Master's Thesis

Estimating Cooling Effect of Urban Green Areas Using Random Forest Prediction Model

Ha-uk Jeong

Department of Urban and Environmental Engineering
(Urban Infrastructure Engineering)

Graduate School of UNIST

2018

Estimating Cooling Effect of Urban Green Areas Using Random Forest Prediction Model

Ha-uk Jeong

Department of Urban and Environmental Engineering
(Urban Infrastructure Engineering)

Graduate School of UNIST

Estimating Cooling Effect of Urban Green Areas Using Random Forest Prediction Model

A thesis/dissertation
submitted to the Graduate School of UNIST
in partial fulfillment of the
requirements for the degree of
Master of Science

Ha-uk Jeong

6. 12. 2018

Approved by

Advisor

Gihyoug Cho

Thesis/Dissertation Title

Ha-uk Jeong

This certifies that the thesis/dissertation of Ha-uk Jeong is approved.

6/12/2018

signature

Advisor: Gihyoug Cho

signature

Jeongseob Kim

signature

Jibum Chung

signature

Abstract

Rapid urban growth has resulted in increases in the amount of impervious surfaces in cities, inducing changes in urban climates, which are generally warmer than their rural surroundings This phenomenon is known as the urban heat island (UHI) effect. For these reasons, the use of green areas has been increasing to reduce the heat island phenomenon in the city. However, many studies have analyzed the UHI intensity using land surface temperature (LST) since it is extremely difficult to accurately measure the surface air temperature for a wide range of urban spaces. Because the LST is a measurement of how hot the land is to the touch, it is different from the thermal condition in which people actually feel. Therefore, this study evaluates the cooling effect of green areas by estimating the air temperature using the Random Forest model, one of the machine learning techniques. 138 AWS temperature data were collected for Seoul city. Land surface temperature was calculated by LandSat 8 OLI image using Surface Energy Balance Algorithm (SEBAL). Also, vegetation, Urban, and Weather related variables were used for prediction. The vegetation variables are Nomalized Difference Vegetation Index (NDVI), Weighted Difference Vegetation Index (WDVI), Soil Adjusted Vegetation Index (SAVI) and Leaf Area Index (LAI). The variables related to the weather were Albedo, elevation, longitude, latitude, and Julian Day. Also, we used the 100m and 400m buffered area within the building area, the building coverage area and the average height of the buildings. In order to analyze the cooling effect of the green space, the temperature distribution was confirmed for each land use in Jung - gu and Jongno - gu in Seoul. As a result of the analysis, Random Forest prediction model showed high prediction performance(R^2 =0.69. RMSE=1.23). The temperature cooling effect of green areas was also confirmed, but the effect was insufficient.

Contents

List of Figures

List of Tables

I. Introduction

Rapid development of the city has increased the impervious area and increased the heat of artificial heat, and caused urban heat island phenomenon where the city temperature is higher than the suburbs (Kim et al., 2017). As a result of urban heat island phenomenon, there are economic and social problems that increase energy use in summer and inconvenience physical activity. In many studies conducted to mitigate these heat island phenomena, the shape of the city has shown that heat island strength can be reduced (Kong et al., 2014). In particular, securing enough green spaces in the city center has been proposed as a planned way to reduce the city's temperature, as well as providing a relaxing space for citizens (Park et al., 2016). Most studies analyzing the effect of temperature reduction on green areas have analyzed the heat island phenomenon using surface temperature. Since the surface temperature depends on the material or characteristics of the surface, it has a limitation that it differs from the actual atmospheric temperature. The atmospheric temperature is directly related to the thermal comfort of the person in the external space. However, since the number of meteorological stations that measure temperature is very limited, it is difficult to utilize it in research to analyze the relationship between microscopic urban environment and microclimate.

On the other hand, the prediction performance of the nonlinear relationship has been improved due to the recent developments of the machine learning technology. In the remote sensing field, studies on the prediction of the atmospheric temperature using the surface temperature have been actively carried out. In the existing remote sensing research, MODIS satellites were used to predict the surface temperature (Yoo et al., 2018). However, MODIS satellites were difficult to utilize in urban areas with a spatial resolution of 1 km. Within the city, there are various physical spatial characteristics within 1 km. Therefore, it is necessary to predict the atmospheric temperature with a high spatial resolution in order to have the urban planning meaning for the physical environment.

Therefore, in this study, two - step study was carried out to estimate the atmospheric temperature by using LandSat 8 OLI satellite and to predict the atmospheric temperature by applying it to the green space. First, a random forest prediction model was developed to estimate the maximum and minimum temperatures. In order to solve the problem of the sample which is the limitation of the existing LandSat satellite research, we obtained the landsat images of various dates and predicted the temperature by considering the date related variables as the fixed effect. In addition, the model was developed by adding variables related to urban development that were not considered in the research of remote sensing field. Second, based on the developed model, we analyzed the effect of temperature reduction on the green areas in the seoul. Existing studies have been conducted using surface temperature due to insufficient atmospheric temperature data. Therefore, in this study, the temperature reduction effect of green space was analyzed using the predicted atmospheric temperature model.

Seoul, Gyeonggi-do and Incheon as the study sites, and the 9 cloudless landSat 8 OLI satellite images, vegetation variables, 138 automatic weather observation data, urban development related variables, Julian day, DEM , Latitude and longitude was used to develop the model. In order to analyze the effect of temperature reduction on green space, Jung-gu, Jongno-gu located in Seoul were selected and the temperature reduction effect was analyzed.

II. Literature Review

2.1 Remote Sensing – Predict Air Temperature

Table 1. Air temperature prediction study – Satellite and Method

Title	Satellite	Method
A statistical framework for estimating air temperature using MODIS land surface temperature data.	MODIS	stepwise regression analysis
Comparison of Multiple Linear Regression, Cubist Regression, and Random Forest Algorithms to Estimate Daily Air Surface Temperature from Dynamic Combinations of MODIS LST Data	MODIS	RF, cubist regression, linear regression
Evaluation of MODIS Land Surface Temperature Data to Estimate Near-Surface Air Temperature in Northeast China	MODIS	All Subsets Regression,
Estimating Daily Maximum and Minimum Land Air Surface Temperature Using MODIS Land Surface Temperature Data and Ground Truth Data in Northern Vietnam	MODIS	Multiple linear regression
Estimating daily air temperature across the Southeastern United States using high-resolution satellite data: A statistical modeling study.	MODIS	mixed linear models, cross-validation, RMSPE
Estimation of daily air temperature based on MODIS land surface temperature products over the corn belt in the US	MODIS	-
Mapping maximum urban air temperature on hot summer days	Landsat5 & 7	RF,SVM,Regression
Estimating air surface temperature in Portugal using MODIS LST data	MODIS	Mixed bootstrap, Jackknife resampling, Model Efficiency Index, RMSE
Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa	MODIS	-
Parameterization of air temperature in high temporal and spatial resolution from a combination of the SEVIRI and MODIS instruments	MSG SEVIRI, MODIS	Regression
Modeling air temperature through a combination of remote sensing and GIS data	Landsat5 & 7, AVHRR, MODIS	Multiple regression, RMSE
Estimation of diurnal air temperature using MSG SEVIRI data in West Africa	MSG SEVIRI	RMSE, MBE

Integrating AVHRR satellite data and NOAA ground observations to predict surface air temperature: a statistical approach	AVHRR	kriging, Multi linear regression, Cross-validation,
Estimating surface air temperatures, from Meteosat land surface temperatures, using an empirical solar zenith angle model	Meteosat	-

In this study, the literature review was divided into two parts. In the first part, literature was conducted in the field of remote sensing. The development of the predicted model of the atmospheric temperature has proceeded in the meteorological field. Many machine learning techniques were used for prediction, but the Random Forest method had the highest prediction performance(Phan Thanh Noi et al., 2017).

In the case of prediction of atmospheric temperature, it is difficult to analyze the relationship with a general statistical model. The relationship of the variables used for temperature prediction is nonlinear. General statistical methods have poor nonlinear predictive performance. Therefore, random forests with high nonlinear relationship prediction performance have been used recently. Many remote sensing studies use MODIS satellites for temperature prediction. The MODIS satellite has a spatial resolution of 1km and has the advantage of photos twice a day. However, in urban studies, spatial resolution of 1 km is difficult to utilize. LandSat satellites provide high resolution data of 30 meters, whereas MODIS satellites have a wide spatial resolution. However, there are disadvantages that it is difficult to use cloudy satellite images and Landsat shoots twice a month. Because of these problems, previous studies mainly used MODIS satellites in the remote sensing field to analyze large area space. However, recent studies have conducted temperature prediction studies based on LandSat satellite data collected over many years (Hung Chak Ho et al., 2014).

When using Landsat satellite, it showed lower R-square than MODIS satellite. However, unlike extracting only the variables at the location where AWS is located in previous researches, we have applied the buffer to the used variables to study the spatial extent of the neighboring variables to observe low root mean square error (RMSE). While the studies using MODIS have been based on satellite data accumulated for many years, LandSat data has a disadvantage that the number of data samples is small. In the case of the atmospheric temperature, the R-square is high in the MODIS data collected over many years because it shows a similar temperature pattern every year for a clear cloudless day. Normalized Difference Vegetation Index (NDVI), reflectivity (Albedo) and elevation (DEM) were mainly used for the variables used for temperature prediction. In most studies, the surface temperature was calibrated and used. In case of LandSat8 OLI, it is possible to calculate surface temperature simply by using Band10 and Band11. However, it has a drawback that it only takes into consideration the brightness of the surface. Energy balance algorithms have been developed and used to solve these problems. In this study, SEBAL (Surface Energy Balance Algorithm for Land) method was used. SEBAL method is using LandSat's various bands, we calculate the brightness-related parameters and various vegetation-related parameters to calculate the surface temperature considering the surface energy characteristics (Bastiaanssen et al., 1998).

