Handbook of Geospatial Approaches to Sustainable Cities

Edited by Qihao Weng in collaboration with Cheolhee Yoo

First published 2024

ISBN: 978-1-032-15481-7 (hbk)
ISBN: 978-1-032-15534-0 (pbk)
ISBN: 978-1-003-24456-1 (ebk)

8

GeoAI for High-Resolution Urban Air Temperature Estimation and Urban Heat Island Monitoring

Cheolhee Yoo and Qihao Weng

DOI: 10.1201/9781003244561-11

The Open Access version of this chapter was funded by Qihao Weng.

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

8 GeoAI for High-Resolution Urban Air Temperature Estimation and Urban Heat Island Monitoring

Cheolhee Yoo and Qihao Weng

INTRODUCTION

Although urban areas constitute merely 3% of the Earth's total surface, they accommodate more than half of the global population today (Seto et al., 2012). This urban population is projected to increase to 2.5 billion by 2050 (Angel et al., 2011). Urban areas are expanding rapidly, replacing existing natural land cover and vegetation with impervious surfaces (Foley et al., 2005). The significant solar absorption and high thermal capacity of these impervious surfaces warm the core areas of cities more than their peripheries, a phenomenon known as the Urban Heat Island (UHI) effect. The UHI effect is well-known to exacerbate environmental issues such as air pollution and drought (Stone, 2005; Im & Eltahir, 2018). More importantly, its negative impact on human health, including issues like heatstroke and mental health, has been extensively revealed, making the issue of UHIs in densely populated cities a growing concern (Basu & Samet, 2002; Tzoulas et al., 2007).

In order to measure the intensity of UHIs, it is first necessary to delineate the urban area (Voogt & Oke, 2003). A common method used in previous studies involves clustering urban areas using impervious surface data and defining the extent of neighboring areas through a buffering approach (Yang et al., 2003; Lu et al., 2011). However, this method struggles to reflect the heterogeneous surface characteristics created by diverse elements within the city, such as buildings and trees, as it represents the urban area as a single cluster. To address this issue, a classification system known as the Local Climate Zone (LCZ) has been proposed (Stewart & Oke, 2012). This standard system divides urban areas into 17 classes based on building characteristics (height, density) and vegetation types. Typically, studies define the UHI intensity as the difference between the average temperature of each LCZ class within the city and the average temperature of a base class, a Low

DOI: 10.1201/9781003244561-11

Plant (LCZD) (Alexander & Mills, 2014). This method allows for the differentiation of thermal environmental characteristics by different urban forms, thus enabling a more nuanced analysis of UHI intensity. Consequently, LCZ is widely utilized in current research studies on urban thermal environmental analysis (Quan & Bansal, 2021; Yang et al., 2019; Eldesoky et al., 2021).

Alongside the delineation of urban areas, another consideration is how urban temperatures are monitored. A common approach is to use satellite-based land surface temperature (LST) data (Chang et al., 2021; Yoo et al., 2022). Many Satellite-based LSTs are public and offer broad coverage to capture comprehensive thermal environmental phenomena within and around cities (Zhou et al., 2018). However, these LST data present several challenges in urban studies. Firstly, multispectral sensor-based satellite sensors cannot obtain values beneath clouds, posing significant limitations, particularly in cities located in humid climate regions that are heavily affected by cloud cover (Zeng et al., 2018; Yoo et al., 2020a).

Another challenge is the resolution trade-off (Xia et al., 2019). Typically, satellite data with higher spatial resolution have lower temporal resolution. For example, LST data based on the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor, which is capable of monitoring urban areas and is mounted on the Aqua and Terra satellites, can obtain LST data four times daily, but its spatial resolution of 1 km limits detailed monitoring of urban temperature distributions. In contrast, Landsat-based thermal data, known to be effective for detailed spatial analysis of urban thermal environments due to its relatively high-spatial resolution (~100 m), has a significant limitation: its temporal resolution is approximately 16 days, precluding frequent temperature monitoring at multiple time points.

Furthermore, satellite-based LST data, which measures surface temperature, typically differs from the air temperature that humans can perceive, rendering it somewhat insufficient for direct use in thermal environment studies related to human exposure to heat risk (Yoo et al., 2018). Specifically, in cities, the spatial distribution of LST and air temperature can significantly vary and their patterns are also different and dynamic over time (Cao et al., 2021), complicating the use of satellite-based surface temperature data in urban thermal environment studies.

Alternatively, another typical method of monitoring urban temperatures is to acquire temperature data based on observation stations, which measure the temperature at a height of approximately 1.5 m above the ground in city-installed weather stations (Meier et al., 2017). These well-maintained stations provide high temporal resolution temperature data, making them effective for analyzing diurnal temperature changes (Waqas & Athar, 2018). However, a major issue with using observation station-based air temperature to monitor the urban thermal environment is the number of stations (Shi et al., 2015). While these stations require consistent and sophisticated management to provide high-quality temperature data, most reliable stations are managed by national institutions, meaning there are typically insufficient numbers of these stations to monitor the detailed temperature distribution within a city accurately.

To seamlessly monitor the temperature distribution in cities using observation station-based air temperature data, various preliminary studies have been conducted, initially involving spatial modeling based on interpolation (Hart et al., 2009; Bárdossy et al., 2021). The air temperature data from observation stations are used to produce a temperature map of the city, primarily through interpolation techniques such as Kriging or IDW. However, these simple interpolation techniques are known to be more suitable for natural-type land surface regions with relatively homogeneous spatial distributions of temperature (Ashiq et al., 2010). Given the complex surface characteristics of cities, their temperature distribution is markedly heterogeneous, which presents limitations for the application of interpolation techniques. To overcome this, more complex statistical models have been used (Ho et al., 2014; Yoo et al., 2018). These models use input predictors reflecting thermal characteristics, and modeling is conducted with the measured air temperature from the observation stations as the target. In particular, various machine learning techniques are being used today for urban atmospheric temperature mapping (Wang et al., 2023). For example, Ho et al. (2014) used the Random Forest machine learning technique to produce a daily maximum air temperature for Vancouver, Canada, with a resolution of 100 m and 16 days, using Landsat-based LST as input data.

Furthermore, dos Santos (2020) used the Gradient Boosting machine learning technique to map the daily air temperature of London with a 1 km resolution. Venter et al. (2020) used the Random Forest machine learning technique to produce daily maximum and minimum temperatures for Oslo, Norway, with a 30 m resolution. Although these techniques, based on machine learning, mapped air temperatures with relatively high accuracy, most have a daily temporal resolution, which still presents limitations for monitoring the urban thermal environment at high spatiotemporal resolution.

Hrisko et al. (2020) produced hourly air temperature based on a Neural Network machine learning technique using GOES-16 LST from a geostationary satellite as input data for major cities in the United States. However, the spatial resolution was as low as 2 km. To date, there have been few studies that have been able to analyze the detailed temperature distribution of cities with a high hourly temporal resolution and discuss the spatiotemporal patterns of UHI phenomena in-depth.

In this chapter, we have devised a research method to map air temperature with a high spatiotemporal resolution of hourly 250 m for Seoul, South Korea. In particular, we propose an effective method using multi-task machine learning techniques to map the temperature of various time frames with high accuracy simultaneously. We will compare and analyze the spatial patterns of air temperature by LCZ class using the created hourly 250 m air temperature and simulate the diurnal cycle by calculating the intensity of the UHI effect.

