

NEAR SURFACE AIR TEMPERATURE ESTIMATION THROUGH PARAMETRIZATION OF MODIS PRODUCTS

E. Khesali ^{1*}, M. R. Mobasheri ¹

¹ Dept. of Photogrammetry and Remote Sensing, Faculty of Geodesy and Geomatics Engineering, K.N. Toosi University of Technology, Tehran, Iran- elahe.khesali@mail.kntu.ac.ir, mobasheri@kntu.ac.ir

Commission IV, WG IV/3

KEY WORDS: Air temperature, Atmospheric parameters, Linear regression, SVR, MODIS, Remote Sensing.

ABSTRACT:

Near-surface air temperature is a key factor in many studies and its spatiotemporal patterns are highly dependent on the ground surface characteristics and vary over time and space. So Land Surface Temperature (LST) is an important parameter for air temperature estimation. In this study, it is tried to model the air temperature by deploying some of the parameters that affect it. The parameters that have been taken into account in this study include land surface temperature, Normalized Difference Vegetation Index (NDVI), Vapor Pressure (VP) and Lifted Index (LI) as a measure of atmospheric stability. To assess the impact of each of these parameters, different linear regression models, were tested. Support Vector Regression (SVR) and hybrid artificial neural network methods were also performed. To model and evaluate the time series data of Georgia State in The United States of America over 1 year have been used. The NDVI, Total Precipitable Water (TPW), LST and LI parameters are products of MODIS. VP is calculated by using a logarithmic model from the TPW. Finally, it was found out that the LST and VP have positive effects, LI has negative and NDVI had a slightly positive impact on the air temperature at 2 meters height. The achieved accuracy in the linear model when all parameters are involved was 2.29°C with a correlation coefficient of $R=0.96$. Next, the SVR model was examined and applied to the linear model taking all parameters into account. It was found that it does not end up to any significant increase in accuracy but certainly increases the computation time. The accuracy of this model was about 2.25°C with a correlation coefficient of 0.96. Finally, a hybrid artificial neural network was examined. It was found that it increases the accuracy but certainly increases the computation time. The achieved accuracy of this model was about 2.14 °C with a correlation coefficient of 0.96.

1. INTRODUCTION

Near-surface air temperature is a key factor in the study of hydrology, ecology and climatological issues and in vegetation processes such as photosynthesis, transpiration, and evaporation (Khesali and Mobasheri 2020).

The spatial-temporal pattern of near-surface air temperature strongly depends on the surface covers as well as parameters such as air pressure, water vapor and to some extent air stability, as they may vary over time and change in different parts of the earth. This makes the air temperature important to be estimated over time and in different locations with different conditions (Mildrexler et al. 2011). Usually, the air temperature is measured in meteorological Synoptic stations all over the world at certain times of the day. However, these measurements are carried out at certain points and by applying some interpolation procedures, one can find the value of these parameters for other points located between stations. Of course, the results of such interpolation show erroneous results that may cause problems in forecasting and numerical modeling.

Clearly, the factors such as surface cover, land use, surface temperature, the amount of ambient water vapor, the total precipitable water, wind speed and direction and few more can affect the air temperature to different degrees (Mildrexler et al. 2011).

With the development of remote sensing technology, satellite data have facilitated the estimation of LST over the entire globe with sufficient precision as well as the acceptable resolution in time and space. The methods introduced in this context can be categorized into three categories:

1) Statistical methods (simple and advanced): These methods estimate the air temperature by establishing a linear relationship between the air temperature and parameters which are believed to have influences on regulating the air temperature.

Yan et al. (2009) developed a statistical algorithm for MODIS data for daytime air temperature (T_a) retrievals in east China. First a statistical regression is applied. Then, for different latitude zones the first guess of T_a is corrected using a series of bias equations. The overall statistics of bias = -0.09°C , $R = 0.96$ and RMSE = 3.23°C shows that the algorithm has a good performance.

Yunus et al. (2012) compared the annual land surface temperature of MODIS and the annual air temperature measured at a 1.5-meter height in weather station having different amounts of vegetation cover. They showed that the relationship between these two temperatures strongly depends on the type of land cover.