In many studies, SVM (Support Vector Machine), Random Forest, Mixed bootstrap, and Cubist regression models were used for temperature prediction, but Random Forest had the highest predictive power. Machine learning has a problem with the black box that does not know the process in the interpretation of the results. Therefore, in order to solve this problem, in the study using the Random Forest, the data preprocessing process was removed before the model development to remove the unnecessary variables. In the model analysis, the cross – validation was used to evaluate the model to improve the explanatory power of the model.

2.2 Green Areas – Cooling Effect

Table 2. Cooling effect study – Temperature data

Title	Temperature Data
Variation Profiles of Temperature by Green Area of Apartments in Gangnam, Seoul	Landsat ETM+
An Analysis of Micro-climate Environmental Changes Followed by Establishment of an Urban Park	ENVI-met
Influence of park size on the park cooling effect-Focused on Ilsan new town in Korea	Landat 8 OLI
Use of remote sensing and geographical information systems to estimate green space surface-temperature change as a result of urban expansion	Landsat TM
Spatial pattern of greenspace affects land surface temperature: evidence from the heavily urbanized Beijing metropolitan area, China	Landsat 5
A study on the Urban Heat Island Phenomenon Using LANDSAT TM Thermal Infrared Data - In the Case of Seoul	Landsat TM
Estimation of land surface temperature-vegetation abundance relationship for urban heat island studies	Landsat 7
Effects of green space dynamics on urban heat islands: Mitigation and diversification	Landsat TM
Effects of green space spatial pattern on land surface temperature: Implications for sustainable urban planning and climate change adaption	Landsat 5

In the second part, literature review was conducted on the effect of the temperature reduction in green areas.

Urban green areas have a temperature-reducing effect, and the larger the size, the more likely it is to affect the temperature change (Park, Sun-Young et al., 2009).

In the case of neighborhood parks, it was found that the effect of reducing the temperature around the park was increased, and that the larger the scale, the more effective the effect (Oh et al., 2005).

Landsat ETM + was used to analyze the temperature change of the green area of the apartment complex in the city center. The higher the NDVI, the lower the surface temperature (Hong et al., 2004).

In the case of Daegu Park, the micro-environment was quantitatively analyzed, and it was confirmed that the micro-environment of the park was different according to temperature and humidity (Kim Dae-wook et al., 2010).

Using the surface temperature of Landsat 8, the temperature reduction effect and the temperature reduction effect according to the size of the park were confirmed in Ilsan new city. It was confirmed that the effect of temperature reduction was shown by the size and distance of the park (Jonghwa Park, 2016).

We analyzed the temperature changes of urban development in Seoul. Landsat TM satellite images were used. Unlike the suburban area, the surface temperature of Seoul was increased as the development progressed (Sin Dong-hoon et al., 2005).

We observed the temperature change of the city center in Beijing, China. Landsat surface temperature was used and when the greenspace was increased by 10%, the surface temperature decreased by 0.86 °C (Xiaoma Li et al., 2012).

Landsat TM was used for the Seoul city to observe the temperature changes due to altitude and land cover. As a result, we could confirm the heat island phenomenon around downtown Seoul and confirmed that the green zone is effective in reducing the temperature. In addition, it was confirmed that there is no temperature reduction effect according to altitude in Seoul (Minho Park, 2001).

Using the Landsat TM temperature data, the effect of temperature reduction according to the type of green space was analyzed. As a result, it was confirmed that the effect of temperature reduction varies depending on the type of greenery (Ranhao., 2017).

However, in most of the studies, the surface temperature was used to study the effect of reducing the temperature in green areas. However, the surface temperature is different from the atmospheric temperature, which directly affects human activity, and the necessity of research through the atmospheric temperature has been suggested. The effect of temperature reduction on green areas has proved its effectiveness in many existing studies. However, the study of cooling effect of green space using atmospheric temperature is insufficient.

III. Method

3.1 Study Area and Sample

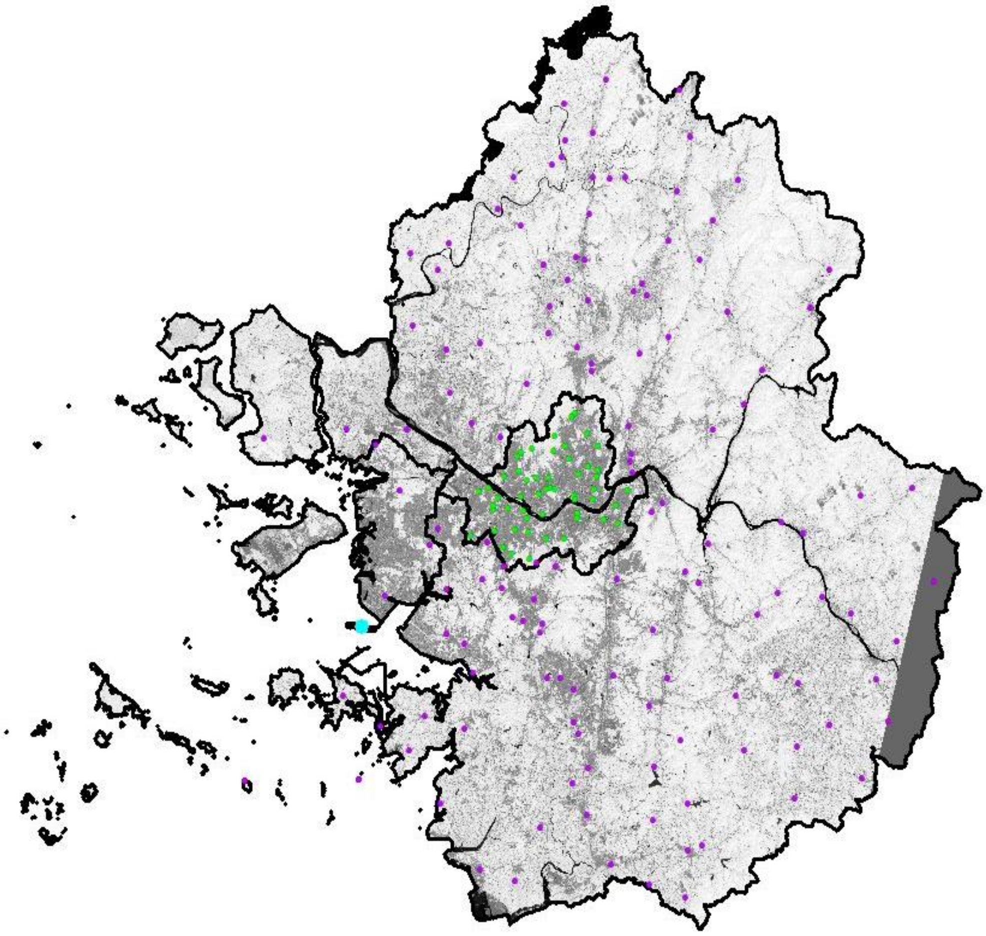


Figure 1. AWS location

The spatial range of the model for estimating the atmospheric temperature is the Seoul metropolitan area (Incheon, Gyeonggi-do). Seoul is the capital of Korea and the administrative district covers an area of 605.21㎢ The area of Incheon is 1,029 ㎢ and the area of Gyeonggi is 10,171㎢ We collected temperature data from 138 automatic weather stations (AWS) located in the metropolitan area. Only inland data are used excluding islands located in the West Sea. We calculated the surface temperature, vegetation and urban development parameters of LandSat8 OLI data for 138 AWS with spatial ranges of 30m, 100m and 400m.

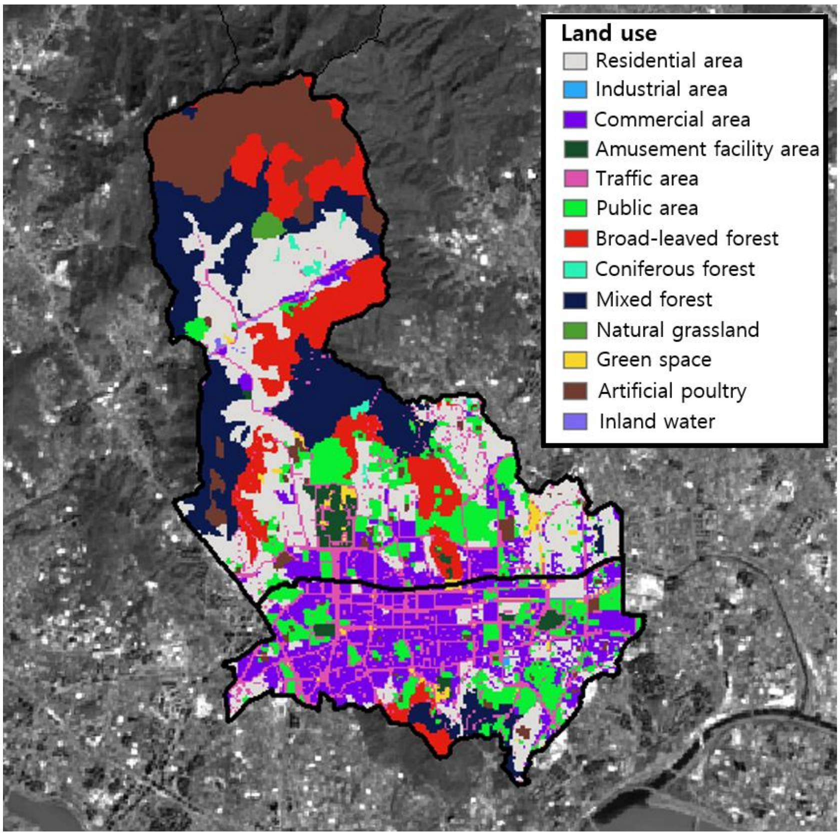


Figure 2. Jung-gu, Jongno-gu Land use

In order to analyze the effect of temperature reduction, the temperature reduction effect was analyzed according to the land use in Jung - gu and Jongno - gu in Seoul. There were 13 land use types in total, but they were classified into six categories of industrial, residential, commercial, public, green, and amusement areas.

