Comparing Model Results for Predicting Temporal Variations of SST and Rainfall Rate

Omar SAWALHA[1, *], Abdallah DWIKAT[1], Sogand MOLA[2], Nura AHMED[1], Melike YETER[3], Zafer ASLAN[1]

[1]Istanbul Aydin University, Faculty of Engineering, Department of Computer Engineering, Florya, 34295, Istanbul, Türkiye

[2]Istanbul Aydin University, Anadolu BIL Vocational School, Computer Programming. Florya, 34295, Istanbul, Türkiye

[3]Istanbul Aydin University, Faculty of Engineering, Department of Industrial Engineering, Florya, 34295, Istanbul, Türkiye

omarsawalha@stu.aydin.edu.tr; abdallahdwikat@stu.aydin.edu.tr, nuraahmed@stu.aydin.edu.tr, zaferaslan@aydin.edu.tr, sogandmola@stu.aydin.edu.tr, melikeyeter@stu.aydin.edu.tr

ABSTRACT

The application of advanced forecasting methods can help in agriculture, fisheries, and disaster management, reducing losses and optimizing the distribution of resources. It can also contribute to climate models; create more effective policy in environmental protection efforts. In this paper, it is made to determine the most suitable model techniques for the study regions are determined by applying advanced machine learning techniques and comparing the findings. The outputs of the research are intended to contribute to climate change studies. The main aim of the study is to effectively forecast and monitor spatio-temporal variations of average Sea Surface Temperature (SST) and monthly/annual total rainfall rate. Machine learning models can offer significant insights.

Long-term annual total rainfall rate (2008-2023) in Antalya and SST data were considered at selected stations from (2007-2023) for Antalya and (2004-2023) for Adana in the Mediterranean Region. Machine learning techniques, including support vector regressions (SVR), Linear Regression (LR), XGBoost, AdaBoost, Random Forests (RF), Decision Trees (DT), and Long-Short-Term Memory (LSTM), provide robust methods for analyzing environmental data. These algorithms can uncover complex relationships between rainfall rate, SST, and other climatic variables, and the study included other weather features. Cross-validation techniques were employed to comprehensively evaluate and compare the performance of the machine learning models. There is sufficient evidence between observations and model results with $\alpha = 1\%$. For simulation of SST temporal variation, the performance of LSTM and RF models is higher in Antalya and Adana than other ML techniques in respectively.

Keywords: SST, Rainfall Rate, Climate Change, Machine Learning Techniques

INTRODUCTION

Estimating the amount of precipitation observed in tropical cyclones is of great importance for warning systems and disaster management. Based on artificial intelligence (AI), important findings are being obtained in this field. Yang and his group (2024) conducted a two-way investigation of artificial intelligence predictions of precipitation observed in tropical cyclones: Modeling the numbering of the satellite image and analyzing predictability. They created a high-accuracy prediction of precipitation amount and precipitation intensity using Global Precipitation Measurement. Through the analysis of rainfall patterns and distributions, emphasis is placed on the reliability and robustness of the rainfall forecast. Types of results: (a) Precipitation amount, and distribution categories differ in predictable success; b) Considering precipitation intensity and viewpoints, precipitation falling on land is less difficult to predict than falling on marine area. It is also stated that the proposed method can predict extreme precipitation lasting 0-120 minutes with an accurate rate of 87%.

The research study conducted by Cramer and his group (2017) includes a comprehensive evaluation of seven different machine learning methods for precipitation estimation in weather forecasting. The rainfall forecast has a chaotic structure. Planning of water resources has a very important place in a climate system that has a direct impact on agricultural and biological systems. In recent years, numerous models were developed to increase the accuracy of precipitation forecasting and understanding of precipitation characteristics. Their predicted success level has increased.

Das and his group (2017) analyzed the application of the random forest algorithm method for estimating the amount of heavy rainfall in a short time. Short-term forecasting of thunderstorms and convective activity has important applications in disaster management, aviation flight planning and air traffic management. Recent studies have focused on machine learning techniques for such problems, which use a combination of observations from different sources such as Numerical Weather Prediction (NWP) models, radar, satellite, and ground surface measurement systems, and produce short-term forecasts in periods of several hours. However, preliminary determination of the importance of individual atmospheric parameters and determination of a large number of input variables, features are one of the important studies in determining machine learning methods to reduce CPU time.

Understanding and predicting changes in Sea Surface Temperature (SST) and rainfall rates are becoming increasingly vital in a world grappling with climate change. These two variables are not just abstract climate indicators; they directly influence our daily lives and economic activities. Accurate forecasts of SST and precipitation can help farmers decide when to plant crops, guide fisheries in managing their resources sustainably, and allow for timely preparation in the face of natural disasters like floods and droughts. Beyond these immediate benefits, precise predictions of these variables provide critical input for shaping policies on environmental conservation and climate adaptation, which are fundamental to achieving sustainable development goals.

