# Statistical Relationship between Satellite observed Land Surface Temperature and in-situ measured Surface Air Temperature over the Indian Region: An Exploratory Study

Utkarsh Tyagi, Ujjwal K. Gupta
Space Applications Centre, Indian Space Research Organisation
Email: utkarsh@sac.isro.gov.in

(Received on 16 Jan 2024; in final form 5 June 2024)

***DOI***: *https://doi.org/10.58825/jog.2024.18.2.142*

**Abstract:** This paper presents an investigation of spatio-temporal estimation of daily minimum (Tmin) and maximum (Tmax) surface air temperature using satellite (INSAT and MODIS) derived Land surface temperature and in-situ observational data over the Indian region. Least Absolute Shrinkage and Selection Operator (LASSO) regression technique is used to identify influencing neighboring stations. To capture spatial and temporal variability of surface air temperature of a particular IMD station, Linear Regression and Auto-Regressive Integrated Moving Average, which is known as ARIMA, are used respectively. These models are statistically ensembled using stack generalization. Obtained station-by-station relationship is validated on an independent test data using Root Mean Squared Error (RMSE) to check the validity of the model under consideration. Results show ensembled model outperform (has lowest RMSE) the traditional methods for prediction of Tmin and Tmax with RMSE within a range of [1, 2] range for most of the regions and seasons.

**Keywords**: Land Surface Temperature, Weather stations, ARIMA, LR, Ensembled Machine Learning

## 1. Introduction

Surface Air Temperature (SAT), which is measured at 2m from land surface, is a crucial biophysical parameter that plays a critical role in vegetation distributions, phenology, and growth (Benavides et al. 2007; Stahl et al. 2006). The maximum daytime temperature is related to wildfire on hot and sunny days (Aldersley et al. 2011; Litschert et al. 2012). An increase in minimum night temperature slows down plant respiration rates and reduces overall crop yield (Hatfield et al. 2011). Therefore, detailed study and analysis of the spatial variability of air temperature is crucial.

Air Temperature is usually measured through in-situ sensors and available for point locations. These sensors at weather stations are typically limited in number and suffer from data losses, network and quality issues. Consequently, data availability and sparsity are major issues to tackle. Climatologists describe seasonal and geographic variations in air temperature with interpolation, gridding and/or averaging techniques like ERA-Interim or ERA5 (Dee et al. 2011), CRUTEM4 that is gridded at $5^{O}$ spatial resolution (Jones et al. 2012) and other bias adjusted datasets. However, there are relatively very few actual air temperature points.

In contrast to weather stations, Satellite products provide Land surface temperature (LST). Satellite observations are processed to derive LST that provides good estimates over spatial grids. It can serve the purpose of filling the gaps and estimating air temperature measurements. There have been multiple attempts to establish a relation between air temperature and satellite derived LST for Africa (Vancutsem et al. 2010); for Europe (Benali et al. 2012); for USA (Oyler et al. 2016); for Asia (Gupta et al. 2018). Multiple important factors affect the thermal environment near the ground and hence SAT (Kawashima et al 2000). Over the years, multiple approaches have been proposed to derive SAT based on remotely sensed data. These approaches include using LST (Benali et al. 2012), temperature-vegetation index (Czajkowski et al. 1997; Goward et al. 1994; Nieto et al. 2011; Prihodko and Goward 1997; Wloczyk et al. 2011; Zhu, Lű, and Jia 2013), statistical approaches (Benali et al. 2012; Florio et al. 2004; Jang, Viau, and Anctil 2004; Kim and Han 2013; Xu, Qin, and Shen 2012; Zhang et al. 2011), and surface energy balance approaches (Pape and Löffler 2004; Sun et al. 2005).

The objective of this paper is to establish a relation between Indian Meteorological Department (IMD) SAT values and satellite derived LST to build a SAT estimation model using proposed ensemble machine learning approaches in spatial and temporal domains.

## 2. Study Area

India is a tropical country extending from the Himalayas in the north to the Indian Ocean in the south. The mainland extends between latitudes 8°4'N and 37°6'N and longitudes 68°7'E and 97°25'E. while the tropic of cancer (23.5°N) bisecting the region into almost half. The physical features of the region are diverse and divided into (1) The Himalayan Mountains (2) The Northern Plains (3) The Peninsular Plateau (4) The Indian Desert (5) The Coastal Plains, and (6) The Islands.