3.2 Variable Description

In this study, we analyzed the data at 9 time points and used the parameters of 52 AWS locations. The number of selected variables is 26, and the total number of constructed samples is 968. In the machine learning analysis, 26 variables were reduced to 14 through preprocessing.

Table 3. Variable description

Type	Variable	Description
Temperature data	Korea Meteorological Administration(°C)	KMA stations maximum and minimum temperature data
Satellite-related variables	LST 30m, LST100m, LST 400m (°C)	LandSat 8 OLI Land suface temperature(SEBAL method)
	Albedo	Reflectance: the degree of reflection of light
	NDVI 30m, 100m, 400m	Nomalized vegetation distribution: Index indicating vegetation distribution
	WDVI 30m, 100m, 400m	Vegetation index corrected vegetation height (soil - vegetation)
	SAVI 30m, 100m, 400m	Soil adjust vegetation index: Vegetation index which minimizes the effect of soil brightness
	LAI 30m, 100m, 400m	Leaf area index: Leaf area index of land area
Urban-related variables	Building Area 100m, Building Area 400m*	Average Building Area
	Building Height 100m, Building Height 400m*	Average height of buildings
	Impervious Area*	Impervious area
Weather-related variables	Julian day	Index that integrates dates with an interval of 1
	Latitude, Longitude(°)	Latitude, longitude
	Elevation(m)*	DEM: Altitude data

Remote Sensing Variables

LandSat8 satellite data were used as remote sensing variables. The LandSat8 satellite is a satellite managed by the US Geological Survey (USGS) and has a spatial resolution of 30m * 30m. It has a total of 11 bands and it captures about twice a month but sometimes it does not provide it according to the weather conditions. To collect the variables, we collected satellite images of cloudless summer days at nine different times from 2013 to 2016. In the satellite images, the surface temperature (SEBAL-LST) using the energy balance algorithm was calculated to construct the variables. The albedo, normalized difference vegetation index (NDVI), Weighted Difference Vegetation Index (WDVI), the Soil Adjusted Vegetation Index (SAVI), and the Leaf Area Index (LAI) as variables.

Albedo

The reflectivity is a value that reflects the light and differs depending on the characteristics and the material of the land. A total of 6 bands were used and LandSat8 images were used after Radiometric Calibration using ENVI 4.1.

$$Albedo = \frac{(((B2 \times 0.3) + (B3 \times 0.277) + (B4 \times 0.233) + (B5 \times 0.143) + (B6 \times 0.036) + (B7 \times 0.012)) - (0.03))}{0.75 + 2 \times (10^{-6})) \times (2011)}$$

$$Bn : Band\ Number$$

Equation 1

Normalized Difference Vegetation Index, NDVI

Nomalized vegetation indices were calculated using ENVI 4.1, and LandSat8 images with radial calibrations were calculated using Band4 and Band5. It is a variable indicating the distribution of vegetation.

$$NDVI = \frac{P_{NIR} - P_{RED}}{P_{NIR} + P_{RED}}$$

Equation 2

Weighted Difference Vegetation Index, WDVI

Regular vegetation index was calculated using ENVI 4.1 and WDVI was calculated by Radiometric Calibration using LandSat8 images using Band3 and Band4. WDVI is a vertical vegetation index which indicates characteristics according to vegetation height.

$$WDVI = NIR - aRED$$

Equation 3

Soil Adjusted Vegetation Index, SAVI

The soil support vegetation index is calculated using NDVI and WDVI calculated above. It is a variable that shows characteristics according to the characteristics of soil, and it is a variable used to distinguish between grassland and inland.

$$SAVI = \frac{(NIR - RED)(1 + L)}{(NIR + RED + L)}$$

$$L = 1 - 2 \times a \times NDVI \times WDVI$$

Equation 4

Leaf Area Index, LAI

The percentage of leaf area in the Vegetation areas. Calculated from ENVI 4.1 and using the previously calculated SAVI.

$$LAI = (11 \times (SAVI^3))$$

Equation 5

Surface Emmisivity

It is an index representing the emissivity of the earth surface. It is a variable used to calculate the surface temperature in the SEBAL algorithm.

$$E_{NB} = 0.97 + 0.0033 \times LAI$$

Equation 6

Longwave & Shortwave radiation

Longwave radiation, and shortwave radiation. It is a variable used to calculate the surface temperature in the SEBAL algorithm.

$$R_o = (((L_{\text{사용}} - 0)/1) - ((1 - E_{\text{사용}}) \times 1))$$

$$L_{\text{사용}} = \mathit{Themal\ Band}\text{의 } \mathit{Radiance}\text{ 보정값}$$

Equation 7

Land Surface Tempeature, LST

The surface temperature was calculated using the SEBAL algorithm. ENVI 4.1 program. It is a method to compute the surface temperature by balancing the energy of the land through satellite image correction. Consider solar radiation and radiation and use several vegetation variables.

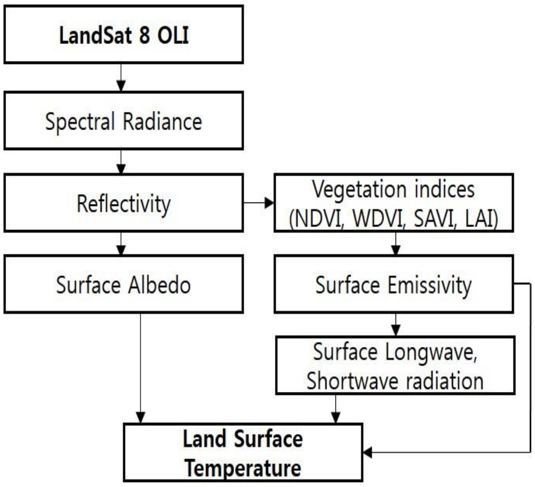


Figure 3. SEBAL Algorithm

The equation for calculating the surface temperature may seem similar in terms of using Band10 and Band11 provided by USGS, but it is calculated by including the various variables calculated above in the equation.

$$LST = ((B10 + B11)/2) - 273.15$$

Equation 8

$$B10 = \frac{1321.08}{\mathrm{Ln}\left(\frac{E_{NB} \times 774.89}{R_{0-\text{[illegible]}1}} + 1\right)}$$

Equation 9

$$B11 = \frac{1201.14}{\mathrm{Ln}\left(\frac{E_{NB} \times 480.89}{R_{0-\text{[illegible]}2}} + 1\right)}$$

Equation 10

Urban and meteorological variables

The impervious area ratio was used by processing the land coverage of the environmental spatial information service. The building area, building height, and elevation are based on data provided by the National Space Information Portal. Julian day is a conversion of monthly dates into one interval. On a clear day, Julian day was used as a predictor variable since the temperature shows a similar pattern every year. The Julian day calculations were done through the Julian day calendar and calculator provided by NASA.

IV. Analytical Method

4.1 Study Design

To develop a model for temperature prediction, we obtained a total of 26 variables and a total of 968 samples for 9 time points. Satellite, urban, vegetation, and meteorological variables were collected for each year. Satellite, urban, and meteorological variables were used as independent variables through preprocessing. The temperature of the meteorological station was used as a dependent variable. The cross validation method was used for model verification and the predictive performance was evaluated based on several developed models. Based on the most reliable model, the temperature reduction effect of the green areas was evaluated.

