Modeling Air Temperature in Forested Areas using Machine Learning

[1]MASSOUD FOROOSHANI, [1,2]ALEXANDER GEGOV, [3]NICK PEPIN, [1]MO ADDA
[1]School of Computing The University of Portsmouth, Portsmouth, UK
[2]English Language Faculty of Engineering, Technical University of Sofia, BULGARIA
[3]School of the Environment, Geography and Geosciences, The University of Portsmouth, UK

Abstract-Air Temperature is a fundamental measure of the Earth's climate but is only measured at fixed locations. Land surface temperature can be measured widely using satellites. To estimate air temperature (Ta) from the surface temperature (Ts) measured on the forested slopes of Kilimanjaro, four models with unique sets of inputs were tested using five machine learning algorithms. The RMSE for each model was compared with a benchmark model. Models and algorithms were ranked according to their RMSE (Root Mean Square Error) The models and algorithms reliability and consistency ranking were calculated. The best model and algorithm were determined. Novel models results were compared with the benchmark model. All models outperformed the benchmark model in the consistency ranking while three out of four models outperformed the benchmark model in the reliability ranking. Thus machine learning improves the estimation of air temperature in this forested environment.

Keywords— Machine learning, air temperature, surface temperature, Kilimanjaro

Received: Agust 17, 2021. Revised: April 12, 2022. Accepted: May 9, 2022. Published: June 3, 2022.

1 Introduction

Kilimanjaro is the largest and highest free-standing mountain in the world. It lies approximately on the equator with a base lying below 1000m and the summit at 5895m. The landscape is naturally divided into several vegetation zones according to elevation. The forest zone extends from 1800 to 3000 m [1] and is the focus of this study. Other zones on the mountain (from highest elevations downwards) include the summit ice-fields, alpine desert, moorland, giant heather, cultivated land and finally the urban zone. Many of these ecosystems have attracted much attention because of their high impact on local and regional climate change.

The core problem associated with current climate change is the build up of carbon dioxide in the atmosphere. Forests play two unrivaled roles in this respect. First, they remove around 30% of carbon emissions released into the atmosphere due to fossil fuel burning, and second they store large reserves of carbon, amounting to double the amount of carbon in the atmosphere [2].

Cloud forests such as those on the lower slopes of Kilimanjaro also play another important role more locally in encouraging cloud formation, collecting cloud water and distributing it around the local watershed, enriching the surrounding ecosystem and providing a habitat for rare species [3].

Air temperature is one of the most important variables in the quantification of climate change [4][5][6] and many studies have suggested that mountain regions are warming faster than other locations. This phenomenon of elevation dependent warming (EDW) has been the subject of much research [4][7][8]. The increase in air temperature during the past decades has not only led to the retreat of glaciers on the upper slopes of Kilimanjaro but has also contributed towards wild fires that have destroyed nearly one third of Kilimanjaro's forest cover [3]. Arguably this has a more extensive overall impact on the whole Kilimanjaro ecosystem than retreating glaciers. This highlights the importance of obtaining reliable estimates of air temperature in the forested zone of Kilimanjaro.

The standard method to measure air temperature is directly at weather stations at 2m height above the surface. There are problems with this approach. The measurement is valid only for the precise location of the weather station and not a large area. Mountain regions in particular are often inaccessible and suffer from a lack of weather stations. The uneven distribution of stations, changes in instrument exposure times, and the lack of long time series and continuous records at all stations, are some of the other problems with weather station data. This data is therefore not always available and has limited spatial coverage.

The introduction of satellites has made it possible to measure the temperature of the Earth's surface over large areas. These data are nearly always available and have extensive spatial coverage in contrast with air temperature measurements that are limited to weather stations.

Using surface temperature measured by satellites (Ts) to estimate air temperature (Ta) is therefore an ongoing focus of research in climate change studies [1][9][10][11]. There are differences between the two variables. Surface temperature is highly dependent on the surface type and changes rapidly in space and time as the surface heats and cools in response to solar radiation. The air temperature shows more stability and, although measured at a fixed point, could be argued to be representative of the local mean temperature.

