ISPRS Journal of Photogrammetry and Remote Sensing 137 (2018) 149–162

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Estimation of daily maximum and minimum air temperatures in urban landscapes using MODIS time series satellite data

Cheolhee Yoo [a], Jungho Im [a,*], Seonyoung Park [a], Lindi J. Quackenbush [b]

[a] School of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan, South Korea
[b] Department of Environmental Resources Engineering, State University of New York College of Environmental Science and Forestry, Syracuse, NY, USA

ARTICLE INFO

Article history:
Received 6 August 2017
Received in revised form 25 January 2018
Accepted 25 January 2018

Keywords:
Land surface temperature
Air temperature
Random forest
MODIS

ABSTRACT

Urban air temperature is considered a significant variable for a variety of urban issues, and analyzing the spatial patterns of air temperature is important for urban planning and management. However, insufficient weather stations limit accurate spatial representation of temperature within a heterogeneous city. This study used a random forest machine learning approach to estimate daily maximum and minimum air temperatures (T_{max} and T_{min}) for two megacities with different climate characteristics: Los Angeles, USA, and Seoul, South Korea. This study used eight time-series land surface temperature (LST) data from Moderate Resolution Imaging Spectroradiometer (MODIS), with seven auxiliary variables: elevation, solar radiation, normalized difference vegetation index, latitude, longitude, aspect, and the percentage of impervious area. We found different relationships between the eight time-series LSTs with T_{max}/T_{min} for the two cities, and designed eight schemes with different input LST variables. The schemes were evaluated using the coefficient of determination (R^2) and Root Mean Square Error (RMSE) from 10-fold cross-validation. The best schemes produced R^2 of 0.850 and 0.777 and RMSE of 1.7 °C and 1.2 °C for T_{max} and T_{min} in Los Angeles, and R^2 of 0.728 and 0.767 and RMSE of 1.1 °C and 1.2 °C for T_{max} and T_{min} in Seoul, respectively. LSTs obtained the day before were crucial for estimating daily urban air temperature. Estimated air temperature patterns showed that T_{max} was highly dependent on the geographic factors (e.g., sea breeze, mountains) of the two cities, while T_{min} showed marginally distinct temperature differences between built-up and vegetated areas in the two cities.

1. Introduction

Air temperature measured at 2 m above the ground in urban landscapes is closely related to various urban problems such as the urban heat island effect (Krüger and Emmanuel, 2013; Li et al., 2004), air pollution (Aw and Kleeman, 2003; Katsouyanni et al., 1993), and human mortality (Liu et al., 2011; Basu and Samet, 2002). In particular, monitoring and prediction of urban maximum and minimum air temperatures (T_{max} and T_{min}, respectively) are crucial in populated areas as they are directly related to fatal disasters such as heat waves and tropical nights (Gabriel and Endlicher, 2011; Romero-Lankao et al., 2012). With large populations and complex infrastructure, tiny temperature changes within a city may significantly affect both human and natural environments (Hondula et al., 2012; Schuster et al., 2014). Therefore, it is important to understand and monitor the spatio-temporal patterns of urban air temperature.

Air temperature is generally measured at weather stations with high temporal resolution and accuracy. In a city, natural and artificial materials are mixed at various spatial scales and respond differently to incoming solar radiation, which in turn affects air temperature (Hart and Sailor, 2009). However, the number of weather stations in urban areas is limited and thus the spatial distribution of air temperature determined through spatial interpolation of *in situ* measurements may be insufficient to provide detailed temperature variation at local scale (Courault and Monestiez, 1999). In addition, approximately three-quarters of the world's largest cities are located very close to coastlines, which complicates accurate spatial interpolation of station-based temperature data (Vogt et al., 1997).

Satellite data provide global land surface information, such as land surface temperature (LST), vegetation indices, elevation, and land cover, which enable estimation of urban air temperature based on the assumption that such land surface information is

* Corresponding author.
E-mail address: ersgis@unist.ac.kr (J. Im).

https://doi.org/10.1016/j.isprsjprs.2018.01.018

highly related to urban air temperature. Many studies have focused on the use of remote sensing-derived LST to estimate air temperature with various approaches such as the Temperature-Vegetation indeX (TVX) method, energy balance models, and statistical and machine learning approaches.

The TVX method assumes that the surface temperature of an infinitely thick, fully vegetated canopy is close to its surrounding air temperature. This technique focuses on the negative correlation between Normalized Difference Vegetation Index (NDVI) and LST (Nemani and Running, 1989), so that air temperature can be obtained by using the maximum NDVI of the study area under investigation (Prihodko and Goward, 1997; Zhu et al., 2013; Stisen et al., 2007). Energy balance models are based on thermodynamics, which calculate the radiation balance from latent, sensible, and soil heat fluxes, considering that the atmosphere and the surface exchange energy and mass (Sun et al., 2005; Zhang et al., 2015). Many statistical analyses, including simple and multiple linear regressions, have been employed to estimate air temperature using LST (Vogt et al., 1997), or with auxiliary variables such as NDVI, solar radiation, and elevation (Shi et al., 2016; Chen et al., 2015; Xu et al., 2012). Recent studies have used machine learning approaches such as random forest (RF), Cubist, support vector machine (SVM), and neural networks to estimate air temperature from remote sensing data (Zhang et al., 2016; Moser et al., 2015; Ho et al., 2014). Machine learning has proven to be flexible in areas with complicated and heterogeneous landscapes for estimating air temperature from LST and additional variables (Noi et al., 2017; Ho et al., 2014).

Two Moderate Resolution Imaging Spectroradiometer (MODIS) sensors onboard Terra and Aqua satellites provide four LST datasets (two daytime and two nighttime) per day. Studies have investigated air temperature estimation using various combinations of these four MODIS LST datasets (Noi et al., 2017; Zhang et al., 2016). They commonly found that estimation accuracy varied by both model and the data used in the combination.

LST datasets have been widely used in urban application studies (Weng, 2009; Zhang et al., 2017; Zhang and Li, 2018; Song and Wu, 2018). However, it is difficult to estimate air temperature solely using remote sensing within urban areas, which typically have a smaller dynamic temperature range than a larger region that includes a mixture of rural and urban areas with a multitude of land cover types. A strong linear relationship between LST and air temperature within a city might not be valid (Voogt and Oke, 2003). Micro-scale advection from neighboring areas with complex surface materials and different land use deteriorates the significance of the relationship between LST and air temperature in urban landscapes (Stoll and Brazel, 1992). Since the TVX method is suitable for areas with high vegetation coverage, it is not appropriate for estimating air temperature in urban regions that typically have low vegetation cover. Consequently, the linear relationship between LST and NDVI is not always valid and many factors including seasonality, soil moisture, land cover, and vegetation phenology affect the relationship (Sandholt et al., 2002; Vancutsem et al., 2010). Quantitative energy balance modeling in an urban area is difficult because of heterogeneous surface materials and complex urban landscapes with varying amount of incoming fluxes (Harman, 2003; Grimmond et al., 2010). Therefore, in order to estimate urban air temperature using satellite images, it is necessary to apply more advanced methods, including multiple regression models using various time series LSTs, machine learning approaches, or other sophisticated algorithms (Ho et al., 2014; Bechtel et al., 2014, 2017; Agathangelidis et al., 2016; Keramitsoglou et al., 2016).

Ho et al. (2014) used two machine learning techniques—SVM and RF—as well as ordinary least squares regression to map the urban T_{max} of Vancouver, Canada, from Landsat satellite data, resulting in a root mean square error (RMSE) of 2.31 °C. However, Landsat satellite imagery has limitations for monitoring urban air temperature due to its low temporal resolution with a 16-day revisit. Bechtel et al. (2014) applied multiple regression methods to estimate air temperature in Hamburg, Germany using LST time series data with 15-min temporal resolution from the geostationary Spinning Enhanced Visible Infra-Red Imager instrument onboard the Meteosat Second Generation satellite (MSG-SEVIRI). To estimate urban air temperature at a particular time, LSTs collected at earlier times were integrated (RMSE = 1.5–1.8 °C). Moreover, Bechtel et al. (2017) modified the method suggested by Bechtel et al. (2014) to improve the air temperature estimation by considering global radiation of specific time together with dense time series of LSTs for 29 urban and rural stations in Germany (RMSE <2.0 °C for most stations). Agathangelidis et al. (2016) estimated air temperature in Athens, Greece from MSG-SEVIRI images, by applying multi-temporal approaches, using polynomial regression and artificial neural networks. Second order polynomial regression combining concurrent LST and air temperature observations with LST from the previous 1–8 h resulted in RMSEs of 1.0–1.2 °C, with artificial neural networks giving a slightly better result. Using more LST data generally yielded higher model performance because air temperature is influenced by ground temperature from the previous hours (Bechtel et al., 2014, 2017; Agathangelidis et al., 2016). Keramitsoglou et al. (2016) developed an operational real-time nowcasting module for monitoring air temperature at interpolated 1 km resolution with MSG-SEVIRI images. They retrieved vertical profiles of air temperature based on a physical algorithm using MSG-SEVIRI LST and evaluated model performance using air temperature observations measured in 15 cities in Europe and North Africa (RMSE = 2.3 °C). Since the time when daily T_{max} and T_{min} occur does not typically correspond to satellite data acquisition time, it is very challenging to estimate them from satellite-derived LST data. This present study therefore investigated how multiple LSTs measured at different times can be effectively combined to estimate the T_{max} and T_{min} in two urban regions with different climate conditions.

The objectives of this research were to: (1) examine the relationship between multi-temporal LSTs and daily T_{max} and T_{min} in urban regions with different climate conditions; (2) develop machine learning-based models to estimate urban T_{max} and T_{min} using selected LSTs and other satellite-derived data; and (3) investigate the spatial distribution of the daily T_{max} and T_{min} during summer time in urban landscapes. Moderate Resolution Imaging Spectroradiometer (MODIS) LST time series data were used to estimate daily T_{max} and T_{min} in two mega cities—Los Angeles, USA and Seoul, South Korea. NDVI, elevation, solar radiation, latitude, longitude, aspect and percentage of impervious area were used as auxiliary data. Random forest, a widely used machine learning approach, was used to estimate daily T_{max} and T_{min}.