STUDY AREA AND DATA

Study Area

The city chosen for this study is Seoul, the capital city of South Korea, located in the northwestern part of the country. Geographically divided by the Han River into northern and southern regions, Seoul is encircled by four prominent mountains.

The city center is approximately 60 km from South Korea's western coastline and covers an area of around 605 square kilometers. Seoul is characterized by its high population density, with a populace of approximately 9.98 million as of 2022.

According to the Köppen–Geiger climate classification system, Seoul is characterized by a humid subtropical climate. Its summers are marked by elevated temperatures, high humidity, and substantial rainfall, primarily due to the influence of the East Asian Monsoons. Seoul is widely acknowledged as a city susceptible to heatwaves and the UHI phenomenon. Consequently, monitoring urban temperatures, particularly during the summer season, is essential. However, the effectiveness of this activity is significantly constrained by the prevalent cloud cover, which restricts the applicability of satellite-based LST. Given these conditions, Seoul emerges as an appropriate region for the high-resolution temperature mapping that constitutes the goal of the current study.

The research period for this study was from 2018 to 2020. For analytical purposes, the study duration was segregated into two distinct seasons: the cold season (November–April) and the warm season (May–October).

SATELLITE AND IN-SITU METEOROLOGICAL DATA

This study involved the use of data obtained from MODIS Terra and Aqua satellites. The wide application of MODIS data in observing atmospheric, oceanic, and terrestrial phenomena is well-recognized in the scientific community. For this study, specific MODIS products, LST and Reflectance—corresponding to the h28v05 region for Seoul—were procured for the period of 2018 to 2020 from the Reverb ECHO portal (http://reverb.echo.nasa.gov). We utilized daily LST data products MYD11A1 (from Aqua) and MOD11A1 (from Terra), both of which have a spatial resolution of 1 km. These LST data were generated using a generalized split-window algorithm, as outlined in the studies by Wan and Dozier (1996). The Terra MODIS satellite provides LST data collected at 10:30 am solar local time (daytime) and 10:30 pm (nighttime). In contrast, Aqua MODIS offers LST data collected at local solar times of 1:30 pm (daytime) and 1:30 am (nighttime). For this study, both daytime and nighttime LST data from Terra, along with the nighttime LST data from Aqua, were collected.

MODIS Reflectance data in the Red and Near-Infrared (NIR) bands were obtained from a MOD09GQ product featuring a spatial resolution of 250 meters. Digital surface model (DSM) data, at a spatial resolution of 30 meters, was sourced from the Advanced Land Observing Satellite (ALOS) Global Digital Surface Model. The study also utilized the Global Man-made Impervious Surface (GMIS) dataset, acquired at an equivalent 30-meter resolution. To procure high-resolution LST data for the input parameter for air temperature modeling, a total of 32 clear-sky daytime scenes from Landsat 8 LST and three nighttime scenes from ASTER LST were collected. The full list of satellite-based data used is described in Table 8.1.

Hourly air temperature readings from 2018 to 2020 were collated from a variety of weather stations situated across Seoul. These encompassed 43 Automatic Weather Stations (AWS) and Automated Surface Observing Systems (ASOS) managed by the Korea Meteorological Administration, 37 stations operated by

TABLE 8.1

Description of Satellite-Based Data Products with Their Source and Resolution Used in This Study

Products	Source	Resolution
Surface Reflectance (Red/NIR)	Clear-sky MODIS MOD09GQ	250 m daily
Terra Daytime LST (TD)	Clear-sky MODIS MOD11A1 (from Terra)	1 km daily, 10:30 AM
Aqua Nighttime LST (AN)	Clear-sky MODIS MYD11A1 (from Aqua)	1 km daily, 01:30 AM
Terra Nighttime LST (TN)	Clear-sky MODIS MOD11A1 (from Terra)	1 km daily, 10:30 PM
Digital Surface Model (DSM)	ALOS Global Digital Surface Model	30 m
Impervious surface (IMP)	Impervious percentage from Global Man-made Impervious Surface	30 m
Landsat daytime HLST	Landsat8 Level2 LST (32 clear-sky scenes)	100 m 16-day
ASTER nighttime HLST	ASTER Level2 LST (3 clear-sky scenes)	90 m 16-day

a Korean public institution, and three stations maintained by the Forest Meteorological Observatory. The cumulative data from these sources served as the reference dataset for this study. The spatial distribution of these weather stations is illustrated in Figure 8.1.

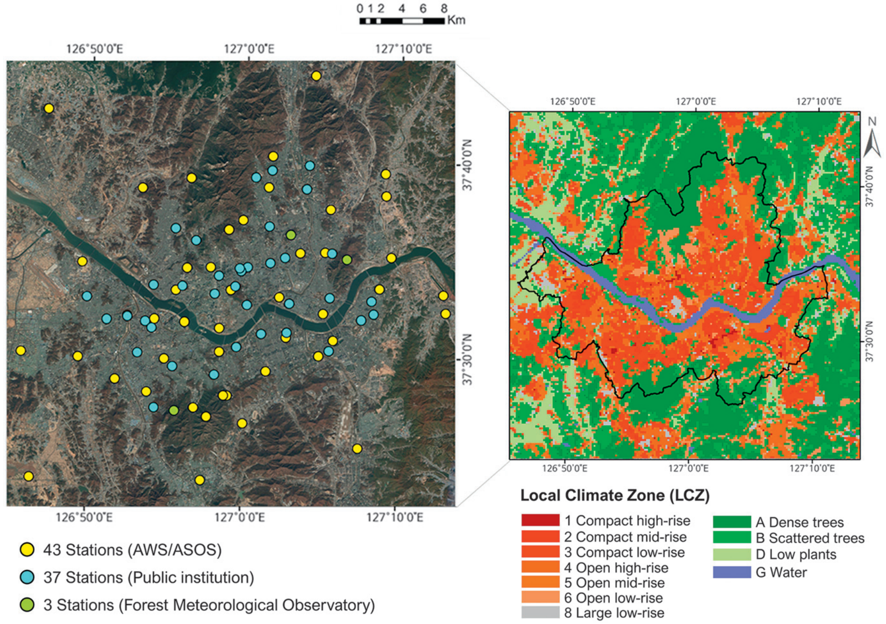


FIGURE 8.1 Study areas with the location of weather stations (left) and Local Climate Zones (LCZ) in Seoul (right).

LOCAL CLIMATE ZONE MAP

This study adopted the LCZ framework to examine the UHI phenomenon in Seoul at varying temporal intervals, such as hourly increments. A comprehensive LCZ map of Seoul, at a resolution of 50 meters, was developed by integrating several data sources, including Sentinel-2A reflectance, Landsat 8 LST, and building datasets. This LCZ map was generated based on the Convolutional Neural Network (CNN) deep learning method, as outlined by Yoo et al. (2020b). Upon map production, the LCZs were resampled to enhance their resolution to 250 meters, using a majority voting approach. The LCZ framework comprises seventeen distinct categories: ten urban (LCZ1 through LCZ10) and seven natural (LCZA through LCZG). The comprehensive list of LCZ classes is depicted in the legend of Figure 8.1.