To model the non-linear and complex relationship between Ta and Ts, machine learning algorithms are a promising option compared with other statistical methods and are investigated in this paper. The next sections will cover past studies, methodology, data collection, data analysis, results and conclusions.

2 Past Studies

2.1 Climate Change

There have been many attempts to derive air temperature from the surface temperature in different environments. These include [9] in the Arctic, [10] in Canada and Alaska, [11] in Russia and China, [4] on the Tibetan Plateau in western China, [5] in Portugal, [1] and [6] in Africa. Not all of these have specifically focused on high mountain environments where the difference between air and surface temperature can become instantaneously large due to intense radiation at high elevations. They also cover a wide range of different vegetation zones including forests, deserts and snow covered landscapes. In all cases it is most common to build regression models to estimate air temperature from surface temperature. Although regression models are a solid framework for modeling and have been widely applied in the references above, the introduction of new machine learning algorithms to the research environment in recent years presents an alternative approach that needs to be evaluated.

2.2 Machine Learning

The application of machine learning algorithms in climate science and weather forecasting goes back to the works of [12] and [13] who investigated the application of Expert Systems (ES) and Artificial Neural Network (ANN) respectively.
Machine learning has also been applied to the prediction of air temperature from surface temperature but in a limited way. The research papers [14], [15], [16], [17], and [18] all use ANN (Artificial Neural Networks) for this purpose. However, other machine learning algorithms including ANFIS (Adaptive Neuro Fuzzy Systems) have been so far restricted to weather forecasting applications and have not been used to estimate air temperature from surface temperature in a climate context. These past research examples also commonly used variable types other than Ta and Ts to estimate air temperature. The combination of a wide variety of machine learning algorithms with the core variables could present a simple but equally efficient approach to the estimation of air temperature from surface temperature.

2.3 Summary

Past research on the application of machine learning algorithms in the estimation of air temperature is limited to a few algorithms. This research therefore will evaluate the application of several machine learning algorithms using only the two core variables, namely surface temperature (Ts) and air temperature (Ta) to present a novel and simple but efficient approach to the estimation of air temperature from surface temperature.

3 Research Methodology

Modeling of large scale, complex, non-linear, ill-defined, and uncertain systems such as climate change systems has been a prime concern for a long time. The application of machine learning (ML) algorithms such as fuzzy systems and neural networks have opened a path for more ML algorithms to be tested and used in this field. Five main algorithms were employed in this study (described below).

3.1 ANFIS (Adaptive Neuro Fuzzy System)

ANFIS is an implementation of a FIS (Fuzzy Inference System) on top of the architecture of an ANN (Artificial Neural Network) combining the power of a fuzzy rule base with the learning capability of neural networks. For a discussion see [19].

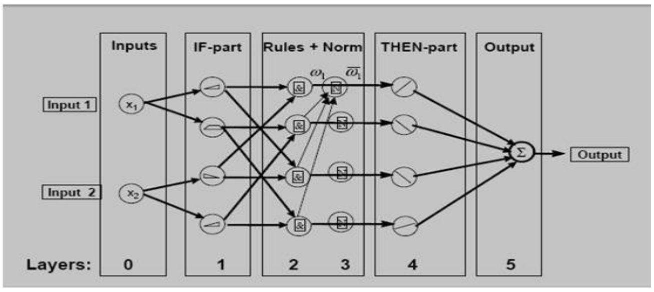

figure 1: ANFIS architecture [20]

3.2 Linear Regression

Linear regression is modelling of the relationship between one or more linear independent variables to predict a dependent variable. The basic regression model for one independent variable is in the form of

$$\boldsymbol{y}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 + \boldsymbol{X}_i + \boldsymbol{\epsilon}_i \quad (1)[21]$$

where $\boldsymbol{y}_i$ is the response variable in the $\boldsymbol{i}^{th}$ trial

$\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ are parameters

$\boldsymbol{X}_i$ is a known constant (the value of the independent variable in the $\boldsymbol{i}^{th}$ trial)

$\boldsymbol{\epsilon}_i$ is a random error

$\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ are called regression coefficients.