2. Study area

Two urban megacities with different climates, Los Angeles and Seoul, were selected for this study (Fig. 1). Los Angeles, the second-most-populated city in the USA (34°03′N and 118°15′W), is composed of both flat and hilly regions. The eastern end of the Santa Monica Mountains, stretching from downtown Los Angeles to the Pacific Ocean, separates the Los Angeles Basin from the San Fernando Valley. The city covers an area of 1290 km^2 and is bounded by the San Gabriel Mountains to the east. The population of Los Angeles in 2016 was estimated to be about 4 million. Los Angeles has a Mediterranean climate, according to Koeppen-Geiger climate classification system, with dry summers and moist winters. Precipitation is concentrated in late fall, winter, and

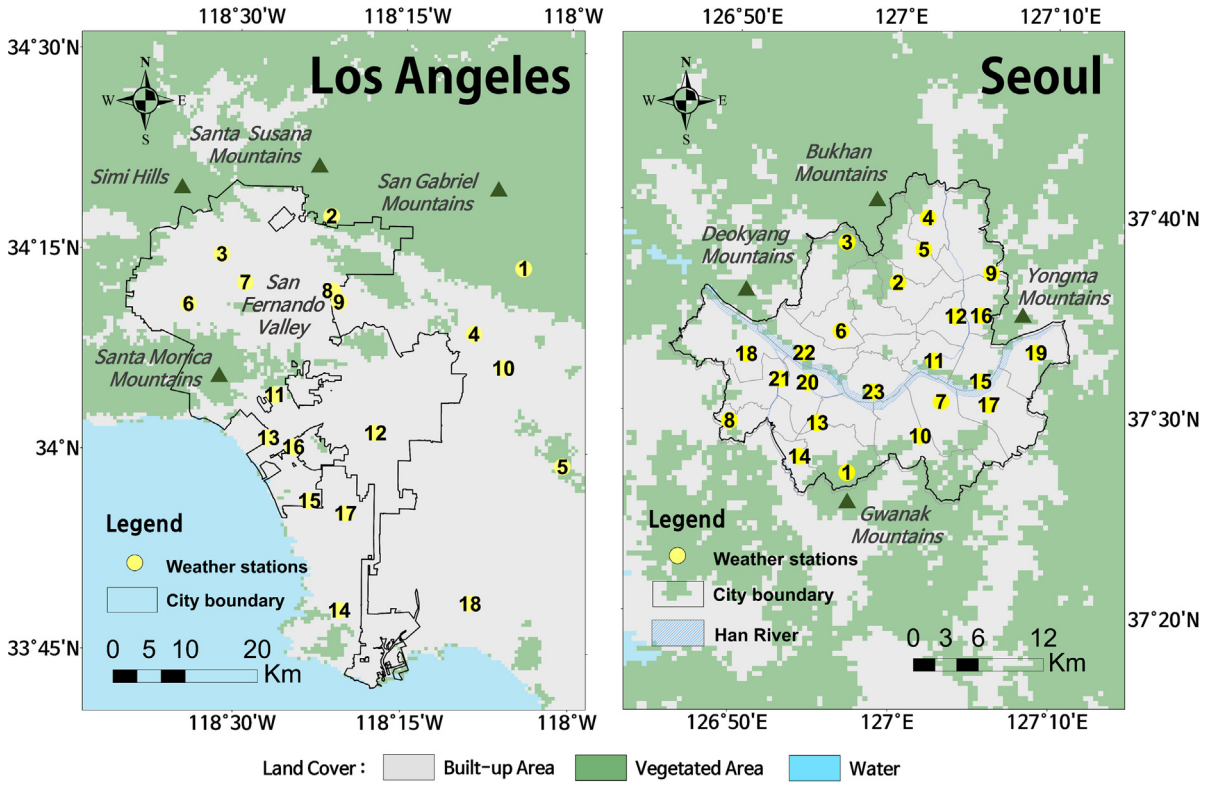

Fig. 1. Study areas and the location of weather stations for Los Angeles (Left) and Seoul (Right). The black line indicates the city boundary. Stations are numbered in order of decreasing elevation (the larger number, the lower elevation). Land cover is aggregated from MODIS MCD12Q1 land cover.

spring, with much less rain in the summer (~0.04 in. during July and August). The average T_{max} and T_{min} are 29.1 °C and 17.8 °C in August, which is the warmest month in Los Angeles.

The other study area is Seoul (37°33′N and 126°58′E), which is situated in the northwest of South Korea. Seoul is geographically divided into northern and southern parts by the Han River and is bordered by four distinct mountains. The center of the city is located 60 km inland from the west coast of South Korea. Seoul has an area of 605 km^2 with very high population density (i.e., the estimated population in 2016 was about 10 million). Seoul has a humid subtropical climate according to the Koeppen-Geiger climate classification system. Summers in Seoul are especially hot and humid because of East Asian monsoons, and have the relatively high amount of summer precipitation (~15.54 in. during July and August). The average T_{max} and T_{min} are 29.6 °C and 22.4 °C, respectively, in August, which is the warmest month in the city.

3. Data and methods

3.1. Satellite data

MODIS Terra and Aqua data were used in this study. MODIS data have been widely utilized to observe atmosphere, ocean and land processes in the Earth. MODIS products—LST and NDVI (h08v05 for Los Angeles and h28v05 for Seoul)—from 2006 to 2016 in hot summer days (July and August) were downloaded from reverb echo (http://reverb.echo.nasa.gov). Daily LST data (MYD11A1 for Aqua and MOD11A1 for Terra), which have 1 km spatial resolution, were used. The MODIS LST data were produced using a generalized split-window algorithm (Wan and Dozier, 1996; Li et al., 2013b). Terra MODIS provides LST data observed at 10:30 am solar local (daytime) and 10:30 pm (nighttime). Aqua MODIS also provides two LST collections at local solar times of 1:30 pm (daytime) and 1:30 am (nighttime). NDVI data were obtained from a 16-day composition of MOD13A2 with 1 km spatial resolution.

The elevation of the two cities was retrieved from Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) with 30 m spatial resolution (https://earthexplorer.usgs.gov). Land cover type data were available from Fine Resolution Observation and Monitoring of Global Land Cover (FROM-GLC), which is a 30 m spatial resolution global land-cover map produced using Landsat Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) data (http://data.ess.tsinghua.edu.cn). A global land cover map (FROM-GLC-Hierarchy) with 250 m spatial resolution created from FROM-GLC (Yu et al., 2014), was obtained to calculate the percentage of impervious surfaces in this study. MODIS land cover data (MCD12Q1) was also used for classifying built-up and vegetated area within the two cities.

3.2. Satellite-derived variables

From the Terra and Aqua satellites, daytime and nighttime LSTs were obtained (LST_{TD}, LST_{TN}, LST_{AD}, LST_{AN}; see Table 1 for designations). In consideration of the lag time of heat transport from ground level land surface temperature to 2 m air temperature (Jin and Mullens, 2014), land surface temperatures taken the day before (LST_{TBD}, LST_{TBN}, LST_{ABD}, LST_{ABN}) were also used. Thus, a total of eight LSTs were taken as variables for daily air temperature estimations in this study (Table 1).

The auxiliary variables to be used in conjunction with the LSTs were selected based on previous studies of air temperature estimation (Table 1); NDVI, elevation, solar radiation, latitude, and longitude have frequently been used in the literature (Xu et al., 2012; Ho et al., 2014; Janatian et al., 2017; Yan et al., 2009). We also used aspect and the percentage of impervious area as additional input variables. Area solar radiation was calculated using the ArcGIS Solar Analyst tool (Esri) to get daily solar radiation. Incoming solar

Table 1
Description of variables used in this study.

Variable type	Acronym (unit)	Description
Land surface temperature (LST)	LST_{AN} (°C)	Aqua LST taken on the day (1:30 am; night)
	LST_{TD} (°C)	Terra LST taken on the day (10:30 am; day)
	LST_{AD} (°C)	Aqua LST taken on the day (1:30 pm; day)
	LST_{TN} (°C)	Terra LST taken on the day (10:30 pm; night)
	LST_{ABN} (°C)	Aqua LST taken the day before (1:30 am; night)
	LST_{TBD} (°C)	Terra LST taken the day before (10:30 am; day)
	LST_{ABD} (°C)	Aqua LST taken the day before (1:30 pm; day)
	LST_{TBN} (°C)	Terra LST taken the day before (10:30 pm; night)
Auxiliary variable (α)	Sol (wh/m^2)	Daily incoming solar radiation
	NDVI	Normalized Difference Vegetation Index
	Lat (°)	Latitude
	Lon (°)	Longitude
	Elev (m)	Elevation
	Asp	Transformed Aspect
	Imp (%)	Percentage of impervious areas
Air temperature	T_{max} (°C)	Maximum air temperature
	T_{min} (°C)	Minimum air temperature

radiation of each pixel is computed by inputting elevation and day of year (DOY) into ArcGIS. Latitude and longitude values were extracted from the information contained in the MODIS tiles. Aspect has been shown to affect air temperature (McCutchan and Fox, 1986). To facilitate incorporation in analysis, the transformed aspect by Beers et al. (1966) was used as an input variable in this study (Eq. (1)).

$$\text{Transformed aspect} = \text{Cos}(45^\circ - \text{Aspect}) + 1 \quad (1)$$

The extent of impervious areas was used to represent the urbanization of the region (Piyoosh and Ghosh, 2017; Hao et al., 2016; Arnold and Gibbons, 1996). We calculated the percentage of impervious surface areas within each $1 \times 1\ km^2$ pixel using the FROM-GLC land cover data that has a 250 m spatial resolution. Daily solar radiation (Sol) was assumed to have the same temporal pattern by year. NDVIs of a 16-day composite were assumed to have the same daily values during the 16-day cycle. Elevation (Elev), aspect (Asp), longitude (Lon), latitude (Lat), and percentage of impervious areas (Imp) were used as static variables. In this study, these seven auxiliary variables were together called "α".

3.3. In situ meteorological data

Daily T_{max} and T_{min} during July and August from 2006 to 2016 were obtained from weather stations in Los Angeles (i.e., National Climatic Data Center (NCDC) stations) and Seoul (i.e., Automatic Weather Stations (AWSs) operated by Korea Meteorological Administration) and used as reference data. Since Los Angeles has a relatively small number of stations, additional stations located near the city boundary were also included in this study. A total of 18 NCDC station around Los Angeles and 23 AWS stations in Seoul were obtained for the analysis.

3.4. Methods for estimating air temperatures

Fig. 2 shows the summarized process flow of our proposed methodology. We compared each of eight LSTs (LST_{TD}, LST_{TN}, LST_{AD}, LST_{AN}, LST_{TBD}, LST_{TBN}, LST_{ABD} and LST_{ABN}) with *in situ* T_{max} and T_{min} for the two cities to design schemes that used different combinations of LSTs. Based on the correlation results between each LST and the *in situ* data, a total of eight test schemes (S1 to S8) were developed to estimate daily T_{max} and T_{min}. The first scheme (S1) has the LST that resulted in the highest correlation with *in situ* measurements; the second scheme (S2) uses the two LSTs that yielded the first and second highest correlation with *in situ* data, with subsequent schemes adding in LSTs in successive correlation order. In this way, the last scheme (S8) uses all eight LSTs. The seven auxiliary variables (α) were added for each scheme (refer to Table 3 in Results and Discussion for scheme summary).

The machine learning approach random forest (RF), was used to estimate the air temperatures. RF has been widely used for classification and regression in remote sensing applications (Amani et al., 2017; Kim et al., 2015; Lu et al., 2014; Naidoo et al., 2012; Park et al., 2016; Rhee et al., 2014; Richardson et al., 2017; Rodriguez-Galiano et al., 2012; Sonobe et al., 2017). The RF approach has recently begun to be used for the estimation of air temperature (Zhang et al., 2016; Ho et al., 2014). RF is an algorithm based on classification and regression trees (CART; Breiman, 2001), which uses a recursive binary split approach to reach final nodes in a tree structure. RF generates a large set of CARTs (typically 500–1000 trees) based on randomly selected subsets through bootstrapping from both training samples for a tree and predictor variables at each node of the tree. This bootstrapping-based randomization approach overcomes existing problems in the CART method, including sensitivity to the training samples and overfitting (Belgiu and Drăguţ, 2016; Kim et al., 2014; Lee et al., 2017; Im et al., 2016; Yoo et al., 2012). To reach a final decision, RF aggregates the outputs from many trees through averaging or weighted averaging for regression. In the RF models for the eight schemes (S1 to S8), LSTs and seven auxiliary variables (α) were used as independent variables, and *in situ* daily T_{max} or T_{min} was used as a dependent variable. The RF was implemented using R statistical software through the 'randomForest' add-on package with default model parameter settings, except for the number of trees (ntree). In this study, we used 1000 trees, which is a widely selected ntree value in remote sensing applications such as estimating air temperature (Xu et al., 2014; Li et al., 2013a; Park et al., 2016).

