USING MODIS LST PRODUCTS TO ESTIMATE AIR TEMPERATURE

Ali Rahimikhoob
Department of Irrigation and Drainage Engineering, College of Aburaihan, University of Tehran , Pakdasht, Iran.
Email: akhob@ut.ac.ir

KEY WORDS: Satellite, ANN model; MODIS-Terra sensor; Iran

ABSTRACT: Air temperature at 2-m height (Ta) above ground surface is very important parameter for many areas of research and applications in various engineering fields. Air temperature is also a key parameter for monitoring the environment and predicting crop yield**.** Generally, a meteorological station provides air temperature as a synoptically representative value. The density of the station network is normally not sufficient and the estimation of Ta with an acceptable level of accuracy has been hindered by lack of high-resolution temporal and spatial data. However, the use of satellite data to estimate Ta at ground level has become an effective way. In this study, artificial neural network (ANN) with one hidden layer were used to estimate Ta in Khuzestan province (in the South West of Iran) using satellite remotely sensed land surface temperature (Ts) data acquired by the MODIS-Terra sensor. The input variables for the ANN model were the daytime and night time MODIS Ts, extraterrestrial solar radiation and Julian day. A total of 365 images of MOD11A1 T_s product for the year 2007, covering the area of this study were collected from the Land Processes Distributed Active Archive Center (LP DAAC). The results of this study showed that mean daily Ta can be estimated with acceptable levels of the statistical indicators from MODIS data and from the two geographic parameters using the ANN model. Root mean square error and R^2 for the comparison between observed and estimated Ta for the tested data are 1.9 ^{o}C and 0.97, respectively.

1. INTRODUCTION

Air temperature (Ta) near the Earth's surface is very important parameter for many areas of research and applications in various engineering fields. Daily mean air temperature is employed with humidity, wind speed and solar radiation to estimate evapotranspiration. Ta is also a key parameter for monitoring the environment, predicting crop yield and climatic changes. Ta is generally measured directly in weather stations with temperature sensors mounted in the shelter located 2m above the ground. Generally, a weather station provides air temperature as a synoptically representative value. The air temperature has temporal and spatial variations. The density of the station network (number of station per unit area) is frequently insufficient to represent the spatial distribution of air temperature at detailed spatial scales due to lack of high-resolution data, especially in developing countries. According to Lozada and Sentelhes (2008), one of the main problems in conducting agrometeorological studies is the lack of temperature data. However, the development of satellite remote sensing technology provides an opportunity to obtain such high-resolution data. Satellite remote sensing is an attractive tool for obtaining information about the ground surface for detecting regional variations of various surface phenomena. But most of the phenomena such as air temperature are not investigable by satellites directly; therefore the models with input data adapted to satellite images should be designed and compiled. There are two methods for estimating the air temperature from satellite images. One is the surface temperature/spectral vegetation index (TVX) method developed by Czajkowski *et al.* (1997) and Prihodko and Goward (1997). This method is based on the correlation between land surface temperature (Ts) and normalized difference vegetation index. The other is an empirical equation derived from statistical results for estimating air temperature. Vogt *et al.* (1997) applied a simple linear regression to estimate daily mean (Ta) and maximum air temperature (Tmax) based on NOAA-AVHRR Ts and good agreements were obtained. The accuracy of the air temperature derived by the TVX and the empirical methods always depends on the accuracy of the retrieved Ts. The objective of this study was to examine the potential of ANN model for estimating daily mean air temperature (Ta) using Ts data acquired by the MODerate Resolution Imaging Spectroradiometer (MODIS)/Terra sensor.