METHODS

VARIABLE PROCESSING

The 250 m MODIS Red and Near-Infrared (NIR) reflectances were used as input variables. In addition, the Normalized Difference Vegetation Index (NDVI) was calculated by measuring the difference between the visible red and near-infrared bands. The 30 m Digital Surface Model (DSM) was bilinearly resampled to 250 m, and the slope was computed from the DSM. The 30 m impervious surface percentage (Imp) data, high-resolution Landsat 8, and ASTER clear-sky LSTs (HLST) were also resampled to 250 m.

The Euclidean distance from the centroid of the study area point (DCP), a representation of the urban thermal heat phenomena concentrated in the city center, was computed at a resolution of 250 m. Daily solar radiation was estimated using the ArcGIS Solar Analyst tool (Esri). The incoming solar radiation for each pixel was computed by inputting the elevation and the day of the year (DOY) into ArcGIS. For this study, DOY values were transformed to a range from −1 to 1 over a one-year period, using a sine function to account for seasonality (i.e., setting the peak of summer as 1 and the depth of winter as −1), and 250 m pixel-level DOY were used as input variables.

The study used MODIS 1 km Terra daytime (TD) and Aqua daytime (AD) LST, as well as Aqua night (AN) LST. All the input variables and their corresponding acronyms are presented in Table 8.2.

AIR TEMPERATURE ESTIMATION

The method for estimating 250 m air temperature consisted of two steps. The first step involved downscaling the MODIS 1 km LST to 250 m LST using a Random Forest (RF) model with various input parameters (also called kernels in downscaling). The second step entailed estimating air temperature at an hourly resolution of 250 m, based on a multi-task machine learning model. The flow chart, encapsulating step 1 and step 2, is illustrated in Figure 8.2.

TABLE 8.2
Variable Used for High Spatiotemporal Air Temperature Mapping in This Study

Acronym	Variable	Resolution
Red	Band1 from MODIS Reflectance	250 m daily
NIR	Band2 from MODIS Reflectance	250 m daily
NDVI	Normalized Difference Vegetation Index from MODIS Reflectance	250 m daily
DSM	ALOS Global Digital Surface Model	250 m (Resampled)
Slope	Slope by DSM	250 m
Aspect	Aspect by DSM	250 m
IMP	The impervious surface percentage from GMIS	250 m (Resampled)
HLST	Clear-sky daytime (Landsat8) / nighttime (ASTER) high-resolution LST	250 m
DCP	Euclidean Distance from the Centroid of the study area Point	250 m
DOY	Sine function transformed Day of Year	250 m
SOLAR	Area Solar radiation within a day	250 m
TD, AD LST	MODIS 1 km Terra Daytime and Aqua Daytime LST	1 km
AN LST	MODIS 1 km Aqua Nighttime LST	1 km

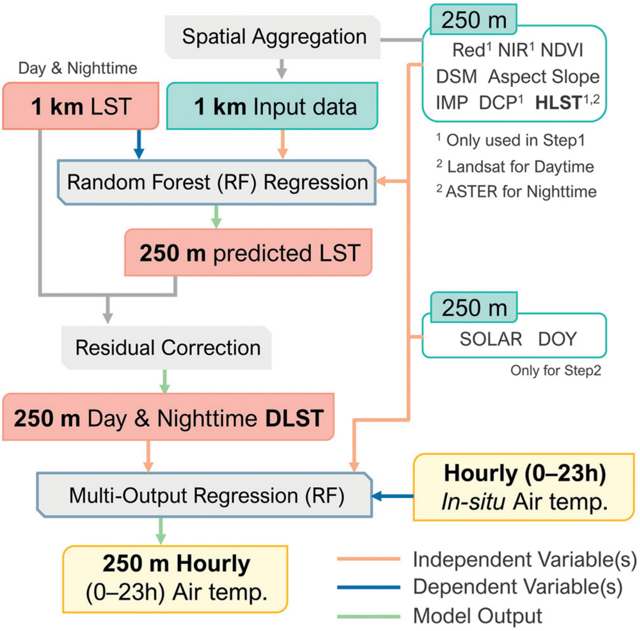


FIGURE 8.2 Flowchart of the study, divided into two main parts: Land Surface Temperature (LST) downscaling (Step 1) and air temperature estimation (Step 2).

Step 1: LST Downscaling

The first step involves downscaling 1 km MODIS LST to a resolution of 250 m. Initially, high-resolution input data, encompassing Red, NIR, NDVI, DSM, Slope, IMP, DSM, and HLST, were resampled to 1 km, matching the resolution and grids of MODIS LST. Subsequently, a downscaling model was established separately for each TD, AD, and AN MODIS LST using a Random Forest classifier. Random Forest is a machine learning algorithm rooted in the principles of Classification and Regression Trees (CART), as proposed by Breiman (2001). This algorithm utilizes a recursive binary split approach to navigate the nodes within a tree structure until terminal nodes are reached. To accomplish this, RF generates a large array of CARTs (usually between 500–1000 trees). These trees are constructed using subsets randomly selected through a bootstrapping method from training samples for a tree and predictor variables at each tree node. This bootstrapping-based randomization strategy effectively addresses the issues intrinsic to the CART method, including its sensitivity to training samples and a propensity for overfitting. RF concludes its decision-making by assimilating the outputs from the multitude of trees through an averaging process for regression.

We constructed a model for each date of each LST image using the Random Forest (RF) machine learning algorithm. This algorithm, based on Classification and Regression Trees (CART) as delineated by Breiman (2001), employs a recursive binary split methodology to navigate through the nodes within a tree structure until it reaches terminal nodes. To accomplish this, RF generates a large array of CARTs, typically between 500–1000 trees. These trees are constructed using subsets that are randomly selected through a bootstrapping method from both training samples for a tree and predictor variables at each tree node. This bootstrapping-based randomization approach effectively addresses the issues inherent to the CART method, such as its sensitivity to training samples and tendency towards overfitting (Belgiu & Drăguţ, 2016). RF concludes its decision-making by integrating the outputs from the multitude of trees through either an averaging or weighted averaging process for regression.

A critical input parameter in this step is the HLST variable, which is directly linked with the target MODIS LST. However, we only had 32 daytime scenes for Landsat LSTs and three nighttime scenes for ASTER LSTs. We calculated the Structural Similarity Index Measure (SSIM) between each MODIS LST and all 32 Landsat LSTs (for daytime) and three ASTER LSTs (for nighttime) after normalization. This was to identify the HLST that could best represent the similar LST pattern of the target MODIS LSTs. Based on the SSIM, the HLST with the highest similarity was selected and used as an input variable for each target date's MODIS LST downscaling model. After constructing the RF model, the original 250 m input data were inputted to produce the predicted 250 m LST. A residual correction process originally suggested by Kustas et al., 2003 was applied to produce the DLST finally. This residual correction process generally uses the following steps: 1) aggregating the simulated high-resolution LST (LSThigh) to the original MODIS LST scale, 2) subtracting the aggregated LSThigh (LSTag) from original MODIS LST (called residuals; Δ), 3) resampling the residuals to LSThigh scale and adding these residuals to the LSThigh for generating MODIS DLST.

