Research Article

Air Temperature Modeling Based on Land Surface Factors by the Cubist Method (Case Study of Hamoun International Wetland)

Farhad Zolfaghari,1 Hasan Khosravi,2 and Shahram Khalighi Sigaroudi2

1Higher Education Complex of Saravan, Saravan, Iran
2Faculty of Natural Resources, University of Tehran, Karaj, Iran

Correspondence should be addressed to Farhad Zolfaghari; zol.farhad@gmail.com

Received 27 July 2023; Revised 11 December 2023; Accepted 29 December 2023; Published 8 January 2024

Copyright © 2024 Farhad Zolfaghari et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The drying up of Hamoun International Wetland (HIW) and the loss of vegetation in this area have led to an increase in ambient temperature. This research examines the changes in the surface of HIW and its role in air temperature (Tair) using data on land surface temperature (LST), vegetation, wind speed, and relative humidity. The Cubist regression model (CRM) is used to simulate the effects of land surface factors (LSFs) on Tair. Four microsites with different plant cover percentages were selected for this purpose. After data collection, 75% of the data were used for modeling and 25% of the data were used for model testing. The results showed that CRM has adequate performance for estimating Tair. The assessment of the relationship between land surface temperature (LST) and Tair at 2 meter height showed that there was a high correlation coefficient between 0.86 and 0.91 in the different microsites. The results of using CRM for estimating Tair showed that this model can estimate air temperature from independent parameters of LST, wind speed, vegetation percentage, and relative humidity with a correlation coefficient of 0.98. In this model, the LST, relative humidity, and vegetation percentage were entered with values of 100%, 93%, and 83% respectively. Wind speed was not included in the model because the measurements were constant and less than 4 m/s throughout the period (no changes).

1. Introduction

Plant cover and climate change are interrelated at various temporal and spatial scales [1]. Land degradation is the primary causal factor of climate change and the rise of air temperature. Land surface temperature (LST) rises significantly in the lake areas that were dried by evaporation, indicating a high correlation between the land-use change and LST. Due to the high surface temperature, the area is vulnerable to natural hazards, diseases, severe droughts, and reduced soil productivity [2]. These areas have also been observed to be covered with salt crusts, which may increase the risk of desertification. The Normalized Difference Built-up Index (NDBI), the percentage of vegetation (PV), the Normalized Difference Vegetation Index (NDVI), altitude, and wind speed have been identified as important factors for urban heat island (UHI) intensity during the day and at night [3]. The paper suggests that expanding green spaces and reducing anthropogenic emissions in the megacity’s center could be a mitigation plan for diminishing LST and UHI effects [3]. In addition, studies conducted on the influence of greenery on mitigating UHI have indicated that all green spaces help urban areas adapt to the impact of UHI regardless of whether they are parks, street trees, or green roofs [3].

This study assesses the relationship between vegetation degradation and air temperature in the case of the Hamoun International Wetland (HIW). Land cover change plays an important role in the climate system [4]. Numerous studies show differences in the response of vegetation temperature when air temperature changes by only a few degrees Celsius [5]. Vegetation temperatures are often warmer than the air during the day and cooler than the air at night [6]. Changes in land cover and climate have interactive effects [7]. Human
activities have significantly altered 42–68% of the global land surface by converting natural vegetation into crops, pastures, and woods for harvesting purposes [8]. There are multiple types of land cover change including deforestation, afforestation, irrigation, and destruction of the wetland. Temperature extremes are known to be changing globally [9]. Plant cover change supplies land surface temperature that enables evaporative cooling to suppress high-temperature extremes and may affect low-temperature extremes. The results of studies in the Hamoun International Wetland Basin using microscale exponential models with different climatic scenarios (RCP4.5 and RCP8.5) show that the seasonal and annual average minimum and maximum temperatures for various periods are rising [10]. According to a study on temperature trends in Iran, the average temperature at 94% of the 31 synoptic stations investigated exhibited an increasing trend, with the temperature increase being significant at the 0.05% level in approximately 70% of these stations [11]. The research also indicated that hot and desert areas experienced a more pronounced rise in air temperature than other regions [11]. The study’s findings suggest that Iran’s climate has become hotter and drier over the past six decades, and if the current trend of global warming persists, this issue is expected to be exacerbated in the future [11]. A study that utilized the SHArP generator to model maximum and minimum temperature data found a strong and statistically significant correlation between the data generated by the model and the observational data. The correlation coefficient ranged from 0.8 to 0.95. The study’s results suggest that the SHArP generator is an effective tool for modeling temperature data [12].