This paper covers time series analyses of monthly average sea surface temperature and monthly total rainfall rate variation at two study areas, Antalya and Adana.

MATERIAL and METHODS

Details on study area and methodology have been presented at this part of the paper.

Study area and data

Table 1 shows, coordinates of the two pilot areas. Antalya and Adana observatory stations are both in the Mediterranean Area and under air-sea-land interactions in Türkiye.

Table 1. Geographical descriptions of the study area

Study Area	**Latitude**	**Longitude**
Antalya	36.6719-36.9219°N	30.5883-30.8383°E
Adana	36.4414-36.6914 °N	35.2590-35.5090 °E

Input variables for the study period in Adana are listed in Table 2, https://marine.copernicus.eu/, and Aqua/AIRS L3 (2024). Daily air temperature, dew-point temperature, relative humidity, wind speed, precipitation, and SST values have been considered in Antalya.

Table 2. Adana Input variables (period for all data is 2004 to 2023)

Data	**Unit**	**Source**
Net short-wave radiation flux	W m2	GLDAS Catchment Land Surface Model L4 daily
Relative Humidity Ascending	percent	AIRS/Aqua L3 Daily Standard Physical Retrieval

Relative Humidity Descending	percent	AIRS/Aqua L3 Daily Standard Physical Retrieval
Air Temperature at surface Ascending	ºK	AIRS/Aqua L3 Daily Standard Physical Retrieval
Air Temperature at surface Descending	ºK	AIRS/Aqua L3 Daily Standard Physical Retrieval
Cloud Fraction	Percentage	MODIS Level 3 Atmosphere Gridded Product (Platnick et al, 2015)
Ice Cloud Effective Particle Radius	microns	MODIS Level 3 Atmosphere Gridded Product
Sea Surface Temperature	ºK	SST_MED_PHY_L3S_MY_010_042
Snow Precipitation Rate	Kg / m^2	GLDAS Noah Land Surface Model L4
Net long wave radiation flux	W m2	GLDAS Noah Land Surface Model L4
Near Surface Air Temperature	ºK	GLDAS Noah Land Surface Model L4
Specific Humidity	Kg / kg	GLDAS Noah Land Surface Model L4
Total Precipitation	kg /m^2	GLDAS Noah Land Surface Model L4
Air Temperature Ascending	ºK	Aqua/AIRS L3 Daily Standard Physical Retrieval
Air Temperature Descending	ºK	Aqua/AIRS L3 Daily Standard Physical Retrieval

This report includes a full analysis of the dataset collected for the city of Antalya, Türkiye (Table 2). The dataset includes Air Temperature, Dew Point, Humidity, Wind Speed, Pressure, and Sea Surface Temperature. All data were converted into SI Units and collected daily between 1 January 2007 and 31 December 2023. Min-Max normalization was used in some activities, such as the Wavelet part, where we converted the data into a Min-Max Scalar or 0-1 range and applied Wavelet for that. Also, it is common to normalize the data in cases using the LSTM Deep Learning model to avoid overfitting. However, if the forecasting includes predicting data out of the test range it can be complicated to re-transform the data into the original as the model will be built to forecast on 0-1 range, Han J. and M. Kanber, (2006): Other additional features collected later are (Cloud Fraction, Long-Wave Radiation Flux, and Short-Wave Radiation Flux). Those features are between 2008 and 2023. In those comparisons, we exclude 2007 from the previous features to align with those features as well.

In this study, we focus on the Mediterranean regional hotspot for climate variability—by analyzing long-term data from three key locations: Antalya and Adana. The region’s diverse climatic conditions present a unique challenge and an opportunity to test the capabilities of different predictive models. Using historical SST and rainfall data spanning over the study period (2007-2023 and 2004-2023), it is aimed to compare various machine learning models to forecast temporal variations in these critical climate variables. Beyond traditional models, Machine Learning models allow us to break down time-series data into different frequency components, revealing the influence of events ranging from.