Since different numbers of LSTs were used for the models, the number of available samples differed by model mainly due to clouds. In order to appropriately evaluate the models, the same samples were used to calibrate and validate the models. Thus, the same samples were determined for all eight models and used for calibration and validation. Ten (10)-fold cross-validation was conducted to evaluate the performance of the eight schemes for each air temperature. The samples were separated into ten subsets and each subset was used to validate the model that was developed using the other nine subsets. This procedure was repeated in all ten subsets for every scheme. The coefficient of determination (R^2) and root mean square error (RMSE) calculated using 10-fold cross-validation results were used to compare the performance of the schemes. Normalized RMSE (i.e., RMSE divided by the range of observations; nRMSE) was also used when different models with different samples were compared in terms of performance. The scheme that resulted in the highest accuracy was identified and used to investigate the spatial distribution of T_{max} and T_{min} for Los Angeles and Seoul. All available samples were used to map the spatial distribution using the best scheme, since the larger the number of samples, the more robust the developed model is (Stisen et al., 2007).

RF measures relative variable importance using out-of-bag samples when a variable is randomly permuted (Breiman, 2001). For each tree, the mean squared error (MSE) of the out-of-bag portion of samples is recorded. Then, the same process is conducted when

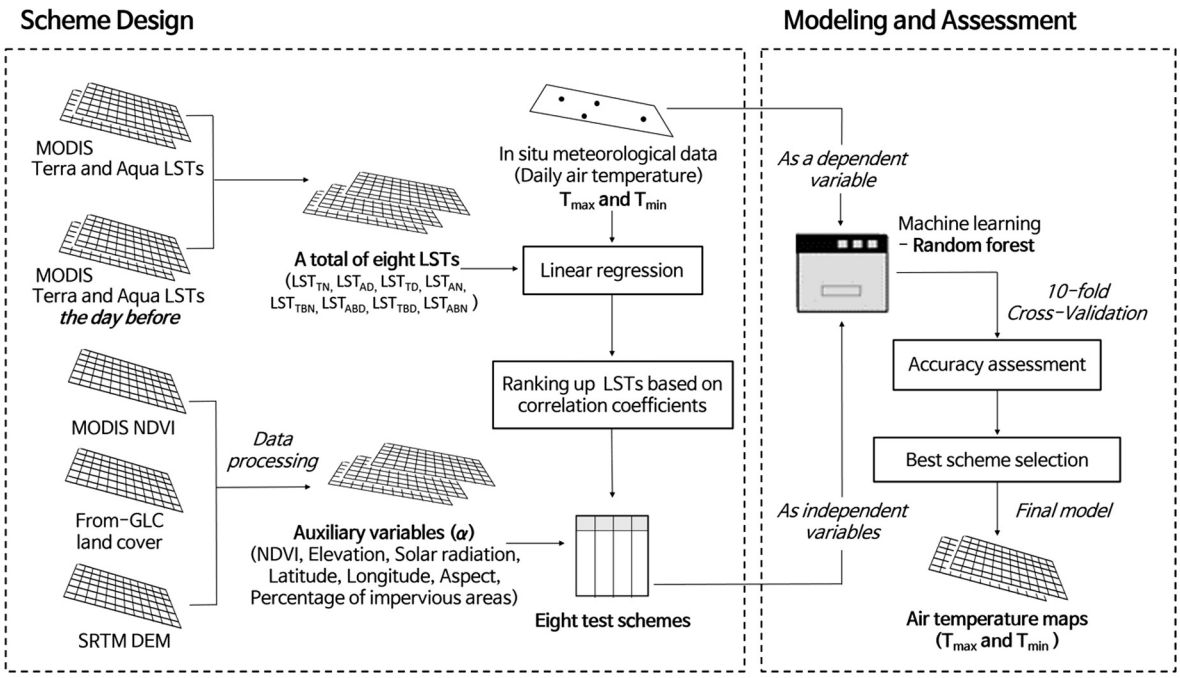

Fig. 2. Flowchart of this study. Procedures are divided into two main parts: scheme design (left) and modeling and assessment (right).

each predictor variable is perturbed. The differences between the two MSEs for all trees are averaged and normalized by the standard deviation of the differences (i.e., increased percentage of mean squared error (%incMSE)). The higher the %incMSE of a variable, the greater the contribution of the variable is. We averaged variable importance through ten iterations of the best schemes for T_{max} and T_{min} estimation in the two cities to ensure the reliability of the results.

4. Results and discussion

4.1. Comparison of air temperatures and LSTs

Fig. 3 shows the variations of MODIS-derived LSTs and T_{max} and T_{min} measured at weather stations for the two cities. Los Angeles has relatively high LSTs in the daytime (Fig. 3b, c), compared to Seoul (Fig. 3h, i). This may be due to more incoming solar radiation and less vegetation cover in dry Los Angeles than Seoul (Table 2). Generally, NDVI and daytime LST show an inverse relationship in the summer. In humid Seoul, wet soils in summer have higher heat capacity than the dryer soils of Los Angeles due to the proportional relationship between soil moisture and heat capacity (Oke, 2002). Therefore, the daytime LSTs of Los Angeles could increase more in a short time than those of Seoul. In the daytime, when sunlight reaches the surface, LSTs (LST_{TD}, LST_{AD}) show a large temperature variation between the stations, compared to the nighttime LSTs (LST_{TN}, LST_{AN}) for both cities. Daytime LSTs were shown to be more unstable than nighttime LSTs due to incoming solar energy (Yang et al., 2017; Zeng et al., 2015). It should be noted that the difference between daytime LSTs and nighttime LSTs is much higher in Los Angeles than Seoul. This suggests a surface heat-trapping phenomenon in Seoul, which may result from the higher heat capacity of the relatively humid soil (Price, 1980).

In Los Angeles, T_{max} is significantly lower than the daytime LSTs. In Seoul, however, T_{max} is similar to the daytime LSTs. One possible reason may be the different summer climate types between the two cities; with cloudy summers, LST and air temperature are closer in Seoul (Gallo et al., 2011; Good, 2016). At both cities, T_{min} is similar to LST_{AN}—which is the lowest daily LST variable—and the temperature variation among stations is insignificant. With no incoming solar radiance at night, the surface heat output exceeds input so the surface cools enough to avoid significant temperature variation among stations. By the same principle, temperature variation of LST_{AN} is more moderate than that of LST_{TN}. A sufficiently cooled surface (i.e., LST_{AN}) shows little spatial variation among the stations at night.

The T_{max} variation among the stations (Fig. 3a) shows a tendency similar to the daytime LSTs (Fig. 3b, c), and the T_{min} variation (Fig. 3d) tends to be similar to that of nighttime LSTs (Fig. 3e, f) in Los Angeles. However, in Seoul, both T_{max} and T_{min} (Fig. 3g, j) have similar variations among the stations as the nighttime LSTs (Fig. 3k, l). Seoul is more densely populated than Los Angeles. More anthropogenic heat sources and air pollutants cause the air to be warm enough throughout the entire city (Shahmohamadi et al., 2011). These factors could influence the reduced variation of both T_{max} and T_{min} among the stations in Seoul compared to Los Angeles (Fig. 3a, d, g, j). These results played an important role in estimating air temperatures in this study.

The correlation between each LST and T_{max} and T_{min} for both cities was examined (Fig. 4). For T_{max} in Los Angeles, LST_{TN} shows the highest correlation ($r = 0.56$). The daytime LSTs, LST_{AD} and LST_{TD}, have a slightly lower r value with T_{max} (both $r = 0.53$). The T_{min} of Los Angeles shows a significantly higher correlation of LST_{AN} ($r = 0.7$), followed by the other nighttime LSTs, LST_{TN}, LST_{ABN}, and LST_{TBN}. The daytime LSTs showed much lower correlation than LST at night ($r = 0.1–0.15$).

In Seoul, T_{max} has the highest correlation with LST_{AN} ($r = 0.66$), followed by LST_{ABN} ($r = 0.58$). The r value between LST_{AD} and T_{max} is the lowest among the LSTs. The decrease in T_{max} variation in Seoul (Fig. 3g) seems to be the primary cause. In the case of T_{min} in Seoul, LST_{ABN} showed the highest correlation ($r = 0.75$), followed by LST_{AN} ($r = 0.66$). It is difficult to identify precisely why LST_{ABN} has a higher correlation with T_{min} than LST_{AN} has, but there could be a time-lag-effect at Seoul caused by surfaces with high thermal capacity retaining heat and continuing to warm the air for a long time. This result illustrates the significance of LST variables of

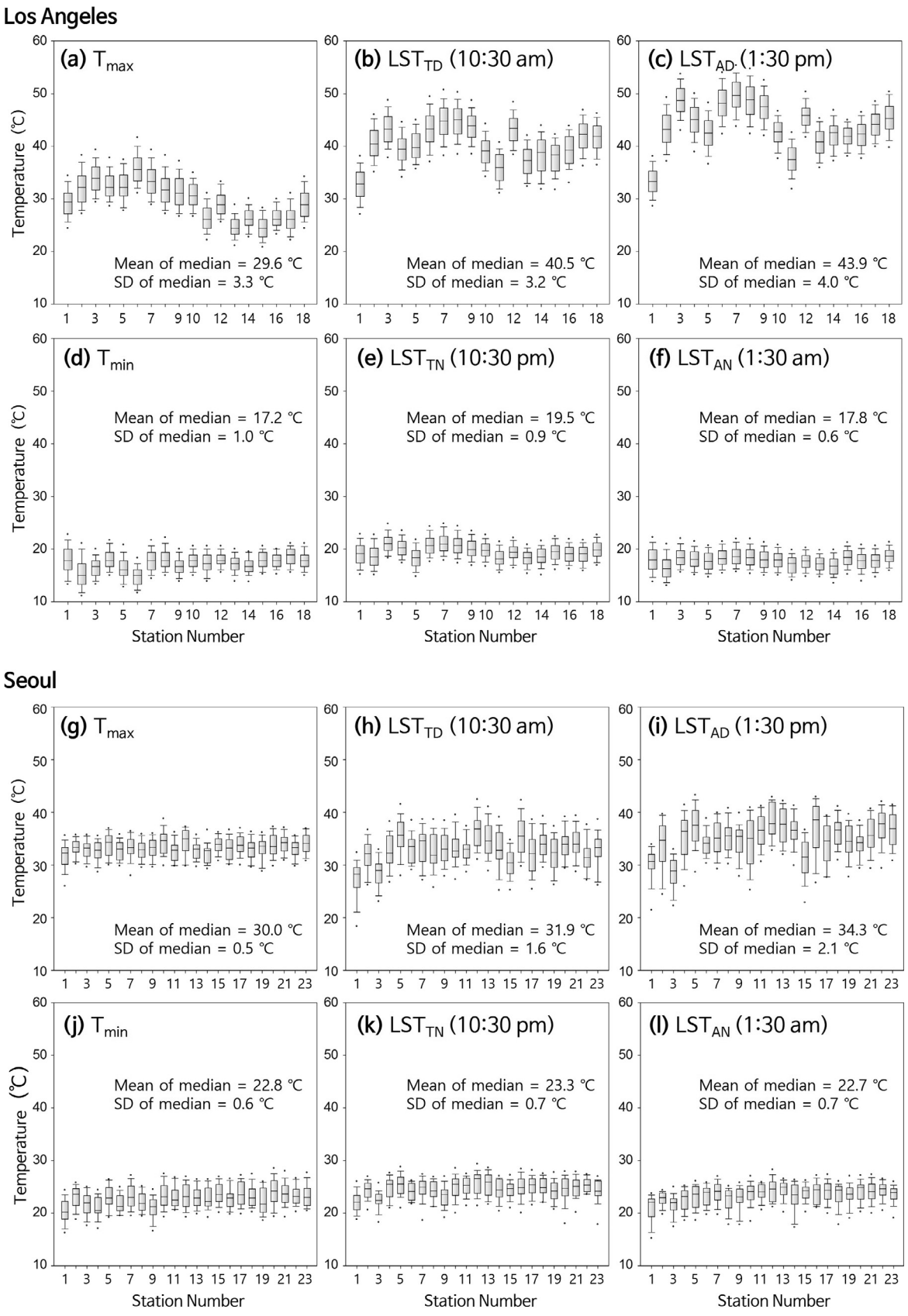

Fig. 3. Boxplots of four LSTs (LST_{TD}, LST_{TN}, LST_{AD}, LST_{AN}) and observed T_{max} and T_{min} for Los Angeles and Seoul for each station. Refer to Fig. 1 for station numbers (18 stations in Los Angeles and 23 stations in Seoul).