2. MATERTIALS AND METHODS

2.1 Study area and data

The area under study was Khuzestan province, which lies between latitudes 29.95^{o}N and 32.9^{o}N and between longitudes 47.6^{o}E and 50.6^{o}E. Khuzestan province is in the southwest of Iran, borders Iraq and the Persian Gulf, and covers an area of 63 238 km2. On the basis of the Koppen climate classification, Khuzestan province is categorized as having an arid climate. Air temperature observations on all acquisition dates of the images used in this study were obtained from 29 meteorological ground stations distributed throughout the province. The spatial distribution of selected stations is shown in Figure 1. The stations belong to the meteorological organization of Iran and cover most of Khuzestan province. The measurements included daily minimum and maximum air temperature (T_{max} and T_{min}, respectively). Daily air mean temperature (Ta) is recorded by calculating the average of the daily maximum and minimum temperature records. A description of the different weather stations and the annual mean weather data are given in Table 1.

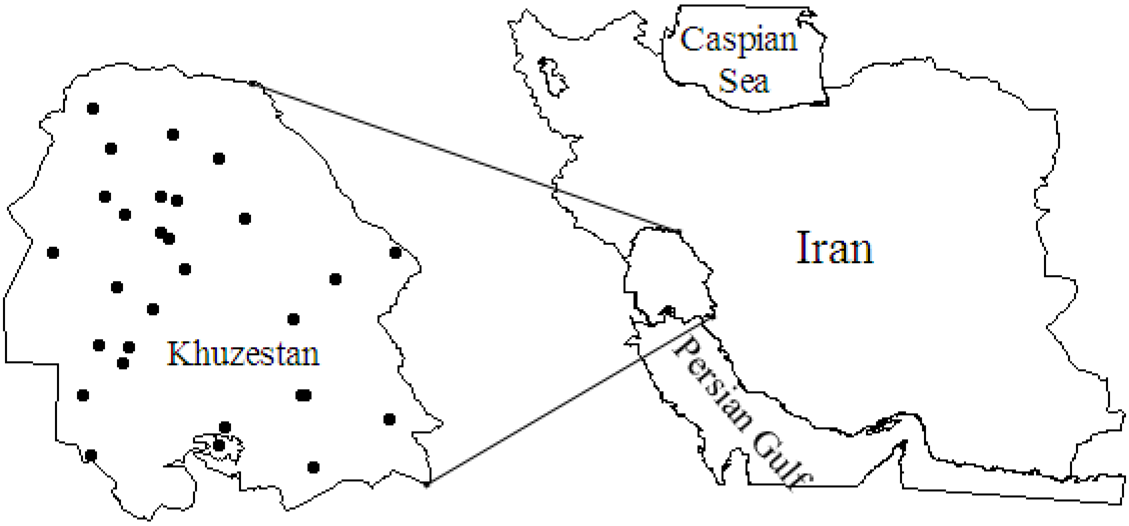

Figure 1- Study area and spatial distribution of the weather stations

2.2 MODIS images

In this study, Ts of day and nighttime data prepared by the MODIS sensor was used as an input to the ANN model. The MODIS sensor, on board of the Terra and Aqua satellites can provide images in 36 spectral bands between 0.62 and 14.385 μm with spatial resolutions of 250 m, 500m and 1 km in various bands. MODIS sensors have a 1 km spatial resolution in the thermal infrared (TIR). Terra and Aqua satellites were launched on 18 December 1999 and on 4 May 2002, respectively (Seemann *et al.*, 2003). These satellites orbit the earth in two distinct sun-synchronous paths. Terra passes from north to south across the equator at about 10:30 a.m. and p.m. (local solar time) at each day, while Aqua passes in the opposite direction from south to north over the equator at about 1:30 a.m. and p.m. (local solar time) at each day.