Xu et al. (2014) developed a model to estimate the daily maximum T_a by Aqua MODIS data and meteorological data in British Columbia, Canada. They selected nine parameters including land surface temperature (LST), modified normalized difference water index (MNDWI), normalized difference vegetation index (NDVI), albedo, latitude, longitude, altitude, solar radiation and distance to the ocean as predictors. The random forest model with mean absolute error, MAE = 2.02°C and $R^2 = 0.74$ had better result than the linear regression model. Chen et al. (2016) presented an algorithm for daily maximum air temperature estimation using MODIS data. This algorithm has two steps: In the first step, the daytime air temperature is estimated based on MODIS products at the Aqua and Terra passing time. Then, through a sinusoidal daytime T_a variation

model the maximum Ta is derived from daytime Ta estimations. Results show that using the proposed algorithm, the spatial distribution of maximum Ta can be retrieved properly. The RMSE of this model is varying from 1.62 to 2.33 K and the correlation coefficient is varying from 0.95 to 0.98.

2) A method based on Temperature Vegetation index (TVX): this method is based on the assumption that vegetation canopy temperature to a good extent approximates near-surface air temperature i. e., the bulk temperature of a very dense canopy is close to air temperature (Prihodko et al. 1997). The TVX method shows a negative correlation between air temperature and NDVI. Also, the TVX proves various attribution to soil moisture, latent heat flux and canopy resistance to transpiration, and (Prihodko et al. 1997). The principle of the TVX method is to find a strong negative correlation between NDVI and LST assuming a uniform atmospheric forcing and similar moisture conditions for a limited region (Stisen et al. 2007). Prihodko et al. (1997) have used TVX for estimation of air temperature and a strong correlation ($R= 0.93$) was claimed between satellite estimates and in situ measured air temperature with a mean error of 2.92°C.

3) Energy-balance approach: this approach has a complete physical basis. The method is based on the assumption that the incoming net radiation flux together with anthropogenic heat flux (if available) equals the sum of the soil heat flux, sensible heat and the fluxes due to the latent heat of evaporation. The method that is working on these principles, requires inputs such as incoming/outgoing short and longwave radiation, relative humidity, precipitation, substrate texture/roughness type, vegetation species, wind speed, air pressure, and elevation. This makes these methods difficult to be used in near-real-time operational applications over vast regions, (Zakšek and Schroedter-Homscheidt 2009).

In this study, it is tried to investigate the parameters affecting air temperature and choose the most influencing ones. The parameters taken into account in this study are LST, surface vegetation cover through NDVI, VP and atmospheric stability using LI. LST, NDVI, TPW, and LI are daily products of MODIS. VP is extracted from TPW using a linear logarithmic model.

2. MATERIAL AND METHODS

2.1 Region of study

Statistical modeling of a variable needs a suitable data population as well as a suitable distribution of input/ output parameters. For this, the data used in this work includes data collected in the State of Georgia, located between the latitude of 30°31' to 35° north, and longitude of 81° to 85°53' west with an area of 154077 square kilometres in the south part of the United States. In 2012, 2600 farms covering 1.3 million hectares under cultivation were in this state. The reason for this selection was because of the availability of the collected data mostly in the cultivation zones. Of course, the research method is such that the results can be used anywhere in the world. Each sensor is scanned at a one-second frequency and the data are summarized at 15-minute intervals as well as at midnight. All weather stations monitor air temperature, vapor pressure deficit, wind speed and direction, relative humidity, solar radiation, precipitation and soil temperature at 2, 4 and 8 inches (Hoogenboom et al. 2003).

In this study, data (Air Temperature and Vapor Pressure) collected in 10 AEMN stations during one year (2015) are used for modeling and evaluation. Two-thirds of the data are used for modeling and another one-third are used for evaluation.