Methodology

However, predicting these dynamic and complex climatic variables is no simple task. Traditional statistical methods often fall short when it comes to capturing the intricate, non-linear interactions between SST, rainfall, and other environmental factors. This is where machine learning steps in as a game-changer. With its ability to analyze vast datasets and uncover hidden patterns, machine learning offers powerful tools for modeling, RF, Xgboost, Adaboost, and LR models for predicting climate variations. Techniques such as SVM, LSTM, and DT have shown great promise in deciphering the complexities of environmental data, providing new insights into climate dynamics that were previously beyond reach. Shalev-Shwartz, S., Ben-David, S. (2014). A decision tree is a predictor, $h: X \rightarrow Y$, that predicts the label associated with an instance x by traveling from a root node of a tree to a leaf (Figure 1). For simplicity, we focus on the binary classification setting, namely, $Y = \{0, 1\}$, but decision trees can be applied to other prediction problems as well. At each node on the root-to-leaf path, the successor child is chosen based on a splitting of the input space. Usually, the splitting is based on one of the features of x or a predefined set of splitting rules. A leaf contains a specific label.

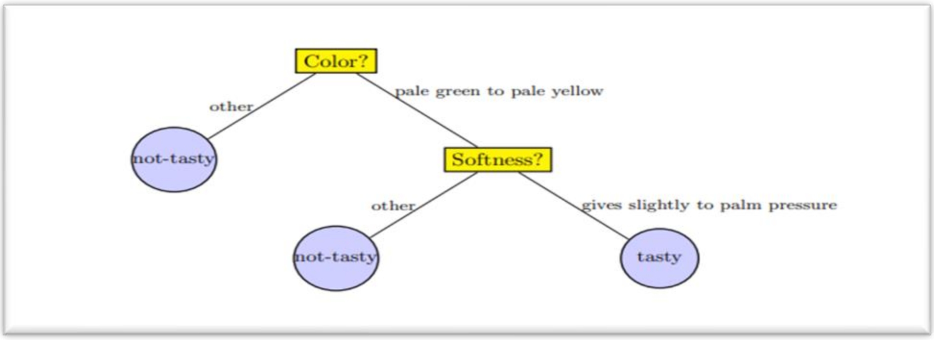

Figure 1. Decision tree algorithm

A random forest is a classifier consisting of a collection of decision trees, where each tree is constructed by applying an algorithm A on the training set S and an additional random vector, θ, where θ is sampled from some distribution. The prediction of the random forest is obtained by a majority vote over the predictions of the individual trees. The support vector classifier described so far finds linear boundaries in the input feature space. As with other linear methods, we can make the procedure more flexible by enlarging the feature space using basis expansions such as polynomials or splines. SVMs can be adapted for regression with a quantitative response, in ways that inherit some of the properties of the SVM classifier.

The essence of this paper lies in the comparative analysis of these models to determine which approaches offer the most reliable predictions for future climatic conditions. By applying rigorous evaluation metrics and cross-validation techniques, we think it is not only advance the field of climate modelling but also provide practical tools for climate resilience. Our findings aim to support decision-makers, scientists, and communities in better preparing for a future shaped by climate change, offering insights that bridge the gap between complex data science and tangible, real-world impact.

ANALYSES

This part of the paper covers the results of statistical analyses of variables and results of modeling.

Statistical Analyses

This report includes a complete analysis of the dataset collected for the city of Antalya, Turkey. The dataset includes Air Temperature, Dew Point, Humidity, Wind Speed, Pressure, and Sea Surface Temperature.

All data were converted into SI Units and collected daily between 1 January 2007 and 31 December 2023. Min-Max normalization was used in some activities, such as the LSTM part, where we converted the data into a Min-Max Scalar or 0-1 range and applied LSTM modelling for that. Also, it is common to normalize the data in cases using the LSTM Deep Learning model to avoid overfitting. However, if the forecasting includes predicting data out of the test range, it can be complicated to re-transform the data into the original as the model will be built to forecast on a 0-1 range (Han J. and M. Kamber, 2006). Other additional features collected later are Cloud Fraction, Long-Wave Radiation Flux, and Short-Wave Radiation Flux. Those features are between 2008 and 2023. In those comparisons, we exclude 2007 from the previous features to align with those features as well.

After revealing the data overview results, Machine learning models applied to the data will be visualized into results in the end by clarifying the best model results for forecasting Weather atmospheric features.

Before clearing the seasonal trend for each feature individually, a histogram was applied on each feature with a red line to show the mean value, which is helpful to understand the distribution of the dataset for each feature (Statistical Analyses, 2007-2023).

The Correlation Coefficient matrix table in the heat map results below shows the relationship between each weather feature and the other.