A total of 365 images of MOD11A1 *T*s product for the year 2007 from Terra-MODIS, covering the area of this study were collected from the Land Processes Distributed Active Archive Center (LP DAAC) (http://lpdaac.usgs.gov/). This product offers daytime and nighttime Land Surface Temperature data stored on a 1 km spatial resolution and gridded in the sinusoidal projection (Wang *et al.*, 2005). The MODIS Ts is derived from two TIR channels, i.e. 31 (10.78–11.28 μm) and 32 (11.77–12.27μm) using the split-window algorithm which corrects for atmospheric effects and emissivity using a look-up table based on global land surface emissivity in the TIR (Vancutsem *et al.*, 2010). The images were reprojected from the original Sinusoidal to the Universal Transverse Mercator coordinate system with WGS84 datum (UTM zone 39N) using the software MODIS Reprojection Tool (MRT). MRT exports resulting images to standard GIS data formats such as GeoTIFF, ArcInfo GRID or Imagine .IMG. The data in these formats were then analyzed using ArcGIS software. The locations of the ground weather stations used in this study were identified on the images, and the day and nighttime *T*s values at the Terra-MODIS products are extracted for those pixel locations.

2.3 Artificial Neural Network (ANN)

In this study, an ANN of the multilayer perceptron (MLP) type with one input layer, one hidden layer and one output layer was used for estimating Ta. MLP networks consist of units (neurons) arranged in layers (input, hidden and output layer) with only forward connections to units in subsequent layers. The number of nodes in the input and the output layers depends on the number of input and output variables, respectively.

In this study, the day and nighttime Terra-MODIS Ts (in ^{o}C), extraterrestrial solar radiation (in MJm^{-2} d^{-1}) and Julian day (number of days in the year starting 1 January) were selected as inputs to the ANN model for estimating daily mean air temperature. The extraterrestrial solar radiation and Julian day were used to reflect diurnal difference and seasonal variation in variation of surface air temperature, respectively. The extraterrestrial solar radiation was calculated as a function of the local latitude and Julian data using the equations provided by Allen et al. (1998). Since the purpose of this study was the estimation of Ta, the ANN has only one output variable.

The performance of the ANN depends on the number of nodes in the hidden layer. Because no specific guidelines exist for choosing the optimum number of hidden nodes for a given problem, this network parameter is often optimized using a combination of empirical rules and trial and error.

In this paper, the log sigmoid activation function is used for both hidden layer and output layer. This function is the most commonly used activation function. In this study, a backpropagation (BP) algorithm was employed to train our MLP neural network. Levenberg–Marquardt (LM), a second-order nonlinear optimization technique, was chosen from the various BP training algorithms available for use in this study. Lack of generalization is caused by overfitting. The network has memorized the training examples, but it has not learned to generalize new situations. Here, the LM algorithm was used with an early stopping criterion to improve the network training speed and efficiency. The accuracy of the networks was evaluated for each epoch in the training through mean squared error. For the criterion, all the data were divided into three sets. The first set is the training set for determining the weights and biases of the network. The second set is the validation set for evaluating the weights and biases and for deciding when to stop training. The last data set is for testing the weights and biases to verify the effectiveness of the stopping criterion and to estimate the expected network operation on new data sets. In this study, to train the ANN model, data from the 17 weather stations (station numbers 1 to 17 in Table 1) with a wide range of latitudes, longitudes, and elevations were collected into one group to produce a model for regional T_a estimation that could be applied for different locations in the Khuzestan province. This train data set had a total of 3511 patterns and was divided at random, with 70 % being reserved to train the ANN and 30% being used to validate the training. The remaining 12 stations (station numbers 18 to 29 in Table 1) were used to test the create model. The test data set had a total of 2398 patterns that were not used for training the model.

In order to suit the consistency of the model, all source data were normalized in the range 0.0 - 1.0 and then returned to original values after the simulation using:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