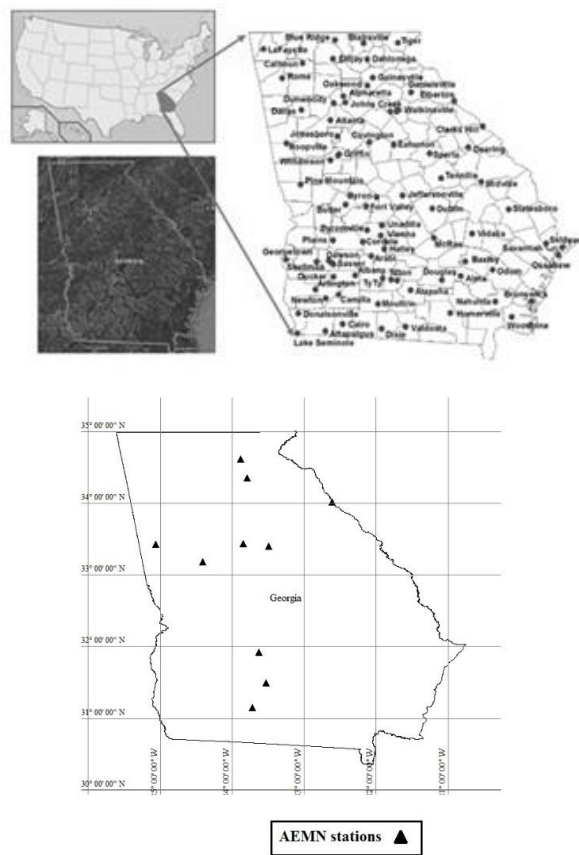


Figure 1. Regions of study

2.2 Satellite Data

Satellite images used in this study are MODIS products collected onboard of Terra and Aqua platforms. These products consist of both daytime and nighttime LST (MOD11 and MYD11) with 1°C accuracy, LI (MOD07 and MYD07), TPW (MOD05 and MYD05) with 7-13% uncertainty, Cloud_Fraction (MOD06 and MYD06) and NDVI (MOD13) for 1 year. The images are corrected for the atmospheric effects and noise and are georeferenced all by the producer.

2.3 Research Methodology

The method of this study consists of several steps shown in Fig.2.

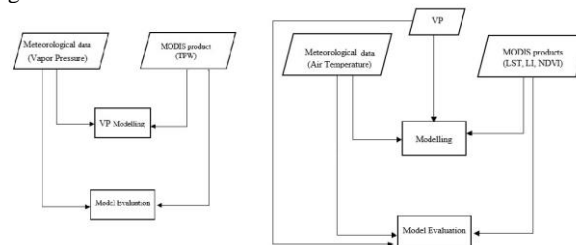


Figure2. Research Flowchart

2.4 Vapor Pressure modeling

The vapor pressure is the pressure exerted by the water vapor molecules present in the atmosphere. This pressure is a component of the total air pressure. As the air temperature rises,

the water vapor pressure increases (KNMI, 2000). First, the relationship between Vapor pressure and TPW is analyzed and vapor pressure is modeled. In this section, data (Vapor Pressure) collected in 10 AEMN stations during one year (2015) are used for modeling and evaluation. Two-thirds of the data are used for modeling and another one-third is used for evaluation. First, the scatterplot of TPW against VP is presented in Fig.3. As can be seen in Fig.3, a linear logarithmic model with an RMSE of 0.38 kpa and a correlation coefficient of 0.86 is the best model. The logarithmic relationship between these two parameters is as follows:

$$VP = 1.13 * \ln(TPW) + 0.91 \quad (1)$$

where VP = Vapor pressure
 TPW = Total Precipitable Water

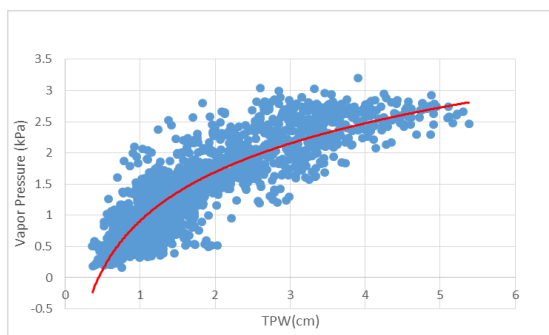


Figure 3. Scatterplots of TPW against VP

2.5 Linear Regression Model

The method is based on regression between air temperature measured in predefined meteorological stations at 2m heights and parameters extracted from MODIS images. The important constraint in this method is the stability of the atmosphere. This is defined by the positive values of the Lifted Index in the relevant MODIS product. The accuracy of this index is estimated to be about 0.5° C (Borbás et al. 2011). The other parameters used in this modeling are LST and NDVI all were MODIS products.