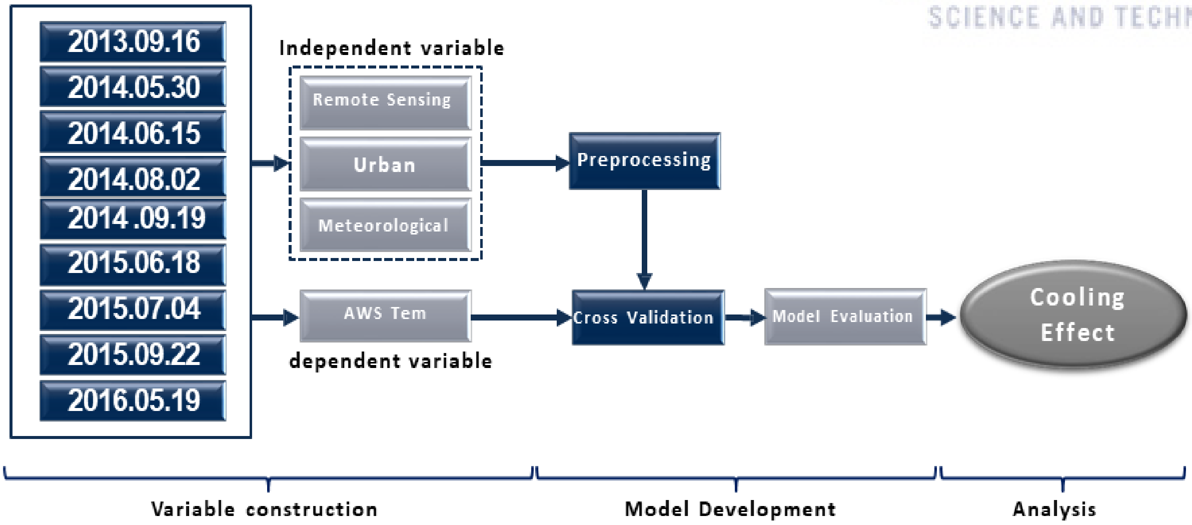


Figure 4. Study flow

4.2 Preprocessing(Decrease variable dimension)

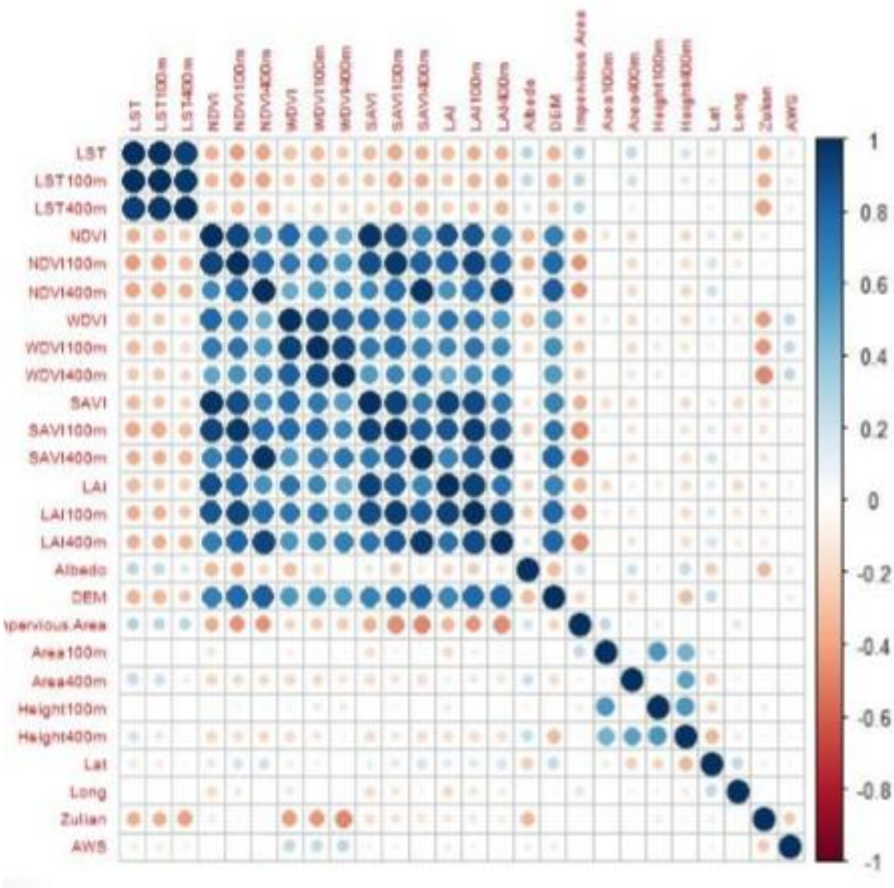


Figure 5. Variable Preprocessing

Machine learning has a problem of over fitting, which has a high prediction performance but an excessively high model performance. Therefore, it is necessary to select and process variables before model building in machine learning. For the preprocessing, we used the CARET (Classification and Regression Training) tool of R package. Preprocessing was performed on 25 variables to be used for model training, and 14 variables were selected through this process. The criteria for selecting variables were one of the variables with correlation of 75% or more. As shown in Figure 5, the variables with strong color are highly correlated. Correlated variables were selected or removed by the CARET tool and used in the model.

4.3 Linear Regression

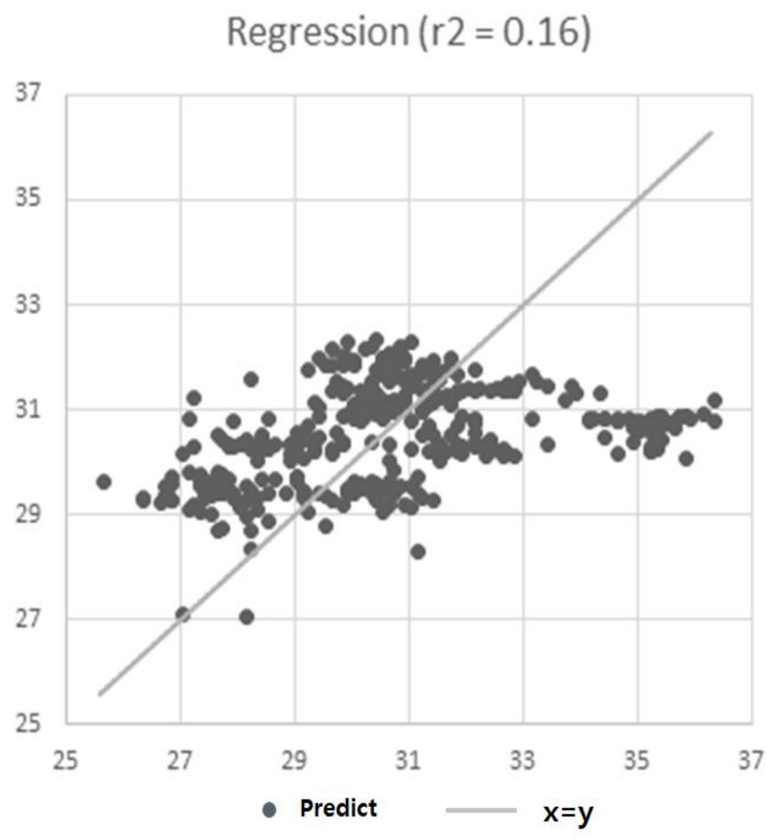


Figure 6. Regression prediction model

The prediction performance was evaluated by regression model before machine learning. Regression model showed 0.16 prediction performance, and prediction performance was low. In many existing remote sensing researches, it is confirmed that the performance of temperature prediction through regression model is remarkably low. It can be confirmed that there is a limit to the temperature prediction by the existing linear analysis method. Many previous studies have used modified regression models or machine learning methods to solve the nonlinear relationship between surface temperature and Air temperature.

4.4 Random Forest

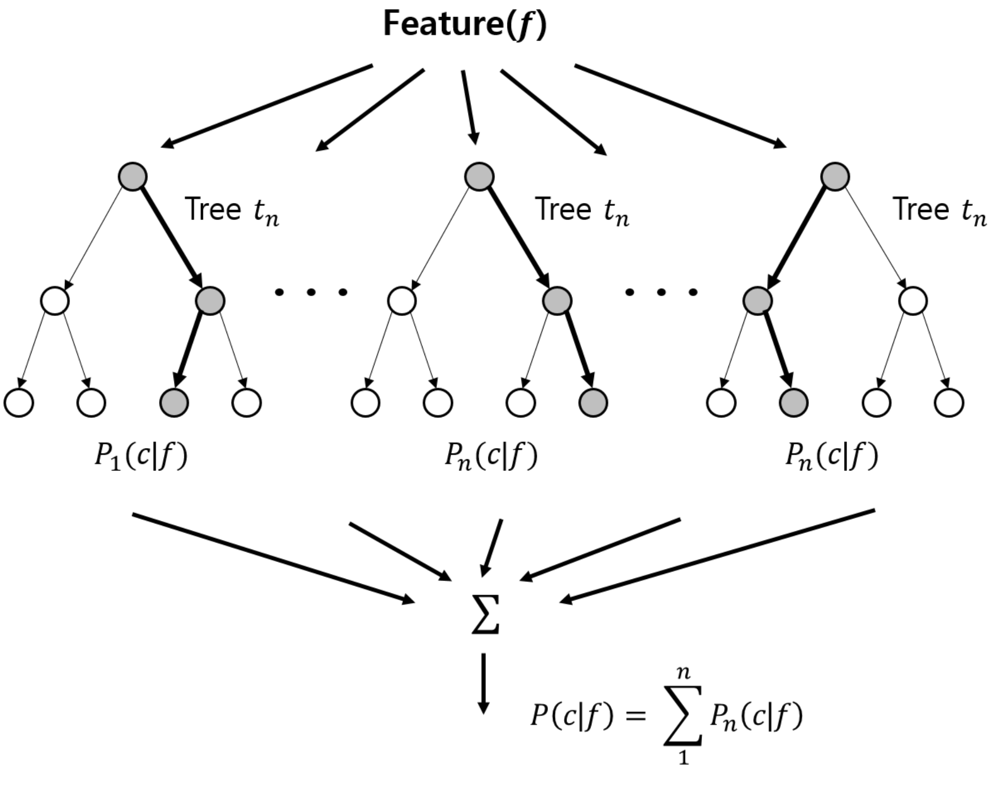


Figure 7. Random forest Algorithm

Random Forest is one of the methods of machine learning. It creates several decision tree through random variable selection and outputs the predicted value by ensemble learning method. The Random Forest has higher prediction performance than other machine learning techniques. A total of 26 variables were used for model construction after preprocessing. The model was verified with 10-fold Cross validation, which is a cross validation method, for model verification after confirming the result with Train 70% and Test 30%.

