International Institute of Information Technology Bangalore

Reading Elective

Project Report

Remote Sensing Analysis: Estimating Land Surface Temperature (LST) and Modeling with Various Indices

---

Sagar - IMT2020134

Dheeraj Sairam - IMT2020533

Chakradhar - IMT2020021

(under guidance of Rahisha ma'am)

1 Introduction

Land Surface Temperature (LST) is a significant parameter in climate studies, weather forecasting, and environmental monitoring as it measures the temperature of the Earth's surface from space. It plays a vital role in various fields such as urban heat island studies, forest fire management, and water resource management. LST can be calculated using thermal infrared remote sensing data, which involves converting thermal infrared radiation into temperature using empirical algorithms. Various factors such as solar radiation, emissivity, and atmospheric conditions can affect LST measurements. The temperature difference between different land surfaces can be calculated using LST data, which provides valuable information on heat distribution. It helps identify areas experiencing heat stress and aids in agriculture, urban planning, and emergency response.

Landsat 8, MODIS on Aqua and Terra, ASTER on Terra, and VIIRS on Suomi-NPP and NOAA-20 are satellites with thermal infrared bands that can be used to calculate LST. We used Landsat 8's Thermal Infrared Sensor (TIRS) for our study. Google Earth Engine (GEE) was an essential tool for calculating LST as it provided easy access to large amounts of satellite imagery data, making it faster and more efficient than traditional methods. Furthermore, GEE produced more precise and detailed results, which is useful in various applications. It has an extensive collection of remote sensing data from multiple satellites, making it simpler to access and process satellite data for LST calculation. GEE is also user-friendly and interactive, allowing for easy code writing and testing without the need for additional software.

Land Surface Temperature (LST) is a crucial variable that can help in the detection of land use and land cover changes such as urbanization and desertification. In this study, we aimed to evaluate the potential of Landsat-8 Thermal Infrared Sensor (TIRS) bands 10 and 11 in calculating LST in a watershed with varied land surface characteristics. We utilized established mathematical algorithms and equations to map LST using Landsat-8 band 10 with the help of Google Earth Engine (GEE). In addition to this, we calculated various indices to establish a relationship between LST and these indices. Based on this relationship, we developed a model to predict LST values at different locations and times.

The use of GEE made it easier to access and analyze large amounts of satellite imagery data. Our analysis of the relationship between LST and indices revealed that LST is sensitive to vegetation and soil moisture, which can be beneficial in detecting changes in land use and land cover. Our study focused on a watershed with varied land surface characteristics, which makes the findings applicable to regions with similar characteristics. The model developed in this study using GEE for LST calculation has the potential to be a valuable tool for predicting LST values at different locations and times. This is particularly useful for applications such as agriculture, urban planning, and climate studies. Our findings highlight the effectiveness of using GEE for LST calculation, demonstrating its power and flexibility for researchers and practitioners in various fields.

2 Materials and Methods

Our study aimed to analyze the land surface temperature (LST) in various cities across India by selecting numerous grid points covering different regions of the country. We conducted a detailed analysis of each grid point by calculating different indices that represent various environmental and climatic conditions. We used Landsat-8 Thermal Infrared Sensor (TIRS) band 10 to calculate the

LST values of these grids and establish a relationship between the various indices and LST values.

To collect data for our analysis, we selected the months of March, April, and May in the year 2019. To derive the indices, we utilized the spectral bands of Landsat 8, specifically the visible, near-infrared, shortwave infrared, and thermal infrared bands. We calculated the Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Soil-Adjusted Vegetation Index (SAVI), Modified Normalized Difference Water Index (MNDWI), Normalized Difference Moisture Index (NDMI), Bare Soil Index (BSI), and Normalized Difference Built-up Index (NDBI) using established formulas commonly used in remote sensing studies to analyze land surface characteristics.

Each index represents different characteristics of the land surface. For example, NDVI represents the density and health of vegetation, with higher values indicating healthier and denser vegetation. EVI is also a vegetation index, but it is more sensitive to canopy structure and has a higher sensitivity to changes in vegetation than NDVI. SAVI is similar to NDVI, but it accounts for the soil background in the calculation. MNDWI is an index that can be used to detect the presence of water bodies. NDMI represents the water content in vegetation and soil. BSI can be used to detect areas with little to no vegetation cover, such as barren land or bare soil. NDBI is used to detect built-up areas, such as urban or industrial zones.