$\boldsymbol{\beta}_1$ is the slope of the regression line.

$\boldsymbol{\beta}_0$ is the Y-intercept of the regression line.

3.3 Polynomial Regression

Polynomial multiple regression models are special cases of the general linear regression models that can have more than one independent variable and variables can take various powers. The general form for one independent variable in second order is:

$$\boldsymbol{Y}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{X}_i + \boldsymbol{\beta}_2 \boldsymbol{X}_i^2 + \boldsymbol{\epsilon}_i \quad (2)[22]$$

3.4 Support Vector Machine

SVM is one of the most popular ML algorithms, developed by [22]. It was packaged as LIBSVM library by [23] to make application easier.

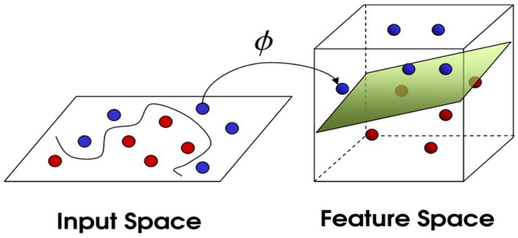

figure 2: Support Vector Machine [24]

SVM maps the input vectors into a high dimensional feature space Z through non-linear mapping chosen a priori. In this space a linear decision surface is constructed with special properties that ensures high generalization ability of the network.

3.5 Simple Regression Tree

Regression trees are a type of decision tree that targets continuous variables. This algorithm builds a tree to predict the output from various inputs. In the recursive partitioning mode. The space is continuously divided into smaller areas that contain a simple model, and therefore the global model has two parts, the recursive partitioning and the simple model. The regression tree uses a tree to represent the recursive partitioning in which each cell or terminal node contains a simple model. The model in each node is a constant estimate of the output.

If the points; $(\boldsymbol{X}_1, \boldsymbol{Y}_1), (\boldsymbol{X}_2, \boldsymbol{Y}_2), \ldots (\boldsymbol{X}_c, \boldsymbol{Y}_c)$ are all the samples belonging to the leaf-node I. Then the model for I is:

$$\hat{\boldsymbol{y}} = \frac{1}{\boldsymbol{C}} \sum_{i=1}^{c} \boldsymbol{y}_i \quad (4)[25]$$

4 Data Collection and Analysis

4.1 Data

The full data-set consists of air and surface temperatures recorded at 22 sites across Kilimanjaro between 990 and 5803 m above sea level [26]. It has been used before by [1] in a preliminary comparison of air and surface temperatures across the mountain. In this study four sites within the forest zone were selected, one on the north-ease slope and three on the south-west slope of the mountain. The range of elevation is from 1890 to 2745m.

The air temperature (Ta) at each site is recorded using an automatic data loggers (Hobo U23-001) installed in a radiation shield at 2 m above ground level. Observations were recorded as an instantaneous value every 30 minutes.

The Surface temperature (Ts) is retrieved from the Terra satellite and consists of the MODIS product MOD11A2 which provides an 8-day mean surface temperature at 1km by 1km resolution. The mean time of the satellite overpasses is 1030 local solar time (day) and 2230 local solar time (night).

For comparison with Ts the mean air temperatures were taken at 1030 and 2230 EAT (East African Time) averaged over the same 8 day periods as the surface temperature were used.

4.2 Variables

Five variables were defined, four of which represented day (1030) and night (2230) air and surface temperatures. The novel variable ΔTs was defined as the difference between day and night surface temperatures (and is a proxy for solar radiation). Four variables were used as input and one variable was used as output (TaD).