Correlation Coefficient or the R-value between each feature and the other helps to understand the relationship between the features, which can be useful not only to understand this relationship but also to use the appropriate features later for applying machine learning models (Figure 3). 0.7 and more can be considered as a strong correlation, between 0.5 and 0.7 as medium, and less than 0.5 as weak. Also, even if the value is negative, as shown below, between the Air Temperature and Pressure, while the values seem to be around -0.60, which is a useful result to be used to predict the pressure or to predict the air temperature based on the other feature. Additionally, having a stronger correlation coefficient between features is very helpful in modelling, as it increases the accuracy of the model in this way. In some cases, removing multi-collinearity can be a useful step to have an explanation of the model built, which is by removing the duplicated correlations with the specific target and keeping the stronger one. For example, if we predict Air Temperature as the target, the Dew point, and Short-Wave radiation flux have a strong correlation with Air Temperature, however, there is a strong correlation already between the flux and the Dew Point the less correlation our side can remove the feature, it will be helpful for interpretation.

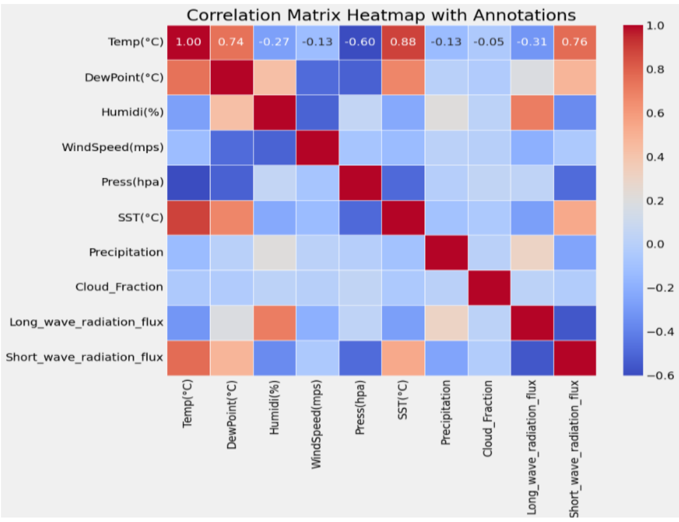

Figure 2. Correlation coefficient matrix table in heat map, (Antalya)

Table 3. Descriptive statistics in Antalya

	Temperature (°C)	**DewPoint (°C)**	**Humidity (%)**	**WindSpeed(mps)**	**Press(hPa)**	**SST (°C)**
Count	6209.000	6209.000	6209.000	6209.000	6209.000	6209.000
Mean	19.600	10.701	61.614	3.324	1006.460	22.57
Std	7.278	7.351	16.704	1.491	6.294	4.52
min	1.389	-16.278	15.200	0.179	914.325	15.30
25 %	13.444	6.667	49.600	2.369	1002.371	18.10
50 %	19.000	11.000	64.700	2.906	1005.758	22.10
75 %	26.167	15.889	74.200	3.755	1009.144	27.00
max	37.778	26.833	100.000	12.249	1026.076	31.00

These statistics summarize and describe the main features of a dataset, giving a quick overview of its central tendency, variability, and distribution (Table 3). (Count (data size), mean, standard deviation, min, quartiles, 25%, 50%, 75%, and max).

Analyses of Sea Surface Temperature in Antalya

Figure 3 shows the daily average SST (ºC) Plot from 2007 to 2023 in Antalya. In recent years, maximum SST values show a slightly increasing trend.

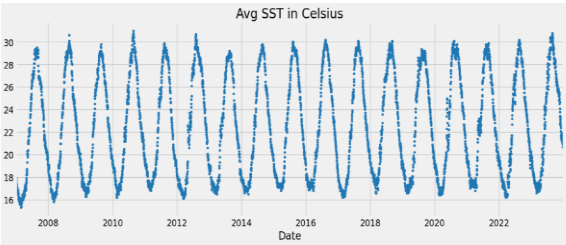

Figure 3. Daily average SST

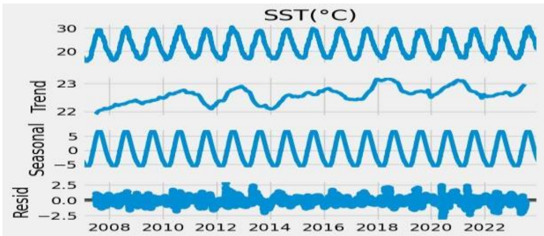

Figure 4. Seasonality trend study

Figure 4 presents the Seasonality Trend in Antalya from 2007-2023, based on daily average SST (Celsius). SST values show an increasing trend in the last decade. Auto and Partial Correlation functions for SST, (ACF/PACF) – (2007-2023) in Antalya are given in Figure 5, (Hastie et al, 2009). Yearly Boxplots (2007-2023) in Antalya based on Daily Average SST (ºC) have been presented in Figure 6.