By analyzing the relationship between these indices and LST values across different grid points, we can gain valuable insights into land use/land cover changes, urban heat islands, and other environmental factors that affect LST. Our study involved an extensive and comprehensive analysis of various grid points across India, providing a broader and more inclusive picture of LST trends in the country. The developed model using GEE can be useful in predicting LST values at different locations and times, which can be crucial for various applications including agriculture, urban planning, and climate studies.

using Band 4 and Band 5 of Landsat 8, calculate the Normalised Difference Vegetation Index(NDVI) using the below formulae:

$$NDVI = \frac{(Band5 - Band4)}{(Band5 + Band4)}$$

using Band 6 and Band 5 of Landsat 8, calculate the Normalized Difference Built-up Index(NDBI) using the below formulae:

$$NDBI = \frac{(Band6 - Band5)}{(Band6 + Band5)}$$

Using Band 3 and Band 5 of Landsat 8, calculate the Normalized Difference Water Index(NDWI) using the below formulae:

$$NDWI = \frac{(Band3 - Band5)}{(Band3 + Band5)}$$

using Band 2, Band 4, and Band 5 of Landsat 8, calculate Enhanced Vegetation Index(EVI) using the below formulae:

$$EVI = \frac{(2.5 * (Band5 - Band4))}{((Band5 + 6 * Band4 - 7.5 * Band2) + 1)}$$

using Band 5 and Band 6 of Landsat 8, calculate the Normalized Difference in Moisture Index(NDMI) using the below formulae:

$$NDMI = \frac{(Band5 - Band6)}{(Band5 + Band6)}$$

using Band 3 and Band 6 of Landsat 8, calculate the Modified Normalized Difference Water Index (MNDWI)using the below formulae:

$$MNDWI = \frac{(band3–Band6)}{(Band3 + Band6)}$$

using Band 4 and Band 5 of Landsat 8, calculate Soil Adjusted Vegetation Index (SAVI) using the below formulae:

$$SAVI = \frac{(1.5 * (band5–Band4))}{(Band5 + Band4 + 0.5)}$$

using Band 2, Band 4, Band 5, and Band 6 of Landsat 8, calculate Bare Soil Index (BSI) using the below formulae:

$$BSI = \frac{((Band4 + Band6) - (Band2 + Band5))}{((Band4 + Band6) + (Band2 + Band5))}$$

The proportion of vegetation and emissivity are two important parameters used in remote sensing analysis. To calculate emissivity, the proportion of vegetation needs to be determined first. And to calculate the proportion of vegetation, the normalized difference vegetation index (NDVI) needs to be calculated.

Once the NDVI is calculated, the proportion of vegetation can be calculated using the following formula:

$$P_v = (NDVI - NDVI_{min})/(NDVI_{max} - NDVI_{min})$$

where NDVImin and NDVImax represent the minimum and maximum values of NDVI, respectively. The resulting value of the proportion of vegetation ranges from 0 to 1, with higher values indicating greater vegetation density.

Finally, the emissivity can be calculated using the following formula:

$$Emissivity = 0.004 * P_v + 0.986$$

where PV represents the proportion of vegetation. The resulting value of emissivity ranges from 0 to 1, with higher values indicating greater emissivity.

The process of retrieving the Land Surface Temperature (LST) from LANDSAT 8 data was carried out in the Google Earth Engine (GEE) platform. The process involved several steps that are briefly described below.

1. The first step was to convert the digital number (DN) values of the Thermal Infrared (TIR) bands to at-sensor radiance. This conversion was done using the conversion coefficients provided by the instrument calibration. At-sensor radiance is the radiance measured by the sensor on the satellite and is affected by the distance from the Earth, solar angle, and atmospheric effects.

   $$L\lambda = M_L * Q_{cal} + A_L–O_i$$

   where the band-specific multiplicative rescaling factor is denoted by M_L, while Q_{cal} represents the image of band 10. The band-specific additive rescaling factor is represented by A_L, and O_i is the correction factor for band 10.

