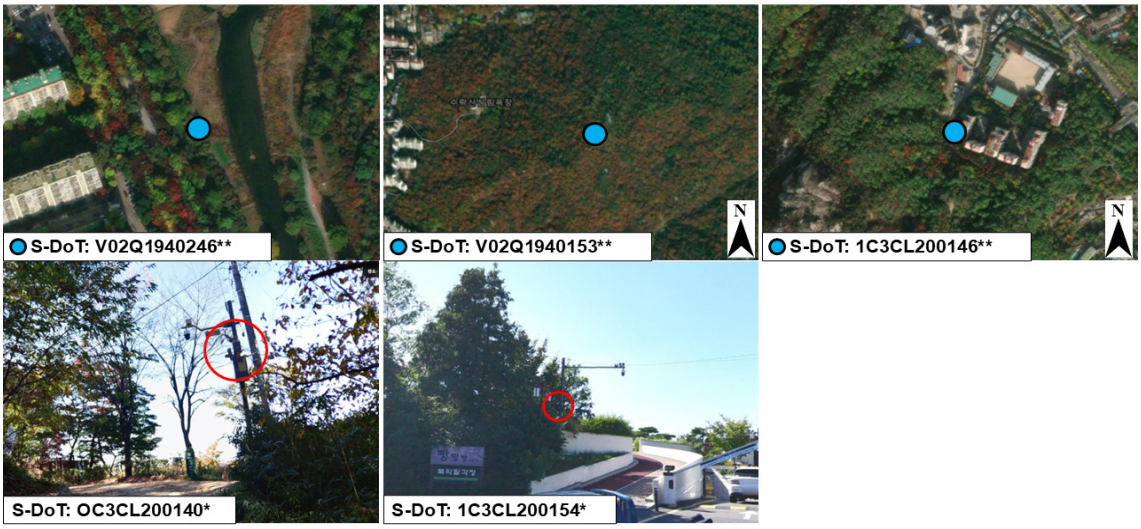


Fig. 11. Installation environment of S-DoT sensors in dense forest with negative ⊿TX. *indicates the installation environment confirmed by Naver Street View, and the red circle indicates the location of the S-DoT sensor. **indicates the surrounding environment of the S-DoT sensor confirmed by aerial photographs

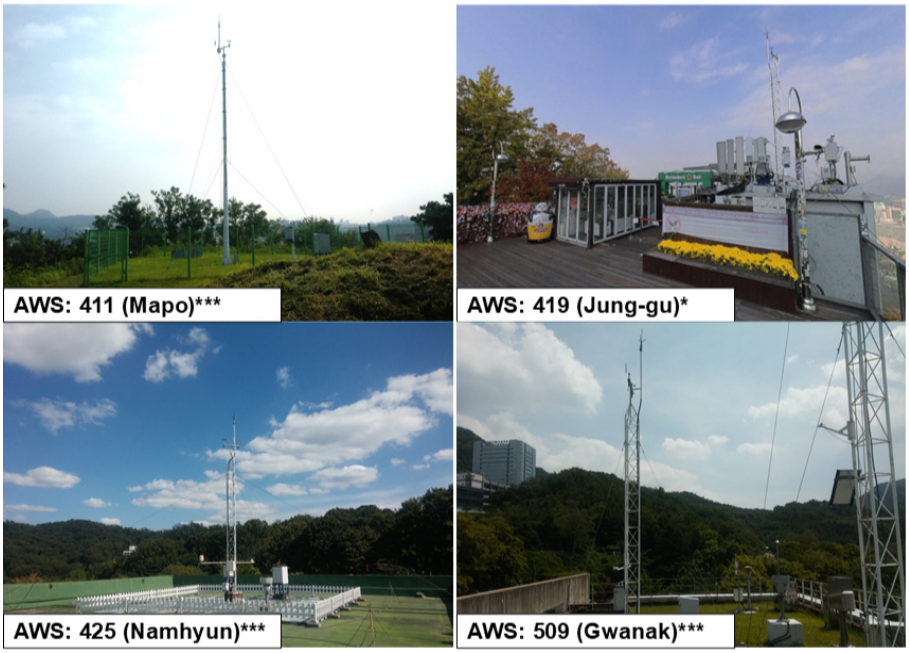


Fig. 12. AWS sensor installation environment within dense forest. *indicates the installation environment confirmed on Naver Street View. ***indicates the installation environment confirmed on the KMA website

the tree canopy. These factors led us to believe that the S-DoTs would be installed at a height of about 3–4 m above the ground, and that their measurements would be affected by the cooling effect of the shade trees in summer when they were installed in forest cover.