It is believed that the air temperature at the proximity of the surface is strongly influenced by surface temperature, surface vegetation cover, air stability, and the air water vapor content. The effect of all the above-mentioned parameters on Tair was investigated and is shown in (Fig.4).

As can be seen in Fig. (4), VP and LST three have positive, LI negative relationship with air temperature while NDVI shows the poor relationship. To account for these parameters, all of them were combined in a regression equation as is shown by equation (2).

$$T_{air} = a * LST + b * NDVI + c * VP + d * LI + e \quad (2)$$

where T_{air} = Air Temperature
 a, b, c, d = coefficients
 LST = Land Surface Temperature
 NDVI = Normalized difference vegetation index
 VP = Vapor Pressure
 LI = Lifted Index

To calculate coefficients a, b, c, d and e, those pixels containing weather stations are selected in MODIS images where Tair is routinely measured at 2m height at these weather stations.

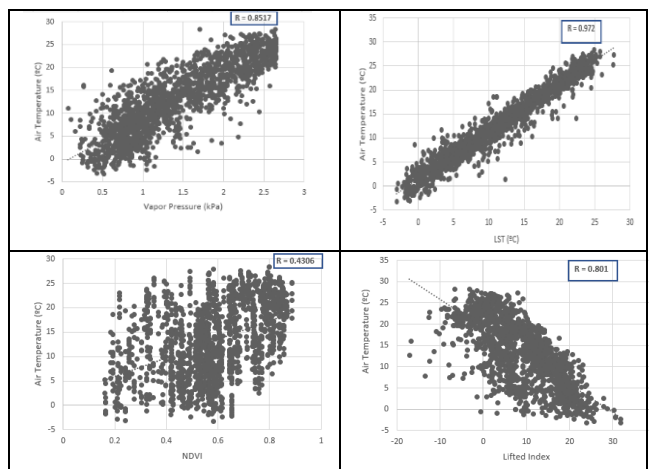


Figure 4. Scatterplots of air temperature Vs different parameters

2.6 Method of Support Vector Regression (SVR)

In classical statistics, classification and regression methods are based on some limiting hypotheses in which the probability distribution models and/or probability density function are known (Park et al., 2001). In practice, in most cases, there is not enough information available regarding the probability distribution function of the variables under study. In these cases, the methods that work well without using probability distribution functions, are preferred. One of these methods is SVR. The special privilege of SVR is the minimization of the operation error in it. This is contrary to the classical algorithms and linear regression methods where the absolute magnitudes of errors are minimized in (Park et al., 2001).

In the SVR method, different kernels such as linear, squares, Gaussian and polynomials are being employed. Usually, the Gaussian radial kernel function performs better. This function is of the form:

$$K(x, y) = \exp\left(-\frac{\|x_i - y_i\|^2}{2\sigma^2}\right) \quad (3)$$

where σ = standard deviation

In building a robust model of SVR, the required parameters must be calculated through an optimization method with acceptable accuracy. These parameters are of the type of kernel function, kernel function parameter σ^2 , the adjusting parameter C, precision parameter ϵ related to the maximum error.

2.7 Hybrid Method of Regression-BNN

Neural networks are computational systems that receive an n-dimensional input vector by “n” available neurons in their input layer and transform it into another m-dimensional output vector at their output layer which consists of “m” output neurons (Khesali et al. 2015). This method focuses on developing a hybrid regression-Artificial Neural Network (ANN) technique. The ANN method applied here is the Back-propagation Neural Network (BNN) and the regression model applied is linear. The BNN is mostly used for the classification of satellite images. BNN consists of the input and output layers and one or more hidden layers connected in a feed-forward way. Each layer consists of a number of neurodes. Every unit feeds the units in the next layer (Khesali et al. 2016). In this part, a BNN and a linear regression model are fused to form a hybrid model. At first, using the training data, BNN is trained. Then, a linear

regression between the output of the network and the expected data is setup. The flowchart of this method is shown in Fig.5.