Step 2: Air Temperature Estimation

Upon generating the 250 m daytime and nighttime DLST in step 1, step 2 entails the estimation of hourly high-spatial resolution air temperatures through a multi-output regression based on RF. To mitigate the issues of cloud contamination in LST, three combinations of different input variables were constructed (Figure 8.3). Combination1 incorporates AN and TD as input variables alongside auxiliary variables (i.e., NDVI, DSM, Slope, Imp, SOLAR, and DOY). Combination 2 employs TD and AN along with auxiliary variables. Finally, combination 3 simultaneously uses AN, TD, and TN, along with auxiliary variables. The final air temperature map is then produced by merging the three combinations. When TD, AN, and TN are all under clear-sky conditions, combination 3 is used. However, if only TD and AN are under clear-sky conditions, combination 1 is applied. Alternatively, if only TD and TN are under clear-sky conditions, combination 2 is used. If all three LSTs are under cloudy conditions, that date is deemed unsuitable for generating an air temperature map. In step 2, the Leave-One-Station Out Cross-Validation (LOOCV) is repeated for weather stations. The models' performance is evaluated using the coefficient of determination (R^2) and the Root Mean Square Error (RMSE).

URBAN HEAT ISLANDS MONITORING

Upon the completion of the 250 m hourly air temperature mapping, the UHI phenomena in Seoul were evaluated using an LCZ classification map. The UHI intensity was determined by calculating the difference in air temperatures between regions classified as LCZ2 (compact mid-rise) and LCZD (low plants). The diurnal cycle of the UHI was subsequently examined separately for both the warm and cold

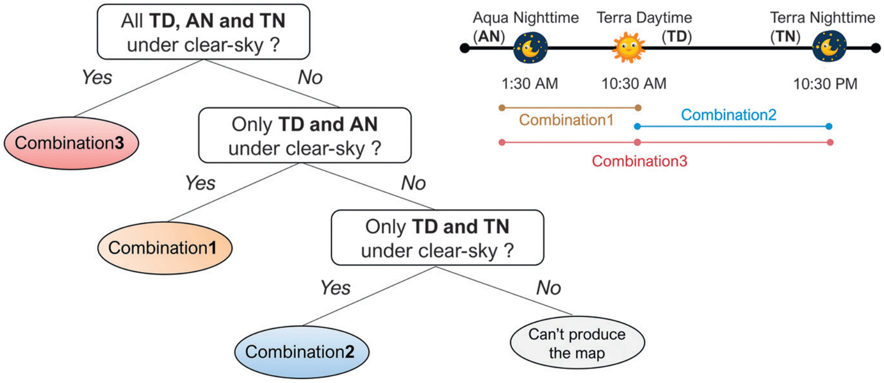


FIGURE 8.3 Combination rule for merging three combinations to create the final air temperature map.

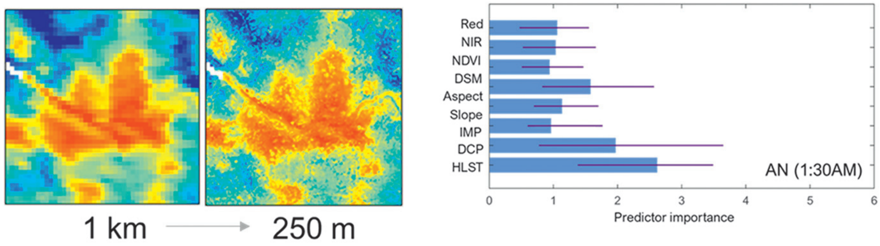
seasons. For a more detailed identification of the UHI, UHI intensity was also calculated based on the difference between LCZ6 (open low-rise) and LCZD.

RESULTS AND DISCUSSION

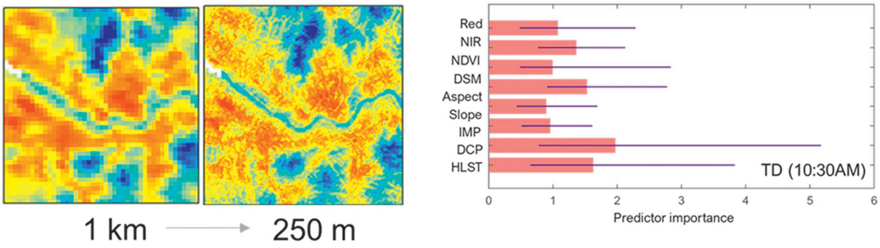
LST Downscaling Results

Figure 8.4 represents the distribution of the average normalized LST over three time slots (1:30 AM, 10:30 AM, 10:30 PM) during the study period, obtained from 1 km MODIS LST and downscaled to 250 m LST (DLST), along with the average variable importance and error bar from Random Forest modeling. The 250 m

(a) AN (1:30 AM; 113 Scenes)



(b) TD (10:30 AM; 159 Scenes)



(c) TN (10:30 PM; 101 Scenes)

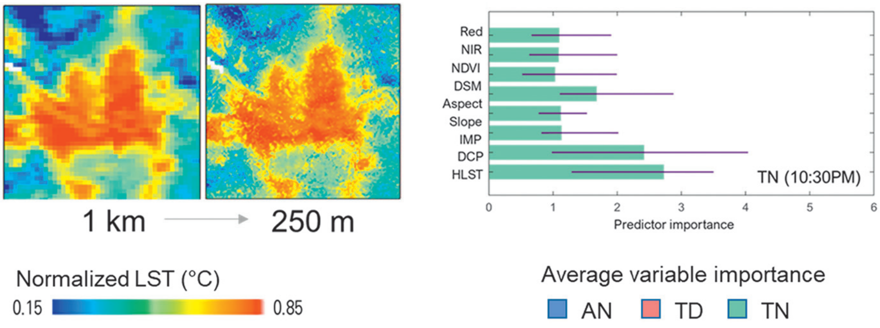


FIGURE 8.4 Averages of normalized 1 km LST and downscaled LST (DLST) with Random Forest (RF) variable importance for AN (a), TD (b), and TN (c) separately.

DLST, downscaled from 1 km, exhibits a similar spatial pattern of overall temperature, but the detailed urban heat environment pattern appears more pronounced. The downscaled technique based on the Random Forest used in this study has been sufficiently validated for its accuracy in previous studies (Hutengs & Vohland, 2016; Ebrahimy et al., 2021).

Upon examining the variable importance, the most important factor in the downscaling model of 1:30 AM AN turned out to be HLST. This suggests that high-resolution LST from other dates similar to the target date can contribute to greater accuracy in downscaling nighttime surface temperatures due to their similar spatial distributions. The second most important variable was DCP, indicating that the distance from the urban center emerged as a significant variable. This implies that urban heat island phenomena exhibit a dominant pattern within the city. For instance, as seen in Figure 8.1, the center of Seoul distinctly shows higher temperatures compared to the surrounding areas. This phenomenon has led to the DCP variable becoming a crucial factor in the detailed representation of temperature patterns. The DSM appeared as the third most important factor in AN downscaling. This interpretation suggests that, given Seoul's geographical features surrounded by mountains, elevation data has been significantly used to simulate the distribution of lower temperatures in the peripheral regions.

At 10:30 AM (TD), the most crucial variable importance was revealed to be DCP rather than HLST. It is apparent that the rank of HLST's importance is lower during the daytime (TD) than at night (AN), which can be attributed to the larger daily variation in the spatial distribution of temperature due to the influence of solar incoming (Luintel et al., 2019). As a result, the number of HLST images with similar thermal environment distributions to the target date LST and the similarity of their distributions appear relatively less significant during daytime. In contrast, DCP and DSM still emerge as the variables with the highest importance. During the day, high temperatures can be observed in urban centers, particularly in the northern region of the river. On the other hand, the surrounding mountainous areas exhibit lower temperatures due to the influence of high altitudes, even during the day.