4.5 10 fold Cross Validation

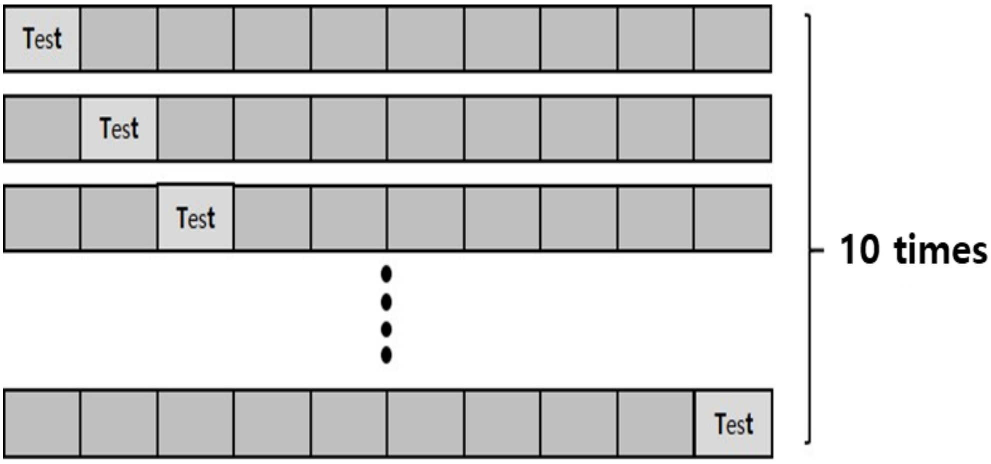


Figure 8. 10-fold crossvalidation

It is an analysis method used mainly when the number of data is small. The data used in the machine learning analysis is divided into 9 : 1 ratios and the data is verified 10 times. Once all the data is in the Test set, it can compensate for the disadvantages of random sampling.

V. Results

5.1 Random Forest Models Performance(Max & Min temperature)

The model results were divided into two cases. The first A case is a model with Julian day and the second B case is a model without Julian day. For each model, Seoul temperature correction model, and Seoul and Gyeonggi Incheon models were analyzed.

Table 4. Air & LST Temperature by date

Date	Julian day	Average : AWS	Average : LST400m
2016.05.19	140	32.20	33.71
2014.05.30	150	32.08	32.06
2014.06.15	166	28.74	28.66
2015.06.18	169	30.94	24.90
2015.07.04	185	30.23	29.96
2014.08.02	214	35.01	24.98
2013.09.16	259	27.66	28.63
2014.09.19	262	27.78	27.30
2015.09.22	265	30.33	27.90

In case of temperature correction model, AWS temperature difference is corrected to LST by Julian day to compensate difference between AWS temperature and LST temperature for each Julian day. For models including Seoul, Gyeonggi and Incheon, only 5 dates were used because of the capture date of Landsat satellite. The temperature distribution and the temperature reduction effect of the green area were confirmed using the final selected B2 model.

Model A-1, Model Performance(Include date, Temperature correction)

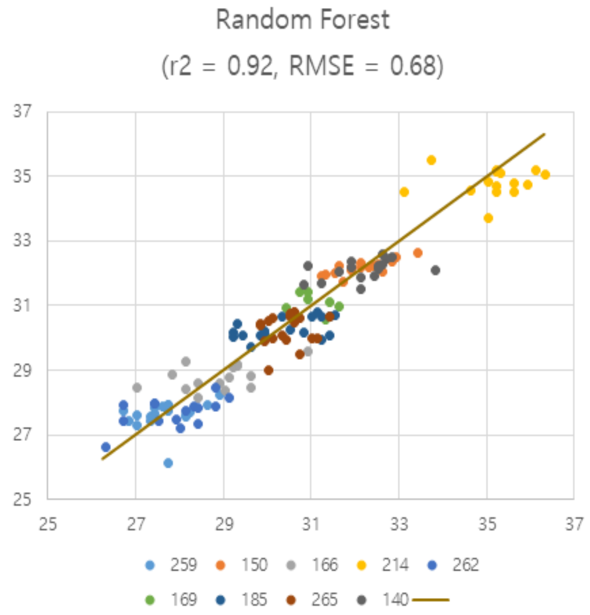


Figure 9. A1 Max Temperature

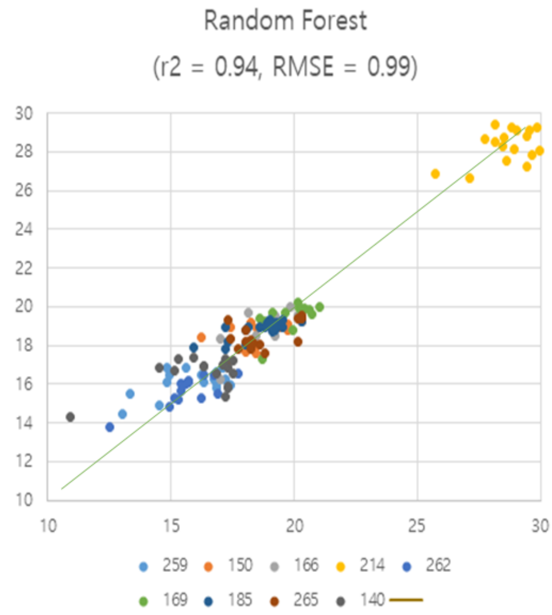
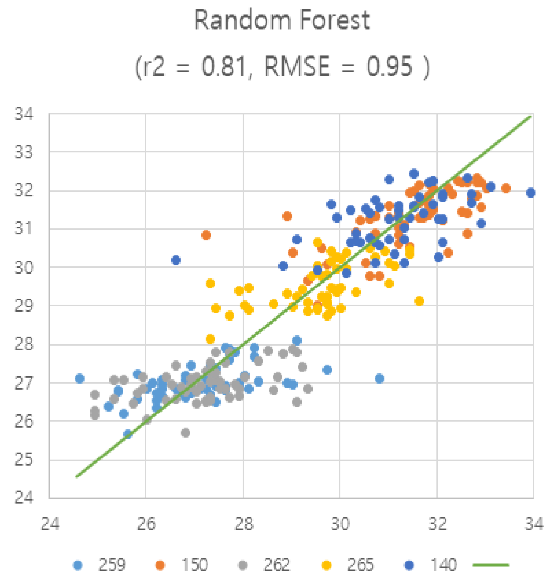


Figure 10. A1 Min Temperature

High prediction performance. However, we can confirm that it is clustered by Julian day. Also, in the minimum temperature model, Julian day 214 (Aug) confirmed that different date and temperature difference are greatly predicted.

Model A-2, Model Performance(Include date, Temperature correction, Regional expansion)



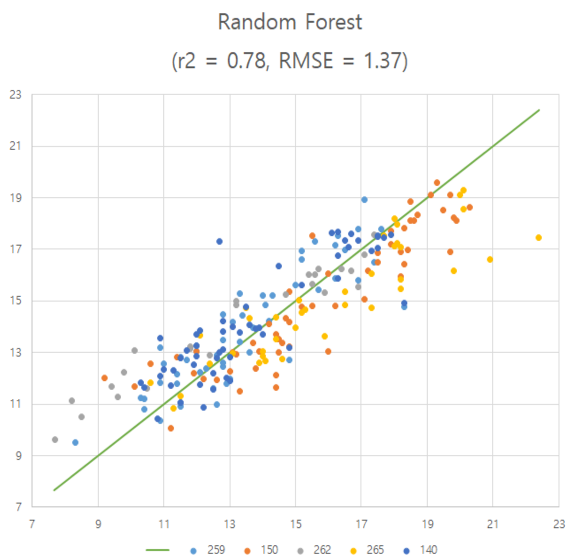


Figure 11. A2 Max Temperature **Figure 12. A2 Min Temperature**

And showed good predictive performance and RMSE. However, we can confirm that cluster tendency is strong by date.

Model B-1, Model Performance(Exclude date, Temperature correction)

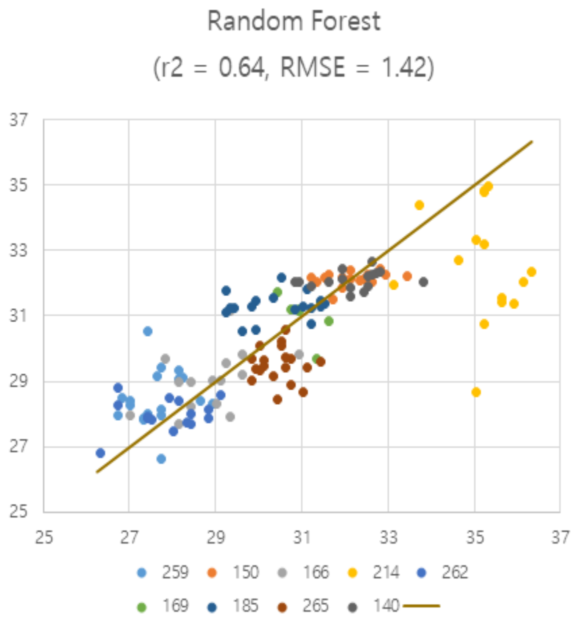


Figure 13. B1 Max temperature

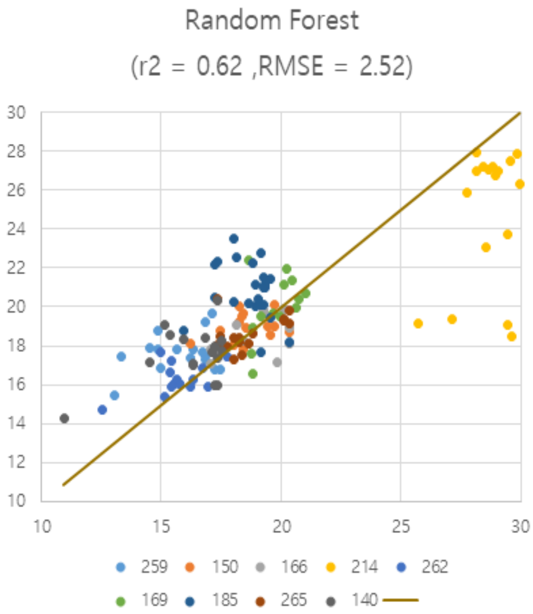


Figure 14. B1 Min Temperature

It can see that the forecasting performance is lower than Model A with date. Also, we can confirm that the result was still clustered by date.