2. The next step involved converting the at-sensor radiance to at-surface radiance by applying atmospheric correction methods. This step accounted for the absorption and scattering of the radiation as it passed through the atmosphere. The atmospheric correction method used in this study was the Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) algorithm. FLAASH is a physics-based algorithm that corrects atmospheric effects by modeling the radiative transfer of light through the atmosphere.

$$B_T = K_2/ln[(K_1/L_\lambda) + 1]–273.15$$

   where the effective temperature at the satellite in absolute units is represented as B_T. The conversion constants for thermal radiation, K_1 and K_2, are specific to each band and are provided in the metadata. To obtain the temperature in Celsius, the radiant temperature is adjusted by subtracting the absolute zero (-273.15°C).

3. Finally, the brightness temperature was converted to LST using a temperature-emissivity separation method, which separates the effects of temperature and emissivity on the observed brightness temperature. Emissivity is the ability of a surface to emit thermal radiation and is dependent on the surface material.

$$T = B_T/(1 + \lambda(B_T/\rho)ln\epsilon\lambda)$$

   where the LST (in Celsius) is represented by T_s, while the at-sensor brightness temperature (in Celsius) is represented by B_T. The wavelength of emitted radiance is denoted by λ (which is equal to 11.5 µm), while $\epsilon\lambda$ represents the emissivity calculated from the proportion of vegetation (P_v).

Using Google Earth Engine, you collected data for each grid point in your study area in India. For each grid point, we calculated various indices such as BSI, EVI, MNDWI, NDBI, NDMI, NDVI, SAVI, NDWI, and NDBAI. These indices capture different aspects of the vegetation, water content, built-up areas, and other relevant features in the region. Additionally, we also obtained Land Surface Temperature (LST) values for each grid point. After calculating the indices and LST values, you exported this data as a CSV file. Each row in the CSV file represents a grid point, and the columns contain the calculated indices values along with the corresponding LST value.

With the exported CSV file containing the indices and LST values, you will proceed to find the best model for calculating the LST. This involves training and evaluating various regression models using the collected data. We will divide the dataset into a training set and a test set. The training set will be used to train the models, and the test set will be used to assess their performance. Each model will be trained using the input indices values (BSI, EVI, MNDWI, NDBI, NDMI, NDVI, SAVI, NDWI, NDBAI) as the independent variables and the corresponding LST values as the dependent variable.

After training the models, you will evaluate their performance using evaluation metrics such as R2, MAE, MSE, and others. These metrics will help assess how well each model fits the data and how accurately it predicts LST values on the test dataset. By comparing the performance of different models using the evaluation metrics, you can determine which model works best for calculating LST values. The model with higher R2, lower MAE and MSE values, and good overall performance will be selected as the best model.

In addition to training the models and evaluating their performance, we will also analyze how each method handles the task of estimating LST values. This will provide insights into the strengths and weaknesses of each model and help us understand which approach works best for our specific dataset.

By employing a range of models and evaluation metrics, we aim to develop an accurate and reliable LST estimation method for the cities in India.

3 Analysis and Results

In order to determine the most appropriate model for calculating Land Surface Temperature (LST), we conducted a comprehensive analysis using the Google Earth Engine platform. Our analysis involved calculating several indices including BSI, EVI, MNDWI, NDBI, NDMI, NDVI, SAVI, NDWI, and NDBAI. These indices provide valuable information about vegetation health, water content, urbanization, and other important factors related to LST.

To perform the analysis, we considered data from a specific time period, ranging from January 3, 2019, to May 31, 2019. For each grid within the dataset, we computed the average values of these indices. This allowed us to obtain a representative measure for each index across the study area.

Additionally, we also calculated the LST values using the available data. Once we completed the calculations, we exported the results as a CSV file. This file contains the averaged, minimum, and maximum index values and corresponding LST values for each grid in the dataset.