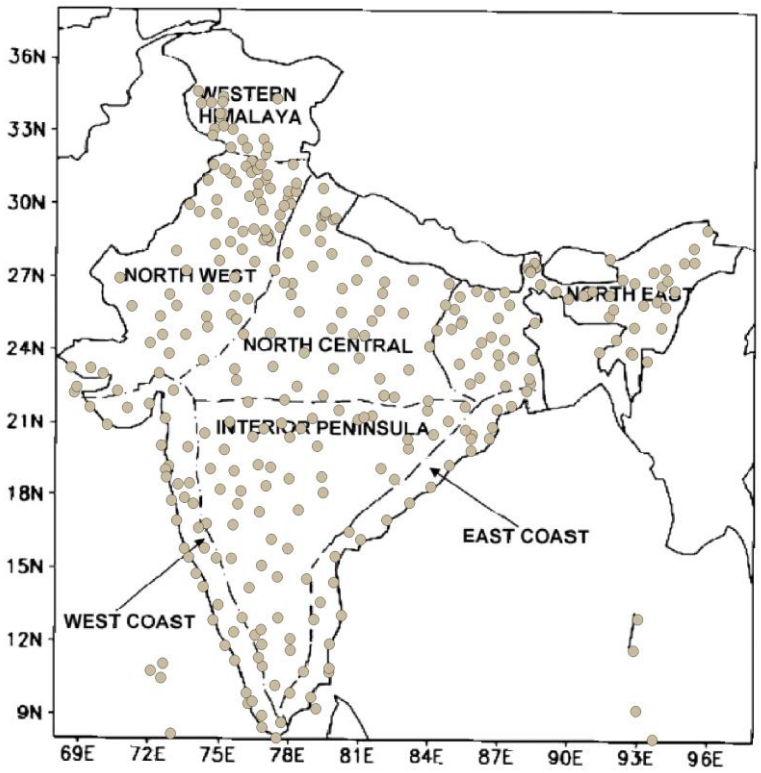


**Figure 1. Distribution of 332 IMD stations in seven homogeneous regions of India (Source: IITM)**

The climate of the country is divided into four seasons viz. winter (January and February), summer (March, April and May), monsoon (rainy) season (June to September), and a post-monsoon period (October to December) by IMD (IMD Met Glossary). These seasons are referred as S1, S2, S3 and S4 respectively in this paper. Indian Institute of Tropical Meteorology (IITM), Pune divided India into seven homogeneous regions namely Western Himalaya (WH), North West (NW), North Central (NC), North East (NE), Indian Peninsula (IP), West Coast (WC) and East Coast (EC). These regions are as shown in Figure 1.

## 3. Datasets

**3.1. Surface Air Temperature (SAT):** Indian Meteorological Department (IMD) provides surface air temperature measured two meters above the ground from 332 ground stations daily. The geographic distribution of these stations is non-uniform throughout India. The daily maximum and minimum values of air temperature are used in this study for the duration of March 2017 to February 2021. The completeness and availability of the record varies with stations and hence techniques are used to avoid stations with a limited set of observations. Distribution of these stations are as depicted in Figure 1.

**3.2. Land Surface Temperature (LST):** Satellite based land surface temperatures are taken from MODIS and INSAT-3D Imager. MODIS sensors are mounted on two polar-orbiting platforms Terra and Aqua that provide LST twice in a day each. MODIS sensors have 36 visible and infrared channels which operate between the spectral ranges of 0.4 µm and 14.4 µm (**Wan et al., 2015**). The equatorial crossing local times of the Terra satellite are about 10:30 and 22:30, whereas these are 01:30 and 13:30 for the Aqua satellite. As LST observations around 13:30 and 01:30 local time are closer to the daily maximum LST and daily minimum LST of a given location respectively (**Sharifnezhadazizi et al., 2019**), LST observations (MYD11C1) from the Aqua MODIS were used in this study. MYD11C1 product provides daytime and nighttime LST at a spatial resolution of 5 km.

INSAT-3D is Indian Earth observation satellite launched by ISRO on July 26, 2013. Its Imager payload is used to derive LST (**Pandya et al 2014**). This half hourly L2B product is available starting from midnight at a resolution of 4 km. These 48 acquisitions are used to compute daily max and min LST over the Indian region.

## 4. Methodology

In this paper, a study is carried out for estimation of SAT from INSAT and MODIS LST products incorporating spatial and temporal analysis for the Indian region. Input is maximum and minimum surface air temperatures from weather stations distributed non-uniformly across the country and daytime and nighttime LSTs from MODIS and INSAT-3D Imager.

### 4.1. Data Preparation

**Surface Air Temperature Data from IMD**

The daily minimum and maximum surface air temperature (SATmin and SATmax respectively) of an IMD station is first checked for completeness. Sometimes, the SAT values for some of the stations are missed. To train a model, it is crucial to have a good amount of data with fair completeness. To ensure completeness the study has selected only those IMD stations which have at least 70% data availability in the study duration.

**Land Surface Temperature Data from INSAT-3D**
The INSAT-3D LST data acquired from Meteorological & Oceanographic Data Archival Centre (MOSDAC) comes in both hierarchical data format (HDF) and georeferenced tagged image format (GeoTiff) format. The GeoTiff formatted data is used in this paper to compute daily LSTmin and LSTmax for all IMD stations, which are qualified in completeness check, as discussed above. IMD station locations are collocated with the satellite data pixel in which they are falling in. Broadly, collocation criteria has two categories: (1) those that keep the data on their original grid and select the closest matches and (2) those that use interpolation and aggregation techniques (e.g., regridding, resampling and kriging) to bring both datasets onto the same grid. Later approach is more suitable when intermittent and highly variable parameters are considered (e.g., convective precipitation) (**Loew et al 201)**. Former approach is chosen in this study because of less regional variability of the geophysical field (SAT). Studies have been carried out for the characterization of the spatial representativeness of in situ, point-like observations for the validation of satellite data products (**Roman et al 2009**). This study uses point observations as surrogate indicator of average LST. Limited weather stations data may yield poorer interpolated SAT values that may affect the estimation model. Limitation of this approach is that it may behave poorly with extreme dynamic variations of land use (local traffic conditions etc.)

**Land Surface Temperature Data from MODIS**
The MODIS data (MYD11C1) is downloaded from an automated program for study period in GeoTiff format. The data is processed further and daily LSTmin and LSTmax is extracted for collocated locations with IMD stations, which are qualified in completeness criteria same as INSAT-3D.