Lastly, at 10:30 PM (TN), a similar pattern of variable importance to the nighttime (AN) is observed, with HLST having the highest importance, followed by DCP and DSM. The similarity in importance between AN and TN can be attributed to the reduced influence of solar incoming at night, leading to comparable spatial temperature patterns. In summary, HLST, DCP, and DSM demonstrate high importance across all three time periods; however, the importance of HLST is observed to be higher at night compared to during the day.

HOURLY AIR TEMPERATURE ESTIMATION ACCURACY ASSESSMENT

The Leave-One-Out Cross-Validation (LOOCV) results for each hour of the three combinations are compared in Figure 8.5 All three combinations exhibit high temporal performance (R^2) for all hours and an acceptable error of approximately 1.0 °C RMSE. For combination 1, the early-night (0–3 h) temporal performances are superior to those of late-night (20–23 h). This could be because the input nighttime DLST is captured at early-night (AN; 1:30 AM). Conversely,

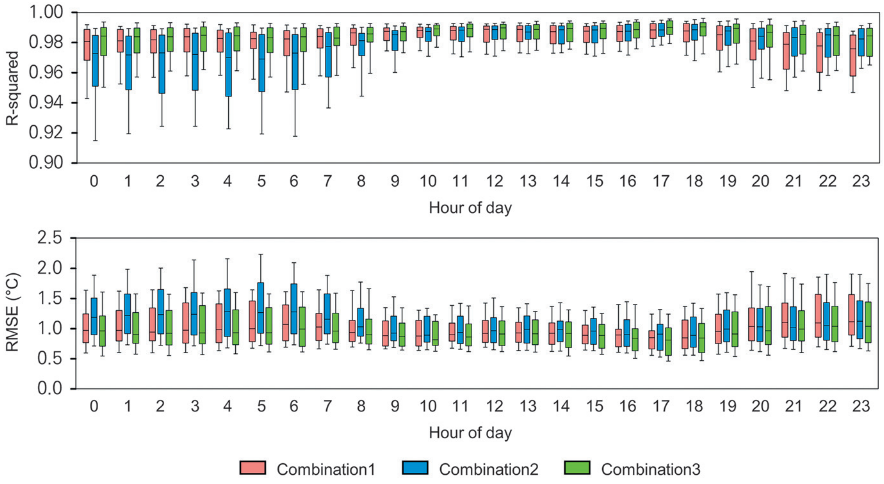


FIGURE 8.5 Leave-One-Out Cross-Validation (LOOCV) results for all hours for the three combinations with a unified sample size according to Combination 3.

combination 2 displays better temporal performance during late-night than early-night, but the difference is significantly larger than that of combination 1. This might be due to the incoming solar heat flows moving from surface to air, making it more challenging to estimate early-night air temperature from late-night LST (Chen et al., 2019).

Combination 3 demonstrates the best performance in terms of R^2 and RMSE among the three combinations, suggesting that using multiple LSTs could produce better air temperature estimation results. These findings are consistent with previous studies, which indicate that incorporating LSTs from multiple time periods improves the accuracy of air temperature estimation (Yoo et al., 2018). Additionally, the R^2 values for daytime are relatively higher than those for nighttime in Figure 8.5. This can be attributed to the larger variation range of temperature during daytime due to the influence of solar incoming, leading to a higher presence of extreme temperature values. Consequently, a high correlation is observed when these values are accurately simulated.

Figure 8.6 presents the all-hour average RMSE distribution based on LOOCV for each station by merging the three combinations. Most stations exhibit a low RMSE below 1.2 °C, particularly in urban core areas. The stations with the highest error are located in mountainous regions, where it is likely challenging to estimate the lower temperature values due to high altitudes. As more stations are established in mountainous areas in the future, the accuracy in these regions is expected to improve.

Nonetheless, when compared to the accuracy of previous studies, the RMSE demonstrates high performance. This is because the range of air temperature estimation error in benchmark studies lies between 1–2 K RMSE (Venter et al., 2020; Hrisko et al., 2020; Ho et al., 2014), indicating that the accuracy of the present study is quite encouraging. In summary, the LOOCV validation results imply that the hourly air temperatures are well estimated within urban areas, and the

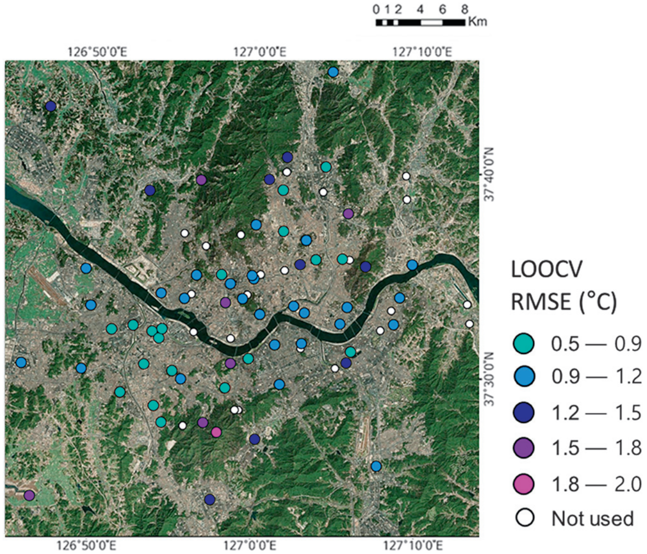


FIGURE 8.6 Average hourly Root Mean Square Error (RMSE) for each station from LOOCV validation results for the three combinations.

entire distribution of air temperatures in the study regions can be represented by the constructed model.

HOURLY AIR TEMPERATURE MAPPING ANALYSIS

Figure 8.7 illustrates the distribution of average normalized air temperature during four representative time periods (5 AM and 10 AM, 3 PM, and 8 PM) for both

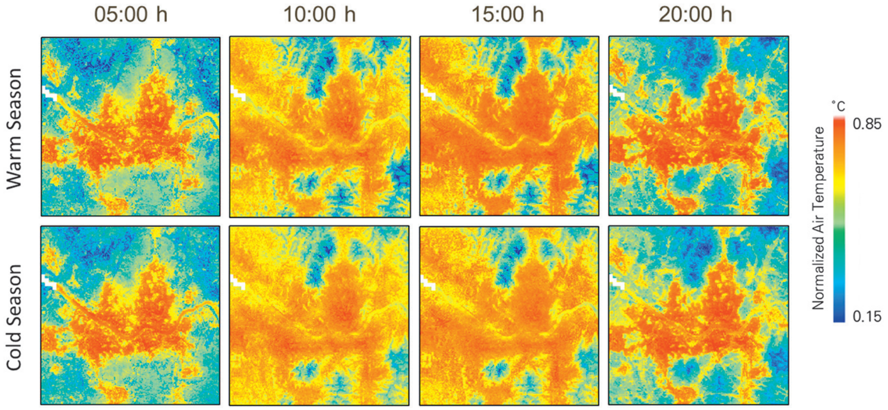


FIGURE 8.7 Final hourly 250 m air temperatures mapped and averaged for warm and cold seasons after normalization, with representative times at 5-hour intervals. Only 95% clear-sky days used: 53 days for Warm Season (May-Oct) and 44 days for Cold Season (Nov-Apr).

warm and cold seasons. For the warm season, a phenomenon of higher temperatures in the city center can be observed at 5 AM, early in the morning. A distinct temperature gradient is evident, with temperatures decreasing as one moves away from the city center. This is a result of the UHI effect, which is more pronounced during early morning hours when the influence of the sun is reduced. Vegetated areas experience a rapid drop in temperature, while impervious surfaces and human anthropogenic heat sources in urban areas retain heat (Sturiale & Scuderi, 2019).