Table 2
Pixel values of the mean (standard deviation) calculated from the NDVI and daily incoming solar radiation (Sol) averaged in study periods (July-August 2006–2016) within city boundary.

City	NDVI/built[a]	NDVI/veg[b]	Sol (wh/m^2)
Los Angeles	0.29 (0.10)	0.44 (0.11)	5541 (94)
Seoul	0.43 (0.13)	0.65 (0.14)	5325 (57)

[a,b] 'Built' and 'veg' indicate the built-up and vegetated areas classified within the city boundary.

the previous day, which have not been used for estimating air temperature in previous studies. In addition, the specific time—1:30 am for LST_{AN} and LST_{ABN}—shows more correlation with both T_{max} and T_{min} in Seoul than LST for the other times, which suggests it can be applied to future studies, such as forecasting short-term temperature using satellite images.

4.2. Scheme performance

The eight schemes (S1 to S8) that use correlation analysis to estimate T_{max} and T_{min} at two different cities are listed in Table 3. Due to cloud-contaminated pixels, the sample size decreases as more LST variables are added. When all eight LSTs are used (S8), the number of samples is 3123 in Los Angeles and only 163 in Seoul, which is particularly impacted by the high cloud cover rate that is common during the humid summer. In order to appropriately compare the results, the same samples corresponding to S8 were used for all schemes.

Table 4 summarizes the 10-fold cross-validation results. In Los Angeles, S5 shows the highest performance for estimating T_{max}, and S8 shows the highest performance for T_{min}. Comparing S1 (using only one LST) with the best performing schemes, S5 yielded a higher increase in performance for estimating T_{max} (ΔR^2 = +0.147 and ΔRMSE = −0.63 °C) than S8 for T_{min} (ΔR^2 = +0.05 and ΔRMSE = −0.16 °C). The higher improvement for estimating T_{max} seems to relate to the much larger dynamic range of T_{max} than T_{min}. In Seoul, S3 produced the best results for estimating both T_{max} and T_{min}. The use of multiple LSTs improved T_{max} and T_{min} estimation. However, there is a limit to improving performance by using more LSTs. Model performance reached a peaked then decreased with additional LSTs in the case of estimating T_{max} in both Los Angeles (S6 to S8) and Seoul (S4 to S8), and T_{min} in Seoul (S4 to S8). The added LSTs that decreased the model accuracy might have redundant or contradictory information.

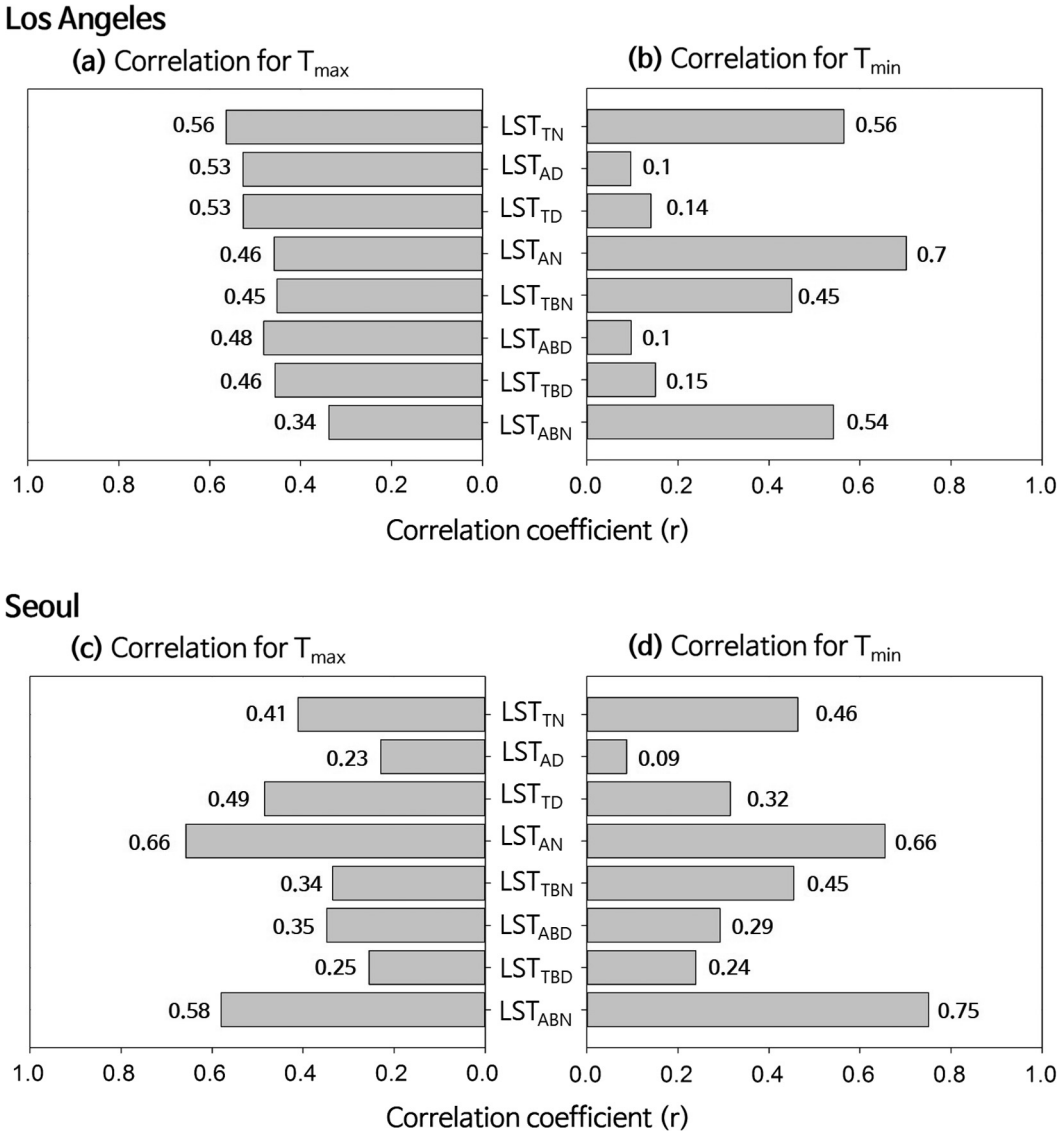

Fig. 4. Correlation coefficients (r values) between the eight LSTs and T_{max} and T_{min} in (a) and (b) Los Angeles and (c) and (d) Seoul.

Table 3
List of eight schemes (S1-S8) with sample size for estimating T_{max} and T_{min} of (a)–(b) Los Angeles and (c)–(d) Seoul. The first scheme (S1) has the LST that had the highest correlation with *in situ* measurements; the second scheme (S2) uses the two LSTs that yielded the first and second highest correlation with *in situ* data, with subsequent schemes adding LSTs in increasing correlation order to the last scheme (S8), which uses all eight LSTs. The seven auxiliary variables (α) were added for each scheme.

Scheme	Input variables	Sample size (N)
(a) T_{max} in Los Angeles		
S1	LST_{TN} + α	7655
S2	LST_{TN} + LST_{AD} + α	6542
S3	LST_{TN} + LST_{AD} + LST_{TD} + α	6004
S4	LST_{TN} + LST_{AD} + LST_{TD} + LST_{ABD} + α	5405
S5	LST_{TN} + LST_{AD} + LST_{TD} + LST_{ABD} + LST_{AN} + α	4250
S6	LST_{TN} + LST_{AD} + LST_{TD} + LST_{ABD} + LST_{AN} + LST_{TBD} + α	3968
S7	LST_{TN} + LST_{AD} + LST_{TD} + LST_{ABD} + LST_{AN} + LST_{TBD} + LST_{TBN} + α	3646
S8	LST_{TN} + LST_{AD} + LST_{TD} + LST_{ABD} + LST_{AN} + LST_{TBD} + LST_{TBN} + LST_{ABN} + α	3123
(b) T_{min} in Los Angeles		
S1	LST_{AN} + α	6087
S2	LST_{AN} + LST_{TN} + α	5699
S3	LST_{AN} + LST_{TN} + LST_{ABN} + α	4373
S4	LST_{AN} + LST_{TN} + LST_{ABN} + LST_{TBN} + α	4159
S5	LST_{AN} + LST_{TN} + LST_{ABN} + LST_{TBN} + LST_{TBD} + α	3815
S6	LST_{AN} + LST_{TN} + LST_{ABN} + LST_{TBN} + LST_{TBD} + LST_{TD} + α	3436
S7	LST_{AN} + LST_{TN} + LST_{ABN} + LST_{TBN} + LST_{TBD} + LST_{TD} + LST_{ABD} + α	3294
S8	LST_{AN} + LST_{TN} + LST_{ABN} + LST_{TBN} + LST_{TBD} + LST_{TD} + LST_{ABD} + LST_{AD} + α	3123
(c) T_{max} Seoul		
S1	LST_{AN} + α	3545
S2	LST_{AN} + LST_{ABN} + α	1427
S3	LST_{AN} + LST_{ABN} + LST_{TD} + α	717
S4	LST_{AN} + LST_{ABN} + LST_{TD} + LST_{TN} + α	649
S5	LST_{AN} + LST_{ABN} + LST_{TD} + LST_{TN} + LST_{ABD} + α	335
S6	LST_{AN} + LST_{ABN} + LST_{TD} + LST_{TN} + LST_{ABD} + LST_{TBN} + α	302
S7	LST_{AN} + LST_{ABN} + LST_{TD} + LST_{TN} + LST_{ABD} + LST_{TBN} + LST_{TBD} + α	216
S8	LST_{AN} + LST_{ABN} + LST_{TD} + LST_{TN} + LST_{ABD} + LST_{TBN} + LST_{TBD} + LST_{AD} + α	163
(d) T_{min} in Seoul		
S1	LST_{ABN} + α	3476
S2	LST_{ABN} + LST_{AN} + α	1427
S3	LST_{ABN} + LST_{AN} + LST_{TN} + α	1193
S4	LST_{ABN} + LST_{AN} + LST_{TN} + LST_{TBN} + α	950
S5	LST_{ABN} + LST_{AN} + LST_{TN} + LST_{TBN} + LST_{TD} + α	533
S6	LST_{ABN} + LST_{AN} + LST_{TN} + LST_{TBN} + LST_{TD} + LST_{ABD} + α	302
S7	LST_{ABN} + LST_{AN} + LST_{TN} + LST_{TBN} + LST_{TD} + LST_{ABD} + LST_{TBD} + α	216
S8	LST_{ABN} + LST_{AN} + LST_{TN} + LST_{TBN} + LST_{TD} + LST_{ABD} + LST_{TBD} + LST_{AD} + α	163

'α' includes NDVI, elevation, solar radiation, latitude, longitude, aspect, impervious area (%).