All the data during the study period is divided into training and testing data such that data from March 2017 to February 2020 would be training data and from March 2020 to February 2021 would be testing data.

**4.2 LASSO regression based Neighbouring Stations Selection**
Once data is prepared, set of neighbouring stations that affect the temperature of an IMD station under study is identified based on LASSO regression on daily SAT values of all stations lying within the range of a given distance. This will tell the neighbouring stations and optimal distance for a station affecting its SAT value. Therefore, LASSO regression is applied at distance of every 20 km for all stations one at a time.

*4.2.1. LASSO regression*
The Least Absolute Shrinkage and Selection Operator (LASSO) regression is a linear statistical method which aims to minimize residual sum of squares subject to the sum of the absolute value of the coefficients being less than constant (**Tibshirani, 1995**). Due to this constraint, it tends to produce some of the coefficients as zero and explains the data based on the remaining coefficients. Thus, it produces models that work like subset selection. The LASSO estimator can be given by (Fonti and Belitser, 2017):

$$\hat{\beta}(\lambda) = \text{argmin}\,(\frac{\|Y - X\beta\|_2^2}{n} + \lambda\|\beta\|_1)$$

where $\|Y - X\beta\|_2^2 = \sum_{i=0}^{n}(Y_i - (X\beta)_i)^2$, $\|\beta\|_1 = \sum_{i=1}^{k}|\beta_i|$ and $\lambda \geq 0$ is parameter that controls strength of the penalty. Higher the value of penalty parameter, higher would be shrinkage. LASSO improves both prediction accuracy and interpretability by combining the goodness of ridge regression and subset selection. It solves the problem of multicollinearity by choosing only one predictor in case of high correlation among a group of predictors.

Therefore, LASSO at each step will not only choose a list of significant and independent neighbouring stations but also give the model accuracy which can be compared with other model accuracies generated at step of 20 km distances. Based on all such model accuracies, the best model is chosen based on coefficient of determination ($R^2$) value that gives a subset of IMD stations as neighbouring set (*Nset*) and the distance up to which temperature of other IMD stations affect temperature.

**4.3. Capturing Spatial Information using Linear Model**
Given the extent to which the temperature of each station is influenced by its neighbouring stations, a minimum residual sum of square linear model is applied on training data to estimate minimum and maximum surface air temperature. The linear model applied for each station is expressed as:

$$\widehat{SAT}_{IMD} = \sum_{i=1}^{p}\alpha_i\, LST_i^{INSAT} + \sum_{i=1}^{p}\beta_i\, LST_i^{MODIS} + \sum_{i=1}^{p}\gamma_i\, LST_i^{INSAT} * LST_i^{MODIS} + \varepsilon$$

where, i is i[th] station out of $p$ stations which includes members of *Nset* and the station under consideration; $\alpha, \beta$ and $\gamma$ are model coefficients of INSAT, MODIS and product of INSAT and MODIS; and, $\varepsilon$ is the error term. In the above linear model, the product of INSAT-LST and MODIS-LST is added as a dummy variable to eliminate any correlation between temperature values of these satellites. A separate linear model is applied for each season. Thus, the proposed linear model estimates SAT based on the inter-relationship of satellite derived LST values between neighbouring stations and the target station using spatial domain information.

**4.4 Capturing Temporal Information using ARIMA Model**
The Autoregressive Integrated Moving Average (ARIMA) model is a statistical forecasting model which forecasts values of a variable based on linear function of past several values of it and random errors. Thus, it can be represented by:

$$y_t = \sum_{i=1}^{p}\Phi_i\, y_{t-i} + \sum_{j=1}^{q}\psi_j\, \varepsilon_{t-j} + \varepsilon_t$$

Here, $y_t$and $\varepsilon_t$ are the forecast value and random error at time period t, respectively; $\Phi_i$ (i = 1,2,3...,p) and $\psi_j$ (j =

1,2,3...,q) are model parameters of the autoregressive and moving average part of ARIMA respectively; p and q are non-negative integers and often referred as orders of the model.

The development of the ARIMA model is a three-step approach. Firstly, time series should be a stationary time series. A stationary time-series has the property that its mean and autocorrelation structure remain constant throughout the time. The stationarity is generally checked by unit root test (Dickey Fuller Test), ADF test and KPSS test. The data is transformed using log, power or differencing to remove trend and heteroscedasticity from time series.

Once data is stationary, it is used to estimate ARIMA model parameters, p and q. One approach to estimate it is using Box and Jenkin methodology (Box et al. 2015) and AIC values. Box and Jenkin method uses autocorrelation function (ACF) and partial autocorrelation function (PACF) to estimate the maximum value of q and p respectively. ARIMA model is developed on training data at each combination of p and q starting from 0 to maximum value. The models are evaluated based on Akaike Information Criteria (AIC) values and the order with lowest AIC value is chosen for final model development. The ARIMA model is developed at each IMD station and model parameters and coefficients are estimated based on training data. The idea is that it will estimate value based on seasonal changes happening at the station.