Variables		
Input/output	**Variable**	**Description**
output	TaD	Air temperature of day
inputs	TaN	Air temperature of night
	TsD	Surface temperature of day
	TsN	Surface temperature of night
	ΔTs	Solar radiation =TsD - TsN

TABLE.1 VARIABLES

4.3 Models

Using a benchmark model in machine learning is a standard way of evaluating/comparing the performance of novel models with an accepted standard. The benchmark model is applied to our research data and results compared with the results from the novel models. The benchmark model simulation was based on research presented in [27], in which ANFIS was used to predict air temperature. The air temperature was used as input and output. The benchmark simulation used TaN as input and TaD as output.

Four different sets of inputs as four novel models were evaluated for the first time to estimate daytime Ta. Different combinations of these variables each have a meaning in the context of climate change studies (see table 2).

Models			
Model	**Acronym**	**Inputs**	**Output**
Model-1	m1	TsN, TaN, TsD	TaD
Model-2	m2	TsN, TsD	TaD
Model-3	m3	TaN, ***ΔTs***	TaD
Model-4	m4	***ΔTs***	TaD

Table.2 Models

4.4 K-fold Cross Validation

The selection of 4-fold cross validation as a performance metric was based on the minimum of data rows available for one-fold.

4.5 Data Sets

The following naming conventions and descriptions were used:

- The testing data set contained 20% of the main data set and its objective was to test the generalizability of the trained and cross validated model with unseen data.
- The learning data set contained 80% of the main data set from which the training (75%) and checking (25%) data sets were selected for 4-fold cross validation to prevent overfitting of the model. The average RMSE was calculated and used as the main performance metric for each model.

4.6 Data Preprocessing

Requirements that determined data per-processing include:

- Two software were used. MATLAB ANFIS GUI [28] needed a special data preparation process. KNIME analytical platform [29] used the same data files prepared for MATLAB.
- Machine learning analysis stages of training, checking, and testing needed different data sets prepared for each stage.
- K-fold cross validation: 4-fold cross validation selected regarding the minimum number of data rows needed for each fold. Data needed to be prepared for each fold individually.
- Variables needed to be extracted from the main data files.
- Novel models with different inputs needed separate data sets.

5 Simulation Results

5.1 Models RMSE

Table 3 contains the RMSE (between observed and predicted Ta) for the four novel models (m1-m4) and the benchmark model (bm) using each of the five algorithms. Figures are the average RMSE of the 4-fold cross validation. The RMSE unit is Celsius degrees and should be interpreted in the context of the climate change studies in which ‘’ errors generally fall in the 2–3 °C range while the level of precision generally considered as accurate is 1–2 °C [30]. These accuracy ranges were regarded in interpreting the results.

Model 4 is universally the poorest in performance (RMSE between 3.5 and 4.7˚C) meaning that ΔTs (solar radiation) as a sole input can not be used to estimate the air temperature in the forest zone. The other three models tend to be fairly similar and RMSE is usually between 2 and 3˚C (greyed). (see appendices, table 3)

5.2 The Best Model

The best model in Table 3 is model-2 combined with the ANFIS algorithm. This model gained the average (4-folds) RMSE = 2.0314 which is in the acceptable accuracy range while fold-3 (fm2fold3) of this model gave and RMSE = 1.9899 as the best model in the testing stage with unseen data which is in the ideal accuracy range (1 – 2 ℃) Testing data in figure 3 is presented with blue dots where the FIS (Fuzzy Inference System) output is presented with red asterisks. (see appendix, figure 3)

The correlation between model-2 inputs (TsN, TsD) and the output (TaD) in the best model is presented in figure 4. The smooth surface suggests a strong correlation between inputs and the output (see appendix, figure 4)

5.3 Model Ranking

To compare the various model and algorithm combinations in more detail they were ranked from best performing (R1) to worst (R25) in Table 4. The following points can be concluded:

- Model-1 was the best performing model for three algorithms, although it does contain the most inputs.
- The best overall combination was model-2 combined with ANFIS.
- The best algorithm averaged across all models was ANFIS.
- LIBSVM and Simple regression trees tended to perform relatively poorly overall (see appendix, table 4)

5.4 Model Reliability and Consistency Ranking

Table 5 summarizes the reliability and consistency rankings for each model. To determine model reliability the mean ranking was used. To determine model consistency the range in the ranking (difference between best and worst ranks) was used. A lower mean ranking presents higher reliability. A lower variation in rankings means higher consistency.