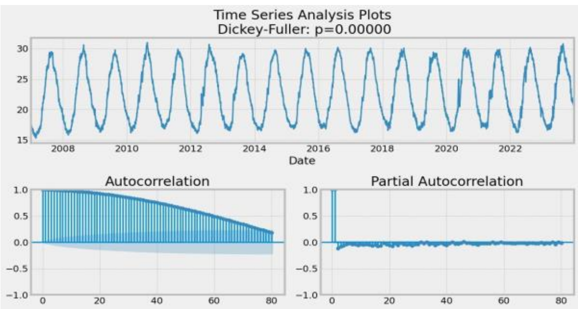

Figure 5. Auto and Partial Correlation functions SST (ºC)

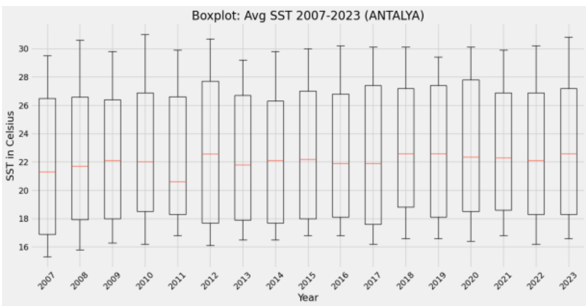

Figure 6. Annual SST, boxplots

The frequency histogram based on Daily Avg SST (ºC) from 2007 to 2023 in Antalya is given in Figure 7. There is a bimodal frequency distribution for SST data in Antalya.

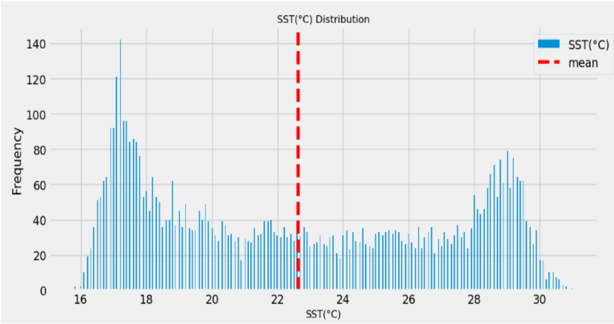

Figure 7. Frequency histogram of SST

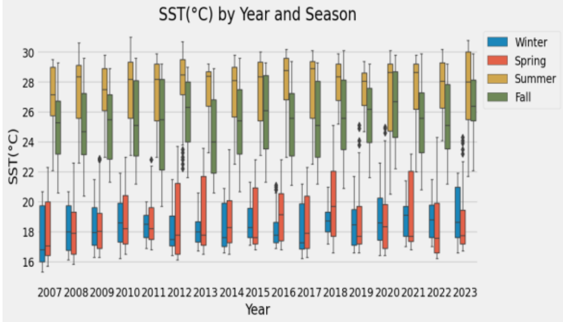

Figure 8. Seasonal variations of SST in Antalya

Boxplots show some seasonal variations of Q1, Q2, Q3, and Q4 for SST (ºC) in Antalya (Figure 8). There are important variations in SST values in the three seasons, except winter. Heat-map after year and month has been shown in Figure 9, 2007-2023 – Antalya- Daily Average SST (ºC). The periodicity of warm SST years is increasing in recent years.

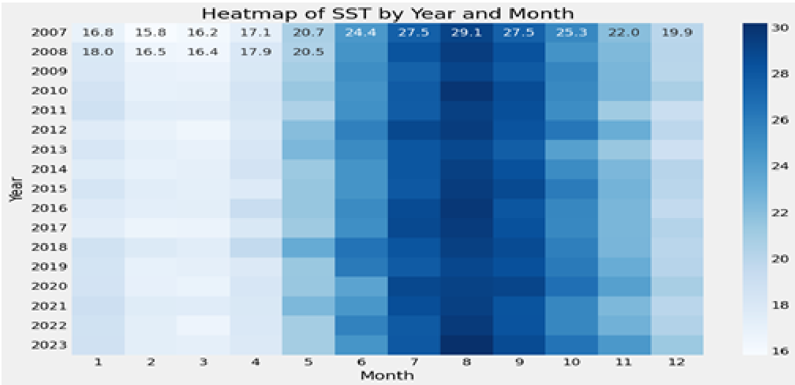

Figure 9. Heat-map after year and month

Analyses of Sea Surface Temperature in Adana

The frequency distribution of SST is very similar to the same distribution in Antalya (Figure 10).