**4.5 Ensembling of models using Stacked generalization**

Predictions of minimum and maximum surface air temperature (SAT) based on spatial information (Linear model) and temporal information (ARIMA model) are finally combined to make an overall spatio-temporal prediction using stacked generalization. Stacked generalization is a method to optimally combine multiple statistical models into an ensemble model that will (theoretically, at least) lead to more accurate predictions than each of the individual ensemble members (Hooker et al. 2018). Based on this, the actual observation value is trained on the predicted value of each observation from all the remaining observations using different models resulting in a stacked generalized model. The specific form of model is given below:

$$\widehat{SAT}_{SG} = \eta_1 \widehat{SAT}_{LR} + \eta_2 \widehat{SAT}_{ARIMA}$$

where, $\widehat{SAT}_{SG}$ , $\widehat{SAT}_{LR}$ and $\widehat{SAT}_{ARIMA}$ are values estimated by seasonal stacked generalization, linear regression and ARIMA model; $\eta_1$ and $\eta_2$ are non-negative coefficients of stacked generalized model. Again, this model is applied to each season as defined by IMD. Thus, at the end, we get four stack generalized models, for each season, developed on training data to estimate SAT of each IMD station.

## 5. Results & Discussion

In the data preparation process, as discussed in subsection 4.1, a data completeness check is applied. Out of 332 stations only 250 IMD stations are selected based on it. Therefore, only these stations are used for further investigation of interrelationship of SAT and satellite derived LST.

Since this study incorporates datasets from different sources, data availability becomes highly important. Hence, season wise combined data availability is analyzed. Combined data availability depicts the data availability of both satellites (MODIS and INSAT3D) and IMD stations ground data for specified seasons. Table 1 shows the number of stations whose data was available for the specific seasons. The seasons are defined as per the IMD classification in section 2.

Spatial relation amongst the stations is captured using LASSO Regression. LASSO is used to compute the neighboring set of stations from candidate models with highest $R^2$. Regression is computed for each station at every 20 km distance. In this manner, multiple LASSO regression models are generated for each station. Out of these models, the neighbouring set of the best model i.e. with maximum coefficient of determination is chosen for further processing. Figure 2 shows the number of stations against the optimal distance at which their neighboring set is selected. The figure 2 shows that out of 250 stations, SAT of 60% to 80% stations can be best explained by SAT of neighbouring IMD stations lying within the radius of 350 to 450 kms for *Tmax* whereas it is 580 to 800 kms for *Tmin*. Difference in optimal distance for *Tmin* and *Tmax* can be explained by data and its availability at respective stations.

Based on LASSO finalized best neighbouring stations, corresponding LST data is retrieved from MODIS and INSAT-3D images and a linear regression model is built season wise on training data. To capture temporal domain information of IMD stations, ARIMA model is developed on SAT data of each IMD station. Individual ARIMA model is fitted using optimal order values of *p, d, q* with minimum AIC values. Distribution of order of these ARIMA models is shown in Figure 3. For *Tmin* and *Tmax*, most of the stations have ARIMA model of order ARIMA (2,0,1). Some stations have different ARIMA order for *Tmax* and *Tmin*. Later, these spatial and temporal information are combined using non-negative least square (NNLS) stack generalization method. The weight of the linear regression model (spatial information) is defined as $\boldsymbol{\eta_1}$ whereas the weight of the ARIMA model (temporal information) is defined as $\boldsymbol{\eta_2}$. Average and standard deviation of seasonal weights are tabulated in the Table 2 below. This shows that in general for all seasons weightage to temporal information is more than spatial information considering selected stations. ARIMA weights are higher for winter and monsoon seasons than other seasons.

To understand behavior of trained spatial, temporal and ensembled models on training data to estimate SAT, a suitable performance measuring criterion is chosen. In this case, performance is measured based on average Root Mean Square Error (RMSE) when trained models are applied on testing data and RMSE values are averaged for all selected IMD stations. The results shown in the Table 3 shows that for both temperature parameters (*Tmin* and *Tmax*), ARIMA outperformed LR for all seasons. LR

model shows the highest RMSE for monsoon (*Tmin*) and winter seasons (*Tmax*) whereas the most erroneous season for ARIMA model is the summer season (*Tmin* and *Tmax*). Similarly, the post-monsoon season surpasses other seasons in LR for both *Tmax* and *Tmin*. In the case of ARIMA models, model performance is best for the monsoon and post-monsoon seasons in *Tmin* and *Tmax* respectively. The ensemble model has lower RMSE values as compared to both and follows the same seasonal error variations as ARIMA.

**Table 1. Number of stations with season wise combined availability of in situ and LST data**

| Seasons | Tmin (#stations) | Tmax(#stations) |
|---|---|---|
| S1 only | 1 | 0 |
| S1 and S2 both | 41 | 10 |
| S1, S2 and S3 | 3 | 1 |
| S1, S2, S3, S4 (All) | 40 | 140 |
| S1, S2 and S4 | 111 | 62 |