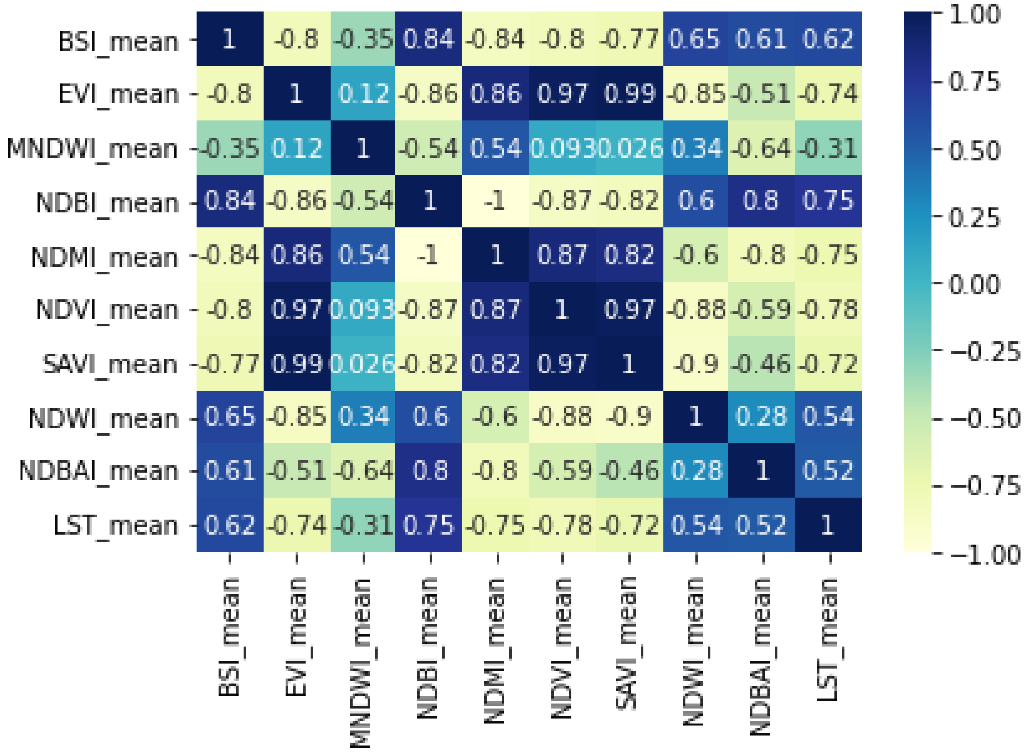
The exported CSV file serves as input for further analysis aimed at identifying the best model for accurately estimating LST. By examining the relationships between the indices and LST, we can determine which model provides the most reliable and accurate predictions of LST for the given dataset and time period. This information is valuable for various scientific, environmental, and practical purposes.

To begin the analysis of the CSV file, our first step is to generate a correlation matrix. This matrix provides valuable insights into the relationships between different variables in the dataset. By examining the correlations, we can identify the independent variables that have a significant impact on the dependent variable (LST in this case).

The correlation matrix showcases the pairwise correlations between all variables included in the CSV file. It presents a comprehensive overview of how each variable is related to others, allowing us to assess the strength and direction of these relationships. Positive correlations indicate that variables tend to increase or decrease together, while negative correlations imply an inverse relationship.

Using the information derived from the correlation matrix, we can make informed decisions regarding the selection of independent variables for constructing the LST model. Variables with strong correlations to LST can be considered potential predictors, as they are likely to have a substantial influence on the LST values.

By carefully examining the correlation matrix, we can identify which independent variables exhibit a strong correlation with LST, and therefore, are the most suitable candidates for inclusion in our model.



Upon analyzing the correlation matrix derived from the CSV file, it becomes evident that Land Surface Temperature (LST) exhibits a significant correlation with most of the indices, indicating a strong relationship. However, one index, MNDWI, does not show a notable correlation with LST.

The positive correlations observed between LST and various indices imply that these indices can serve as reliable predictors for estimating LST values. The positive correlation suggests that as the values of these indices increase, the LST values also tend to increase, and vice versa. This relationship highlights the influence of factors such as vegetation health, water content, and urbanization on surface temperature.

On the other hand, the lack of a substantial correlation between LST and MNDWI indicates that changes in MNDWI values do not correspond closely with changes in LST. This finding suggests that the water content captured by the MNDWI index may not be a significant driver of LST variations during the specified time period and dataset.

Considering the correlations between LST and the other indices, we can select the independent variables that exhibit positive correlations as potential predictors for building the LST model. These variables can provide valuable information for accurately estimating LST and capturing the underlying patterns and dynamics of surface temperature.