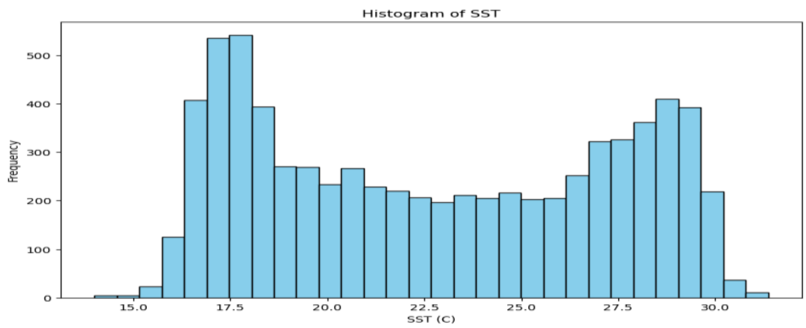

Figure 10. Frequency distribution of daily average SST values in Adana.

Descriptive statistics of daily average SST values observed in Adana have been presented in Table 4. In addition to mean, standard deviation, minimum, and maximum values, cross-correlation coefficients amongst variables are also presented. Higher correlation coefficients have been defined between surface air temperature and surface fluxes.

Table 4. Descriptive statistics in Adana

VARIABLE	count	mean	std	min	max	CORR Sst
SST (°C)	7305	22.7742	4.52316	13.9936	31.3657	1
RelHum_A (%)	7305	60.1770	10.9132	23.25	102	-0.09887
RelHum_D (%)	7305	72.1932	10.6077	26.125	106	-0.06957
SurfAirTemp_A (ºK)	7305	294.613	6.82921	276.75	311.187	0.87635
SurfAirTemp_D (ºK)	7305	291.400	6.00077	270.625	304.062	0.89322
Temperature_A (ºK)	7305	293.017	6.13093	271.187	306.812	0.87912
Temperature_D (ºK)	7305	291.564	6.2193	271.937	306	0.89699
ShortWaveRadFlux (W/m^2)	7305	171.435	66.4488	7.97736	291.503	0.48074
LongWaveRadFlux (W/m^2)	7305	-80.498	23.0849	-148.154	-14.1805	-0.52917
SpecificHumidiy (% normalized)	7305	0.00971	0.00433	0.00163	0.02189	0.80296
TotalPercipitation	7305	1.99E-05	7.10E-05	0	0.00139	-0.15934
Snowf	7305	1.71E-09	1.46E-07	0	1.25E-05	-0.01558
NearSurfaceAirTemperature (ºK)	7305	293.31	6.70851	274.681	308.303	0.89408
CloudEffectiveRadiusIce	7305	34.0025	9.23726	5.18	59.93	-0.0803
CloudFraction (%)	7305	0.44325	0.35627	0	1	-0.38675

Modelling of SST and Precipitation in Antalya

Machine Learning algorithms for Antalya applied on both precipitation and on SST from 2008-2023. 2008-2021 is the training dataset, and 2021-2023 is the testing part. Different features have been used depending on the statistical part, such as the correlation coefficient values between the feature and the target.

Modeling of SST by using Decision Tree Technique in Antalya, (2008-2023)

After data preprocessing, correlation values were found. The features that strongly have a positive correlation with SST are Average Air Temperature, Dew Point, and Short-wave radiation flux. Those features were applied as input into the model, and the output is SST. By using Average Air Temperature, Dew Point, and Short-wave Radiation Flux as input features, the decision tree can effectively model the relationship with SST. While decision trees might not always outperform more sophisticated models, they offer a good balance between simplicity and performance, especially for initial modeling and feature selection. It can align with both simplicity, which DT is easy to understand and interpret, and efficiency, which they can handle both numerical and categorical data efficiently, making them versatile for various types of features. In addition to that, applying some time features can help the model understand the trend in seasons. The two approaches were applied, and we compared the results for both. Error values are defined as follows: Mean Absolute Error (MAE) = 1.3559, Mean Squared Error (MSE) = 3.8671, Root Mean Squared Error (RMSE) = 1.9665, and R-squared (R^2) = 0.7970. After adding the time, additional features are the day of the week, quarter, month, year, day of the year, day of the month, week of the year, weekday, and season. The model showed significant improvement in its efficiency after adding the time features, by helping the model understand the pattern of the dataset over the years. The performance metrics for this model provided better results: Mean Absolute Error (MAE): 0.692, Mean Squared Error (MSE): 0.840, Root Mean Squared Error (RMSE): 0.916, R-squared (R^2): 0.955.

Linear Regression for SST in Antalya (2008-2023)

During the Analysis part, a strong relation seems to exist between the SST values and it in the previous years. This helps to use one-year-earlier data of SST to forecast the next year of SST values. However, it might not be as reliable as other machine learning models due to the changes that happen during the year, with effects like rain or climate change, but achieving an accurate Linear Regression model and understanding this relationship can help monitor and understand those changes that occur from one year and another. After building the model by shifting a 1-year data and assigning the input as SST-1 which is the previous year and the SST is the target for its next year. The results for the model in the test set between 2021-2023 in Antalya show a performance with Mean Squared Error (MSE): 1.008, Root Mean Squared Error (RMSE): 1.004, R-squared (R^2): 0.947.