These findings for negative ⊿TX indicate that street-level sensors have successfully fulfilled their primary objective—capturing the cooler temperatures experienced by pedestrians beneath the tree canopy in dense forested areas with steep slopes. However, considering the strategic deployment of a limited number of sensors to identify areas experiencing severe summer heat and mitigate its adverse effects on public health, deploying additional street-level sensors in such locations may not be the most effective approach. Because existing studies have already demonstrated that forest areas are relatively cooler than urban areas, benefiting from evapotranspiration and tree shade (Li et al., 2024; Park et al., 2021). Furthermore, our study showed that when street-level sensors are installed under tree canopy, they tend to show lower temperatures than reference stations in these areas. We suggest that placing street-level sensors in hotter areas of the city, especially those with positive ⊿TX spatial characteristics, would be more effective.

4.2. Impact of Prediction Uncertainty on ⊿TX

As shown in the scatterplots in Fig. 9, the large variance of ⊿TX across different values of the variables suggests that factors beyond sensor placement environments may also contribute to the observed variations. Therefore, to gain a deeper understanding of other potential drivers of ⊿TX, we considered the impact of prediction uncertainty based on the TX estimation results presented in Sections 3.1.1 and 3.1.2. Zumwald et al. (2021) defined two types of prediction uncertainty: (1) estimated prediction uncertainty, which arises when similar configurations of predictors map to different target values in the training data and is estimated by the prediction algorithm; and (2) extrapolation uncertainty, which occurs when the model predicts temperatures based on predictor configurations not observed during training.

To evaluate estimated prediction uncertainty, we

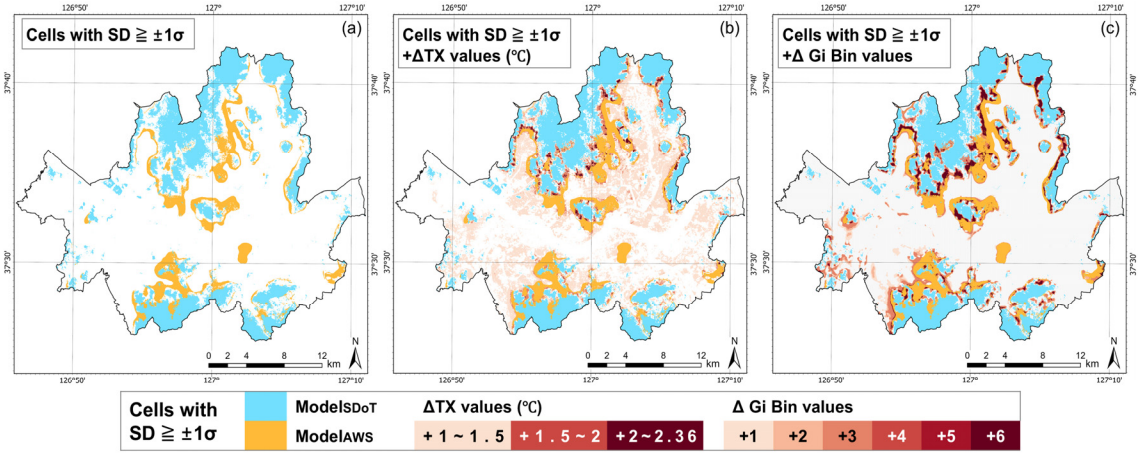


Fig. 13. Cells where the standard deviation of the estimates from five models exceeds ±1σ. (a) shows cells where the standard deviation of the five estimates of each model is greater than ±1σ, (b) shows their positional relationship with ΔTX values, and (c) with ΔGi Bin values