A study that utilized MODIS-Aqua data to analyze land surface temperature (LST) variations over Iran and the modified nonparametric Mann–Kendall test to examine trend significance and Sen’s slope to calculate changes in the rate and direction found that environmental factors, including land cover and elevation, have significant effects on LST trends. Inland waters and swamps experienced positive daytime trends during the warm months of the year and negative nighttime trends during cold months. Mountainous regions experienced severe positive daytime variations in cold months and similar changes in warm months. Changes in land cover and use have resulted in LST variations, with greater daytime than nighttime variations occurring at far higher rates for four land covers: waterbodies, cropland, urban areas, and barren land [13].

To evaluate the effects of land use/land cover (LULC) changes from the drying of Toshka Lakes on land surface temperature (LST), Landsat series Thematic Mapper (TM) and Operational Land Imager (OLI) satellite images were used to estimate LST from 2001 to 2019 using remote sensing and geographic information system (GIS) techniques [2]. The results indicated that the dried areas of the lakes were converted to bare soil and covered by salt crusts, leading to a decrease in the lakes by about 1517.79 km² and an average increase in LST by about 25.02°C from 2001 to 2019. The mean annual LST increased considerably by 0.6°C/year from 2009 to 2019. A strong negative correlation between LST and the Toshka Lakes area (R-square = 0.98) estimated from regression analysis implied that Toshka Lakes drying considerably affected the microclimate of the study area [2].

A study that utilized Sentinel/MODIS data and Google Earth Engine to examine the impact of urban heat islands (UHIs) on Greater Cairo’s land surface temperature (LST) found that the annual average UHI intensity was higher in Egypt’s megacity at night (3.27 K) than during the daytime (0.50 K). Sentinel-3 SLSTR thermal images were utilized to derive annual and seasonal day-night LST during the research period. The seasonal day-night UHI intensities have the highest values during the winter, followed by the summer. Significant influencing factors concerning UHI intensity varied between the daytime and nighttime. NDBI, PV, NDVI, altitude, and wind speed were the driving factors that were influential during the day and at night [3].

The regulation of Tair is an important topic of interest in recent studies. Monitoring LST in different plant covers is a quantitative method that can help us better understand how plants respond to their environment. This method can also guide decision-making regarding Tair management on dry land. Charney’s studies (1975) were among the first to investigate the effect of land surface change on climate [14]. Earth-atmospheric thermology referred to several studies [15–19], which examined the effect of land use change and land cover on Tair, precipitation, and fluxes. Changes in the land surface have important impacts on air temperature and their relevance for climate-change projections [20]. A better understanding and complement of the adequate processes would significantly help reduce uncertainties in studying future climate scenarios. Although plants have evolved mechanisms to achieve some degree of self-cooling ability, plant self-cooling ability is still poorly understood [5].

A study that utilized machine learning (ML) to evaluate and forecast urban Tair in Yerevan, Armenia, while accounting for intricate terrain characteristics concluded that remote sensing is an effective approach for estimating the distribution of Tair in areas where a dense network of weather stations is not available [21].