Table 1- Summary of weather station sites used in this study

Station Number	Station Name	Station Code	Lat. (°N)	Lon. (°E)	T_x (°C)	T_n (°C)
1	Baghemalak	BM	31.53	49.88	29.1	17.7
2	Bandar	BA	30.55	49.15	32.4	18.6
3	Ahwaz	AH	31.33	48.67	33.3	19.2
4	Omidieh	OM	30.77	49.67	35.1	18.3
5	Kesht	KE	31.80	48.77	32.3	17.2
6	Hafttape	HA	32.08	48.35	32.1	17.1
7	Debal	DE	31.08	48.50	34.1	19.3
8	Minab	MI	31.97	48.48	31.9	17.3
9	Masjed	MA	31.93	49.28	32.1	19.6
10	Shoshtar	SH	32.05	48.83	33.0	20.3
11	Salman	SA	30.98	48.47	34.2	19.4
12	Abadan	AB	30.37	48.25	33.8	18.9
13	Molasani	MO	31.60	48.88	32.1	18.1
14	Amirkabir	AM	31.10	48.30	34.3	19.5
15	Behbahan	BE	30.60	50.23	32.7	17.3
16	Bostan	BO	31.72	48.00	33.3	16.5
17	Daryache	DA	30.43	49.10	32.4	18.6
18	Dehdez	DH	31.72	50.27	283	17.1
19	Dezful	DZ	32.40	48.38	32.1	16.3
20	Hamidieh	HM	31.48	48.43	33.0	18.9
21	Hendijan	HE	30.28	49.73	35.7	19.3
22	Hosinieh	HO	32.67	48.27	32.7	16.0
23	Izeh	IZ	31.85	48.72	28.4	13.9
24	Keshtosanat	KS	32.08	48.72	32.8	19.8
25	Lali	LA	32.33	49.10	30.1	17.5
26	Mirza	MR	30.77	48.20	34.7	19.7
27	Omidieh	OI	30.77	49.65	34.7	18.7
28	Ramhormoz	RA	31.27	49.60	32.7	19.7
29	Sardasht	SR	32.50	48.80	29.1	16.5

2.4 Statistical analysis

Ta estimates from the ANN method were compared with the actual data using simple error analysis and linear analysis with the following parameters: coefficient of determination (R^2), mean absolute error (MAE) and root mean square error (RMSE).

3. RESULTS AND DISCUSSION

Before training ANN model for estimating Ta, the minimum and maximum air temperatures measured from the trained stations were compared with daytime and nighttime MODIS Ts derived from single pixels located over the stations. Figures 2 and 3 show these comparisons. As seen in Figure 2, there is a close relationship between Tmax and daytime Ts with the high values of coefficient of determination ($R^2 > 0.84$) for all sites. R^2 ranges from 0.84 to 0.93 and has an average value of 0.90. As seen from the fit line equations in Figure 2 (assume that the linear equation is $Y = mX + c$) in the scatter plots the slope of straight line (m) varies between 0.72 and 1.22 with an average of 0.90. The results in Figure 3 also show that there is a close relationship between Tmin and nighttime Ts for all sites. The R^2 changes from 0.84 to 0.94 with an average of 0.91. The slope of straight line varies between 0.81 and 1.05 with an average of 0.92. The results suggest that the daily mean air temperature could be computed with an appropriate method using daytime and nighttime Ts data.