Model B-2, Model Performance(Exclude date, Temperature correction, Regional expansion)

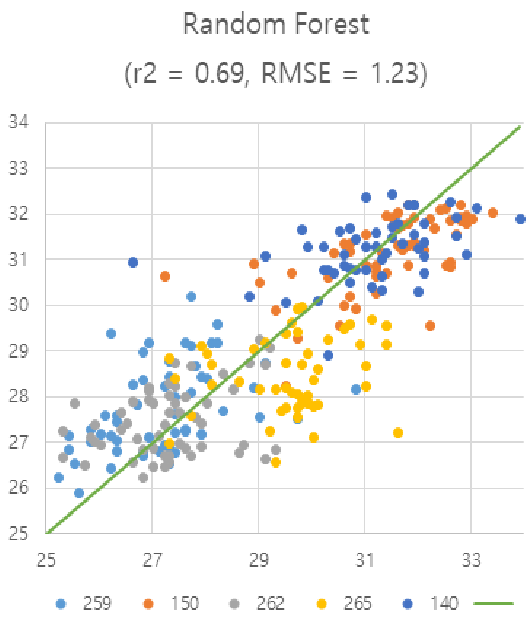


Figure 15. B2 Max Temperature

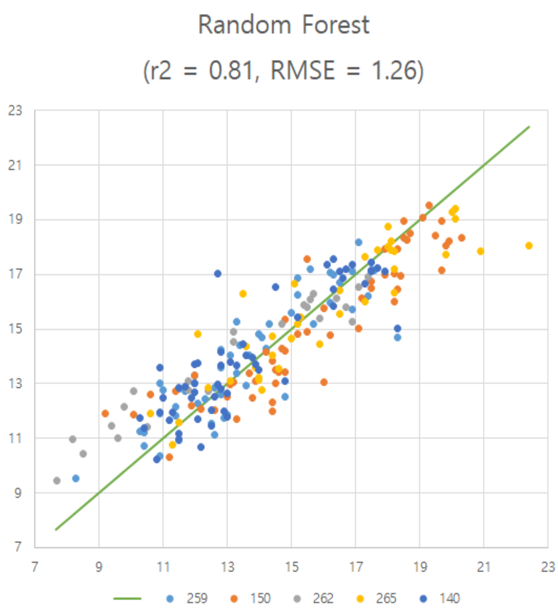


Figure 16. B2 Min Temperature

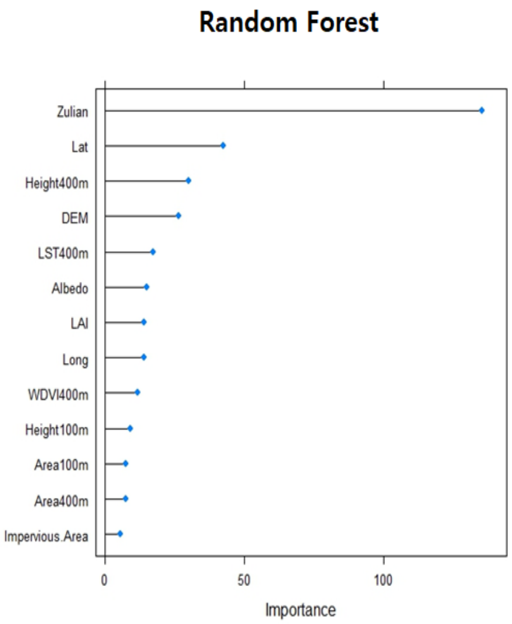
The prediction performance is lower than that of the A models including the dates. However, it can be seen as a model that does not cluster by date.

5.2 Sub conclusion -Model Perfomance

In general, Julian day was included in the model, which showed high prediction performance. However, due to the tendency to cluster by date, there seems to be a problem of overfitting. B models that do not include dates are less accurate than A models overall, but are not clustered by date. In particular, the B-2 model showed good prediction performance and a good forecast distribution. Due to the characteristics of Landsat, it was difficult to acquire images in summer. LandSat is taken once every 15 days and can not be used in the cloud images. In addition, the atmospheric temperature was higher than other dates in August, but the surface temperature was lower. For this reason, there was a tendency to cluster by date in the model for Seoul. However, in the case of Gyeonggi-do & Incheon, the model was not able to obtain a cloudless August image. Because of this, models that have expanded to Gyeonggi & Incheon have a uniform temperature distribution,

5.3 Random Forest Variable Importance

Model A-1, Variable Importance(Include date, Temperature correction)



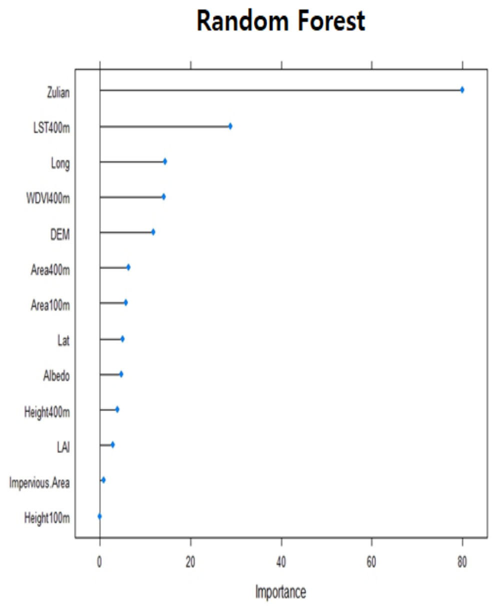
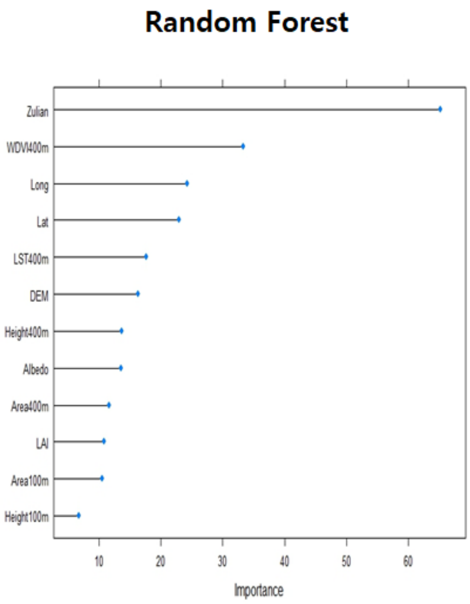


Figure 17. A1 Max Temperature **Figure 18. A1 Min Temperature**

It was found that the tendency of the results to be determined by Julian day was strong. The degree of dependence of the model on the Julian day is so large that the A1 model is clustered by date.

Model A-2, Variable Importance(Include date, Temperature correction, Regional expansion)



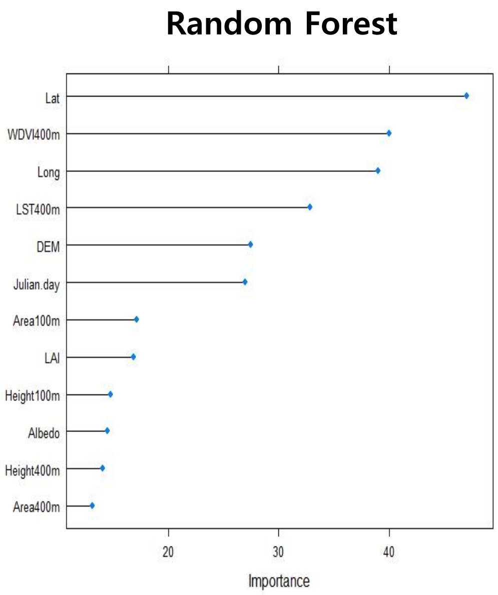


Figure 19. A2 Max Temperature **Figure 20. A2 Min Temperature**

The maximum temperature tends to be determined by Julian day. However, the minimum temperature was more likely to be affected by latitude. In the case of the minimum temperature, the degree of reliance on the Julian day was smaller than the previous model and it was confirmed that it depends on several variables.