A study conducted in the HIW region observed that from 1983 to 2015, the average maximum temperature increased by 2.5°C, and the average minimum temperature increased by 1.8°C in the Sistan region [22]. The study’s results indicate that the average increase in temperature in the Sistan region is greater than the global average increase in temperature. The drying up of HIW and the loss of vegetation in this area seem to have a close relationship with the increase in ambient temperature. In this research, the aim is to model Tair using the CRM in HIW and investigate the role of changes in the wetland surface on air temperature. Based on this hypothesis, the role of LST, vegetation cover, wind speed, and relative humidity in Tair is investigated based on CRM. CRM and minimum LSF can be used to produce temperature data with acceptable reliability in areas without meteorological stations. This approach can help fill the gaps caused by the lack of data in dry areas and far from compensated availability using modeling. The study’s results suggest that this approach can produce air temperature (Tair) data with acceptable reliability.
2. Study Area and Data

The Sistan plain with an extension of about 15197 km² is located in the northeast of Sistan and Baluchestan province (Iran), at the end of the Helmand River Basin and near the Afghanistan border (see Figure 1). This plain has specific geomorphological unevenness despite relative homogeneity. Sistan climate is xerothermic according to all climatic classifications. The Sistan region is located at 21° 20' of northern latitudes and 61° 29' of east longitudes. The average altitude of the plane is 490 m. The annual average rainfall is 61.4 mm/year, while the annual average temperature is 21°C [23]. The average relative humidity is 38 percent [24]. Potential evaporation and transpiration are between 4196 mm and 5700 mm [25]. Most rainfall was concentrated in the fall and winter seasons (more than that indicated in the present study), and species were classified according to their growth habit. The climate of the region is based on the Domarten desert [26].

3. Methodology

Variables that reflect the land characteristics were evaluated and can be influenced by vegetation that has a positive or negative feedback on the climate near the ground. Therefore, in this study, to investigate and evaluate the effect of land surface characteristics on Tair, two microsites in the Sistan region with different plant cover percentages were considered in addition to the synoptic station of Zabol and Zahak. Microsite A with 65% vegetation and 35% bare soil in HIW and microsite B with 20% vegetation and 80% bare soil in abandoned agricultural land were selected. The distance between these microsites is about 20 kilometers, and the difference in height between them is less than 10 meters [27–29]. Measurements were carried out in the autumn and December with wind speeds of less than 4 m/s and a clear sky without clouds [30]. The microsites were also chosen so that the location of the measuring equipment in the center of a circle would be approximately the same as the land surface coverage of up to 200 meter radius [31, 32].

Land surface temperature (LST) and a height of 2 meters above the ground temperature (Tair), wind speed, and relative humidity were measured simultaneously at 5 days for three hours in each of the synoptic stations and microsites. To equalize the measurements, two screen boxes with 55 × 65 × 55 cm (see Figure 2) were constructed similar to the screenshot mounted at the synoptic station in Zabol.

Also, before installing equipment in microsites, all thermometers were first calibrated in zero-degree water. In the next step, to investigate the possible error of the thermometers and equipment, the readings and accuracy of them at the synoptic station of Zabol station were investigated for 24 hours. Because the thermometers used were of the same type as the standard thermometers in the Zabol synoptic station, there was no difference in the measurement between the equipment. LST was measured by using a noncontact laser thermometer. The wind speed was measured at 2 meters by manual wind speed, and the relative humidity was measured using a digital humidity meter mounted at a height of 2 meters inside the display box. To increase the accuracy of the data after three hours, the average of each hour was calculated. To measure the canopy percentage in each microsite, a 30 × 30 meter plot was considered at the center of the equipment deployment site, and the vegetation percentage was measured by the deployment of two transect diagonals (see Figure 3).