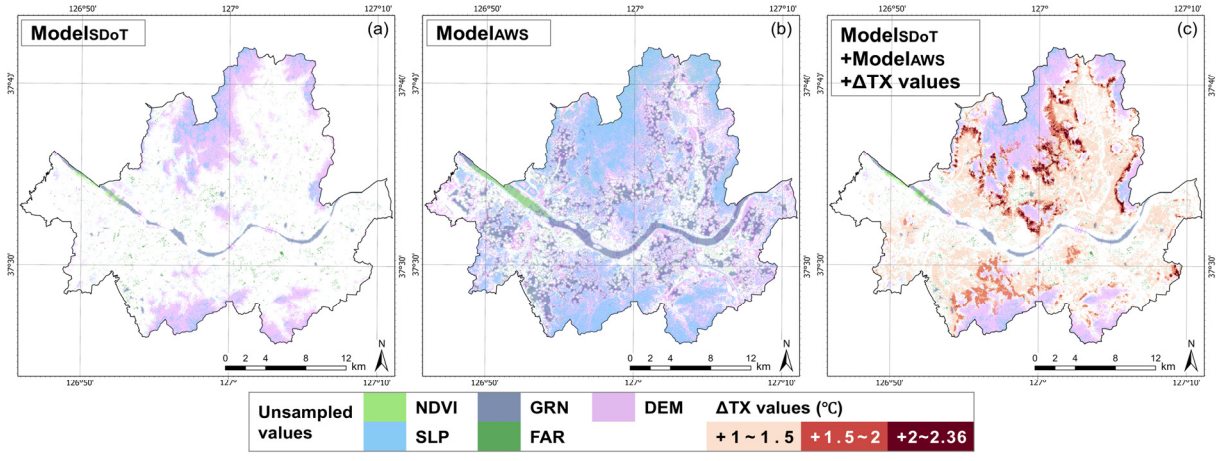


Fig. 14. Spatial relationship between areas of unsampled values and large positive ΔTX for predictors NDVI, GRN, DEM, SLP, and FAR. (a) shows the extrapolated areas of the predictors, (b) shows the overlay of these areas with large positive ΔTX, and (c) shows the overlay with positive ΔGi Bin values

calculated the standard deviation of predictions from the five models generated during the five-fold cross-validation as shown in Fig. 13. We highlighted cells where the standard deviation of these estimates exceeded ±1σ across Seoul. The results indicated that the uncertainty of $Model_{AWS}$ tends to be larger in urban areas with positive ΔTX and spatial characteristics related to that value.

Next, Fig. 14 shows the spatial extent of values that were not sampled in the training data for each of the five variables—NDVI, GRN, DEM, SLP, and FAR—in order to identify areas where extrapolation uncertainty may occur. In the case of $Model_{AWS}$ shown in Fig. 14(b), values absent from the training dataset were distributed throughout Seoul. However, as shown in Fig. 14(c), when these were combined with the $Model_{SDoT}$ samples, it became evident that the areas where positive ΔTX occurred and those with extrapolation uncertainty could be more comprehensively covered.

Through our consideration of prediction uncertainty, we confirmed that positive ΔTX includes estimation error due to the uncertainty of the reference stations, and that the uncertainty in this monitoring can be reduced by

increasing the number of observation stations. The low spatial density and the installation environment constrained by strict installation regulations of the reference stations resulted in low sample diversity in $Model_{AWS}$, and the resulting estimation uncertainty is considered to influence ⊿TX. These results emphasize that while increasing the number of sensors through urban IoT deployment is an important strategy, it is especially crucial to strategically place sensors in areas with spatial characteristics that are not covered by the reference station network. Doing so allows for a more efficient enhancement of sample diversity and improves the overall certainty of urban temperature monitoring.

4.3. The Effectiveness of Urban IoT Sensor Network

The results of this study suggest that street-level sensors can offer critical insights into high summer temperatures, particularly those associated with heat-related illnesses and mortality, which are not captured by reference stations alone. Furthermore, we found that incorporating an IoT sensor network improves the reliability of temperature monitoring compared to relying solely on the existing reference station network. As shown in Table B2, even reference temperatures, which are generally cooler than street-level temperatures in urban areas, averaged over 31℃ across Seoul. On an exceptionally hot day (August 13, 2024), the average pred-TX_{AWS} exceeded 36℃. These temperatures are high enough to cause heat stroke, death, and other serious health risks.

However, Avg-pred$TX_{S\text{-}DOT}$ was even higher, by +0.69℃ on average, with a maximum difference of +2.36℃ than Avg-predTX_{AWS}. This indicates that adding a street-level sensor network increases the likelihood of detecting critical thermal risks with higher certainty. Moreover, these positive ⊿TX values and prediction uncertainties were especially pronounced in areas close to green spaces and characterized by compact, low- to mid-rise, primarily residential urban forms. These areas are where citizens live and experience daily environmental conditions, so the significance of knowing with greater certainty high summer temperatures that cannot be detected by reference stations alone is emphasized. Even in neighborhoods near green areas, street-level temperatures can reach dangerous levels, which underscores the need for evidence-based responses based on measurements.

As previous studies have suggested (Honjo et al., 2015; Lam et al., 2021; Sakthivel and Sengupta, 2025; Shi et al., 2021), our findings support the notion that urban IoT sensor networks can complement reference stations and play an important role in filling the gap in hazard assessment due to high temperatures.

In particular, we propose that placing street-level sensors near the boundary between green spaces and urban land cover, where the positive ⊿TX trend is stronger as described in Section 4.2, can more effectively compensate for the limitations of the reference station network. This approach is especially valuable in developing adaptive strategies for mitigating street-level heat exposure during summer.