- Models-1 is the best in reliability ranking across all algorithms followed by models 2 and 3. the differences in consistency ranking reflect the differences between different input variables.
- Model-4 did not work well in the forest zone, therefore its high consistency should be seen in the context of RMSE results gained by each algorithm (i.e. it is consistently poor)
- Model-2 and Model-1 have the same ranking variation of 17, but Model-2 has lower boundaries than Model-1 so has been ranked as third best in consistency ranking.
- The benchmark model comes after novel models m1, m2, and m3 in reliability ranking whereas in the consistency ranking is the last.

Models Reliability and Consistency Ranking				
Model	Ranking Average	Reliability Ranking	Ranking Variation	Consistency Ranking
m1	9.2	1	17	4
m2	9.8	2	17	3
m3	10	3	13	2
m4	22.2	5	5	1
bm	13.8	4	20	5

Table.5 Models reliability and consistency ranking

5.5 Algorithm Reliability and Consistency ranking

The same concepts were used to determine the reliability and consistency rankings of each algorithm in table 6. ANFIS came up as the best algorithm in reliability ranking across all models followed by Polynomial regression and Linear regression algorithms. Linear regression is the most consistent. The differences in consistency ranking should be referred as differences between algorithms and models. The most reliable algorithms are not the most consistent in performance.

Algorithms Reliability and Consistency Ranking				
Algorithm	Ranking Average	Reliability Ranking	Ranking Variation	Consistency Ranking
ANFIS	6	1	19	5
Polynomial regression	9.8	2	17	4
Linear regression	15	3	9	1
SVM	15.2	4	16	3
Simple regression tree	19	5	9	2

Table.6, Algorithms Reliability and Consistency Ranking

5.6 Performance Evaluation

The performance of novel models was compared with the benchmark model. Overall the novel models outperformed the benchmark model:

- 100% (four out of four models) better in the consistency comparison
- 75% (three out of four models) better in the reliability comparison

6 Discussion

The forest zone of Kilimanjaro has a generally stable temperature regime with slow changes that make it relatively easy to predict Ta from Ts, in comparison with other environments on the mountain (not shown in this paper) which can experience rapid fluctuations. Therefore both Ts and Ta show considerable memory from day to day and can be used for predicting each other. Models 1 and 2 both work well and both include Ts during the day and night. This implies high coupling between air and simultaneous surface temperatures. A proxy for solar heating alone (model 4) is less successful, both due to the high number of cloudy days, and the fact that temperature is controlled as much by transpiration and latent heat flux in the forest, as it is by direct energy balance.

In the forest, the "surface" temperature is actually strongly influenced by the canopy of the forest (up to 20-30m above ground level) which is measured as the effective surface by the satellites. This canopy temperature is quite well coupled with air temperature within the forest, thus explaining the success of the models which use Ts as a predictor for Ta.

Higher up the mountain where there is much less vegetation, the surface measured by the satellite is much nearer ground level, and it is likely to be decoupled from the air temperature measured at 2 m well above the vegetation. Therefore additional work will be required to transfer these findings to other environments on the mountain and elsewhere.

This research used the zone data to cover the forest area. There are four stations in this area. Three stations are located on the north-east wall and the fourth is located on the south-west wall of Kilimanjaro reviving different levels of solar radiation. Further research can focus on the stations to investigate the impact of the location on the models.