3.1. Cubist Regression Model (CRM). Cubist, a rule-based regression technique developed by Quinlan in 1997, is built upon the foundations of decision tree classifiers such as ID3 and C4.5. Cubist is a commercial product and proprietary and has the least algorithmic documentation [33]. The M5 model expands categorical decision trees to handle continuous classes by placing a multivariate linear model at each leaf. Model trees are more accurate than ordinary regression. An additional technique to improve prediction is to use similar training cases, or instances, to determine the value at the new location. Composite models combining instances and model trees are more accurate than model trees alone. It is assumed that Cubist is a composite model that combines a model tree, reformulated as rules, with the instance-based method. The resulting rules have linear multivariate models in their “then” statements [34]. Cubist does not retrieve one final model like random forests, but a set of rules is associated with sets of multivariate models. Then, a specific set of predictor variables will choose an actual prediction model based on the rule that best fits the predictors. Cubist is a commercial, proprietary product and has the least algorithmic documentation in comparison to linear regression and random forest [35]. In this study, data on land surface temperature (LST), air temperature (Tair), wind speed, vegetation percentage, and relative humidity of the air were collected simultaneously from four microsite studies: synoptic station of Zabol, microsite A, microsite B, and synoptic station of Zahak. The study used 90 series or 450 datasets for analysis, except for the data on the synoptic station of Zahak, which were considered for model testing. The output rules for these data were evaluated (Figures 4 and 5). Finally, the model was tested using 30 series of data (150 datasets) related to the synoptic station of Zahak based on the output rules of the model.

3.2. LST and NDVI Estimation Using Landsat 8 Satellite Sensors. To obtain the Normalized Difference Vegetation Index (NDVI) and the land surface temperature (LST) index, we followed the guide provided on the USGS website for Landsat 8 images [36].

3.3. Statistical Analysis. Statistical data analysis was carried out using the SPSS statistics version 24 software. Data were tested for normality and homogeneity of variations using Kolmogorov–Smirnov and Levene’s tests, respectively. An analysis of the variation between LST and Tair was performed using one-way repeated measures. The effects of microclimate factors on Tair were analyzed using an analysis of CRM. The significance of the relationships between Tair
Figure 1: Location of synoptic stations and microsites in this study.

Figure 2: Sample boxed screenshots (a), located in study microsites: (b) microsite A and (c) microsite B.

Figure 3: Sample of the plot to study the percent of plant cover in microsites.
and LST was assessed using Pearson’s correlation. Regression analysis was used to identify significant relationships between Tair and LST.

4. Results and Discussion

4.1. Normality Test. The normalization test results of the collected data on LST, wind speed, relative humidity, and Tair at 2 m height indicated that the data were almost soft and did not normalize the data. Figure 6 provides a visual representation of these results.

4.2. The Relationship between Tair and LST. To evaluate the suitability of land surface temperature (LST) data for air temperature (Tair) estimation, we conducted a correlation coefficient analysis to test the relationship between Tair and LST of all microsite and synoptic stations (Figure 7).
The results of the analysis at Zabol synoptic station, Zahak synoptic station, microsite A, and microsite B showed that there is a positive and strong correlation between LST and air temperature. Furthermore, we investigated the correlation between surface temperature of the ground and air temperature at a height of 2 meters in four different microsites. The results showed that the Zahak station had the highest correlation coefficient of 0.909, while the Zabol station had the lowest value of 0.869. In addition, we found that Sig was less than 0.01 in all sites, indicating statistical significance (see Table 1).

4.3. CRM. The result of the Cubist regression model for 90 series of LST data, Tair, wind speed, vegetation percentage, and relative humidity for measured data at the synoptic station of Zabol and Zahak, HIW (microsite A), and abandoned land (microsite B) is shown in Figures 4 and 5.