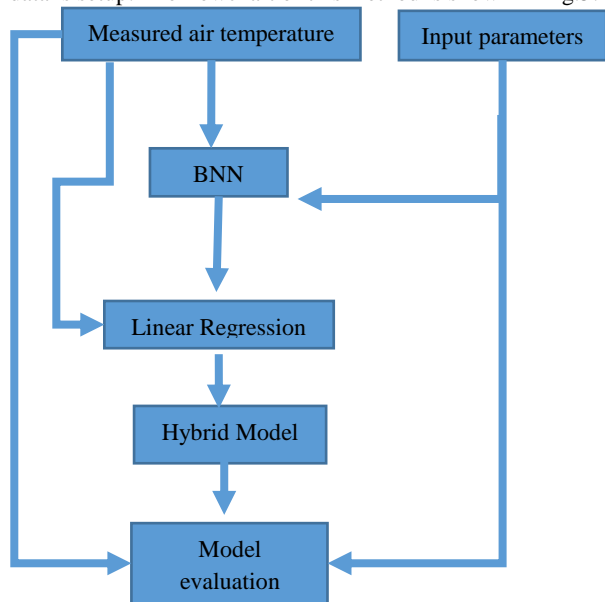


Figure 5. The flowchart of the proposed Hybrid Method

3. RESULTS AND ANALYSIS

As can be seen in Fig.4, individual parameters of LST, VP, and LI showed an acceptable correlation with the air temperature. Among these, LST and VP demonstrate a positive while LI has a negative linear correlation with the air temperature. However, the effects of NDVI and LST are evident although the NDVI itself doesn't show a good correlation with air temperature.

3.1 Results of linear regression

The aim was to find a linear regression that can relate air temperature to different combinations of parameters such as LST, VP, LI and NDVI. Twenty-nine different combinations were tested. 8 of these who had the highest correlation (R) and the lowest RMSE values are shown in table (1) while the plot of the best model against in site measured values is shown in (Fig. 6).

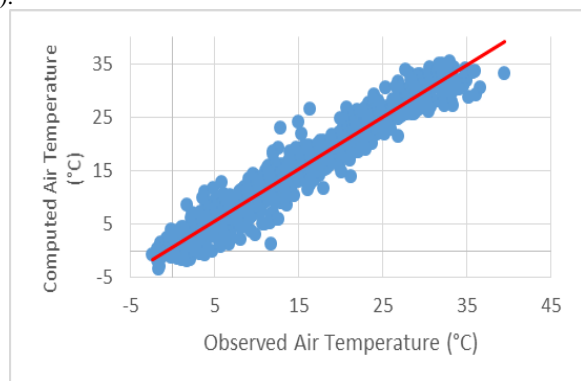


Figure 6. Scatterplots of computed air temperature against observed air temperature for a best linear model

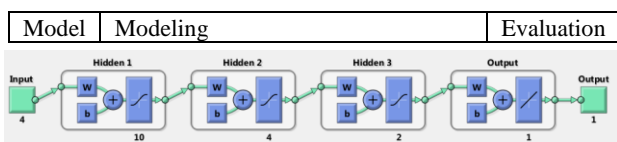


Figure 7. The best structure of the ANN

no.	Equation	R	RMSE(C)
1	$T_{air}=1.0073LST-0.1309$	0.95	2.8
2	$T_{air}=0.731LST+3.379VP-0.569$	0.96	2.32
3	$T_{air}=0.793LST-0.179LI+4.679$	0.95	2.59
4	$T_{air}=0.74LST+2.12VP-0.07LI+2.29NDVI+0.16$	0.96	2.29
5	$T_{air}=0.724LST+3.103NDVI+3.106VP-1.852$	0.96	2.32
6	$T_{air}=0.843LST+5.462NDVI-0.99$	0.95	2.70
7	$T_{air}=0.72LST+2.932VP-0.065LI+0.838$	0.96	2.31
8	$T_{air}=0.72LST+2.932LI-0.065NDVI+0.838$	0.96	2.54

Table 1. Different linear regression models

As can be seen in Fig.4, the order of affectivity of the parameters and indices on air temperature is as $LST > VP > LI \gg NDVI$. This means that the LST has the highest influence while the effect of NDVI is least. Also, VP is considerably important in air temperature modulation.