Table 4
Ten (10)-fold cross-validation results (R^2 and RMSE) for eight schemes for two cities. The same samples (corresponding to S8) were used: 3123 for Los Angeles and 163 for Seoul. Best models are highlighted in bold.

	T_{max} in Los Angeles (N = 3123)		T_{min} in Los Angeles (N = 3123)	
Scheme	R^2	RMSE (°C)	R^2	RMSE (°C)
S1	0.695	2.32	0.727	1.35
S2	0.743	2.13	0.761	1.23
S3	0.825	1.77	0.765	1.22
S4	0.829	1.75	0.765	1.22
S5	**0.842**	**1.69**	0.767	1.22
S6	0.839	1.70	0.774	1.20
S7	0.837	1.71	0.775	1.20
S8	0.837	1.71	**0.777**	**1.19**
	T_{max} in Seoul (N = 163)		T_{min} in Seoul (N = 163)	
Scheme	R^2	RMSE (°C)	R^2	RMSE (°C)
S1	0.615	1.12	0.690	1.21
S2	0.642	1.07	0.734	1.12
S3	**0.679**	**1.03**	**0.737**	**1.12**
S4	0.663	1.05	0.736	1.12
S5	0.647	1.07	0.723	1.14
S6	0.659	1.05	0.725	1.15
S7	0.651	1.06	0.728	1.15
S8	0.646	1.07	0.730	1.14

Based on evaluation of the different schemes, we selected S5 and S8 to further investigate the spatial distribution of T_{max} and T_{min}, respectively, for Los Angeles, and S3 for mapping both T_{max} and T_{min} in Seoul. A final RF model was run using all samples. In Los Angeles, R^2 values for estimating T_{max} (Fig. 5a) and T_{min} (Fig. 5b) are 0.850 and 0.777, respectively. In Seoul, R^2 for estimating T_{max} (Fig. 5c) and T_{min} (Fig. 5d) are 0.728 and 0.767, respectively. T_{max} and T_{min} estimation of Los Angeles have 1.7 °C and 1.2 °C cross-validation RMSE, respectively. In Seoul, the cross-validation RMSE values were 1.1 °C and 1.2 °C for T_{max} and T_{min} estimation, respectively. Among the four models, the model estimating T_{max} at Los Angeles resulted in the highest correlation with *in situ* data, but RMSE was also the highest. However, with a lower nRMSE (5.9%) than those of the other air temperature estimations (7–8%), this large RMSE was attributed to the much wider T_{max} range in Los Angeles—i.e., from 19 °C to 48 °C—than the other cases. Nevertheless, all four best performing air temperature models produced R^2 over 0.7 and RMSE values less than 1.7 °C (nRMSE < 8.4%). These performances are comparable with or better than those from the literature (RMSE of 2.31 °C for Vancouver (Ho et al., 2014), RMSE of 1.5–1.8 °C for Hamburg (Bechtel et al., 2014), RMSE of 1–2 °C for Athens (Agathangelidis et al., 2016) and RMSE of 2.3 °C for 15 cities in Europe and North Africa (Keramitsoglou et al., 2016).

LSTs exhibit relatively high variable importance scores in the final RF T_{max} and T_{min} models (i.e., the best performing schemes; Fig. 6). However, it is better to interpret the variable importance together with the multicollinearity between the LSTs which could influence the process of calculating the variable importance (Murray and Conner, 2009). The extent of collinearity among LSTs can be measured using the variance inflation factor (VIF) (Marquaridt, 1970), which quantifies the rate of inflation in the variance of regression coefficients calculated by ordinary least square regression between each variable and the others.

In Los Angeles, the most correlated LST with T_{max} is LST_{TN} (Fig. 4a). The variable importance of LST_{AN}, however, is higher than those of the other four LSTs (Fig. 6a). This might be due to low VIF (2.71) of LST_{AN} (Table 5a). For T_{min} estimation in Los Angeles, the LST_{AN}, which has the highest correlation value with T_{min} (Fig. 4b), and the lowest VIF among LSTs has a distinctively high variable importance (Fig. 6b). In Seoul, the VIF of LSTs are relatively lower than that of Los Angeles, and the order of correlation of each LST (Fig. 4c, d) is consistent with that of the importance scores (Fig. 6c, d).

Among the seven auxiliary variables, it should be noted that solar radiation was identified as a contributing variable in estimating not only T_{max} but also T_{min} for both Los Angeles and Seoul. Since daily solar radiation varies both spatially and temporally (Burgess, 2009), solar radiation might have higher importance than the other auxiliary variables, which were assumed to be static over time. However, variable importance is influenced by many factors, such as model structure (Knudby et al., 2010; Strobl et al., 2007), so interpreting the variable importance demands more attention.

4.3. Analysis of spatial distribution of air temperatures

Average T_{max} and T_{min} maps for July and August 2006 to 2016 were produced for the two cities. Figs. 7 and 8 show air

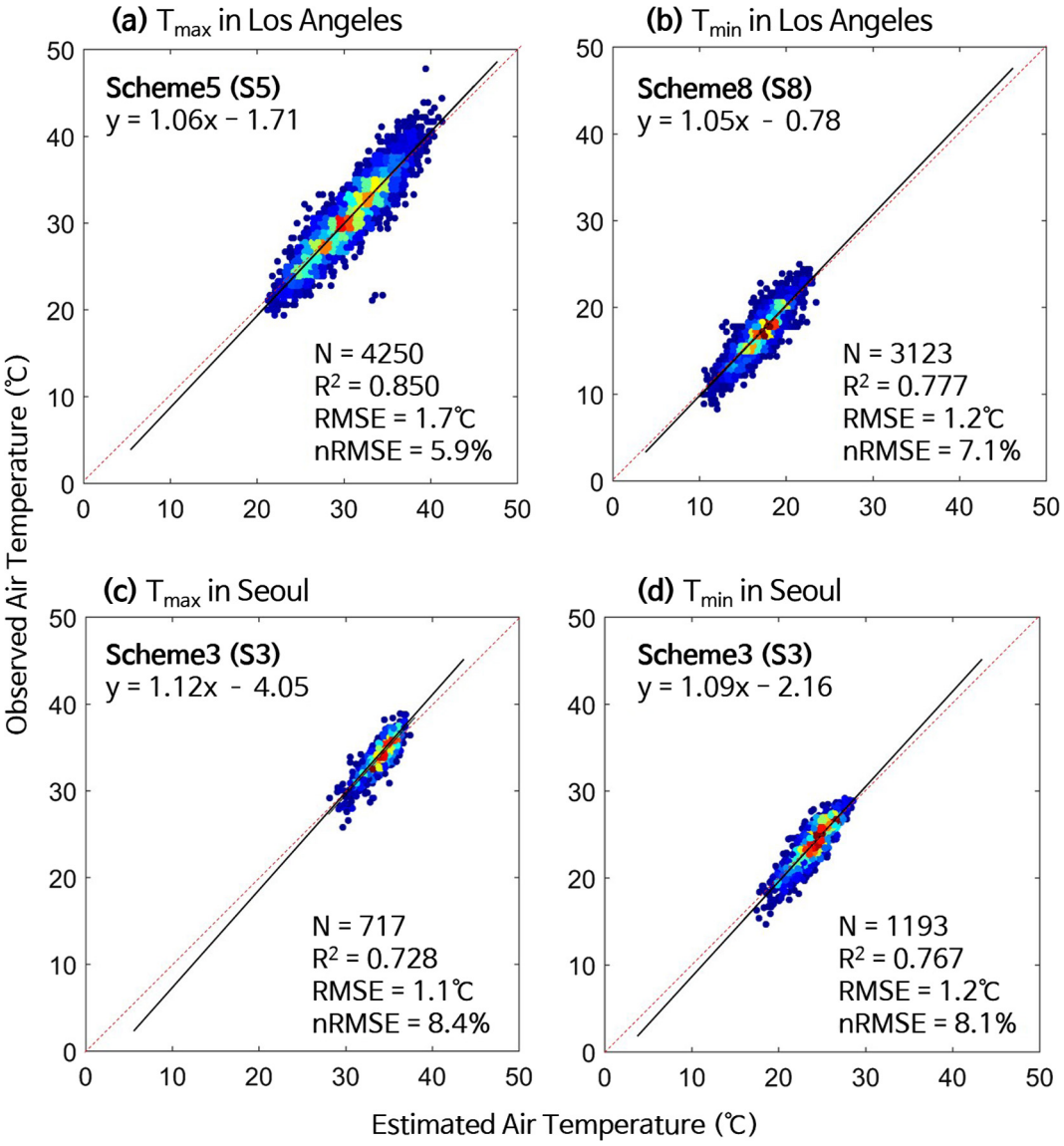

Fig. 5. Density scatter plots between estimated and observed air temperatures from 10-fold cross-validation results based on the best schemes. The color ramp from blue to red corresponds to increasing point density. Black solid lines show the regression line and red dashed lines represent the line of identity (y = x); nRMSE is the normalized root mean square error. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

temperature maps averaged across days with more than 90% non-cloud-contaminated pixels within the city boundaries. The scatter plots show high correlation, which suggests well-simulated spatial distribution of the air temperature in the selected maps. The spatial distributions of T_{max} and T_{min} in Los Angeles are noticeably different. In the case of the T_{max} (Fig. 7a, b), the coastal areas are cooled by a strong sea breeze in the daytime (Simpson, 1994). This causes those regions to have relatively low T_{max} (including stations 13, 15, and 11). However, regions above 34°N latitude—the San Fernando Valley (including stations 6, 3, and 7)—have very high T_{max} above 30 °C. One possible reason is the geographic characteristics of the San Fernando Valley, bounded by the Santa Susana Mountains to the north, the Santa Monica Mountains to the south, and the Simi Hills to the west. The hills and mountains could block the daytime sea breezes from the southwest. Thus, more heat is trapped within the San Fernando Valley—the average observed T_{max} is 34.3 °C at stations 3, 6, and 7); it is known to be even hotter than downtown Los Angeles—the observed T_{max} is 29.1 °C at station 12—in the daytime.

Considering the T_{min} distribution in Los Angeles (Fig. 7c, d), the difference in temperatures between the built-up areas and other regions (e.g., vegetated areas) appears quite clearly. At night, the cold wind coming from the mountain (Smith, 1979) causes the surrounding areas to have much lower T_{min}. From the mountain area to the city center, the T_{min} increases (stations 6, 13, 11, and 3), reaching much higher T_{min} at the center of city (stations 7 and 12), like downtown Los Angeles—except for station 15, located at the international airport near the coast. Due to high energy use at night and the large impervious area at the airport (Edwards, 2005), the region might shows higher T_{min} than the surrounding areas. In summary, for Los Angeles, the cooling effect of the sea breeze is weakened, and more heat is concentrated within the city at night.

