Estimation of daily maximum and minimum air temperature using satellite data in Mega city scale areas: Los Angeles

Cheolhee Yoo[1], Haemi Park, Daehyeon Han, Jungho Im[1]

Abstract: This research attempts to estimate urban maximum and minimum air temperature above 2 meters on the ground. For estimating the air temperature, this research uses the Machine learning approach, Random Forest with several Satellite data. The results show that the estimated Maximum temperature has 0.84 (R2) and 1.7 Celsius temperature (RMSE) and Minimum temperature has almost 0.8 (R2) and 1.2 Celsius temperature (RMSE) by 10-fold cross validation.

Keywords: Urban air temperature, LST, MODIS, Machine learning, RF

1. Introduction

Urban thermal environment is highly related to the various urban issues. Massive energy consumption and land-cover changes from urbanization affect the radiation budget and temperature conditions in urban areas. (Christen et al. 2004) Especially, air temperature is considered as important factor in many fields of urban environment studies including urban heat island (UHI) intensity quantifying (Schwarz et al. 2012), association with human mortality (Liu et al. 2011), amount of energy use (Ca et al. 1998), and air pollution monitoring. (Roberts 2004).

Urban air temperature shows distinctive properties such as spatial heterogeneity that comes from variation of built-up and vegetation densities, different types of buildings and infrastructure in urban areas. Therefore, different spatial patterns of urban air temperature are arranged in the close distances over the whole areas. For identifying spatial distribution of urban air temperature, the data from weather observations has been used for spatial interpolation. Near-surface air temperature is measured at 2 m above the ground by the meteorological station. However, the number of observations is lacking for analyzing the urban air temperature in the entire region. Insufficient in-situ sites can't ensure the representativeness of the complex city. Personal weather stations (PWS) are recently installed by city residents, however the crowdsourced in-situ weather data has the problem of inaccuracy.

Satellite provides remotely sensed data which has high spatiotemporal resolution for environmental research. Remote Sensing (RS) has been therefore used for estimating air temperature from satellite data such as Land surface temperature in various techniques. Nevertheless, most of the preceding researches study on wide regions which areas are greater than 100,000km^2. (Liu et al. 2011), (Xu et al. 2012). In other words, few researches focus on the urban areas for air temperature estimation which sizes are less than 10,000 km^2.

One of the noticeable studies estimating urban air temperature using RS is done by Ho et al. (2014). The land surface temperature with some other variables, derived from Landsat satellite, are processed for estimating the maximum air temperature of the Vancouver,

1. Ulsan National Institute of Science and Technology,
Banyeon-ri, Eonyang-eup, Ulju-gun, Ulsan, Korea
Phone : +82-10-9238-8065
E-mail : yoclhe@unist.ac.kr, ersgis@unist.ac.kr

metropolis of Canada. The results (R^2 = 0.34 and RMSE = 2.31 ℃) show relatively lower value than estimation on larger area, because of complexities of urban environments.

In this study, Using MODIS (Moderate Resolution Imaging Spectroradiometer) data with high temporal resolution, urban maximum and minimum air temperatures are estimated in the city of Los Angeles.

2. Study Area and Data

2.1 Study Area

The study is conducted on the Los Angeles which is the second largest urban city in the United States. The area of the city is 1,302 km^2. The city broadens for 71 km longitudinally and 47 km latitudinally with the average elevation is 86.85 meters.

Los Angeles is located in a wide coastal basin and the three sides are surrounded by huge mountains reaching to the 3,000m elevation. Especially, The Santa Monica Mountains extends from inner town to the Pacific Ocean and the mountain divides the Los Angeles into two parts: San Fernando Valley (upper) and Los Angeles Basin (bottom) which have different climate patterns.

Summer is more significant than other seasons for studying urban temperature phenomenon. (Zhou et al. 2014). For that reason, chosen periods of this study are the July-August in 2006 to 2015.

2.2 Data

1x1 km spatial resolution - Land surface temperature (LST) is got from MODIS (Moderate Resolution Imaging Spectroradiometer). There are two types of LST: MOD11A1 from Terra and MYD11A1 from Aqua spacecraft. 16 days MODIS MOD13A2 is used for NDVI (Normal Distribution Vegetation Index).