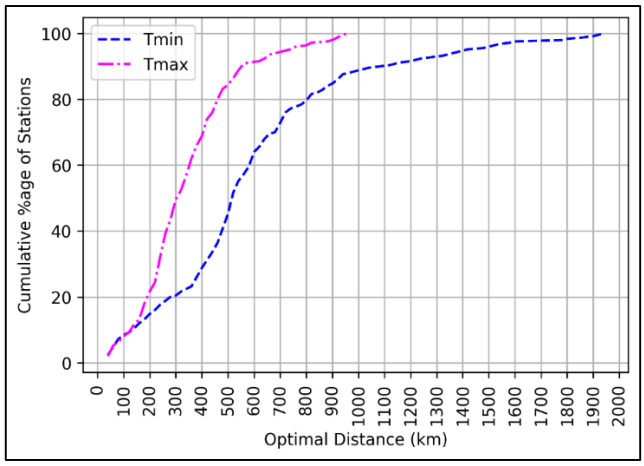


**Figure 2. Optimal Distance at which LASSO finalized Neighbouring Stations**

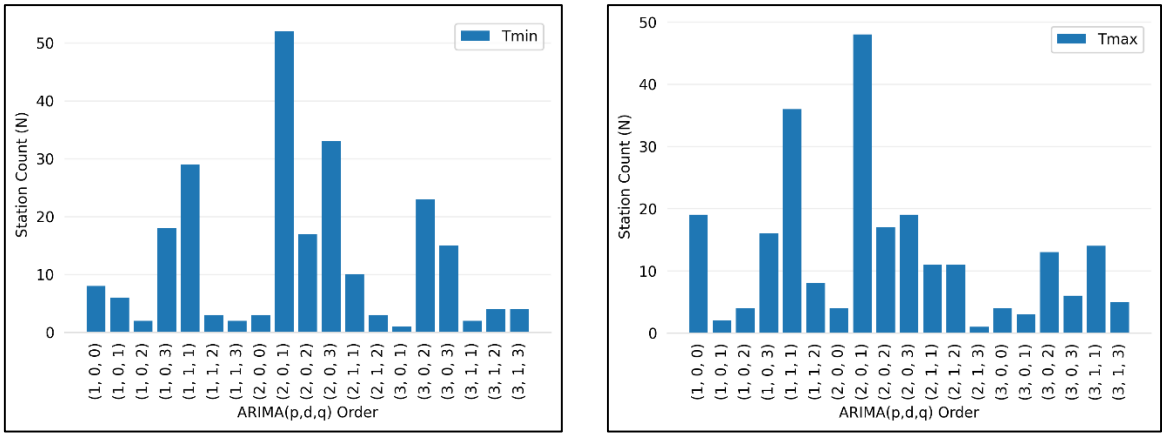


**Figure 3. Distribution of ARIMA (p,d,q) order**

**Table 2. Coefficients of Ensembled Model**

| Seasons | Tmin | | | | Tmax | | | |
|---|---|---|---|---|---|---|---|---|
| | Linear Regression (LR) ($\eta_1$) | | ARIMA ($\eta_2$) | | Linear Regression (LR) ($\eta_1$) | | ARIMA ($\eta_2$) | |
| | Avg. | Std. Dev. | Avg. | Std. Dev. | Avg. | Std. Dev. | Avg. | Std. Dev. |
| S1 | 0.163 | 0.163 | 0.842 | 0.160 | 0.199 | 0.146 | 0.803 | 0.144 |
| S2 | 0.233 | 0.175 | 0.768 | 0.175 | 0.388 | 0.183 | 0.615 | 0.182 |
| S3 | 0.159 | 0.220 | 0.841 | 0.218 | 0.184 | 0.172 | 0.816 | 0.171 |
| S4 | 0.297 | 0.181 | 0.701 | 0.184 | 0.364 | 0.192 | 0.635 | 0.193 |

**Table 3. Seasonal RMSE for Linear regression, ARIMA and Ensemble Models (Tmin and Tmax)**

| Seasons | Tmin | | | Tmax | | |
|---|---|---|---|---|---|---|
| | LR | ARIMA | Ensembled | LR | ARIMA | Ensembled |
| S1 | 3.36 | 1.60 | 1.55 | 3.01 | 1.7 | 1.63 |
| S2 | 2.82 | 1.62 | 1.61 | 2.75 | 1.87 | 1.81 |
| S3 | 3.40 | 1.24 | 1.31 | 2.74 | 1.47 | 1.39 |
| S4 | 2.50 | 1.30 | 1.31 | 2.30 | 1.20 | 1.29 |

The averaged seasonal RMSE gives a summary of proposed methodology and may not give insights of model performance. Thus, a frequency distribution plot of seasonal RMSE is plotted as shown in Figure 4. The RMSE values upto five are plotted. This plot shows that S3, which performs worst in LR for *Tmin*, has less data availability as compared to other seasons due to monsoon season and subsequently has high RMSE values.

In ARIMA, most stations fall in [1, 2] RMSE category for all seasons in *Tmin*. Similar results are observed in *Tmax* except in S4, where RMSE is found to be even better. It can be inferred from the Figure 4 that ensembled method followed almost the same distribution pattern as ARIMA but with lowered averaged RMSE values.