In the case of Seoul, built-up areas and surrounding mountains show apparent temperature differences both in T_{max} and T_{min} (Fig. 8a, c). However, the spatial pattern of temperatures within the city is different between the T_{max} and T_{min}. The eastern part of the urban area (~127°E longitude) has a relatively higher T_{max}

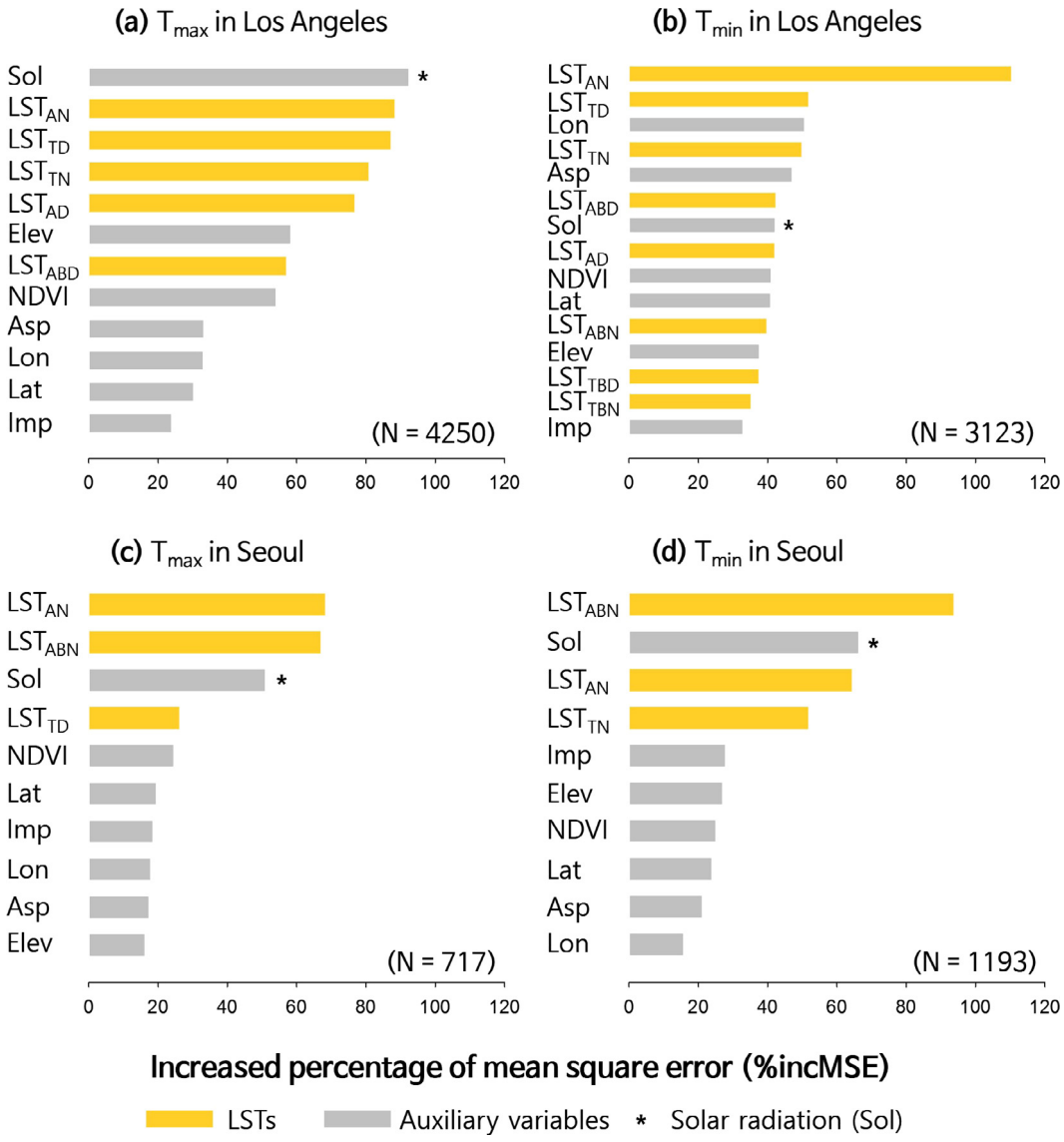

Fig. 6. Relative variable importance (i.e., increase in percentage of mean square error (%incMSE) using out-of-bag samples when a variable was permuted) calculated from the final RF models (i.e., best performing schemes).

Table 5
Variance inflation factors (VIF) among eight LSTs in two cities. Higher VIF of a variable means it has higher multi-collinearity with the other variables in estimating the target variable.

Variance inflation factors (VIF)		
	(a) Los Angeles	(b) Seoul
LST_{TN}	3.62	1.99
LST_{AD}	5.24	1.86
LST_{TD}	4.28	2.49
LST_{AN}	2.71	2.04
LST_{TBN}	3.40	2.01
LST_{ABD}	5.20	2.13
LST_{TBD}	4.50	1.95
LST_{ABN}	2.97	2.83

than the western area (Fig. 8a). The sea breeze blowing from the West Sea of the Korean Peninsula is strong in the daytime, and the western part of Seoul is cooled by the breeze. But the influence of this breeze decreases as it enters the eastern inland (Pokhrel and Lee, 2011). In Seoul, most of the built-up areas have higher T_{min} compared to the surrounding areas (stations 9, 1, 3, and 4) (Fig. 8c, d). This is due to the impervious surfaces holding heat longer in the afternoon compared to other areas, such as mountains; the sea breeze effect disappears at night.

It should be noted that the difference in temperature between built-up and vegetated areas is more apparent in the T_{min} rather than the T_{max} in both cities (Fig. 7b, d, 8b, d). This is consistent with previous studies, showing that urbanization is more likely to affect the T_{min} than the T_{max}, along with studies showing that the urban heat island phenomenon can be identified more at night than during the day in a city (Kalnay and Cai, 2003; Hua et al., 2008; Hamdi and Vyver, 2011). Considering that the T_{max} distribution of the two cities is greatly influenced by both sea breezes and cooling mountain winds at night, we suggest the use of wind-related variables for future urban temperature estimation studies.

4.4. Novelty and limitation

Most existing studies estimating air temperature by combining the MODIS multi-time series LSTs use only four (two daytime and

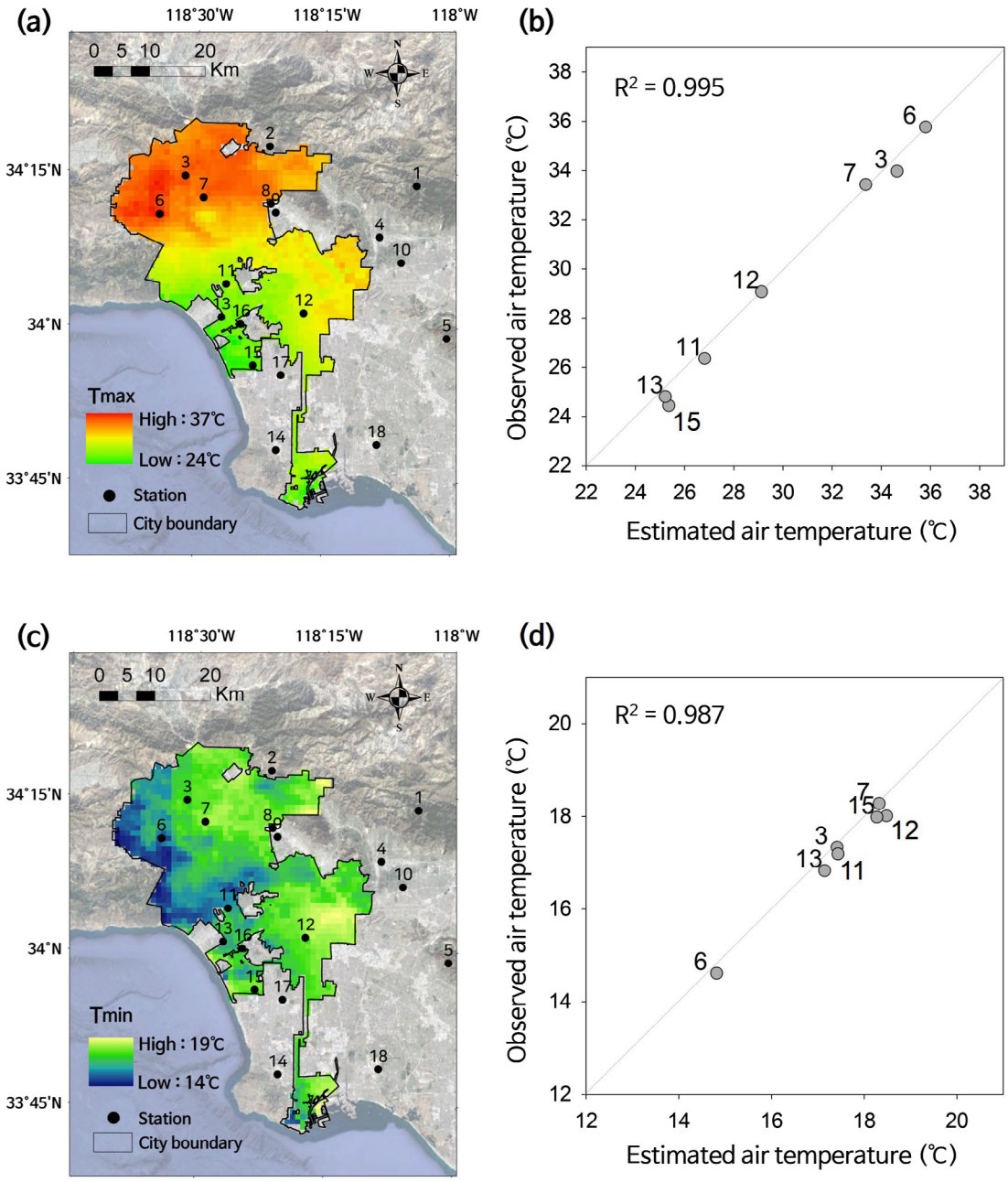

Fig. 7. (a) Map of spatial distribution of average T_{max} and (b) the corresponding scatterplot between the average observed air temperature and average estimated air temperature in Los Angeles. (c) and (d) for T_{min}.

two nighttime) LSTs collected on the target date (Noi et al., 2017; Zhang et al., 2016). This study took four LSTs of the target day and four LSTs from the prior day, using a total of eight MODIS LSTs for daily air temperature estimation. The LSTs of the previous day were important variables for both T_{max} and T_{min} estimation in urban landscapes with different environments. In particular, LST_{ABN} was identified as the most important and second most important variable among eight LSTs for T_{max} and T_{min} estimations, respectively, in Seoul. Therefore, we expect LSTs of the previous day (especially night time LSTs) will be used in future urban air temperature studies. In addition, unlike many of the previous studies that estimated urban air temperature focusing on a single urban study area (Ho et al., 2014; Bechtel et al., 2014; Agathangelidis et al., 2016) or continental scale studies with one station for each city (Keramitsoglou et al., 2016), we investigated two mega cities with quite different climatic and environmental characteristics (i.e., dry Los Angeles and humid Seoul) on a local scale. This suggests that the proposed methodology can be successfully applied to a wide range of urban landscapes with limited *in situ* air temperature measurements.

The research findings from this study can be applied to forecast future air temperature using only LSTs collected the day before. RF models using the LSTs of the day before (LST_{ABN}, LST_{TBD}, LST_{ABD} and LST_{TBN}) with auxiliary variables were tested for both cities (Table 6). Although the overall performance is slightly lower than the previous eight schemes, it is very promising to forecast daily T_{max} and T_{min} using satellite-derived LSTs. In particular, in the case of the T_{min} in Seoul, the performance is very good ($R^2 = 0.767$, RMSE = 1.19 °**C**), which can be used to forecast tropical nighttime lows.