5. Conclusion

This study aimed to quantify high summer temperatures that cannot be detected by reference stations but can be identified by street-level IoT sensors, and to clarify their relationship with spatial characteristics. To achieve this, we generated gridded TX maps and their corresponding temperature differences for 17 days at two observation levels and defined ⊿TX maps for Seoul by averaging the data. We then examined the relationship between ⊿TX and variables representing local landscape and urban structure, yielding the following key findings: (1) The 17-day average ⊿TX reached a maximum of +2.36℃ and a minimum of -0.93℃. (2) The tendency for positive ⊿TX was more pronounced in urban areas with compact low- or mid-rise structures located near green spaces. (3) The tendency for negative ⊿TX was more pronounced in areas with dense forest land cover on steep slopes, suggesting that street-level measurements reflect the influence of tree shade. (4) Large ⊿TX values were also influenced by the prediction uncertainty of the reference

station network.

The most significant contribution of this study is its quantification of undetected, more severe summer high temperatures at the street-level that reference stations alone fail to capture, as well as the identification of their spatial distribution. Especially in compact low- or mid-rise urban areas close to green spaces, there is a strong temperature difference with the street-level that reference stations cannot detect, and this temperature difference and estimation error may have already caused unrecognized heat-related health hazards. Strategically placing these IoT sensors in such areas can fill the undetected temperature gap between vertical levels and support strategies for heat mitigation and adaptation. This information is crucial for decision makers aiming to improve urban heat risk assessment.

However, this study is subject to certain limitations. It was conducted as a case study within a restricted spatiotemporal range in Seoul over only 17 summer days. While AI-based methods for generating gridded air temperature maps using temperature data observed by urban meteorological networks have been explored in many studies, similar models could be developed and applied in other cities where comparable data are available. Nonetheless, the ΔTX values and their characteristics observed in this study may be specific to Seoul's summer conditions, necessitating further research and careful evaluation in different urban contexts.

Despite these limitations, this study is distinguished by its use of high-density public urban IoT sensor networks distributed across Seoul. The findings from this study are expected to be valuable for a broad range of stakeholders, including urban planners, national and local policymakers, and medical professionals, in the context of urban health risk management and adaptation strategies.

Abbreviations

S-DoT, 3rd party sensor operated by Seoul City (Smart Seoul Data of Things); KMA, Korea Meteorological Administration; AWS, Automatic Weather System operated by KMA; TX, Daily maximum temperature; $Obs\text{-}TX_{SDoT}$, Observed TX by S-DoTs; $Obs\text{-}TX_{AWS}$, Observed TX by AWSs; $Model_{SDoT}$, The model trained using $Obs\text{-}TX_{SDoT}$ as label data; $Model_{AWS}$, The model trained using $Obs\text{-}TX_{AWS}$ as label data; $Pred\text{-}TX_{SDoT}$, Daily TX predicted by $Model_{SDoT}$; $Pred\text{-}TX_{AWS}$, Daily TX predicted by $Model_{AWS}$; $Avg\text{-}PredTX_{SDoT}$, $Pred\text{-}TX_{SDoT}$ averaged over 17 days; $Avg\text{-}PredTX_{AWS}$, $Pred\text{-}TX_{AWS}$ averaged over 17 days; ΔTX, Daily ΔTX averaged over 17 days

Acknowledgment

This work was supported by Korea Environment Industry & Technology Institute (KEITI) through "Climate Change R&D Project for New Climate Regime", funded by Korea Ministry of Environment (MOE) [RS-2022-KE002472].

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this work the authors used DeepL and ChatGPT in order to improve the clarity and fluency of the English language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

Beck C, Straub A, Breitner S, Cyrys J, Philipp A, Rathmann J, Schneider A, Wolf K, Jacobeit J. 2018. Air temperature characteristics of local climate zones in the Augsburg urban area (Bavaria, southern Germany) under varying synoptic conditions. Urban Clim 25: 152-166. doi: 10.1016/j.uclim.2018.04.007

Benali A, Carvalho AC, Nunes JP, Carvalhais N, Santos A. 2012. Estimating air surface temperature in Portugal using MODIS LST data. Remote Sens Environ 124: 108-121. doi: 10.1016/J.RSE.2012.04.024

Bowler DE, Buyung-Ali L, Knight TM, Pullin AS. 2010. Urban greening to cool towns and cities: A systematic review of the empirical evidence. Landsc Urban Plan 97(3): 147-155. doi: 10.1016/j.landurbplan.2010.05.006