At 10 AM, after the sun has risen, a different temperature distribution can be observed at 5 AM. Low plants and scattered tree areas around the city are heated up due to the influence of solar incoming radiation. At 3 PM, which is typically the time of day when the maximum air temperature is observed, the entire study area, excluding high-altitude mountainous regions, exhibits high temperatures with relatively low variation compared to nighttime. This temperature pattern changes at 8 PM, when the temperature in peripheral areas rapidly decreases, accentuating the higher temperatures in the urban center. A comparison of the temperature patterns at 5 AM and 8 PM reveals that the temperatures in the city center are generally higher than those in the surrounding areas at early morning and late-night. The study confirms that the UHI phenomenon is more pronounced at night than during the day in Seoul, which is consistent with previous studies on cities with a prominent nocturnal UHI effect (Hu & Xue., 2016; Yoo et al., 2022). In the cold season, similar temporal changes in temperature patterns are observed, although the intensity of warming in the city center is less pronounced than in the warm season. The temperature difference between the city center and the surrounding areas is smaller during the daytime (e.g., 10 AM and 3 PM) and also in the cold season.

By plotting the distribution of temperatures for each LCZ during the four time periods (Figure 8.8), distinct UHI patterns can be observed for both warm and cold seasons at 5 AM. Compact built-up types (LCZ1-3) exhibit the highest

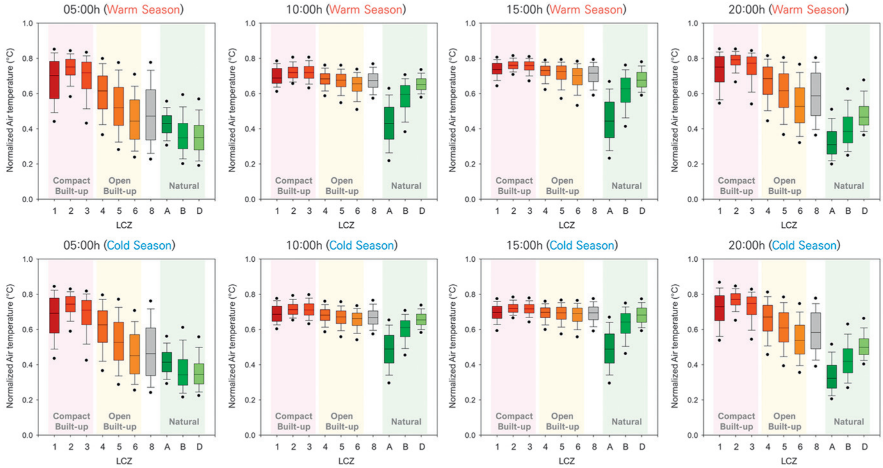


FIGURE 8.8 Boxplots of temperature distribution for hourly 250 m air temperatures using LCZ for each time, shown separately for warm (upper) and cold (lower) seasons.

temperatures, followed by open built-up types (LCZ4-6), and natural types (LCZA, B, and D). Among LCZ1-3, LCZ2 is slightly warmer than LCZ1, likely reflecting the shading effect of tall buildings (Zhou & Chen, 2018). The temperature difference due to height is more pronounced between open built-up types than compact built-up types at night, with lower heights (from LCZ4 to LCZ6) corresponding to lower average temperatures.

In contrast, during the daytime (10 AM and 3 PM), the temperature distribution among most classes does not exhibit significant variation, excluding the mountainous area of LCZA. The temperature difference between compact and open built-up types is not substantial, and natural types exhibit only slightly lower temperatures than urban types. Notably, during the cold season, the low plant natural type (LCZD) exhibits higher temperatures than the open built-up classes at 10 AM.

As night falls, substantial temperature variations are observed between urban and natural types, with a clear distribution pattern within open built-up types: higher building height correspond to higher average temperatures. The study results demonstrate that using LCZs enables a more detailed investigation of the distribution of thermal environments according to different building and vegetation types within a city, and the results suggest distinct differences in these patterns over time.

DIURNAL CYCLE OF UHI INTENSITY

Figure 8.9 displays the diurnal cycle of UHI intensity. For the difference between LCZ2 and LCZD, distinct curved cycles appear, with high UHI intensity in early nighttime that decreases during daytime and increases again in late-nighttime. The high UHI intensity at nighttime occurs because energy absorbed by man-made materials during the day is released slowly at night, resulting in the heating of air (Arnfield, 2003). Comparing the warm and cold seasons in Figure 8.9(a), higher UHI

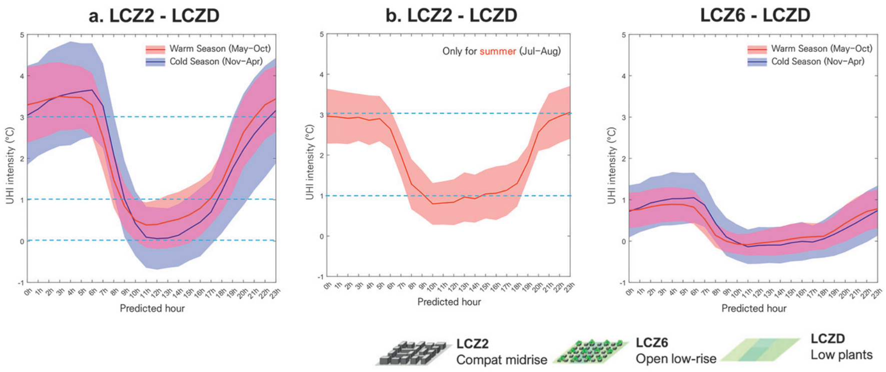


FIGURE 8.9 Diurnal cycle of Urban Heat Island (UHI) intensity between LCZ2 and LCZD for warm and cold seasons (a), summer season only (b), and the difference between LCZ6 and LCZD (c). Only 95% clear-sky days used: 53 days for Warm Season (May-Oct) and 44 days for Cold Season (Nov-Apr).

intensity is observed in early nighttime during the cold season, while daytime UHI is higher in the warm season than in the cold season. Notably, negative UHI intensity appears in the daytime during the cold season (i.e., UCI).

Focusing on the warm season (Figure 8.9(b)), daytime UHI intensity is higher than in other seasons. This result supports previous studies that revealed an association between heatwaves and UHI, where hot summer heatwaves could intensify UHI phenomena (Founda & Santamouris, 2017). In the case of calculating UHI intensity using LCZ6 and LCZD, a distinct UHI cycle between daytime and nighttime is not observed. This suggests that there are different patterns of UHI for each LCZ class, so using detailed air temperatures and LCZs can aid in in-depth UHI assessments.