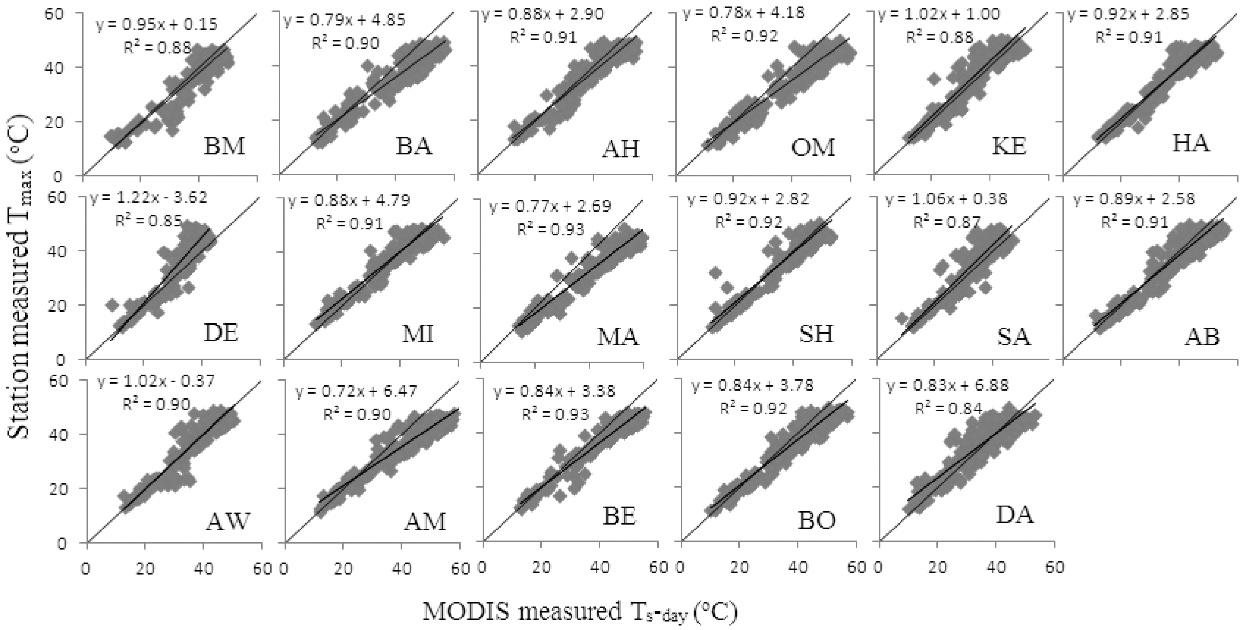

Figure 2- Comparison of station-measured maximum air temperature and daytime MODIS land surface temperature for the trained locations (see Table 1 for weather station codes).

The optimal node number in the hidden layer of the network was determined using a trial and error method by considering the MAE, RMSE and R^2 values from a test sample. In this study, ten ANNs were trained with one to 10 nodes in the hidden layer, and the aforementioned statistical parameters were calculated using only the whole test data set after each training run. Figure 4 shows the effect of changing the number of nodes in hidden layer on the network accuracy. It is clear that this factor had a significant bearing on the network accuracy. The variation of MAE oscillated more than the RMSE and R^2. Based on the three statistical results, the network that employed four nodes in the hidden layer provided the best results, with MAE, RMSE and R^2 values of -0.17 °C, 1.9 °C and 0.96, respectively for testing data.

The developed ANN model was preformed for the data set of test locations. Figure 5 shows scatter plots of the Ta measured values and those estimated by the proposed ANN model for test locations. As seen in this figure, all Ta data appear to be well distributed along the 1:1 line. Good coefficients of determination (R^2) were obtained at each station, with values above 0.92. R^2 ranges from 0.92 to 0.98 and has an average value of 0.96. The statistical summary of Ta estimates for test locations is presented in Table 3. The RMSE values ranges from 1.5 to 2.4 °C. The maximum RMSE was found to be 2.5 °C for Ramhormoz station, while the best result was found to be 1.4 °C for Lali station. The maximum MBE was found to be 1.6 °C, while the best result was found to be 0.0. The maximum R^2 between the measured and ANN-estimated Ta values was found to be 0.98 for Ramhormoz and Sardasht stations, while the minimum R^2 was found as 0.92 for Izeh station. According to the average R^2, RMSE and MBE (0.97, 1.9 °C and -0.2 °C respectively), the ANN model is suitable for converting the day and nighttime MODIS Ts to Ta in Khuzestan province.