Brousse O, Simpson CH, Poorthuis A, Heaviside C. 2024. Unequal distributions of crowdsourced weather data in England and Wales. Nat Commun 15(1): 4828. doi: 10.1038/s41467-024-49276-z

Cerlini PB, Silvestri L, Saraceni M. 2020. Quality control and gap-filling methods applied to hourly temperature observations over central Italy. Meteorol Appl 27(3). doi: 10.1002/met.1913

Chajaei F, Bagheri H. 2024. Machine learning framework for high-resolution air temperature downscaling using LiDAR-derived urban morphological features. Urban Clim 57: 102102. doi: 10.1016/j.uclim.2024.102102

Chapman L, Bell S, Randall S. 2023. Can crowdsourcing increase the durability of an urban meteorological network? Urban Clim 49: 101542. doi: 10.1016/J.UCLIM.2023.101542

Chen G, Shi Y, Wang R, Ren C, Ng E, Fang X, Ren Z. 2022. Integrating weather observations and local-climate-zone-based landscape patterns for regional hourly air temperature mapping using machine learning. Sci Total Environ 841: 156737. doi: 10.1016/j.scitotenv.2022.156737

Chen S, Yang Y, Deng F, Zhang Y, Liu D, Liu C, Gao Z. 2022. A high-resolution monitoring approach of canopy urban heat island using a random forest model and multi-platform observations. Atmos Meas Tech 15(3): 735-756. doi: 10.5194/amt-15-735-2022

Demuzere M, Kittner J, Martilli A, Mills G, Moede C, Stewart ID, van Vliet J, Bechtel B. 2023. Global map of Local Climate Zones (3.0.0) [Data set]. Zenodo; [accessed 2025 Mar 19]. https://zenodo.org/records/8419340.

dos Santos RS. 2020. Estimating spatio-temporal air temperature in London (UK) using machine learning and earth observation satellite data. Int J Appl Earth Obs Geoinf 88. doi: 10.1016/j.jag.2020.102066

Esri. 2025. Optimized hot spot analysis (spatial statistics) ArcGIS Pro 3.4; [accessed 2025 Mar 18]. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/optimized-hot-spot-analysis.htm.

Estévez J, Gavilán P, Giráldez JV. 2011. Guidelines on validation procedures for meteorological data from automatic weather stations. J Hydrol (Amst) 402(1-2): 144-154. doi: 10.1016/j.jhydrol.2011.02.031

Fenner D, Bechtel B, Demuzere M, Kittner J, Meier F. 2021. CrowdQC+—A quality-control for crowdsourced air-temperature observations enabling world-wide urban climate applications. Front Environ Sci 9. doi: 10.3389/fenvs.2021.720747

Fiebrich CA, Morgan CR, McCombs AG, Hall PK, McPherson RA. 2010. Quality assurance procedures for mesoscale meteorological data. J Atmos Ocean Technol 27(10): 1565-1582. doi: 10.1175/2010JTECHA1433.1

Galalizadeh S, Morrison-Saunders A, Horwitz P, Silberstein R, Blake D. 2024. The cooling impact of urban greening: A systematic review of methodologies and data sources. Urban For Urban Green 95: 128157. doi: 10.1016/j.ufug.2023.128157

Ho HC, Knudby A, Sirovyak P, Xu Y, Hodul M, Henderson SB. 2014. Mapping maximum urban air temperature on hot summer days. Remote Sens Environ 154: 38-45. doi: 10.1016/j.rse.2014.08.012

Honjo T, Yamato H, Mikami T, Grimmond CSB. 2015. Network optimization for enhanced resilience of urban heat island measurements. Sustain Cities Soc 19: 319-330. doi: 10.1016/J.SCS.2015.02.004

IBM. 2025. What is random forest? IBM; [accessed 2025 Apr 7]. https://www.ibm.com/think/topics/random-forest#:~:text=Random%20forest%20is%20a%20commonly,both%20classification%20and%20regression%20problems.

IPCC. 2023. Climate change 2023: Synthesis report, summary for policymakers. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, Lee H, Romero J (eds)]. Geneva.

Jaganmohan M, Knapp S, Buchmann CM, Schwarz N. 2016. The bigger, the better? The influence of urban green space design on cooling effects for residential areas. J Environ Qual 45(1): 134-145. doi: 10.2134/jeq2015.01.0062

KMA. 2016. Full text of ground weather observation guidelines. Seoul.

KMA. 2019. Standard specification for automatic weather observation equipment. Seoul.