NOVELTY AND LIMITATION

The novelty of this study lies in the construction of a two-step air temperature mapping framework, which enables the development of high spatiotemporal urban air temperature data. By performing satellite-based LST downscaling in step 1, we overcame the issue of low spatial resolution in satellite-based LST – mostly about 1 km – by downscaling the MODIS daily from 1 km LST to 250 m. We also addressed the main limitation of low temporal resolution in studies estimating air temperature from satellite-based LST by employing multi-task learning techniques to simultaneously estimate air temperatures at multiple times, resulting in hourly air temperature data. Notably, the validation results showed stable estimation performance (RMSE about 1.2 K) across all time frames, and the mapping analysis confirmed the ability to simulate time-specific urban temperature patterns accurately. Multi-task learning is particularly effective for hourly air temperature mapping, as it allows for simultaneous estimation of similar outputs, enabling efficient and rapid output production. The ability to quickly and effectively estimate the air temperature for all 24 hours is a significant novelty. This method can be particularly useful for real-time operating systems that require rapid production of air temperature data, enabling effective air temperature mapping.

In this study, we used the produced hourly urban air temperature map and LCZs to conduct an in-depth analysis of thermal environment patterns (e.g., UHI) based on urban building and vegetation types. The results from the study area of Seoul revealed that the difference in temperature distribution between LCZ classes is greater at night than during the day, with a clear difference in temperature according to height in open-type LCZs at night. Moreover, the diurnal cycle patterns of UHI intensity were identified using LCZs, with distinct curved patterns showing high intensity in the early morning, lower intensity during the day, and decreasing intensity at night. Particularly, we found that slight cooling heat island phenomena appeared during the cold season. This study highlights the benefits of using LCZs in contrast to previous research, which mostly analyzed specific time frames in line with satellite image capture times by directly mapping hourly air temperature and examining time patterns and diurnal cycles for each LCZ class.

A major limitation of this study is its reliance on satellite LST data. Although we proposed a method to analyze the thermal environment of the target date using past HLSTs to overcome issues such as cloud coverage in conventional satellite images, the downscaling accuracy may decrease if there are no similar HLSTs for the target date. Additionally, despite proposing a combination method to address cloud-related issues, there are still dates when mapping could not be performed due to the influence of clouds. We expect that utilizing LST data with gap-filled cloud coverage could further improve this issue (Shiff et al., 2021). As this study is a case study targeting Seoul, the thermal environment patterns revealed, such as high UHI intensity at nighttime compared to daytime, may not be generalized to all cities. Different patterns may emerge depending on the city's climate condition and geographical characteristics. As more cities acquire sufficient and reliable observation station data, we anticipate that more interesting analyzes of city-specific thermal environment patterns will become possible.

CONCLUSION

Urban areas have been experiencing an increase in temperature due to rapid urbanization and climate change, resulting in the phenomenon known as the UHI effect. Notably, the thermal environment within cities exhibits diverse patterns depending on the types of buildings and vegetation, and the distribution and intensity of UHI vary over time. To analyze these patterns, high spatiotemporal resolution temperature data is required, which can monitor extensive urban areas. However, satellite imagery providing LST faces limitations due to the trade-off between spatial and temporal resolutions and cloud contamination issues. Additionally, 1.5 m ground-above air temperature data from observation stations is insufficient for monitoring detailed urban thermal environment distributions due to the limited number of stations within cities.

In this study, we propose a novel two-step algorithm for mapping hourly urban air temperature at a 250 m spatial resolution in Seoul. In step 1, we downscaled the 1 km MODIS LST data to 250 m using a machine learning Random Forest model, employing high-resolution satellite LST data from sources such as ASTER and Landsat as primary input data for downscaling. In step 2, we estimated hourly air temperature from the downscaled 250 m LST data for multiple time periods. To address the cloud contamination issue in satellite-based LST, we proposed a method for combining LST data from various time periods. We also suggested an efficient method for simultaneously calculating 24-hour hourly air temperature using a Random Forest-based multi-task learning estimation model.

The results of the proposed air temperature estimation technique showed high estimation accuracy, with an average RMSE of approximately 1.0 °C across all 24 hours. The generated air temperature maps effectively simulated detailed urban temperature patterns and their variations across different times. By analyzing the thermal environment distribution of urban types using LCZs, we found that in Seoul, higher building density and taller buildings within open built-up type LCZ classes were associated with higher temperatures and distinct UHI patterns during the night. In contrast, during the day, most regions within the city exhibited relatively lower UHI effects due to sufficient incoming solar heat-up. The diurnal cycle of UHI intensity displayed a curved pattern, with higher intensity at night, lower intensity during the day, and increasing intensity again at night. However, the magnitude of this cycle varied depending on which LCZ class was used as a basis for calculating intensity. The hourly air temperature maps produced in this study can be considered valuable resources for enhancing the understanding of urban thermal environments in both spatial and temporal dimensions. In particular, the multi-task learning technique employed for estimating multiple time periods simultaneously is expected to be effectively utilized in practical applications.

REFERENCES

Alexander, P. J., & Mills, G. (2014). Local climate classification and Dublin's urban heat island. *Atmosphere*, 5(4), 755–774.

Angel, S., Parent, J., Civco, D. L., Blei, A., & Potere, D. (2011). The dimensions of global urban expansion: Estimates and projections for all countries, 2000–2050. *Progress in Planning*, 75(2), 53–107.

Arnfield, A. J. (2003). Two decades of urban climate research: A review of turbulence, exchanges of energy and water, and the urban heat island. *International Journal of Climatology: a Journal of the Royal Meteorological Society*, 23(1), 1–26.

Ashiq, M. W., Zhao, C., Ni, J., & Akhtar, M. (2010). GIS-based high-resolution spatial interpolation of precipitation in mountain–plain areas of Upper Pakistan for regional climate change impact studies. *Theoretical and Applied Climatology*, 99, 239–253.

Bárdossy, A., Seidel, J., & El Hachem, A. (2021). The use of personal weather station observations to improve precipitation estimation and interpolation. *Hydrology and Earth System Sciences*, 25(2), 583–601.

Basu, R., & Samet, J. M. (2002). Relation between elevated ambient temperature and mortality: a review of the epidemiologic evidence. *Epidemiologic Reviews*, 24(2), 190–202.

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24–31.

Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.

Cao, J., Zhou, W., Zheng, Z., Ren, T., & Wang, W. (2021). Within-city spatial and temporal heterogeneity of air temperature and its relationship with land surface temperature. *Landscape and Urban Planning*, 206, 103979.

Chang, Y., Xiao, J., Li, X., Frolking, S., Zhou, D., Schneider, A., … & Wu, Y. (2021). Exploring diurnal cycles of surface urban heat island intensity in Boston with land surface temperature data derived from GOES-R geostationary satellites. *Science of the Total Environment*, 763, 144224.

Chen, J., Chu, R., Wang, H., Zhang, L., Chen, X., & Du, Y. (2019). Alleviating urban heat island effect using high-conductivity permeable concrete pavement. *Journal of Cleaner Production*, 237, 117722.

dos Santos, R. S. (2020). Estimating spatio-temporal air temperature in London (UK) using machine learning and earth observation satellite data. *International Journal of Applied Earth Observation and Geoinformation*, 88, 102066.

Ebrahimy, H., Aghighi, H., Azadbakht, M., Amani, M., Mahdavi, S., & Matkan, A. A. (2021). Downscaling MODIS land surface temperature product using an adaptive random forest regression method and Google Earth Engine for a 19-years spatio-temporal trend analysis over Iran. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 2103–2112.

Eldesoky, A. H., Gil, J., & Pont, M. B. (2021). The suitability of the urban local climate zone classification scheme for surface temperature studies in distinct macroclimate regions. *Urban Climate*, 37, 100823.

Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., … & Snyder, P. K. (2005). Global consequences of land use. *Science*, 309(5734), 570–574.

Founda, D., & Santamouris, M. (2017). Synergies between Urban Heat Island and Heat Waves in Athens (Greece), during an extremely hot summer (2012). *Scientific Reports*, 7(1), 10973.

Hart, Q. J., Brugnach, M., Temesgen, B., Rueda, C., Ustin, S. L., & Frame, K. (2009). Daily reference evapotranspiration for California using satellite imagery and weather station measurement interpolation. *Civil Engineering and Environmental Systems*, 26(1), 19–33.

Ho, H. C., Knudby, A., Sirovyak, P., Xu, Y., Hodul, M., & Henderson, S. B. (2014). Mapping maximum urban air temperature on hot summer days. *Remote Sensing of Environment*, 154, 38–45.

Hrisko, J., Ramamurthy, P., Yu, Y., Yu, P., & Melecio-Vázquez, D. (2020). Urban air temperature model using GOES-16 LST and a diurnal regressive neural network algorithm. *Remote Sensing of Environment*, 237, 111495.

Hu, X. M., & Xue, M. (2016). Influence of synoptic sea-breeze fronts on the urban heat island intensity in Dallas–Fort Worth, Texas. *Monthly Weather Review*, 144(4), 1487–1507.

Hutengs, C., & Vohland, M. (2016). Downscaling land surface temperatures at regional scales with random forest regression. *Remote Sensing of Environment*, 178, 127–141.

Im, E. S., & Eltahir, E. A. (2018). Urbanization-induced drought: Anthropogenic land-use change impacts on the regional climate. *Journal of Climate*, 31(24), 10077–10094. doi: 10.1175/JCLI-D-18-0090.1

Kustas, W. P., Norman, J. M., Anderson, M. C., & French, A. N. (2003). Estimating subpixel surface temperatures and energy fluxes from the vegetation index–radiometric temperature relationship. *Remote Sensing of Environment*, 85(4), 429–440.

Lu, D., Moran, E., & Hetrick, S. (2011). Detection of impervious surface change with multitemporal Landsat images in an urban–rural frontier. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), 298–306.

Luintel, N., Ma, W., Ma, Y., Wang, B., & Subba, S. (2019). Spatial and temporal variation of daytime and nighttime MODIS land surface temperature across Nepal. *Atmospheric and Oceanic Science Letters*, 12(5), 305–312.

Meier, F., Fenner, D., Grassmann, T., Otto, M., & Scherer, D. (2017). Crowdsourcing air temperature from citizen weather stations for urban climate research. *Urban Climate*, 19, 170–191.

Quan, S. J., & Bansal, P. (2021). A systematic review of GIS-based local climate zone mapping studies. *Building and Environment*, 196, 107791.

Seto, K. C., Güneralp, B., & Hutyra, L. R. (2012). Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proceedings of the National Academy of Sciences*, 109(40), 16083–16088.

Shi, T., Huang, Y., Wang, H., Shi, C. E., & Yang, Y. J. (2015). Influence of urbanization on the thermal environment of meteorological station: Satellite-observed evidence. *Advances in Climate Change Research*, 6(1), 7–15.

Shiff, S., Helman, D., & Lensky, I. M. (2021). Worldwide continuous gap-filled MODIS land surface temperature dataset. *Scientific Data*, 8(1), 74.

Stewart, I. D., & Oke, T. R. (2012). Local Climate Zones for urban temperature studies. *Bulletin of the American Meteorological Society*, 93(12), 1879–1900. doi: 10.1175/BAMS-D-11-00019.1

Stone Jr, B. (2005). Urban heat and air pollution: An emerging role for planners in the climate change debate. *Journal of the American Planning Association*, 71(1), 13–25.

Sturiale, L., & Scuderi, A. (2019). The role of green infrastructures in urban planning for climate change adaptation. *Climate*, 7(10), 119.

Tzoulas, K., Korpela, K., Venn, S., Yli-Pelkonen, V., Kaźmierczak, A., Niemela, J., & James, P. (2007). Promoting ecosystem and human health in urban areas using Green Infrastructure: A literature review. *Landscape and Urban Planning*, 81(3), 167–178.

Venter, Z. S., Brousse, O., Esau, I., & Meier, F. (2020). Hyperlocal mapping of urban air temperature using remote sensing and crowdsourced weather data. *Remote Sensing of Environment*, 242, 111791.

Voogt, J. A., & Oke, T. R. (2003). Thermal remote sensing of urban climates. *Remote Sensing of Environment*, 86(3), 370–384.

Wan, Z., & Dozier, J. (1996). A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Transactions on Geoscience and Remote Sensing*, 34(4), 892–905.

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 Climate*, 49, 101499.

Waqas, A., & Athar, H. (2018). Observed diurnal temperature range variations and its association with observed cloud cover in northern Pakistan. *International Journal of Climatology*, 38(8), 3323–3336.

Xia, H., Chen, Y., Li, Y., & Quan, J. (2019). Combining kernel-driven and fusion-based methods to generate daily high-spatial-resolution land surface temperatures. *Remote Sensing of Environment*, 224, 259–274.

Yang, J., Jin, S., Xiao, X., Jin, C., Xia, J. C., Li, X., & Wang, S. (2019). Local climate zone ventilation and urban land surface temperatures: Towards a performance-based and wind-sensitive planning proposal in megacities. *Sustainable Cities and Society*, 47, 101487.

Yang, L., Xian, G., Klaver, J. M., & Deal, B. (2003). Urban land-cover change detection through sub-pixel imperviousness mapping using remotely sensed data. *Photogrammetric Engineering & Remote Sensing*, 69(9), 1003–1010.

Yoo, C., Im, J., Cho, D., Lee, Y., Bae, D., & Sismanidis, P. (2022). Downscaling MODIS nighttime land surface temperatures in urban areas using ASTER thermal data through local linear forest. *International Journal of Applied Earth Observation and Geoinformation*, 110, 102827.

Yoo, C., Im, J., Cho, D., Yokoya, N., Xia, J., & Bechtel, B. (2020a). Estimation of all-weather 1 km MODIS land surface temperature for humid summer days. *Remote Sensing*, 12(9), 1398.

Yoo, C., Im, J., Park, S., & Quackenbush, L. J. (2018). Estimation of daily maximum and minimum air temperatures in urban landscapes using MODIS time series satellite data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 137, 149–162.

Yoo, C., Lee, Y., Cho, D., Im, J., & Han, D. (2020b). Improving local climate zone classification using incomplete building data and Sentinel 2 images based on convolutional neural networks. *Remote Sensing*, 12(21), 3552.

Zeng, C., Long, D., Shen, H., Wu, P., Cui, Y., & Hong, Y. (2018). A two-step framework for reconstructing remotely sensed land surface temperatures contaminated by cloud. *ISPRS Journal of Photogrammetry and Remote Sensing*, 141, 30–45.

Zhou, D., Xiao, J., Bonafoni, S., Berger, C., Deilami, K., Zhou, Y., … & Sobrino, J. A. (2018). Satellite remote sensing of surface urban heat islands: Progress, challenges, and perspectives. *Remote Sensing*, 11(1), 48.

Zhou, X., & Chen, H. (2018). Impact of urbanization-related land use land cover changes and urban morphology changes on the urban heat island phenomenon. *Science of the Total Environment*, 635, 1467–1476.