7 Conclusion

The research confirms the reliability of machine learning algorithms (especially ANFIS) to estimate air temperature from satellite-measured surface temperature in a remote forested environment with few measured climate variables. The coupling between air temperature and surface temperature ensures model success in the forested zone of Kilimanjaro. The results could be applicable to other forested areas. Further research however is required to apply this approach to other areas and land-cover types on the mountain, and further afield.

References:

[1] N. Pepin, E. E. Maeda, and R. Williams, 'Use of remotely sensed land surface temperature as a proxy for air temperatures at high elevations: Findings from a 5000 m elevational transect across Kilimanjaro', Journal of Geophysical Research: Atmospheres, vol. 121, no. 17, Art. no. 17, 2016.

[2] J. G. Canadell and M. R. Raupach, 'Managing forests for climate change mitigation', science, vol. 320, no. 5882, pp. 1456–1457, 2008.

[3] A. Hemp, 'Climate change and its impact on the forests of Kilimanjaro', African Journal of Ecology, vol. 47, pp. 3–10, 2009.

[4]N. Pepin, H. Deng, H. Zhang, F. Zhang, S. Kang, and T. Yao, 'An examination of temperature trends at high elevations across the Tibetan Plateau: the use of MODIS LST to understand patterns of elevation-dependent warming', Journal of Geophysical Research: Atmospheres, 2019.

[5] A. Benali, A. Carvalho, J. Nunes, N. Carvalhais, and A. Santos, 'Estimating air surface temperature in Portugal using MODIS LST data', Remote Sensing of Environment, vol. 124, pp. 108–121, 2012.

[6] C. Vancutsem, P. Ceccato, T. Dinku, and S. J. Connor, 'Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa', Remote Sensing of Environment, vol. 114, no. 2, Art. no. 2, 2010.

[7] E. Palazzi, L. Mortarini, S. Terzago, and J. Von Hardenberg, 'Elevation-dependent warming in global climate model simulations at high spatial resolution', Climate dynamics, vol. 52, no. 5–6, Art. no. 5–6, 2019.

[8] D. Guo and H. Wang, 'The significant climate warming in the northern Tibetan Plateau and its possible causes', International Journal of Climatology, vol. 32, no. 12, Art. no. 12, 2012.

[9] M. Urban, J. Eberle, C. Hüttich, C. Schmullius, and M. Herold, 'Comparison of satellite-derived land surface temperature and air temperature from meteorological stations on the pan-Arctic Scale', Remote Sensing, vol. 5, no. 5, pp. 2348– 2367, 2013.

[10] S. Hachem, C. Duguay, and M. Allard, 'Comparison of MODIS-derived land surface temperatures with ground surface and air temperature measurements in continuous permafrost terrain', The Cryosphere, vol. 6, no. 1, pp. 51–69, 2012.

[11] S. Shen and G. G. Leptoukh, 'Estimation of surface air temperature over central and eastern Eurasia from MODIS land surface temperature', Environmental Research Letters, vol. 6, no. 4, p. 045206, 2011.

[12] W. Moninger et al., 'Summary of the First Conference on Artificial Intelligence Research in Environmental Sciences (AIRIES)', Bulletin of the American Meteorological Society, vol. 68, no. 7, Art. no. 7, 1987.

[13] D. W. McCann, 'A neural network short-term forecast of significant thunderstorms', Weather and Forecasting, vol. 7, no. 3, Art. no. 3, 1992.

[14] D. Zhao, W. Zhang, and X. Shijin, 'A neural network algorithm to retrieve nearsurface air temperature from landsat ETM+ imagery over the Hanjiang River Basin, China', in 2007 IEEE International Geoscience and Remote Sensing Symposium, 2007, pp. 1705–1708.

[15] J.-D. Jang, A. Viau, and F. Anctil, 'Neural network estimation of air temperatures from AVHRR data', International Journal of Remote Sensing, vol. 25, no. 21, pp. 4541–4554, 2004.

[16] M. Hayati and Z. Mohebi, 'Application of artificial neural networks for temperature forecasting', World Academy of Science, Engineering and Technology, vol. 28, no. 2, pp. 275–279, 2007.