KMA. 2020. Regional climate characteristics in Korea. KMA; [accessed 2025 Mar 24]. https://www.weather.go.kr/w/climate/statistics/regional-char.do#wnuri-menu.

Lam YF, Ong CW, Wong MH, Sin WF, Lo CW. 2021. Improvement of community monitoring network data for urban heat island investigation in Hong Kong. Urban Clim 37: 100852. doi: 10.1016/J.UCLIM.2021.100852

Lau T-K, Lin T-P. 2024. Investigating the relationship between air temperature and the intensity of urban development using on-site measurement, satellite imagery and machine learning. Sustain Cities Soc 100: 104982. doi: 10.1016/j.scs.2023.104982

Li H, Zhao Y, Wang C, Ürge-Vorsatz D, Carmeliet J, Bardhan R. 2024. Cooling efficacy of trees across cities is determined by background climate, urban morphology, and tree trait. Commun Earth Environ 5(1): 754. doi: 10.1038/s43247-024-01908-4

Lowry WP. 1977. Empirical estimation of urban effects on climate: A problem analysis. J Appl Meteorol 16(2): 129-135. doi: 10.1175/1520-0450(1977)016<0129:EEOUEO>2.0.CO;2

Meier F, Fenner D, Grassmann T, Otto M, Scherer D. 2017. Crowdsourcing air temperature from citizen weather stations for urban climate research. Urban Clim 19: 170-191. doi: 10.1016/J.UCLIM.2017.01.006

Napoly A, Grassmann T, Meier F, Fenner D. 2018. Development and application of a statistically-based quality control for crowdsourced air temperature data. Front Earth Sci (Lausanne) 6. doi: 10.3389/feart.2018.00118

Oke TR. 1982. The energetic basis of the urban heat island. Q J R Meteorol Soc 108(455): 1-24. doi: 10.1002/qj.49710845502

Oke TR, Mills G, Christen A, Voogt JA. 2017. Urban climates. Cambridge University Press; [accessed 2023 Jan 4]. https://www.cambridge.org/core/product/identifier/9781139016476/type/book.

Park J, Kim J-H, Sohn W, Lee D-K. 2021. Urban cooling factors: Do small greenspaces outperform building shade in mitigating urban heat island intensity? Urban For Urban Green 64: 127256. doi: 10.1016/j.ufug.2021.127256

Pigliautile I, Pisello AL. 2020. Environmental data clustering analysis through wearable sensing techniques: New bottom-up process aimed to identify intra-urban granular morphologies from pedestrian transects. Build Environ 171: 106641. doi: 10.1016/J.BUILDENV.2019.106641

Romero Rodríguez L, Guerrero Delgado Mc, Castro Medina D, Sánchez Ramos J, Álvarez Domínguez S. 2024. Forecasting urban temperatures through crowdsourced data from citizen weather stations. Urban Clim 56: 102021. doi: 10.1016/j.uclim.2024.102021

Sakthivel P, Sengupta R. 2025. Spatial bias in placement of citizen and conventional weather stations and their impact on urban climate research: A case study of the Urban Heat Island effect in Canada. Urban Clim 59: 102280. doi: 10.1016/J.UCLIM.2024.102280

Seoul Metropolitan Government. 2020a. S·DoT: Smart Seoul Data of Things. Seoul.

Seoul Metropolitan Government. 2020b. 2020 smart city and informatization implementation plan. Seoul.

Seoul Metropolitan Government. 2020c. Basic information of smart Seoul data of things. Seoul City; [accessed 2025 Mar 9]. https://smart.seoul.go.kr/board/41/1243/board_view.do.

Seoul Metropolitan Government. 2025 Feb 20. Seoul City population density statistics. Seoul Metropolitan Government; [accessed 2025 Mar 24].

https://data.seoul.go.kr/dataList/10790/S/2/datasetView.do.

Shashua-Bar L, Hoffman ME. 2000. Vegetation as a climatic component in the design of an urban street: An empirical model for predicting the cooling effect of urban green areas with trees. Energy Build 31(3): 221-235. doi: 10.1016/S0378-7788(99)00018-3

Shen H, Jiang Y, Li T, Cheng Q, Zeng C, Zhang L. 2020. Deep learning-based air temperature mapping by fusing remote sensing, station, simulation and socioeconomic data. Remote Sens Environ 240. doi: 10.1016/j.rse.2020.111692

Shi R, Hobbs BF, Zaitchik BF, Waugh DW, Scott AA, Zhang Y. 2021. Monitoring intra-urban temperature with dense sensor networks: Fixed or mobile? An empirical study in Baltimore, MD. Urban Clim 39: 100979. doi: 10.1016/J.UCLIM.2021.100979.