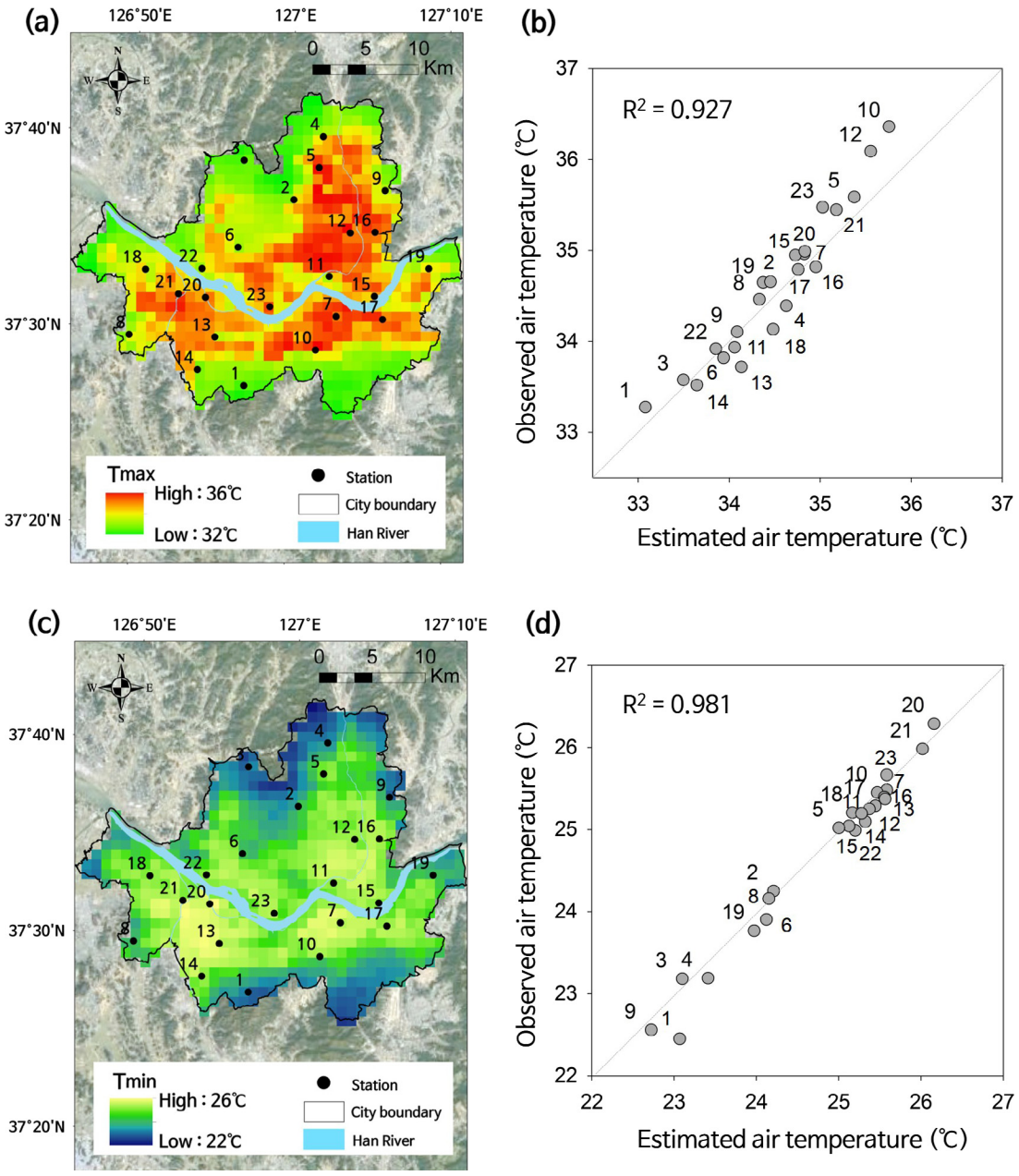

Fig. 8. (a) Map of spatial distribution of average T_{max} and (b) the corresponding scatterplot between the average observed air temperature and average estimated air temperature in Seoul. (c) and (d) for T_{min}.

Table 6
The 10-fold cross-validation results of the models using just LSTs collected the day before with auxiliary variables (LST_{ABN} + LST_{TBD} + LST_{ABD} + LST_{TBN} + α) for forecasting T_{max} and T_{min}. The numbers of samples are 4595 for Los Angeles and 635 for Seoul.

	Air temperature	R^2	RMSE (°C)	nRMSE (%)
Los Angeles	T_{max}	0.691	2.43	8.55
	T_{min}	0.567	1.64	8.89
Seoul	T_{max}	0.649	1.40	11.90
	T_{min}	0.767	1.19	8.01

This study has some limitations. When various LSTs are combined, there are many no-data regions due to cloud contamination, especially in humid areas such as Seoul. This is a really challenging problem when using satellite-derived LSTs for forecasting T_{max} and T_{min}. One solution could be using multi-temporal solar radiations as surrogate variables for contaminated LSTs (Bechtel et al., 2017). Although Los Angeles and Seoul have different climate characteristics, they are located in coastal areas in mid-latitude, which implies that it is necessary to investigate other areas which have completely different geographic characteristics. In addition, there are slight variations of local solar time by pixel and collection date when MODIS data are collected (Duan et al., 2014). Thus, the

normalization of LST_{AN}, LST_{TD}, LST_{AD} and LST_{TN} datasets for consistent local solar time might improve the performance of the proposed models. Because this study was focused on using satellite-based products, other variables such as wind velocity and relative humidity should be examined in future research.

5. Conclusion

In this study, we estimated T_{max} and T_{min} in two mega cities, Los Angeles and Seoul, with different climate and environmental characteristics. Eight MODIS Terra and Aqua LSTs (4 times a day and 4 times the day before), and seven auxiliary variables were used for temperature estimation based on random forest machine learning. Eight schemes with various combinations of LSTs were evaluated and the best performing schemes were used to examine the spatial distribution of T_{max} and T_{min} for Los Angeles and Seoul. Though urban areas have complex landscapes, this study produced very promising results with R^2 higher than 0.7 and RMSE less than 1.7 °C for estimating both T_{max} and T_{min} in the cities. The spatial distribution of T_{max} and T_{min} for Los Angeles and Seoul showed distinct patterns caused by various factors such as sea breeze, surrounding mountains, and urban density.

This study revealed that there were specific LSTs that showed higher correlation with each T_{max} and T_{min} in urban landscapes with different environments. We also found that using LSTs collected the day before, which were not considered in previous studies, was crucial for urban air temperature estimation. It should be also noted that the use of more LSTs does not guarantee better performance in estimating air temperature. Several valuable findings from this study deserve a further study focusing on forecasting (~24 h in advance) daily maximum (heat waves) and minimum (tropical nighttime lows) air temperatures.

Acknowledgements

This research was supported by the Space Technology Development Program and the Basic Science Research Program through the National Foundation of Korea (NRF) funded by the Ministry of Science, ICT, & Future Planning and the Ministry of Education of Korea, respectively (Grant: NRF-2017M1A3A3A02015981; NRF-2017R1D1A1B03028129), and the Korea Meteorological Administration Research and Development Program under Grant KMIPA 2017-7010.

References

Agathangelidis, I., Cartalis, C., Santamouris, M., 2016. Estimation of air temperatures for the urban agglomeration of Athens with the use of satellite data. Geoinform. Geostat.: Overview 4, 2. https://doi.org/10.4172/2327 4581.

Amani, M., Salehi, B., Mahdavi, S., Granger, J., Brisco, B., 2017. Wetland classification in Newfoundland and Labrador using multi-source SAR and optical data integration. GISci. Remote Sens. 54, 779–796.

Arnold Jr., C.L., Gibbons, C.J., 1996. Impervious surface coverage: the emergence of a key environmental indicator. J. Am. Plann. Assoc. 62, 243–258.

Aw, J., Kleeman, M.J., 2003. Evaluating the first-order effect of intraannual temperature variability on urban air pollution. J. Geophys. Res.: Atmos. 108.

Basu, R., Samet, J.M., 2002. Relation between elevated ambient temperature and mortality: a review of the epidemiologic evidence. Epidemiol. Rev. 24, 190–202.

Bechtel, B., Wiesner, S., Zakšek, K., 2014. Estimation of dense time series of urban air temperatures from multitemporal geostationary satellite data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7, 4129–4137.

Bechtel, B., Zakšek, K., Oßenbrügge, J., Kaveckis, G., Böhner, J., 2017. Towards a satellite based monitoring of urban air temperatures. Sustain. Cities Soc.

Beers, T.W., Dress, P.E., Wensel, L.C., 1966. Notes and observations: aspect transformation in site productivity research. J. Forest. 64, 691–692.

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: a review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Burgess, P., 2009. Variation in Light Intensity at Different Latitudes and Seasons, Effects of Cloud Cover, and the Amounts of Direct and Diffused Light. Continuous Cover Forestry Group, Forres, UK (accessed May 2017) http://www.ccfg.org.uk/conferences/downloads/P_Burgess.pdf.

Chen, F., Liu, Y., Liu, Q., Qin, F., 2015. A statistical method based on remote sensing for the estimation of air temperature in China. Int. J. Climatol. 35, 2131–2143.

Courault, D., Monestiez, P., 1999. Spatial interpolation of air temperature according to atmospheric circulation patterns in southeast France. Int. J. Climatol. 19, 365–378.

Duan, S.-B., Li, Z.-L., Tang, B.-H., Wu, H., Tang, R., 2014. Generation of a time-consistent land surface temperature product from MODIS data. Remote Sens. Environ. 140, 339–349.

Edwards, B., 2005. The Modern Airport Terminal: New Approaches to Airport Architecture. Spon Press, New York.

Gabriel, K.M., Endlicher, W.R., 2011. Urban and rural mortality rates during heat waves in Berlin and Brandenburg, Germany. Environ. Pollut. 159, 2044–2050.

Gallo, K., Hale, R., Tarpley, D., Yu, Y., 2011. Evaluation of the relationship between air and land surface temperature under clear-and cloudy-sky conditions. J. Appl. Meteorol. Climatol. 50, 767–775.

Good, E.J., 2016. An in situ-based analysis of the relationship between land surface "skin" and screen-level air temperatures. J. Geophys. Res.: Atmos. 121, 8801–8819.

Grimmond, C., Blackett, M., Best, M., Barlow, J., Baik, J., Belcher, S., Bohnenstengel, S., Calmet, I., Chen, F., Dandou, A., 2010. The international urban energy balance models comparison project: first results from phase 1. J. Appl. Meteorol. Climatol. 49, 1268–1292.

Hamdi, R., Vyver, H., 2011. Estimating urban heat island effects on near-surface air temperature records of Uccle (Brussels, Belgium): an observational and modeling study. Adv. Sci. Res. 6, 27–34.

Hao, P., Niu, Z., Zhan, Y., Wu, Y., Wang, L., Liu, Y., 2016. Spatiotemporal changes of urban impervious surface area and land surface temperature in Beijing from 1990 to 2014. GISci. Remote Sens. 53, 63–84.

Harman, I.N., 2003. The Energy Balance of Urban Areas. University of Reading.

Hart, M.A., Sailor, D.J., 2009. Quantifying the influence of land-use and surface characteristics on spatial variability in the urban heat island. Theor. Appl. Climatol. 95, 397–406.

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 Sens. Environ. 154, 38–45.

Hondula, D.M., Davis, R.E., Leisten, M.J., Saha, M.V., Veazey, L.M., Wegner, C.R., 2012. Fine-scale spatial variability of heat-related mortality in Philadelphia County, USA, from 1983–2008: a case-series analysis. Environ. Health 11, 16.

Hua, L., Ma, Z., Guo, W., 2008. The impact of urbanization on air temperature across China. Theor. Appl. Climatol. 93, 179–194.

Im, J., Park, S., Rhee, J., Baik, J., Choi, M., 2016. Downscaling of AMSR-E soil moisture with MODIS products using machine learning approaches. Environ. Earth Sci. 75, 1120.