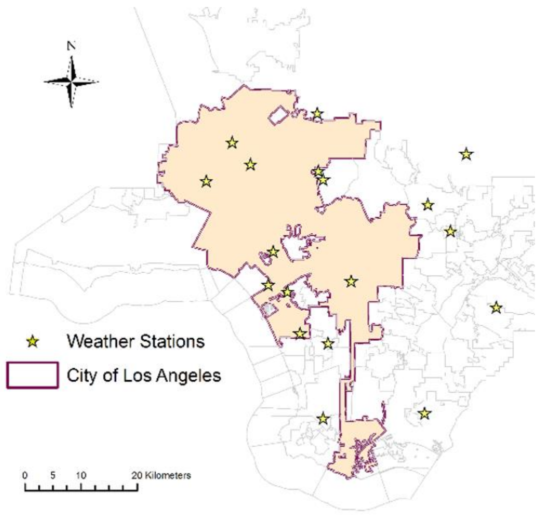

Figure 1. Study area and weather stations

DEM (Digital Elevation Model) data comes from Shuttle Radar Topography Mission (SRTM) and Impervious surface area data can be got from Los Angeles County GIS Data Portal (http://egis3.lacounty.gov/dataportal/gis-data-viewers/).

In-situ maximum and minimum air temperature data are downloaded from NOAA climate data center (http://www.ncdc.noaa.gov/). The 18 weather stations in Los Angeles are used for reference data (Figure1). Due to the few observations within the city, some stations are selected from near the Los Angeles city boundary.

3. Methodology

Multi temporal Land Surface temperature

Terra crosses the equator at 10:30 am/pm and Aqua crosses at 1:30am/pm. As a result, MODIS provides 4 times land surface data in a day. This study uses 8 times (the day and Pre-day) land surface data for estimation of daily air temperature. (Figure2)

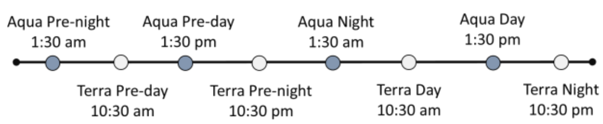

Figure 2. MODIS (Aqua/Terra) product captured time

Auxiliary variables

In addition to the Land surface temperature, 8 auxiliary variables are used for estimation: NDVI, DEM, Latitude, Longitude, Solar radiation, Impervious surface area (%), Distance from Land surface hotspot and Aspect.

NDVI is the index of vegetation "greenness". 16 days average NDVI is provided by MODIS product.

DEM, Latitude and Longitude are the geographic information, correlated with the dynamic temperature characteristic.

Solar radiation is the energy received from the sun and it causes many environmental processes. The incoming solar radiation is calculated by using the ArcGIS tools with DEM data.

There are lots of impervious surface areas in the city and have important role for urban heat island effect.

Land surface hot-spot is found by searching the highest pixel value among that of 10-years average Night time surface temperature (MYD11A1) values in Los Angeles. Distance from the spot is used as variable for estimation.

Aspect is simply processed by ArcGIS tool. However, the value of 360° is same as 0°. For solving that problem, this study transforms the aspect using the equation (1) below, made by Beers et al. (1966).

$$\textit{Transformed aspect} = \cos(45° - \text{Aspect}) + 1 \quad \textbf{(1)}$$

All of the above variables are resampled by 1x1 km pixel sizes to match with the spatial resolution of MODIS data.

Random Forest (RF).

Among the various methods for air temperature estimation, this study applies the statistical model, "Random Forest" which is one of the Machine Learning approaches. Ho et al. (2014) suggested this technique as an urban air temperature estimation.

Figure 3. Random Forest process

RF is an ensemble learning method for classification, regression and other tasks by constructing the multitude of decision trees and yielding the class that is the mean prediction of the individual trees (Figure3).

8-time series land surface temperatures and auxiliary variables above-mentioned are put into the RF process.