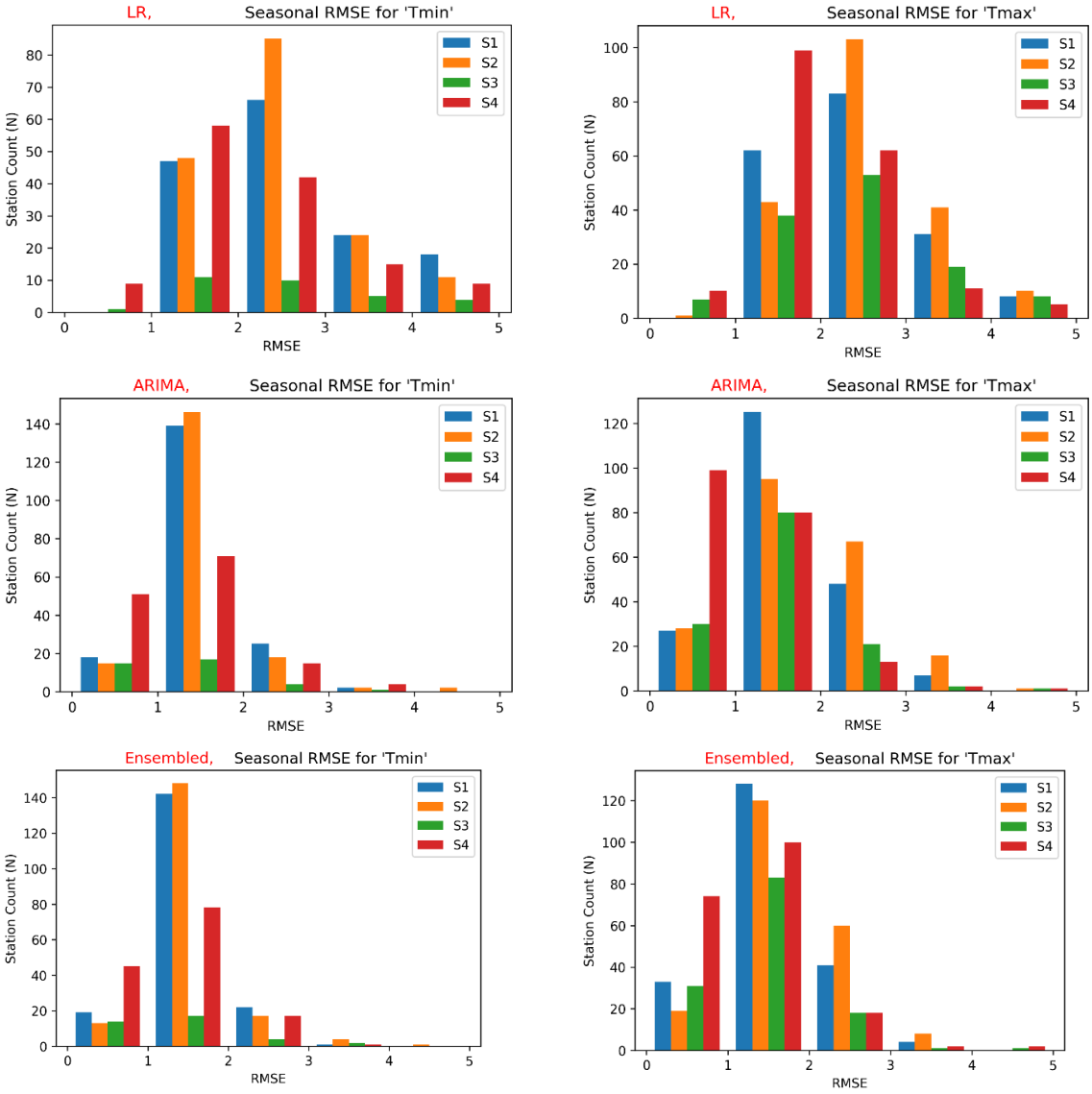


**Figure 4. Seasonal RMSE with respect to number of stations**

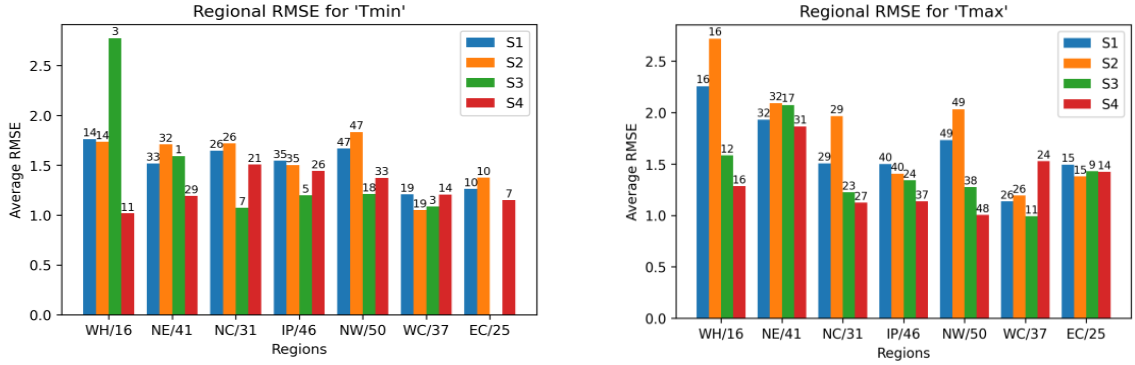


**Figure 5. Regional RMSE with respect to number of stations**

In this paper, a geographical analysis is also done to understand regional variation in estimated temperature values using the proposed ensemble method. The temperature homogeneous regional distribution of the selected 250 stations is: Western Himalayas (WH)-16, North East (NE)-41, North Central (NC)-31, Indian Peninsula (IP)-46, North west (NW)-50, West Coast (WC)-37 and East Coast (EC)-25. The remaining four stations lie outside the Indian mainland and therefore do not fit in any category given by IITM, Pune. The Figure 5 shows temperature homogeneous regional distribution of averaged RMSE of estimated IMD temperatures for each season. This helps to understand variation in RMSE in different regions and seasons. The values on top of the bar are the number of stations from which the average RMSE is calculated. From the Figure 5, it can be inferred that for *Tmin* the RMSE values lie in [1, 2] except WH region in Monsoon season. For *Tmax*, RMSE values lie in [1,2] for all regions except WH and NE. This could be due to mountainous terrain and typical Himalayan mountain climate.