Model B-1 Variable Importance(Exclude date, Temperature correction, Regional expansion)

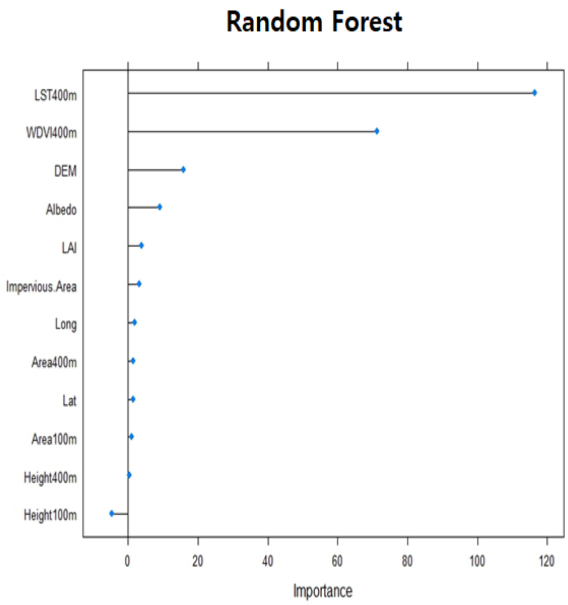


Figure 21. B1 Max Temperature

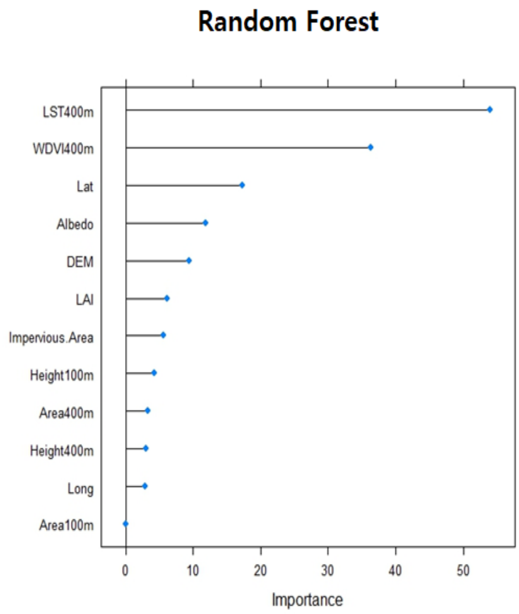


Figure 22. B1 Min Temperature

The degree of dependence on LST as a whole was great. Also, the tendency to rely on WDVI is large, but other variables are less affected.

Model B-2 Variable Importance(Exclude date, Temperature correction, Regional expansion)

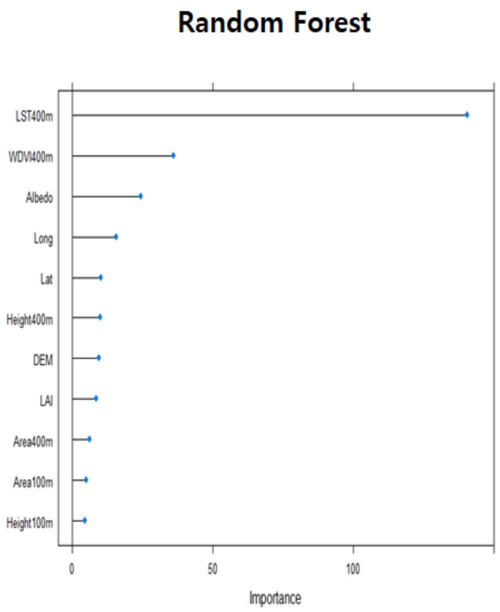


Figure 23. B2 Max Temperature

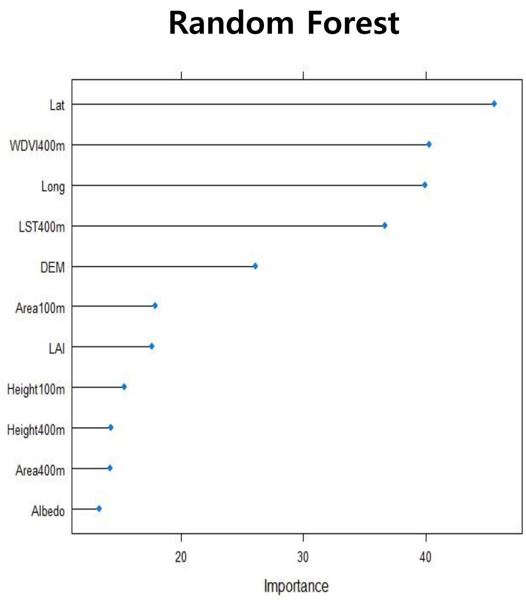


Figure 24. B2 Min Temperature

At the maximum temperature, LST had a large effect, however, latitude had a big impact on the Minimum Temperature.

5.4 Sub conclusion -Variable Importance

If the model contains a date, Julian day has a great influence on the model, so we can confirm that the models of A are overfitting by date. On the contrary, in Model B, the surface temperature was used as an important parameter for prediction. In Model B, Latitude was used as an important variable in Minimum Temperature Prediction.

5.5 Urban Temperature Distribution Using Prediction Model

Model B2 was finally used as a prediction model. The temperature prediction performance was shown to be the most uniform without cluster by date. As shown in the figure below and in the case of predicted temperature for Jung-gu, Jongno-gu, Seoul, it is confirmed that LST and predicted temperature distribution were similar. The predicted temperature showed a temperature difference of 6 ℃. In the case of the surface temperature, a temperature difference of 16 ° C was confirmed. The predicted atmospheric temperature did not show a large difference in the city center, but the LST could confirm the temperature difference even in the city center.

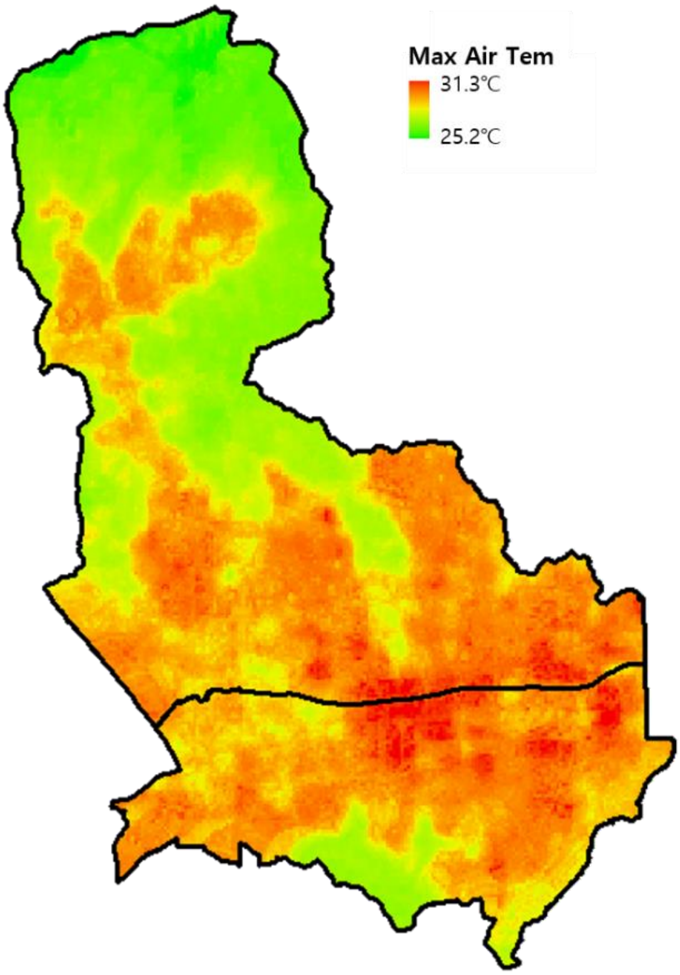


Figure 25. Predict Max-Air temperature(Jung-Gu,Jongno-Gu)

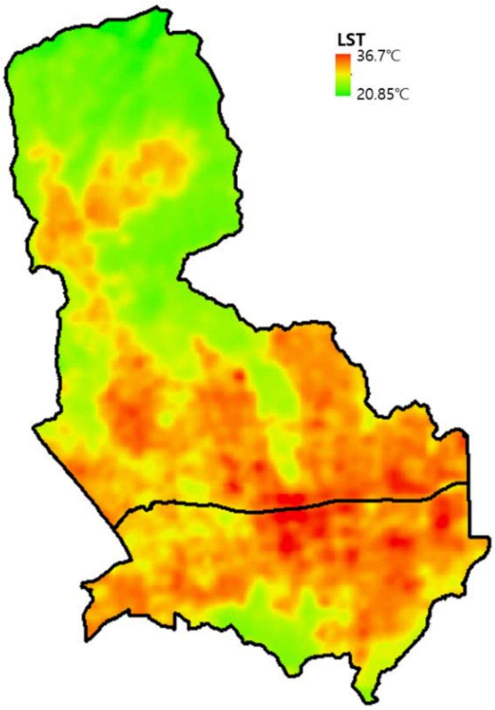


Figure 26. Land Surface Temperature(Jung-Gu, Jong-ro)

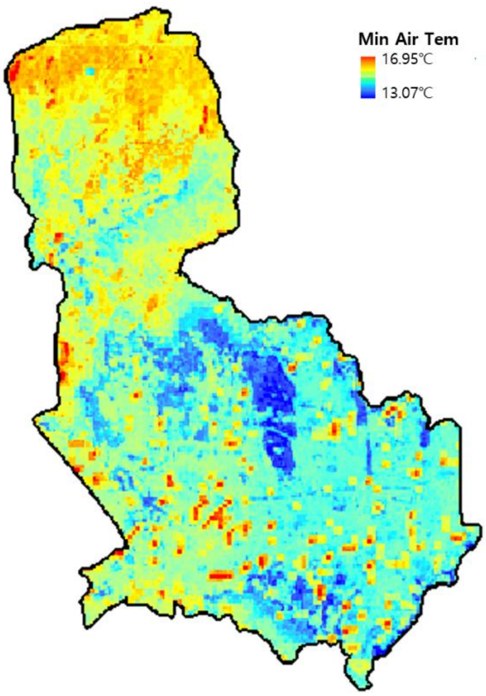


Figure 27. Predict Min-Air temperature(Jung-Gu,Jongno-Gu

Table 5. Temperature by application

LST	Max Air tem	Min Air tem	Type
31.811	29.342	14.372	Industrial
30.315	28.930	14.635	Residential
31.380	29.308	14.660	Commercial
30.808	29.103	14.503	Public
26.795	27.406	14.849	Green
31.178	29.226	14.438	Amusement

When the temperature was compared for Jung - gu and Jongno - gu, it was found that the maximum temperature of the green area was lower than other areas. However, in the case of the minimum temperature, the green area temperature in the city center was lower, but the green area located in the north of Jongno was higher. This is because the longitude, latitude and vegetation index greatly influence the results in the B2 minimum temperature prediction model

In addition, temperature comparisons were made between the six land uses to analyze the cooling effect of green areas. LST and the predicted temperature were found to be low in the green areas. However, it was confirmed that the minimum temperature did not decrease. This is because the temperature decreases depending on the type and location of the green area.