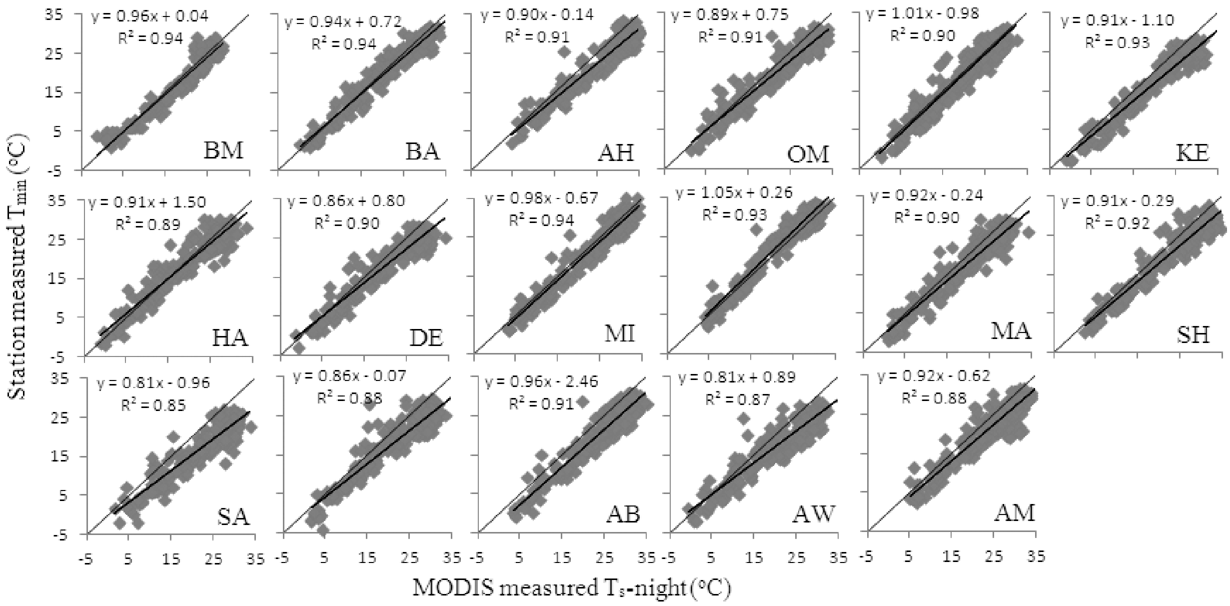

Figure 3- Comparison of station-measured minimum air temperature and nighttime MODIS land surface temperature for the trained locations (see Table 1 for weather station codes).

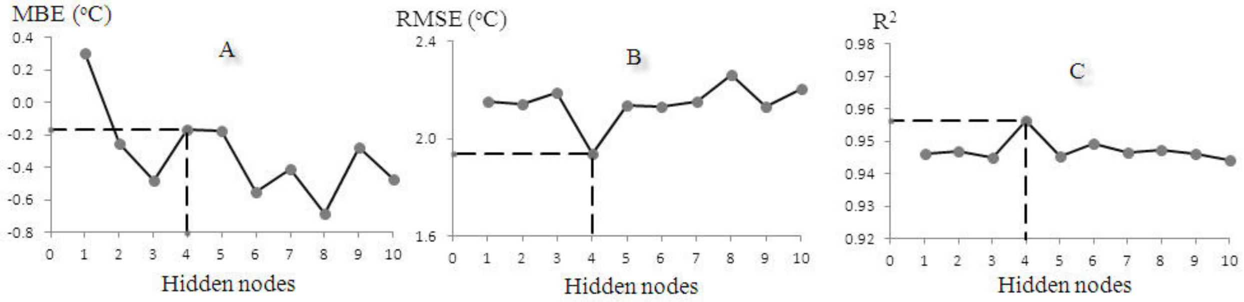

Figure 4- Relation between accuracy of neural network and the number of nodes for estimating Ta. A) MBE, B) RMSE and C) R^2