Stewart ID, Oke TR. 2012. Local climate zones for urban temperature studies. Bull Am Meteorol Soc 93(12): 1879-1900. doi: 10.1175/BAMS-D-11-00019.1

Tsin PK, Knudby A, Krayenhoff ES, Ho HC, Brauer M, Henderson SB. 2016. Microscale mobile monitoring of urban air temperature. Urban Clim 18: 58-72. doi: 10.1016/j.uclim.2016.10.001

Venter ZS, Brousse O, Esau I, Meier F. 2020. Hyperlocal mapping of urban air temperature using remote sensing and crowdsourced weather data. Remote Sens Environ 242: 111791. doi: 10.1016/j.rse.2020.111791

Vuckovic M, Kiesel K, Mahdavi A. 2017. Studies in the assessment of vegetation impact in the urban context. Energy Build 145: 331-341. doi: 10.1016/J.ENBUILD.2017.04.003

Wang H, Yang J, Chen G, Ren C, Zhang J. 2023. Machine learning applications on air temperature prediction in the urban canopy layer: A critical review of 2011-2022. Urban Clim 49: 101499. doi: 10.1016/j.uclim.2023.101499

World Meteorological Organization. 2023. Guidance on measuring, modelling and monitoring the Canopy Layer Urban Heat Island (CL-UHI). Geneva; [accessed 2025 Mar 31]. https://urban-climate.org/wp-content/uploads/2024/06/1292_en.pdf.

Yang J, Yu M, Liu Q, Li Y, Duffy DQ, Yang C. 2022. A high spatiotemporal resolution framework for urban temperature prediction using IoT data. Comput Geosci 159: 104991. doi: 10.1016/J.CAGEO.2021.104991

Yu Y, Li P, Huang D, Sharma A. 2024. Street-level temperature estimation using graph neural networks: Performance, feature embedding and interpretability. Urban Clim 56: 102003. doi: 10.1016/J.UCLIM.2024.102003

Zumwald M, Knüsel B, Bresch DN, Knutti R. 2021. Mapping urban temperature using crowd-sensing data and machine learning. Urban Clim 35: 100739. doi: 10.1016/J.UCLIM.2020.100739

Appendix

Appendix A

Table A1. Overview of the QC process implemented in this study. The raw data used for QC included columns for station serial numbers, location information, hourly observation values, and observation timestamps

QC step	Test name	Description	Source
1	Missing value test	All values are flagged as False, if station location information (long., lat.), temperature value, or observation time data is missing or stored in an improper format.	Added conditions for missing hourly observation values, and observation timestamps to main QC step 1 (m1) of Fenner et al. (2021).
2	Gross-error limit test	All values outside the range [+10℃, +50℃] were flagged as FALSE.	(Cerlini et al., 2020; Estévez et al., 2011; Fiebrich et al., 2010)
3	Robust Z-score range test	All values for which the robust Z-score of the temperature values were not within the critical range of [0.01, 0.95] were flagged as FALSE.	main QC step 2 (m2) of Crowd QC plus v.1.1 (Fenner et al., 2021)
4	Removal monthly ratio test	For each station, we calculated the proportion of values removed by the process up to step 3 per period. For stations where the proportion exceeded the cutoff percentage value, all values during that period were flagged as FALSE. In this study, the test period was set to one month and the cutoff percentage was set to 20%.	main QC step 3 (m3) of Crowd QC plus v.1.1 (Fenner et al., 2021)

Table A2. Proportion of true values for hourly temperature data at each QC step. The total proportion in parentheses indicates a decrease compared to the previous step. Step 1 flagged 3.26% of the total values as False. Step 2 flagged 2.30% of the values that passed step 1 as False. Step 3, which identified relative outliers in the data distribution, flagged 4.09% of the values that passed step 2 as False. This step resulted in the largest data removal in the QC process. Finally, in step 4, 86.75% of the raw data (11,431,942 rows) were left as usable data

Year	Before QC (rows)	The percentage of rows remaining after each QC step				Num. of available stations
		Step 1 (%)	Step 2 (%)	Step 3 (%)	Step 4 (%)	
2020	2,191,988	99.62	99.42	93.8	89.44	819
2021	2,628,155	99.79	99.55	94.85	90.63	1,048
2022	2,746,189	97.92	96.88	92.37	88.05	1,075
2023	2,614,006	97	89.88	87.5	85.53	993
2024	2,985,074	91.04	88.42	84.86	81.59	973
Total	13,177,321	96.74 (- 3.26%)	94.44 (-2.30%)	90.35 (- 4.09%)	86.75 (- 3.6%)	1,125