[17] C. N. Schizas, S. Michaelides, C. S. Pattichis, and R. Livesay, 'Artificial neural networks in forecasting minimum temperature (weather)', in 1991 second international conference on artificial neural networks, 1991, pp. 112–114.

[18] K. A. KUMARI, N. K. BOIROJU, T. Ganesh, and P. R. REDDY, 'Forecasting surface air temperature using neural networks', International Journal of Mathematics and Computer Applications Research, vol. 3, pp. 65–78, 2012.

[19] J.-S. Jang, 'ANFIS: adaptive-network-based fuzzy inference system', IEEE transactions on systems, man, and cybernetics, vol. 23, no. 3, pp. 665– 685, 1993.

[20] Wikipedia, 'File:Anfis.JPG', Sep. 2020, Accessed: Sep. 05, 2020 [Online]. Available: https://en.wikipedia.org/wiki/File:Anfis.JPG

[21] J. Neter, M. H. Kutner, C. J. Nachtsheim, W. Wasserman, and others, 'Applied linear statistical models', 1996.

[22] C. Cortes and V. Vapnik, 'Support-vector networks', Machine learning, vol. 20, no. 3, pp. 273–297, 1995.

[23] C.-J. Lin and C. Chang, 'LIBSVM: a library for support vector machines, 2001', Software available at, vol. 10, no. 1961189.1961199, 2001.
[24] Shehzadex, 'support vector machine.png', WIKIMEDIA COMMONS, Nov.27,2021https://commons.wikimedia.org/wiki/File:Kernel_yontemi_ile_veriyi_daha_fazla_dimensiyonlu_uzaya_tasima_islemi.png
[25] L. Brieman, J. H. Friedman, R. A. Olshen, and C. J. Stone, 'Classification and regression trees', Wadsworth Inc, vol. 67, 1984.
[26] N. C. Pepin, 'Meteorological data for 22 sites across Kilimanjaro'.University of Portsmouth, 2015 2004. https://researchportal.port.ac.uk/en/datasets/dataset-for-kilimanjaro-climate-2012-present
[27] P. Kumar, 'Minimum weekly temperature forecasting using ANFIS', Computer Engineering and Intelligent Systems, vol. 3, no. 5, pp. 1–6, 2012.
[28] MATLAB version R2020a. Natick, Massachusetts: The Mathworks, Inc., 2020.
[29] Kime Analytics Platform version 4.3.4. Zurich, Switzerland: KNIME AG, 2020.
[30] A. Benali, A. Carvalho, J. Nunes, N. Carvalhais, and A. Santos, 'Estimating air surface temperature in Portugal using MODIS LST data', Remote Sensing of Environment, vol. 124, pp. 108–121, 2012.

Appendix*:*

Models RMSE							
Models	Inputs	ANFIS	Polynomial regression	Linear regression	LIBSVM	Simple regression tree	Average RMSE
m1	TsN, TaN, TsD	2.042025	2.12	2.337	2.204	2.973	2.335205
m2	TsN, TsD	2.0314	2.182	2.441	2.262	2.934	2.37008
m3	TaN, ΔTs	2.069875	2.2	2.442	2.23	2.827	2.353775
m4	ΔTs	3.667775	3.689	3.686	4.186	4.758	3.997355
bm	TaN	2.08345	2.214	2.478	4.186	2.88	2.76829
Average RMSE		2.378905	2.481	2.6768	3.0136	3.2744	