As can be seen in table (1), the best prediction model is no. 4 that consists of LST, NDVI, VP, and LI where the accuracy is better than 2.29°C with a correlation coefficient of 0.96. However, models no. 2, 5 and 7 are also performing well in air temperature estimation with an uncertainty of about 2.32 °C.

3.2 Evaluation of SVR approach

Based on the fact that all afore-mentioned parameters have shown some influences on the air temperature, it is decided to involve all of them in the SVR modeling procedure (Juang and Hsieh. 2009). The model was run using two-third of data for training and the rest for evaluation. Based on this, linear kernel and the Gaussian kernel with $\sigma=8$, a value of 4 for cost function and value of $\epsilon=0$ for the threshold due to the precision needed for the accuracy in the air the temperature determination, were selected. The resulting statistics of these two models are shown in table (2).

3.3 Evaluation of Hybrid Method of Regression-BNN

In this part, a backpropagation neural network and a linear regression model are fused to form a hybrid model to be used for air temperature estimation. First, using training data a BNN is trained. The input vector consists of LST, NDVI, LI, and VP. The best structure of the ANN is shown in Fig.7. The output of the network is air temperature. Then, a regression between the output of the network and the measured air temperature in the meteorological station is extracted. The results of the RMSE evaluation indicate that the Hybrid model works better compared to the linear regression and SVR (Table 2). The plots of the best model prediction against site measured values are shown in (Fig. 8).

Modeling		Evaluation	
Model		R	RMSE(°C)
SVR	Linear Kernel	0.96	2.25

SVR	Gaussian kernel)	0.96	2.31
Hybrid ANN	$T_{air} = 0.987 * T_{net} + 0.22$	0.96	2.14

Table2. SVR and Hybrid Neural Network models and their statistical data

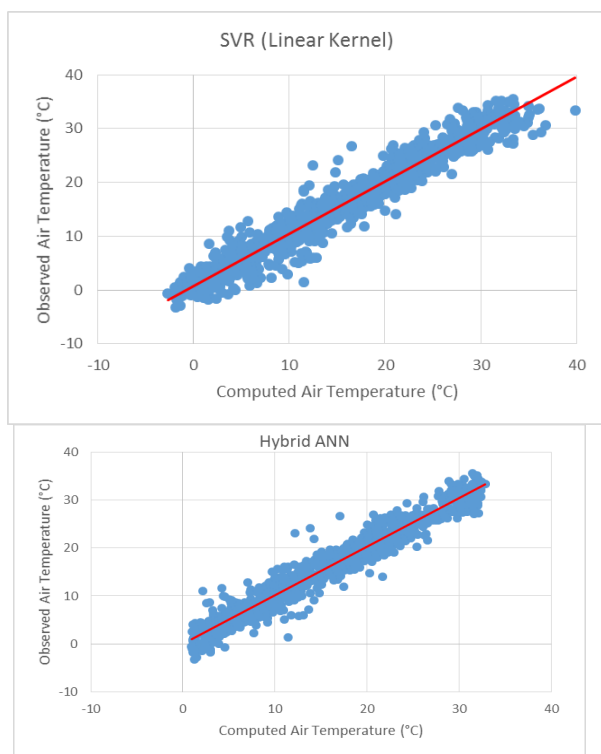


Figure8. Plots of model predicted air temperature against observed ones for SVR and Hybrid ANN models

4. CONCLUSIONS

The air temperature at the proximity of the earth's surface is one of the key variables in hydrological, climatological and agricultural studies. Usually, air temperature is measured at 2m height in weather stations under a shelter called screen while it is believed that this parameter is changing noticeably from point to point and the one that is measured in the weather station cannot necessarily represent the air temperature at the neighboring pixels.

Using parameters and indices such as NDVI, LST, LI (where all of them are products of MODIS) and VP, few linear models, were investigated. The best precision achieved was for model 4 in the table (1).

In addition to the linear model, SVR and Hybrid ANN models were tested too and the precision achieved was found to be of the order of 2.25°C and 2.14°C respectively. Finally, the linear model no.4 in the table (1) and the Hybrid ANN model in the table (2) are suggested by these authors. However, these authors would like to advise that these models must first be evaluated in other geographical and climatological regions before being used.