Janatian, N., Sadeghi, M., Sanaeinejad, S.H., Bakhshian, E., Farid, A., Hasheminia, S.M., Ghazanfari, S., 2017. A statistical framework for estimating air temperature using MODIS land surface temperature data. Int. J. Climatol. 37, 1181–1194.

Jin, M.S., Mullens, T., 2014. A study of the relations between soil moisture, soil temperatures and surface temperatures using ARM observations and offline CLM4 simulations. Climate 2, 279–295.

Kalnay, E., Cai, M., 2003. Impact of urbanization and land-use change on climate. Nature 423, 528–531.

Katsouyanni, K., Pantazopoulou, A., Touloumi, G., Tselepidaki, I., Moustris, K., Asimakopoulos, D., Poulopoulou, G., Trichopoulos, D., 1993. Evidence for interaction between air pollution and high temperature in the causation of excess mortality. Arch. Environ. Health: Int. J. 48, 235–242.

Keramitsoglou, I., Kiranoudis, C.T., Sismanidis, P., Zakšek, K., 2016. An online system for nowcasting satellite derived temperatures for urban areas. Remote Sens. 8, 306.

Kim, M., Im, J., Han, H., Kim, J., Lee, S., Shin, M., Kim, H.-C., 2015. Landfast sea ice monitoring using multisensor fusion in the Antarctic. GISci. Remote Sens. 52, 239–256.

Kim, Y.H., Im, J., Ha, H.K., Choi, J.-K., Ha, S., 2014. Machine learning approaches to coastal water quality monitoring using GOCI satellite data. GISci. Remote Sens. 51, 158–174.

Knudby, A., Brenning, A., LeDrew, E., 2010. New approaches to modelling fish–habitat relationships. Ecol. Model. 221, 503–511.

Krüger, E., Emmanuel, R., 2013. Accounting for atmospheric stability conditions in urban heat island studies: the case of Glasgow, UK. Landscape Urban Plann. 117, 112–121.

Lee, S., Han, H., Im, J., Jang, E., Lee, M.-I., 2017. Detection of deterministic and probabilistic convection initiation using Himawari-8 Advanced Himawari Imager data. Atmos. Meas. Tech. 10, 1859.

Li, M., Im, J., Beier, C., 2013a. Machine learning approaches for forest classification and change analysis using multi-temporal Landsat TM images over Huntington Wildlife Forest. GISci. Remote Sens. 50, 361–384.

Li, Q., Zhang, H., Liu, X., Huang, J., 2004. Urban heat island effect on annual mean temperature during the last 50 years in China. Theor. Appl. Climatol. 79, 165–174.

Li, Z.-L., Tang, B.-H., Wu, H., Ren, H., Yan, G., Wan, Z., Trigo, I.F., Sobrino, J.A., 2013b. Satellite-derived land surface temperature: current status and perspectives. Remote Sens. Environ. 131, 14–37.

Liu, L., Breitner, S., Pan, X., Franck, U., Leitte, A.M., Wiedensohler, A., von Klot, S., Wichmann, H.-E., Peters, A., Schneider, A., 2011. Associations between air temperature and cardio-respiratory mortality in the urban area of Beijing, China: a time-series analysis. Environ. Health 10, 51.

Lu, Z., Im, J., Rhee, J., Hodgson, M., 2014. Building type classification using spatial and landscape attributes derived from LiDAR remote sensing data. Landscape Urban Plann. 130, 134–148.

Marquaridt, D.W., 1970. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. Technometrics 12, 591–612.
McCutchan, M.H., Fox, D.G., 1986. Effect of elevation and aspect on wind, temperature and humidity. J. Climate Appl. Meteorol. 25, 1996–2013.
Moser, G., De Martino, M., Serpico, S.B., 2015. Estimation of air surface temperature from remote sensing images and pixelwise modeling of the estimation uncertainty through support vector machines. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8, 332–349.
Murray, K., Conner, M.M., 2009. Methods to quantify variable importance: implications for the analysis of noisy ecological data. Ecology 90, 348–355.
Naidoo, L., Cho, M.A., Mathieu, R., Asner, G., 2012. Classification of savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a Random Forest data mining environment. ISPRS J. Photogramm. Remote Sens. 69, 167–179.
Nemani, R.R., Running, S.W., 1989. Estimation of regional surface resistance to evapotranspiration from NDVI and thermal-IR AVHRR data. J. Appl. Meteorol. 28, 276–284.
Noi, P.T., Degener, J., Kappas, M., 2017. Comparison of multiple linear regression, cubist regression, and random forest algorithms to estimate daily air surface temperature from dynamic combinations of MODIS LST data. Remote Sens. 9, 398.
Oke, T.R., 2002. Boundary Layer Climates. Routledge.
Park, S., Im, J., Jang, E., Rhee, J., 2016. Drought assessment and monitoring through blending of multi-sensor indices using machine learning approaches for different climate regions. Agric. For. Meteorol. 216, 157–169.
Piyoosh, A.K., Ghosh, S.K., 2017. Semi-automatic mapping of anthropogenic impervious surfaces in an urban/suburban area using Landsat 8 satellite data. GISci. Remote Sens. 54, 471–494.
Pokhrel, R., Lee, H., 2011. Estimation of the effective zone of sea/land breeze in a coastal area. Atmos. Pollut. Res. 2, 106–115.
Price, J.C., 1980. The potential of remotely sensed thermal infrared data to infer surface soil moisture and evaporation. Water Resour. Res. 16, 787–795.
Prihodko, L., Goward, S.N., 1997. Estimation of air temperature from remotely sensed surface observations. Remote Sens. Environ. 60, 335–346.
Rhee, J., Park, S., Lu, Z., 2014. Relationship between land cover patterns and surface temperature in urban areas. GISci. Remote Sens. 51, 521–536.
Richardson, H.J., Hill, D.J., Denesiuk, D.R., Fraser, L.H., 2017. A comparison of geographic datasets and field measurements to model soil carbon using random forests and stepwise regressions (British Columbia, Canada). GISci. Remote Sens. 54, 573–591.
Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS J. Photogramm. Remote Sens. 67, 93–104.
Romero-Lankao, P., Qin, H., Dickinson, K., 2012. Urban vulnerability to temperature-related hazards: a meta-analysis and meta-knowledge approach. Global Environ. Change 22, 670–683.
Sandholt, I., Rasmussen, K., Andersen, J., 2002. A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. Remote Sens. Environ. 79, 213–224.
Schuster, C., Burkart, K., Lakes, T., 2014. Heat mortality in Berlin-Spatial variability at the neighborhood scale. Urban Clim. 10, 134–147.
Shahmohamadi, P., Che-Ani, A., Maulud, K., Tawil, N., Abdullah, N., 2011. The impact of anthropogenic heat on formation of urban heat island and energy consumption balance. Urban Studies Research 2011.
Shi, L., Liu, P., Kloog, I., Lee, M., Kosheleva, A., Schwartz, J., 2016. Estimating daily air temperature across the Southeastern United States using high-resolution satellite data: a statistical modeling study. Environ. Res. 146, 51–58.
Simpson, J.E., 1994. Sea Breeze and Local Winds. Cambridge University Press.
Smith, R.B., 1979. The influence of mountains on the atmosphere. Adv. Geophys. 21, 87–230.
Song, Y., Wu, Changshan, 2018. Examining human heat stress with remote sensing technology. GISci. Remote Sens. 55, 19–37.
Sonobe, R., Yamaya, Y., Tani, H., Wang, X., Kobayashi, N., Mochizuki, K.-I., 2017. Assessing the suitability of data from Sentinel-1A and 2A for crop classification. GISci. Remote Sens. 54, 918–938.
Stisen, S., Sandholt, I., Nørgaard, A., Fensholt, R., Eklundh, L., 2007. Estimation of diurnal air temperature using MSG SEVIRI data in West Africa. Remote Sens. Environ. 110, 262–274.
Stoll, M.J., Brazel, A.J., 1992. Surface-air temperature relationships in the urban environment of Phoenix, Arizona. Phys. Geogr. 13, 160–179.
Strobl, C., Boulesteix, A.-L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinf. 8, 25.
Sun, Y.-J., Wang, J.-F., Zhang, R.-H., Gillies, R., Xue, Y., Bo, Y.-C., 2005. Air temperature retrieval from remote sensing data based on thermodynamics. Theor. Appl. Climatol. 80, 37–48.
Vancutsem, C., Ceccato, P., Dinku, T., Connor, S.J., 2010. Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. Remote Sens. Environ. 114, 449–465.
Vogt, J.V., Viau, A.A., Paquet, F., 1997. Mapping regional air temperature fields using satellite-derived surface skin temperatures. Int. J. Climatol. 17, 1559–1579.
Voogt, J.A., Oke, T.R., 2003. Thermal remote sensing of urban climates. Remote Sens. Environ. 86, 370–384.
Wan, Z., Dozier, J., 1996. A generalized split-window algorithm for retrieving land-surface temperature from space. IEEE Trans. Geosci. Remote Sens. 34, 892–905.
Weng, Q., 2009. Thermal infrared remote sensing for urban climate and environmental studies: methods, applications, and trends. ISPRS J. Photogramm. Remote Sens. 64, 335–344.
Xu, Y., Knudby, A., Ho, H.C., 2014. Estimating daily maximum air temperature from MODIS in British Columbia, Canada. Int. J. Remote Sens. 35, 8108–8121.
Xu, Y., Qin, Z., Shen, Y., 2012. Study on the estimation of near-surface air temperature from MODIS data by statistical methods. Int. J. Remote Sens. 33, 7629–7643.
Yan, H., Zhang, J., Hou, Y., He, Y., 2009. Estimation of air temperature from MODIS data in east China. Int. J. Remote Sens. 30, 6261–6275.
Yang, Y.Z., Cai, W.H., Yang, J., 2017. Evaluation of MODIS land surface temperature data to estimate near-surface air temperature in Northeast China. Remote Sens. 9, 410.
Yoo, S., Im, J., Wagner, J.E., 2012. Variable selection for hedonic model using machine learning approaches: a case study in Onondaga County, NY. Landscape Urban Plann. 107, 293–306.
Yu, L., Wang, J., Li, X., Li, C., Zhao, Y., Gong, P., 2014. A multi-resolution global land cover dataset through multisource data aggregation. Sci. China Earth Sci. 57, 2317–2329.
Zeng, L., Wardlow, B.D., Tadesse, T., Shan, J., Hayes, M.J., Li, D., Xiang, D., 2015. Estimation of daily air temperature based on MODIS land surface temperature products over the corn belt in the US. Remote Sens. 7, 951–970.
Zhang, H., Zhang, F., Ye, M., Che, T., Zhang, G., 2016. Estimating daily air temperatures over the Tibetan Plateau by dynamically integrating MODIS LST data. J. Geophys. Res.: Atmos. 121.
Zhang, L., Weng, Q., Shao, Z., 2017. An evaluation of monthly impervious surface dynamics by fusing Landsat and MODIS time series in the Pearl River Delta, China, from 2000 to 2015. Remote Sens. Environ. 201, 99–114.
Zhang, R., Rong, Y., Tian, J., Su, H., Li, Z.-L., Liu, S., 2015. A remote sensing method for estimating surface air temperature and surface vapor pressure on a regional scale. Remote Sens. 7, 6005–6025.
Zhu, W., Lű, A., Jia, S., 2013. Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products. Remote Sens. Environ. 130, 62–73.
Zhang, X., Li, P., 2018. A temperature and vegetation adjusted NTL urban index for urban area mapping and analysis. ISPRS J. Photogramm. Remote Sens. 135, 93–111.