4. Result and Discussion

Air temperature estimation

By 10-fold cross validation, The R^2 of maximum temperature has higher performance than minimum temperature, however RMSE (eq 2) values are quite the reverse.

NRMSE (eq 2) of two temperatures are almost similar, because the range of maximum temperature (16 - 48℃) is wider than that of minimum temperature (7 – 28 ℃).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(pred_i - obs_i)^2}{n}} \qquad NRMSE = \frac{RMSE}{X_{max} - X_{min}} \times 100\% \quad \textbf{(2)}$$

Variable importance

RF approach provides importance value of the variables (Table2). The maximum temperature of Los Angeles usually occurs at near 1pm, which corresponds to the variable importance that Terra day (10:30am) land surface temperature has the highest value.

Table 1. 10-fold cross validation results

	R^2	RMSE	NRMSE
Max_Temp	0.84	1.74 ℃	5.4 %
Min_Temp	0.80	1.16 ℃	5.6 %

Table 2. Variable Importance

Maximum Temperature

	%IncMSE
Terra_day	70.37967
Solar radiation	64.79402
Aqua_night	63.8386
Terra_night	57.32555
Aqua_day	53.48895
NDVI	43.43176
DEM	39.71741
Aqua_pre.day	38.77145
impercent	37.28824
Terra_pre.day	36.62956
Terra_pre.night	34.09624
Hotspot	32.09887
Aqua_pre.night	30.98787
LON	30.4723
aspect	25.45417
LAT	19.52101

Minimum Temperature

	%IncMSE
Aqua_night	102.6819
Terra_day	48.96459
Terra_night	46.4443
Aqua_day	44.56118
Aqua_pre.day	43.54777
Aqua_pre.night	41.342
LON	41.16532
Hotspot	39.55623
aspect	38.48981
impercent	38.03819
Terra_pre.night	36.25261
Solar radiation	36.11951
Terra_pre.day	34.52336
NDVI	32.37316
LAT	31.45518
DEM	29.24318

By the same reason, the minimum temperature of Los Angeles (near 1pm) matches the highest variable importance with Aqua night (1:30am) land surface temperature. (Table2) shows multi temporal land surface temperatures and other variables affect the maximum and minimum temperature in different order.

5. Conclusion

Though urban areas have complex properties, RF estimates both maximum and minimum urban air temperature using various temporal MODIS land surface temperature and other auxiliary variables.

It is possible to develop integrated monitoring and predicting system of temperature with observed temperature in the future.

6. Reference

Beers, T. W., Dress, P. E., & Wensel, L. C. (1966). Notes and observations: aspect transformation in site productivity research. Journal of Forestry, 64(10), 691-692.

Ca, V. T., Asaeda, T., & Abu, E. M. (1998). Reductions in air conditioning energy caused by a nearby park. Energy and Buildings, 29(1), 83-92.

Christen, A., & Vogt, R. (2004). Energy and radiation balance of a central European city. International Journal of Climatology, 24(11), 1395-1421.

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.

Liu, L., Breitner, S., Pan, X., Franck, U., Leitte, A. M., Wiedensohler, A., ... & Schneider, A. (2011). Associations between air temperature and cardio-respiratory mortality in the urban area of Beijing, China: a time-series analysis.Environmental Health, 10(1), 1.

Roberts, S. (2004). Interactions between particulate air pollution and temperature in air pollution mortality time series studies. Environmental Research, 96(3), 328-337.

Schwarz, N., Schlink, U., Franck, U., & Großmann, K. (2012). Relationship of land surface and air temperatures and its implications for quantifying urban heat island indicators—an application for the city of Leipzig (Germany).Ecological Indicators, 18, 693-704.

Xu, Y., Qin, Z., & Shen, Y. (2012). Study on the estimation of near-surface air temperature from MODIS data by statistical methods. International journal of remote sensing, 33(24), 7629-7643.

Zhou, W., Qian, Y., Li, X., Li, W., & Han, L. (2014). Relationships between land cover and the surface urban heat island: seasonal variability and effects of spatial and thematic resolution of land cover data on predicting land surface temperatures. Landscape ecology, 29(1), 153-167.