TABLE.3 MODELS RMSE

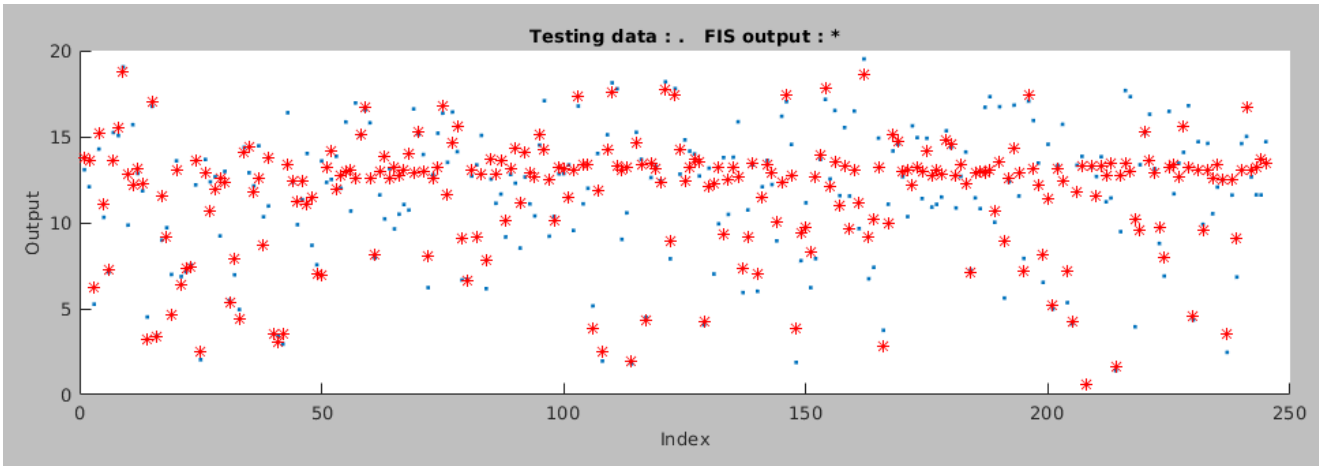

Figure 3: The Best Model Test Results

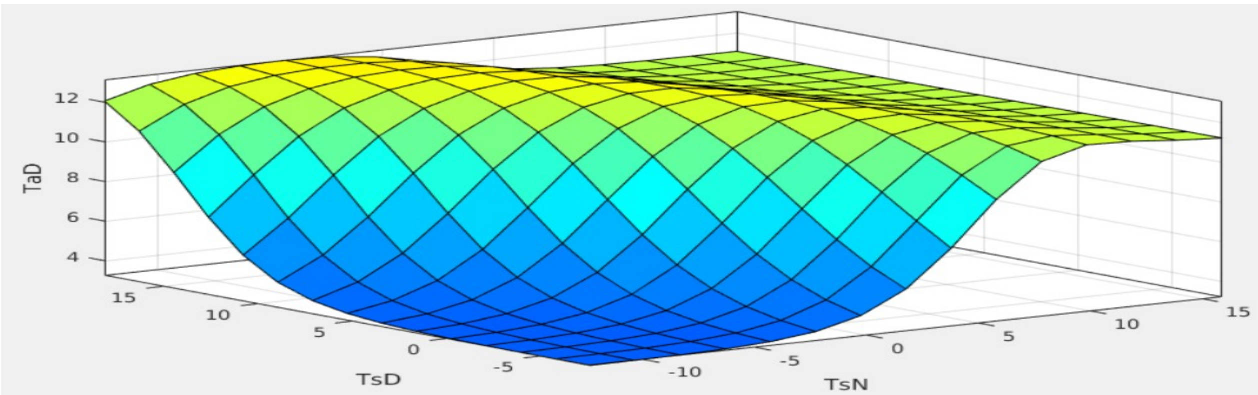

Figure 4: The Best Model Surface Plot

Models Ranking						
Models	Inputs	ANFIS	Polynomial regression	Linear regression	LIBSVM	Simple regression tree
m1	TsN, TaN, TsD	R2	R5	R12	R8	R19
m2	TsN, TsD	R1	R6	R13	R11	R18
m3	TaN, ΔTs	R3	R7	R14	R10	R16
m4	ΔTs	R20	R22	R21	R23	R25
bm	TaN	R4	R9	R15	R24	R17

TABLE.4 MODELS RANKING