Appendix B

Table B1. Model performance by label data type and kernel size. The RMSE per fold represents the RMSE obtained in each of the five tests, while the average of the five folds represents the mean values of the evaluation metrics (RMSE, MAE, R^2) across all five tests

Label data	Kernel size	RMSE per fold (℃)					Average of 5-folds (℃)		
		Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	RMSE	MAE	R^2
S-DoT	1	0.957	1.04	1.025	1.053	1.053	1.025	0.777	0.87
	3	0.95	1.024	1.022	1.054	1.048	1.019	0.772	0.872
	10	0.958	1.03	1.024	1.061	1.048	1.024	0.776	0.871
	20	0.974	1.042	1.021	1.075	1.053	1.033	0.782	0.868
	30	0.963	1.036	1.028	1.065	1.055	1.029	0.78	0.869
	50	0.966	1.037	1.029	1.062	1.055	1.03	0.777	0.869
AWS	1	0.688	0.869	0.81	0.71	0.781	0.772	0.598	0.911
	3	0.721	0.871	0.788	0.716	0.781	0.775	0.595	0.911
	10	0.693	0.868	0.753	0.705	0.817	0.767	0.59	0.912
	20	0.705	0.868	0.769	0.701	0.79	0.767	0.59	0.913
	30	0.693	0.878	0.759	0.689	0.792	0.762	0.582	0.913
	50	0.783	0.857	0.772	0.699	0.792	0.78	0.59	0.91

Table B2. Descriptive statistics for Pred-TX estimated by $Model_{SDoT}$ and $Model_{AWS}$. The bottom row represents TX map statistics averaged over 17 days

Date	Pred-TX_{SDoT} (℃)				Pred-TX_{AWS} (℃)			
	Min.	Max.	Ave.	Std.	Min.	Max.	Ave.	Std.
20-06-15	28.57	32.24	30.24	0.63	28.13	30.15	29.1	0.38
20-07-17	30.91	33.92	32.49	0.61	29.93	32.63	31.45	0.61
20-09-19	26.15	29.45	27.84	0.82	26.13	27.37	26.9	0.35
21-07-20	31.19	35.34	33.51	1	31.51	33.91	32.88	0.67
22-06-21	30.9	35.23	33.32	0.83	31.14	33.62	32.47	0.55
22-07-15	29.55	33.15	31.28	0.74	29.8	32.3	31.16	0.53
22-08-16	30.03	33.11	31.6	0.73	30.23	31.79	31.01	0.4
23-06-16	28.57	32.7	30.43	0.87	28.32	30.35	29.41	0.47
23-06-24	29.38	34.26	32.12	1.03	30.24	32.59	31.55	0.56
23-08-03	33.31	36.85	35.04	0.86	32.93	35.3	34.23	0.67
23-08-19	32.61	36.7	34.8	0.96	33.03	34.82	34.05	0.49
24-06-02	25.22	28.32	26.88	0.81	25.95	27.55	26.72	0.32
24-06-18	30.94	35.11	33.22	0.82	31.16	33.62	32.49	0.54
24-06-26	29.43	34.24	32.34	1.06	30.12	32.31	31.4	0.49
24-08-13	33.73	38.6	36.63	1.2	34.98	36.95	36.06	0.55
24-08-29	32.14	35.85	34.14	0.82	32.78	34.4	33.56	0.42
24-09-30	26.59	30.48	28.78	0.93	27.8	29.05	28.48	0.29
17-day average	30.2	33.49	32.04	0.78	30.41	32.12	31.35	0.44

Table B3. Daily ⊿TX statistics calculated from two models

Date (YY-MM-DD)	Daily ⊿TX (℃)		
	Min.	Max.	Ave.
20-06-15	-0.53	3.28	1.15
20-07-17	-0.65	2.93	1.04
20-09-19	-0.69	2.73	0.94
21-07-20	-1.56	3.15	0.62
22-06-21	-1.29	3.27	0.85
22-07-15	-1.53	2.5	0.12
22-08-16	-0.97	2.33	0.58
23-06-16	-0.95	3.7	1.02
23-06-24	-2.15	3.02	0.57
23-08-03	-1.07	3.25	0.81
23-08-19	-1.27	3.03	0.76
24-06-02	-1.59	1.9	0.16
24-06-18	-1.19	3.08	0.73
24-06-26	-1.99	3.19	0.95
24-08-13	-2.35	2.81	0.58
24-08-29	-1.42	2.58	0.58
24-09-30	-1.63	2.02	0.31
17-day average	-0.93	2.36	0.69