We tried different models for predicting LST, the models are mentioned below:

1. Linear Regression

2. Polynomial Regression

3. Random Forest
4. Support Vector regressor
5. Ridge regression
6. Lasso Regression
7. Kernel Ridge
8. Gradient Boost
9. XG Boost
10. Decision trees
11. MLP Regressor

To optimize the performance of the models for estimating Land Surface Temperature (LST), we conducted a Grid Search Cross-Validation (CV) process to identify the best hyperparameters. The Grid Search CV technique systematically explores various combinations of hyperparameters to find the optimal configuration that yields the highest model accuracy.

After identifying the best hyperparameters for each model, we trained the models using 80 percent of the available data. The remaining 20 percent of the data was reserved for testing the trained models. This approach ensures an unbiased evaluation of the model's performance on unseen data, enabling us to assess its generalization capabilities.

By utilizing the best hyperparameters and training the models on the majority of the data, we obtained the following results, as presented in the table:

Final Results

Model	MAE	MSE	R^2 score	Training R^2 Score	Best Features
Linear Regression	1.731819697942487	6.420445821520478	0.7720133710591022	0.798133218858156	'BSI_mean', 'MNDWI_mean', 'NDVI_mean', 'SAVI_mean', 'NDBAI_mean'
Polynomial Regression	144.92649694963998	1103183.6890271576	0.8586027966802722	1.0	'MNDWI_mean', 'NDBI_mean', 'NDMI_mean', 'NDVI_mean', 'SAVI_mean', 'NDWI_mean', 'NDBAI_mean'
Random Forest Regressor	1.9152262749551359	9.027306164048472	0.7188322358716752	0.9667470908916171	'EVI_mean', 'MNDWI_mean', 'NDVI_mean', 'SAVI_mean', 'NDWI_mean'
Support Vector Regressor	2.0022890137324936	8.792066629250447	0.7187574079451111	0.7439661676293376	'BSI_mean', 'NDVI_mean', 'SAVI_mean', 'NDWI_mean'
Ridge Regression	1.7341918862092043	6.459242117571475	0.7704733248895206	0.7980626314774321	'BSI_mean', 'MNDWI_mean', 'NDVI_mean', 'SAVI_mean', 'NDBAI_mean'
Lasso Regression	2.012480088740237	10.250762771609994	0.6368234559007442	0.7258865650066841	'MNDWI_mean', 'NDBI_mean', 'NDVI_mean', 'NDWI_mean', 'NDBAI_mean'
Kernel Ridge	2.145950234438521	7.749060745836938	0.7909054744528813	0.6885771691795833	'MNDWI_mean', 'NDBI_mean', 'NDMI_mean', 'NDVI_mean', 'SAVI_mean'
Gradient Boost	1.9954853149277236	8.298562003295796	0.7223135311626754	0.9412423497699499	'BSI_mean', 'EVI_mean', 'MNDWI_mean', 'NDMI_mean', 'NDVI_mean'
XG Boost	2.229420693182632	10.159158871125802	0.7281324795044264	0.886475456715604	'MNDWI_mean', 'NDBI_mean', 'SAVI_mean'
Decision Tree Regressor	2.5673934658946633	15.65144895540603	0.6198740508036413	1.0	'NDVI_mean', 'NDWI_mean'
MLP Regressor	1.662999020900779	5.977707022128167	0.8525260247677593	0.9291038355597872	'BSI_mean', 'EVI_mean', 'MNDWI_mean', 'NDMI_mean', 'NDVI_mean', 'SAVI_mean', 'NDBAI_mean'

In this analysis, we aimed to determine the optimal combination of features from a given data set for predicting a target variable. The approach involved systematically evaluating different combinations of columns and training different models for each combination. Performance metrics such as the coefficient of determination (R2 score), mean squared error (MSE) and mean absolute error (MAE) were computed on a separate validation set to assess the models' predictive accuracy. The combination of features yielding the highest R2 score was considered the most effective. This method allowed us to identify the subset of features that exerted the strongest influence on predicting the target variable. The findings provide insights into the key features and their impact on the model's predictive performance.