The primary difficulty in estimating SAT from satellite based LST is the variable emissivity associated with different surface types. Surface type may differ in its water, mineral content and effect of vegetation (Mendelsohn et al 2007). Since SAT is not a function of LST alone. Therefore, RMSE values can be attributed to some extent to other factors that influences SAT including Vegetation density (kawashima et al 2000).

## 6. Conclusions

Surface air temperature (SAT) is a crucial for vast number of applications. In the present study, Spatio-temporal variation for daily *Tmax* and *Tmin* are studied. Spatial variation is modeled using linear regression and time variation are modeled with ARIMA. Satellite derived land surface temperature (LST) from satellites INSAT and MODIS is used to capture spatial variability and counter data availability issues of weather stations. LASSO regression used to derive the optimal distance affecting the observation of a weather station. Since it is data intensive technique, hence data availability and its accuracy is paramount. Results shown reconfirm that as for most of the stations optimal distance is 800km (*Tmin*) and 500km (*Tmax*). The variations in optimal distance can be due to the less data availability for *Tmin*. ARIMA models for different stations has different order but most of the station follows ARIMA (2,0,1) for both *Tmin* and *Tmax*.

These two model were ensembled, which is then evaluated for seasonal & regional performances. The ensembled model shows good result within [1,2] RMSE range for most of the regions and seasons. Western Himalayan & North East region has maximum RMSE because of its terrain, elevation and climate. It can be concluded from the study that *Tmin* and *Tmax* are better predicted by ensembling time series and linear model of satellite observations and spatially related weather station observations.

The study is limited in an aspect that it is data intensive and tries to capture local variations. It can further be improved by stacking Land Use & Land Cover (LULC) variation for the locations. The study evaluates the applicability of point-based model and its possible application for the generation of interpolated gridded products.

**Acknowledgements**

The authors would like to thank Director, Space Application Center for motivating to carry out this study. We are grateful for the support and guidance of Deputy Director, Earth, Ocean, Atmospheric, Planetary Science & Applications Area (EPSA). We also thank Group Director, MOSDAC Research Group (MRG) and Group Director, Vedas Research Group (VRG) for providing resources and valuable suggestions. The authors are thankful to data repositories viz. Meteorological & Oceanographic Satellite Data Archival Centre (MOSDAC, SAC, ISRO), Land Processes Distributed Active Archive Centre (LP-DAAC) and IMD for providing satellite data and insitu meteorological observations respectively.

## References

Aldersley A., S. J. Murray and S. E. Cornell (2011). Global and regional analysis of climate and human drivers of wildfire. Science of the Total Environment, 409(18), pp.3472-3481.

Benali A., A. C. Carvalho, J. P. Nunes, N. Carvalhais and A. Santos (2012). "Estimating Air Surface Temperature in Portugal Using MODIS LST Data." Remote Sensing of Environment 124: 108– 121. DOI:10.1016/j.rse.2012.04.024.

Benavides R., F. Montes, A. Rubio and K. Osoro (2007). Geostatistical modelling of air temperature in a mountainous region of Northern Spain. Agricultural and Forest Meteorology, 146(3-4), pp.173-188.

Box G. E., G. M. Jenkins, G. C. Reinsel and G. M. Ljung (2015). Time series analysis: forecasting and control. John Wiley & Sons.

Czajkowski K. P., T. Mulhern, S. N. Goward, J. Cihlar, R. O. Dubayah and S. D. Prince (1997). "Biospheric Environmental Monitoring at BOREAS with AVHRR Observations." Journal of Geophysical Research 102: 29651–29662. DOI:10.1029/97JD01327.

Dee D.P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, D. P. Bauer and P. Bechtold (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Quarterly Journal of the royal meteorological society, 137(656), pp.553-597.

Florio E. N., S. R. Lele, Y. C. Chang, R. Sterner and G. E. Glass (2004). "Integrating AVHRR Satellite Data and NOAA Ground Observations to Predict Surface Air Temperature: A Statistical Approach." International Journal of Remote Sensing 25: 2979–2994. DOI:10.1080/01431160310001624593.

Fonti V. and E. Belitser (2017). Feature selection-using LASSO. VU Amsterdam Research Paper in Business Analytics, 30, pp.1-25.

Goward S. N., R. H. Waring, D. G. Dye and J. L. Yang (1994). "Ecological Remote-Sensing at OTTER: Satellite Macroscale Observations." Ecological Applications 4: 322–343. DOI:10.2307/1941937.

Gupta U., P. K. Patel, S. Surati and M. Oza (2018). Gridded Temperature generation using INSAT-Land Surface Temperature data and India Meteorological Department temperature data for Indian region. Journal of Geomatics, 12(2).

Hatfield Jerry L., J. B. Kenneth, A. K. Bruce, H. Lewis, Ziskz and C. I. Roberto (2011). Climate Impacts on Agriculture: Implications for Crop Production, Publications from USDA-ARS / UNL Faculty. 1350.

Hooker J., G. Duveiller and A. Cescatti (2018). A global dataset of air temperature derived from satellite remote sensing and weather stations. Scientific data, 5(1), pp.1-11.