VI. Conclusions

The purpose of this study is to construct a prediction model of the atmospheric temperature using 30m high resolution LandSat satellite data and to analyze the temperature reduction effect of the green area using the model results.

For the Random Forest prediction model, the B-2 model showed the best prediction performance (= 0.69, RMSE = 1.23). However, the maximum temperature at day is different within 1 ~ 2 ℃, and it is difficult to make accurate analysis with the RMSE shown in the prediction model. Also, the number of data samples was insufficient (968) to produce sufficiently reliable results. Considering the large number of samples used in studies using existing MODIS, the lack of available samples remains a technical limitation.

The effect of reducing the temperature of the green area was confirmed at a slight level. It is considered that the width of the maximum temperature change during the summer season does not appear to be significant according to spatial characteristics. When compared with land use, the temperature difference between green area and other land use was found in LST. However, when comparing the predicted maximum and minimum temperatures, the cooling effect of the green area was hardly confirmed. Also, because the maximum temperature is greatly influenced by the date or solar radiation, it does not appear to show large differences within the city. In the case of the LST used in the previous studies, the effect of the temperature reduction of the green area can be clearly distinguished because the temperature change occurs at a large difference regardless of the solar radiation. However, it can be confirmed that the actual atmospheric temperature does not show a large difference within the city. The minimum temperature predictions were high variable importance in latitude, vegetation index and longitude. The minimum temperature distribution appeared regardless of LST. For this reason, the cooling effect was not clearly visible.

It is expected that an effective methodology for the analysis of urban environment and thermal environment in a microscopic range will be possible by using a temperature prediction model using Landsat. In addition, this methodology is expected to have high utility for the spatial design guidelines to improve the thermal environment, such as the arrangement of buildings and the shape of the street canyons, as well as the effect of reducing the temperature of green spaces.

REFERENCES

1. Kim, M., Lee, K., & Cho, G. H., 2017, “Temporal and spatial variability of urban heat island by geographical location: A case study of Ulsan, Korea”. Building and Environment, 126: 471-482.

2. Kong, F., Yin, H., James, P., Hutyra, L. R., & He, H. S., 2014, “Effects of spatial pattern of greenspace on urban cooling in a large metropolitan area of eastern China”, Landscape and Urban Planning, 128: 35-47.

3. Cheolhee Yoo, Jungho Im, Seonyoung Park, Lindi J. Quackenbush, 2018, “Estimation of daily maximum and minimum air temperature in urban landscapes using MODIS time series satellite data”, ISPRS Journal of Photogrammetry and Remote Sensing, 137: 149-162.

4. 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.

5. Kim, Jeong-Ho Oh., Deuk-Kyun, Yoon., Yong-Han., 2015, “Variation Profiles of Temperature by Green Area of Apartments in Gangnam, Seoul”, Journal of Environmental Science International, 24(5), 647-655.

6. Eliasson I., Svensson M., 2003, “Spatial air temperature variations and urban land use – a statistical approach”, Meteorological Application 10: 135-149.

7. SUE GRIMMOND, 2007, “An Analysis of Micro-climate Environmental Changes Followed by Establishment of an Urban Park”, The Geographical Society, 173(1), 3-99.

8. Park, Jong-Hwa., Cho, Gi-Hyoug., 2016,“Influence of park size on the park cooling effect-Focused on Ilsan new town in Korea”, Journal of Korea Planning Association, 51(5), 247-261.

9. Dong-hoon Shin, Kyoo-seock Lee, 2005, “Use of remote sensing and geographical information systems to estimate green space surface-temperature change as a result of urban expansion”, Landscape and Ecological Engineering, 1(2), 169-176.

10. Xiaoma Li, Weiqi Zhou, Zhiyun Ouyang, Weihua Xu, Hua Zheng, 2012, “Spatial pattern of greenspace affects land surface temperature: evidence from the heavily urbanized Beijing metropolitan area, China”, Landscape Ecology, 27(6), 887-898.

11. NASA, Julian date Calendar/Calculator, https://www-air.larc.nasa.gov/tools/jday.htm

12. Phan Thanh Noi, Marin Kappas, Jan Degener, 2016, “Estimating Daily Maximum and Minimum Land Air Surface Temperature Using MODIS Land Surface Temperature Data and Ground Truth Data in Northern Vietnam“, remote sensing, 8(12).

13. A. Benali, A.C. Carvalho, J.P. Nunes, N. Carvalhas, A. Santos, 2012, “Estimating air surface temperature in Portugal usin MODIS LST data”, Remote sensing of Environment, 124: 108-121.

14. Liuhua Shi, Pengfei Liu, Itai Kloog, Mihye Lee, Anna Kosheleva, Joel Schwartz, 2016, “Estimating daily air temperature across the southeastern United States using high-resolution satellite data: A statistical modeling study”, Environmental Research, 146: 51-58.

15. Yuan Z. Yang, Wen H. Cai, Jian Yang, 2017, “Evaluation of MODIS Land Surface Temperature Data to Estimate Near-Surface Air Temperature in Northeast China”, Remote Sensing, 9:410.

16. W.G.M. Bastiaanssen, M.Menenti, R.A. Feddes, A.A.M. Holtslag, 1998, "A remote sensing surface energy balance algorithm for land(SEBAL). 1. Formulation", Journal of Hydrology, 212-213:198-212.

17. Lin Liu, Yuanzhi Zhang, 2011, "A study on the Urban Heat Island Phenomenon Using LANDSAT TM Thermal Infrared Data - In the Case of Seoul", Remote Sensing, 1535- 1552.

18. Nasime Janatian, Morteza Sadeghi, Seyed Hossein Sanaeinejad, Elham Bakhshian, Ali Farid, Seyed Majid Hasheminia, Sadegh Ghazanfari, 2016, "A statistical framework for estimating air temperature using MODIS land surface temperature data", INTERNATIONAL JOURNAL OF CLIMATOLOGY, 37(3), 1181-1194.

19. Phan Thanh Noi, Jan Degener, Martin Kappas, 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 sensing, 9(5), 398.

20. Linglin Zeng, Brian D. Wardlow,Tsegaye Tadesse, Jie Shan, Michael J. Hayes, Deren Li, Daxiang Xiang, 2015, "Estimation of daily air temperature based on MODIS land surface temperature products over the corn belt in the US", Remote sensing, 7(1), 951-970.

21. Christelle Vancutsem, Pietro, Ceccato, Tufa, Dinku, Stephen J.Connor, 2010, "Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa", Remote Sensing of Environment, 114(2), 449-465.

22. KlemenZakšek, Marion, chroedter-Homscheidt, 2009, "Parameterization of air temperature in high temporal and spatial resolution from a combination of the SEVIRI and MODIS instruments", ISPRS Journal of Photogrammetry and Remote Sensing, 64(4), 414-421.

23. J. Cristóbal, M. Ninyerola, X. Pons, 2008, "Modeling air temperature through a combination of remote sensing and GIS data", Journal of Geophysical Research:Atmospheres banner, 113(13), 16.

24. Simon Stisen, Inge Sandholt, Anette Nørgaard, Rasmus Fensholt, LarsEklundh, 2007, "Estimation of diurnal air temperature using MSG SEVIRI data in West Africa", Remote Sensing of Environment, 110(2), 262-274.

25. E. N. Florio, S. R. Lele, Y. Chi Chang, R. Sterner&G. E. Glass, 2003, "Integrating AVHRR satellite data and NOAA ground observations to predict surface air temperature: a statistical approach", International Journal of Remote Sensing, 25(15), 2979-2994.

26. M. P. Cresswell,A. P. Morse,M. C. Thomson &S. J. Connor, 2010, "Estimating surface air temperatures, from Meteosat land surface temperatures, using an empirical solar zenith angle model", International Journal of Remote Sensing, 20(6), 1125-1132.

27. QihaoWeng, DengshengLu, Jacquelyn Schubring, 2004, "Estimation of land surface temperature-vegetation abundance relationship for urban heat island studies", Remote Sensing of Environment, 89(4), 467-483.

28. Rangao Sun, Liding Chen, 2017, "Effects of green space dynamics on urban heat islands: Mitigation and diversification", Ecosystem Services, 23, 38-46.

29. Matthew Maimaitiyiming, Abduwasit Ghulam, Tashpolat Tiyip, , Filiberto Pla, Pedro Latorre-Carmona, Ümüt Halik, Mamat Sawut, Mario Caetano, 2014, "Effects of green space spatial pattern on land surface temperature: Implications for sustainable urban planning and climate change adaption", ISPRS Journal of Photogrammetry and Remote Sensing, 89, 59-66.