ACKNOWLEDGMENT

These authors are grateful to Dr. Pam Knox at Georgia university, US meteorology department for her assistance in providing weather data and also many thanks to MODIS administration for their precious products.

REFERENCES

- Chen, Y. Sun, H. and Li, J. 2016. Estimating daily maximum air temperature with MODIS data and a daytime temperature variation model in Beijing urban area. *Remote Sensing. Letters*, 7:9, 865-874.
- Borbas, E. F., Seemann, S. W., Kem, A. N. I. K. O., Moy, L. E. S. L. I. E. ., LI, J., Gumley, L. I. A. M., & Menzel, W. P. 2011. MODIS Atmospheric Profile Retrieval Algorithm Theoretical Basis Document, Collection6.
- http://modis-atmos.GSFC.NASA.gov/MOD07_L2/atbd.
- Hoogenboom, G. and Coker, D.D., 2003. The Georgia automated environmental monitoring network: Ten years of weather information for water resources management. *Georgia Institute of Technology*.
- Jacob, F., Petitcolin, F., Schmugge, T., Vermote, E., French, A., Ogawa, K. 2004. Comparison of land surface emissivity and radiometric temperature derived from MODIS and ASTER sensors. *Remote Sensing of Environment*, 90(2): p. 137-152.
- Juang, C.F. and Hsieh, C. D., 2009. TS-Fuzzy System-based support vector regression. *Fuzzy Sets Syst*. 160.2486-2504.
- Khesali, E., Enayati, H., Modiri, M. and Aref, M.M., 2015. Automatic ship detection in Single-Pol SAR Images using texture features in artificial neural networks. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(1), p.395.
- Khesali, E., Zoej, M.J.V., Mokhtarzade, M. and Dehghani, M., 2016. Semi automatic road extraction by fusion of high resolution optical and radar images. *Journal of the Indian Society of Remote Sensing*, 44(1), pp.21-29.
- Khesali, E. and Mobasheri, M., 2020. A method in near-surface estimation of air temperature (NEAT) in times following the satellite passing time using MODIS images. *Advances in Space Research*, 65(10), pp.2339-2347.
- Mildrexler, D.J., M. Zhao, and S.W. Running. 2009. Testing a MODIS Global Disturbance Index across North America. *Remote Sensing of Environment*, 113(10): p. 2103-2117.
- Mildrexler, D.J., M. Zhao, and S.W. Running, A. 2011. global comparison between station air temperatures and MODIS land surface temperatures reveals the cooling role of forests. *Journal of Geophysical Research: Biogeosciences*. 116(G3): p. G03025.
- Park, Y., Luo, L., Netoff, T. 2011. Seizure prediction with spectral power of EEG using cost-sensitive support vector machines. *Epilepsia*. 52(10): p. 1761-1770.
- Prihodko, L. and S.N. Goward. 1997. Estimation of air temperature from remotely sensed surface observations. *Remote Sensing of Environment*. 60(3): p. 335-346.
- Stisen, S., Sandholt, I., Norgaard, A., Fensholt, R., Eklundh, L. 2007. Estimation of diurnal air temperature using MSG SEVIRI data in West Africa. *Remote Sensing of Environment*, 110(2): p. 262-274.

Xu, Y., Knudby, A., Ho, H.C., 2014. Estimating daily maximum air temperature from MODIS in British Columbia, Canada, *International Journal of Remote Sensing*, Volume 35, Issue 24: p. 8108-8121.

Yan, H., Zhang, J., Hou, Y., He, Y., 2009. Estimation of air Temperature from MODIS data in east China, *International Journal of Remote Sensing*, Volume 30, Issue 23: p. 6261-6275.

Yunus, F., Jaafar, J., Mahmud, Z., Shafie, A. & Idris, R. 2012. Role of environmental factors in modeling of air temperature element in peninsular Malaysia. *International Conference System Engineering and Technology (ICSET)*.

Zakšek, K. and M. Schroedter-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): p. 414-421.

Gago-Silva, A., 2016. GRASS GIS in Grid Environment. doi.org/10.6084/m9.figshare.3188950.