Jang J.-D., A. A. Viau and F. Anctil (2004). "Neural Network Estimation of Air Temperatures from AVHRR Data." International Journal of Remote Sensing 25: 4541–4554. DOI:10.1080/ 01431160310001657533.

Jones P.D., D. H. Lister, T. J. Osborn, C. Harpham, M. Salmon and C. P. Morice (2012). "Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010." Journal of Geophysical Research: Atmospheres, 117(D5).

Kim D. and K. Han (2013). "Remotely Sensed Retrieval of Midday Air Temperature Considering Atmospheric and Surface Moisture Conditions." International Journal of Remote Sensing 34: 247–263. DOI:10.1080/01431161.2012.712235.

Loew A. (2017), Validation practices for satellite-based Earth observation data across communities, Rev. Geophys., 55, 779–817. DOI:10.1002/2017RG000562.

Litschert S. E., T. C. Brown and D. M. Theobald (2012). "Historic and future extent of wildfires in the Southern Rockies Ecoregion", USA. Forest Ecology and Management, 269, pp.124-133.

Mendelsohn R., P. Kurukulasuriya, A. Basist, F. Kogan and C. Williams (2007). "Climate analysis with satellite versus weather station data." Springer Science Climatic Change (2007) 81:71–83. DOI: 10.1007/s10584-006-9139-x

Nieto H., I. Sandholt, I. Aguado, E. Chuvieco, and S. Stisen (2011). "Air Temperature Estimation with MSG-SEVIRI Data: Calibration and Validation of the TVX Algorithm for the Iberian Peninsula." Remote Sensing of Environment 115: 107–116. DOI:10.1016/j.rse.2010.08.010.

Oyler J. W., S. Z. Dobrowski, Z. A. Holden and S. W. Running (2016). Remotely sensed land skin temperature as a spatial predictor of air temperature across the conterminous United States. Journal of Applied Meteorology and Climatology, 55(7), pp.1441-1457.

Pandya, D. B. Shah and R. P. Singh (2014) "LST Baseline Document." INSAT-3D Algorithm Theoretical Basis Development Document, pp. 281-306.

Pape R., and J. Löffler (2004). "Modelling Spatio-Temporal Near-Surface Temperature Variation in High Mountain Landscapes." Ecological Modelling 178: 483–501. DOI:10.1016/j.ecolmodel.2004.02.019.

Prihodko L. and S. N. Goward (1997). "Estimation of Air Temperature from Remotely Sensed Surface Observations." Remote Sensing of Environment 60: 335–346. DOI:10.1016/S0034-4257 (96)00216-7.

Roman, M. O. (2009), "The MODIS (collection v005) BRDF/ALBEDO product: Assessment of spatial representativeness over forested landscapes." Remote Sens. Environ., 113(11), 2476–2498. DOI:10.1016/j.rse.2009.07.009.

Kawashima S., T. Ishida, M. Minomura, T. Miwa (2000). "Relations between surface temperature and air temperature on a local scale during winter nights." Journal of Applied Meteorology, 39 (2000), pp. 1570-1779

Sharifnezhadazizi Z., H. Norouzi, S. Prakash, C. Beale and R. Khanbilvardi (2019). A global analysis of land surface temperature diurnal cycle using MODIS observations. Journal of Applied Meteorology and Climatology, 58(6), pp.1279-1291.

Stahl K., R. D. Moore, J. A. Floyer, M. G. Asplin and I. G. McKendry (2006). Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density. Agricultural and forest meteorology, 139(3-4), pp.224-236.

Sun Y. J., J. F. Wang, R. H. Zhang, R. R. Gillies, Y. Xue and Y. C. Bo (2005). "Air Temperature Retrieval from Remote Sensing Data Based on Thermodynamics." Theoretical and Applied Climatology 80: 37–48. DOI:10.1007/s00704-004-0079-y.

Tibshirani R. (1996). Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), pp.267-288.

Vancutsem, C., P. Ceccato, T. Dinku and S. 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), pp.449-465.

Wan Z., S. Hook and G. Hulley (2015). MYD11C1 MODIS/Aqua Land Surface Temperature/Emissivity Daily L3 Global 0.05Deg CMG V006 [Data set]. NASA EOSDIS Land Processes DAAC. Accessed 2021-07-26 from https://doi.org/10.5067/MODIS/MYD11C1.006

Wloczyk C., E. Borg, R. Richter and K. Miegel (2011). "Estimation of Instantaneous Air Temperature above Vegetation and Soil Surfaces from Landsat 7 ETM+ Data in Northern Germany." International Journal of Remote Sensing 32: 9119–9136. DOI:10.1080/ 01431161.2010.550332.

Xu Y., Z. Qin and Y. Shen (2012). "Study on the Estimation of Near Surface Air Temperature from MODIS

Data by Statistical Methods." International Journal of Remote Sensing 33: 7629–7643. DOI:10.1080/01431161.2012.701351

Zhang W., Y. Huang, Y. Yu and W. Sun (2011). "Empirical Models for Estimating Daily Maximum, Minimum and Mean Air Temperatures with MODIS Land Surface Temperatures." International Journal of Remote Sensing 32: 9415–9440. DOI:10.1080/01431161.2011.560622.

Zhu W., A. Lű and S. Jia (2013). 'Estimation of Daily Maximum and Minimum Air Temperature Using MODIS Land Surface Temperature Products.' Remote Sensing of Environment 130: 62–73. DOI:10.1016/j.rse.2012.10.034.