The first rule for 75 samples with an average of 8.56 and a range of -1 to 23.8 and a standard error of 1.26 was presented with a model. If $LST \leq 23.4^\circ C$, then Tair at 2 meters above the ground level is estimated by

$$Tair = 12.17 + 0.508LST - 0.166 \text{ relative humidity} - 2 \text{ plant cover percent}. \quad (1)$$

The second rule is presented for 6 samples with an average of 21.18 and a data range of 18.4–24.6 and a standard error of 1.91. When $LST \geq 23.4$ and relative humidity $\leq 22$, then air temperature is as follows:

$$Tair = 39.84 - 0.592 \text{ LST}. \quad (2)$$

The third rule is presented for 9 samples with an average of 14.92 and a data range 11.9–17.8 and a standard error of 1.04. When $LST > 23.4$ and relative humidity $> 22$, then air temperature is as follows:
4.3.1. Testing the Rules of the CRM. To test the rule presented by the CRM model, real data from the Zahak synoptic station were selected. The results showed that there is a significant correlation between the actual data and the predicted data for Tair with a correlation coefficient of 0.94 and a P value of 0.01. CRM can be used to estimate Tair at an altitude of 2 meters with high confidence by utilizing the characteristics of LSF, such as land surface temperature (LST), percentage of vegetation cover, wind speed, and relative humidity. The results of the observation and estimated data based on the rules extracted from CRM are shown in Figure 8.

4.4. The Results of Manual LST and Satellite LST. The examination of the relationship between the temperature of manual LST, taken in different uses by a hand-held laser thermometer, and the temperature extracted from Landsat 8 satellite images (satellite LST) in the investigated sites revealed that the Pearson correlation coefficient for 45 investigated points is 0.879 (Figure 9).

4.5. The Results of NDVI and LST. The examination of the relationship between the NDVI and LST in the investigated sites revealed a negative correlation between these indices ($r = -0.773$). The correlation coefficient and the explanation coefficient resulting from the relationship between these two indicators are shown in Figure 10. The results indicate that LST decreases with the increase of the NDVI. In site B and Zabol station, where the amount of vegetation was small, the LST extracted from the Landsat 8 images shows a high value.

4.6. The Role of LSF with Different Coverage Percentages in Tair. The results of the linear regression model between LST and Tair at 2 meters in synoptic stations of Zabol and Zahak and microsites A and B showed that there was a strong correlation between LST and air temperature, with the coefficient of determination ranging from 0.75 to 0.82. Shi et al. found a coefficient of determination of 0.79 between air temperature and LST in their research [37]. They stated that changes in air temperature with altitude, plant cover, and land-use changes are more justifiable than LST alone. The results of using the CRM in this study to estimate Tair at 2 meters height showed that independent parameters of LST, wind speed, vegetation percentage cover, and relative humidity of this model were correlated coefficients with 0.98.
Figure 8: The results of CRM for predicting Tair. (a) The first rule of the model, (b) the second rule, and (c) the third rule of CRM for Tair prediction.

Figure 9: The results of the relationship between manual LST and satellite LST.

Figure 10: The results of the relationship between LST and NDVI extracted from Landsat 8 satellite.
Based on the model output, only 17% of the data to estimate the air temperature do not include vegetation cover in the model. This indicates the key and important role of vegetation in ambient temperature in this arid region. The findings of Sadeqi and Kahya [11] are consistent with the results of this study. They showed that air temperature has increased by 94% of the 31 synoptic stations in Iran [11]. The study also found that hot and desert areas have experienced a significant increase in air temperature compared to other areas. The results of both studies suggest that climate change has a significant impact on air temperature.

In general, it can be said that one of the reasons for the excessive increase in ambient temperature compared to the global average is the loss of vegetation in HIW due to the drying of this international wetland. The results of this study are consistent with the findings of other researchers [2]. They showed that in the areas where lakes dried up due to evaporation, land surface temperature (LST) significantly increased. The study also found a high correlation between land use and LST. The results of both studies suggest that changes in land use and the drying up of wetlands can have a significant impact on air temperature and the environment, because one of the limiting factors of vegetation is water, which significantly decreases natural vegetation with the occurrence of drought [38].