Random Forest for Precipitation in Antalya (2008-2023)

The Random Forest model was used in forecasting the precipitation for the same period (2021-2023). The training data used daily basis data of total precipitation data in the city of Antalya with mm/day unit. The features used for building the model are Humidity (%), Cloud Fraction, Long Wave Radiation Flux, Short wave radiation flux, Air Temperature, and Precipitation as the target. The Precipitation of last year's features seems to have little effect on Precipitation for the next year. However, exploring the use of different features in each model can help find the best way to build the model. After automating the part of tuning the hyper parameter of the Random Forest Model, we got the best parameters as min samples split=5, min samples leaf=8, and estimators=1000. After error analyses: Mean Absolute Error (MAE): 2.499, Mean Squared Error (MSE): 35.977, Root Mean Squared Error (RMSE): 5.998, R-squared (R^2): 0.166. The model with a MAE of 2.5 and MSE of 35% seems to be getting the patterns, but it still needs more improvement.

LSTM Long-Short Term Memory for SST in Antalya (2008-2023)

The LSTM Deep Learning model was used in forecasting the SST for the same period with the same test size (2021-2023). The training data used daily basis data of SST data in the city of Antalya in Celsius. Min-Max Scalar normalization was used for a range of 0-1 to ensure reliability and avoid over-fitting in such a sensitive model, to be over-fitted like LSTM. A Multivariate LSTM model was used with the features of Air Temperature, Dew Point, Short-Wave radiation flux, and SST as our target. The metric results of the test set are: Mean Absolute Error (MAE): 0.350, Mean Squared Error (MSE): 0.246, Root Mean Squared Error (RMSE): 0.496, R-squared (R^2): 0.987. During the tuning part, it is assumed that given the past 10 days' observations, we need to forecast the next 5 days' observations. The model built was to Sequence Model with one encoder layer and one decoder layer.

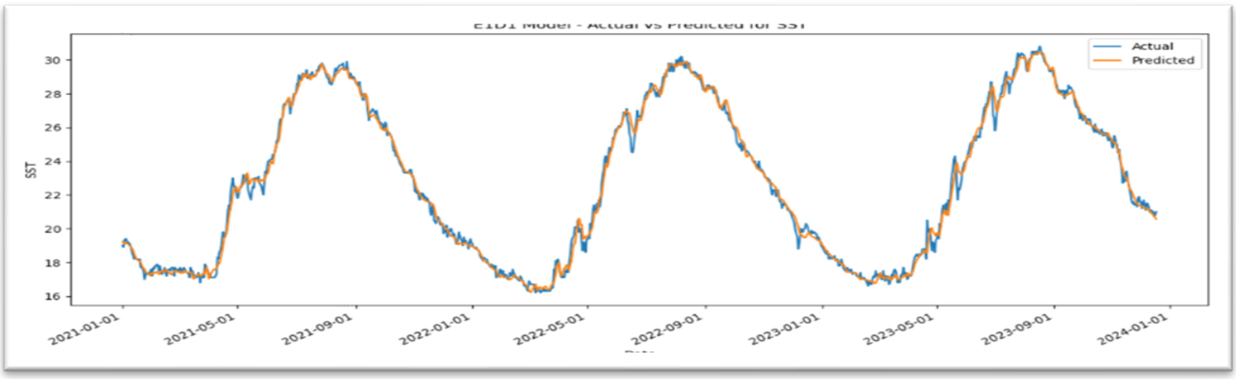

Figure 11. Modelling of SST values by using LSTM Methodology in Antalya

The results and the metrics results for the LSTM model were embarrassing, and the model shows its ability to accurately forecast SST in the test set, even with specific patterns and periods (Figure 11). The LSTM model results were the best and highest reliability for building SST scenarios.

LSTM Long-Short Term Memory for Precipitation in Antalya (2008-2023)

The LSTM Deep Learning model was used in forecasting the precipitation for the same period with the same test size (2021-2023). The training data used daily basis data of Precipitation (mm) in Antalya. The result of the LSTM in the test set (2021-2023) using the LSTM model E1D1 Metrics for the test set: Mean Absolute Error (MAE): 2.822, Mean Squared Error (MSE): 37.405, Root Mean Squared Error (RMSE): 6.116, R-squared (R^2): 0.0411. Above is the result, the past 3 days of precipitation used to forecast the next 4 days. The features used in this model are four: Humidity, Cloud Fraction, Long-Wave radiation flux, and Precipitation which is our target. Regarding the positive correlation coefficient value with Precipitation, those were the features selected.