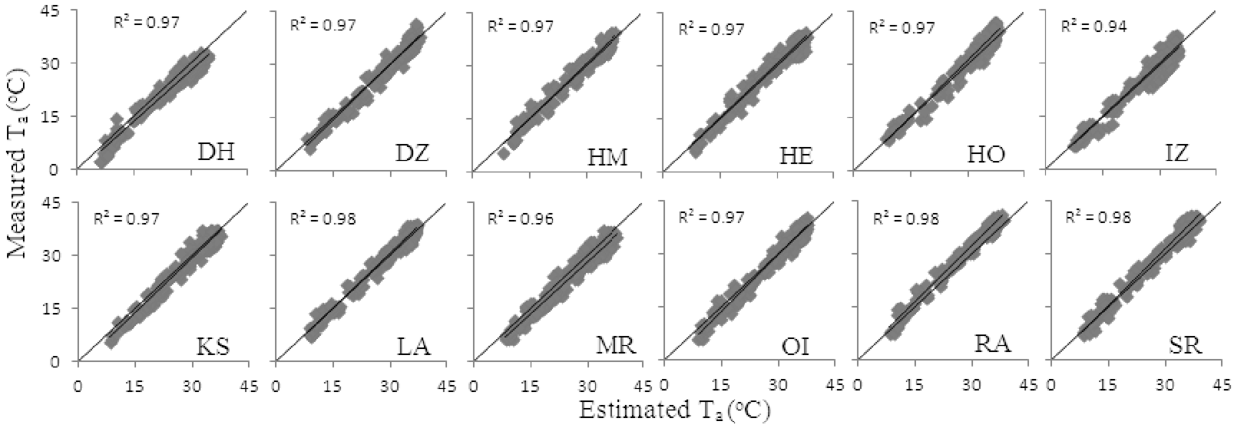

Figure 5- Scatter plots between measured and ANN-estimated Ta for the test data set (see Table 1 for weather station codes)

Table 2- Statistical summary of T_a estimates for test locations

Station Code	R^2	RMSE (^oC)	MBE (^oC)
DH	0.97	2.2	1.6
DZ	0.97	1.5	0.1
HM	0.97	1.6	-0.5
HE	0.97	1.7	-0.7
HO	0.97	2.1	-1.2
IZ	0.94	1.9	0.3
KS	0.97	1.7	0. 7
LA	0.96	1.4	-0.4
MR	0.97	2.3	1.5
OI	0.97	1.7	0.0
RA	0.98	2.5	-2.2
SR	0.98	2.2	-1.5
Average	0.97	1.9	-0.2

4. CONCLUSIONS

In this research, an alternative method for the operational assessment of daily Ta distribution at a regional scale was evaluated. The method is based on the ANN model, which relies primarily on Ts data from the MODIS sensor. The day and nighttime Terra-MODIS Ts, extraterrestrial solar radiation and Julian day were used as inputs with Ta as output. The proposed approach was applied to Khuzestan province. The study demonstrated that modeling of mean daily air temperature is possible through the use of ANNmodel with a precision of about 1.9 °C from Ts data.

References:

Czajkowski, K.P., Mulhern, T., Goward, S.N., Cihlar. J., Dubayah, R.O., Prince, S.D. 1997. Biospheric environmental monitoring at BOREAS with AVHRR observations. Journal of Geophysical Research, 102, pp. 29651–29662.

Lozada, B.L., Sentelhes, PC. 2008. Modelos de estimaci´on de las temperaturas promedios de la m´ınima, maxima y media diaria para la Regi´on andina de Venezuela. Agronom´ıa Tropical, 58(2),pp. 141–153.

Prihodko, L., Goward, S.N. 1997. Estimation of air temperature from remotely sensed surface observations. Remote Sensing of Environment, 60(3), pp. 335–346.

Seemann, S.W., Li, J., Menzel, W.P., Gumley, L.E. 2003. Operational retrieval atmospheric temperature, moisture, and ozone from MODIS infrared radiances. Journal of Applied Meteorology, 42, pp. 1072–1091.

Vogt, J., Viau, A.A., Paquet, F. 1997. Mapping regional air temperature fields using satellite derived surface skin temperatures. International Journal of Climatology, 17, pp. 1559–1579.

Wang, K., Wan, Z., Wang, P., Sparrow, M., Liu, J., Zhou, X., Haginoya, S. 2005. Estimation of surface long wave radiation and broadband emissivity using Moderate Resolution Imaging Spectroradiometer (MODIS) land surface temperature/emissivity products. Journal of Geophysical Research, 110 (11), pp. 1-12.