It is necessary to reclaim it as an important natural reserve and regulate the temperature conditions for the life of the inhabitants of these areas. Vegetation loss is also a significant factor in the increase in dust storms in recent years due to the drying up of HIW [39]. The results of modeling the determination of Tair using CRM showed that the change in the percentage of plant cover without considering the type of vegetation up to 83% has a direct effect on Tair. This indicates that due to the reduction of water entering the lagoon and the destruction of vegetation in HIW, the trend of the relative increase in ambient temperature in this area will be more intense. Given the impact of vegetation on air temperature (Tair) at a height of two meters in the study area, it is imperative to pay close attention to compliance with the environmental status of this HIW for the restoration and protection of plant species and green cover in this region.

4.7. The Relationship between Remote Sensing Data and Field Data. The relationship between temperature values extracted from the Landsat thermal band and the surface temperature calculated by the laser thermometer is shown in Figure 9. The study examined the sites under investigation, where the ground sites were measured simultaneously and were close together. The temperature differences in the sites were observed, indicating that the surface temperature in any area of the Earth’s surface depends on the characteristics and properties of the surface. The environmental conditions are different and variable, which can affect the surface temperature. At site B, the relationship between the temperature of the ground sites and the temperature extracted from the Landsat image was lower because the percentage of the coating was low and the points were not homogeneous.

The study also found that there is a strong positive correlation between the temperature of the manual LST and the temperature extracted from Landsat 8 satellite images (satellite LST) in the investigated sites. The Pearson correlation coefficient for 45 investigated points is 0.879, which indicates a strong positive linear relationship between the two variables. These results align with previous studies [40, 41]. The study conducted in the HIW region found that the LST-NDVI correlation was negative, as shown in Figure 10. The study also found that on sites with higher NDVI values, LST was lower for Landsat 8 data, which is consistent with the other findings [42].

5. Conclusion

In this research, we examined the possible responses of LSF to Tair using CRM. The primary conclusion of this study is that LSF can lead to significant impacts on LST and Tair in different micsites with different plant covers. This study proved that a very high accuracy of Tair estimation ($R^2 = 0.96$) could be achieved with a combination of LST, percent of vegetation cover, wind speed, and relative humidity. It can be seen that CRM shows the highest accuracy of Tair estimation, which is consistent with another study [35]. According to the research, any changes in Tair are closely related to changes in plant cover percent in LSF, similar to a percentage of vegetation cover, wind speed, and relative humidity. The examination of the relationship between NDVI and LST in the investigated sites revealed a negative correlation between these indices ($r = -0.773$). The correlation coefficient and the explanation coefficient resulting from the relationship between these two indicators are shown in Figure 10. The results indicate that LST decreases with the increase of the NDVI. In site B and Zabol station, where the amount of vegetation was small, the LST extracted from the Landsat 8 images shows a high value. It seems that allocating the minimum water resources to green water or green vegetation in the same arid regions due to the high intensity of evaporation in these areas can provide better living conditions.

5.1. Findings and Limitations. The study shows that the Cubist regression model (CRM) can be used to simulate the effects of land surface factors (LSFs) on air temperature (Tair) with adequate performance. CRM can estimate Tair from independent parameters of LST, wind speed, vegetation percentage, and relative humidity with a high correlation coefficient of 0.98. The study also found that LST with 100% and relative humidity and vegetation percentage, respectively, with 93% and 83% entered the model, as wind speed did not enter the model because throughout the period, the measurements were constant and less than 4 m/s (no changes). The research also highlights the importance of considering the changes in the surface of HIW and its role in Tair using data on LST, vegetation, wind speed, and relative humidity. The study shows that the drying up of HIW and the loss of vegetation in this area have increased ambient temperature. The study results can be used to develop
strategies to mitigate the impact of climate change on wetlands and other ecosystems. The study only examines the changes in the surface of HIW and its role in Tair using data on LST, vegetation, wind speed, and relative humidity. This study did not consider other factors that could influence Tair, such as cloud cover, precipitation, solar radiation, air pollutants, and air quality, because they were not accessible. Further research is needed to validate the findings and to identify other factors that may affect Tair in the wetland.

Data Availability

The original data used to support the findings of this study are included in the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References