Modelling of SST and Precipitation in Adana

Four models (SVR, XGBoost, Ada Boost, RF) were applied to predict SST using both historical SST data and 9 other predictors, Several other features were extracted from the period to better capture patterns, data was split into training, validation, and testing.

Figure 12 shows actual and for model results in 2022.

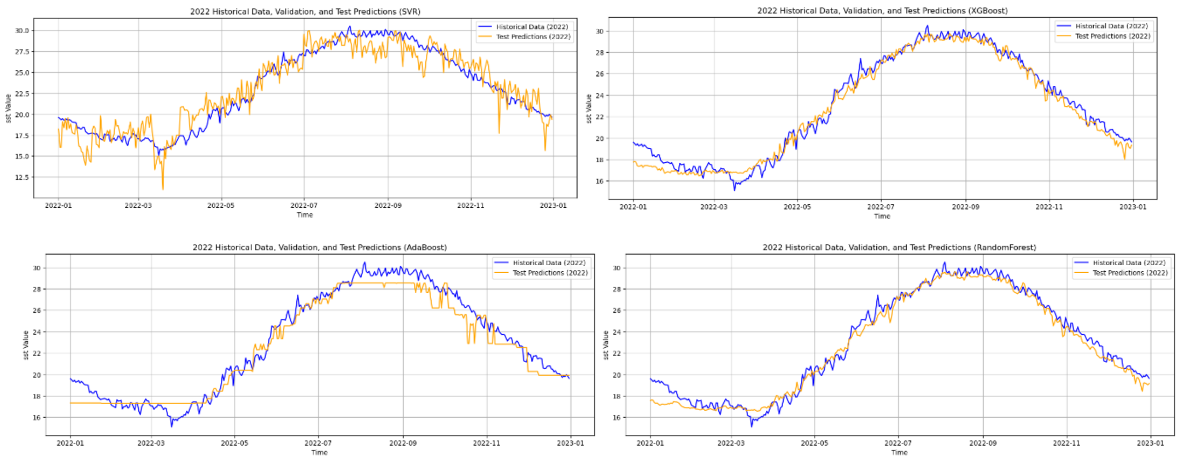

Figure 12. Modelling of SST values by using four different methodologies (SVR, XGBoost, AdaBoost, RF) in Adana. Lines show actual data and model forecasting, (Adana)

Figure 13 shows future scenarios of SST based on RF from 2023 to 2024 in Adana. Model outputs follow annual circulation and maximum SST values gradually show an increasing trend in recent years.

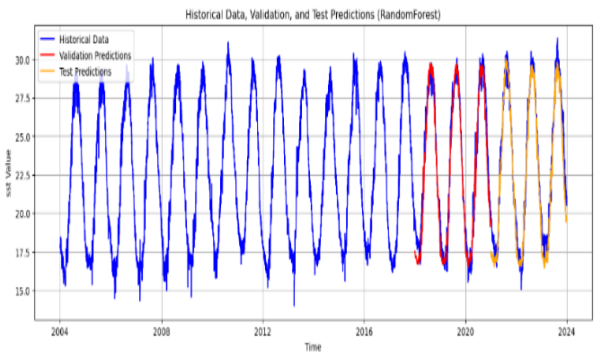

Figure 13. Modelling of SST after RF in Adana

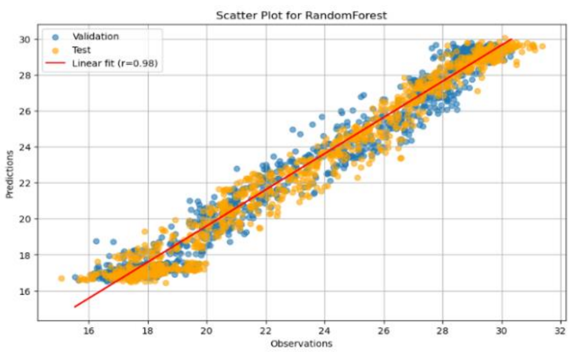

Figure 14. Relation between observed SST values and RF Model output in Adana

After Figure 14, there is sufficient evidence between the SST observation and the model results. Table 5 explains the error and correlation analysis of observations and model results.

Table 5. Error analyses of four different models for the temporal variation of SST simulation

	Validation RMSE	**Validation MSE**	**Validation R2**	**Test RMSE**	**Test MSE**	**Test R2**
XGBoost	0.938	0.881	0.957	0.931	0.868	0.957
RF	0.888	0.789	0.961	0.919	0.845	0.958
AdaBoost	1.049	1.101	0.946	1.170	1.369	0.933
SVR	1.591	2.531	0.876	1.711	2.93	0.857

CONCLUSION

Random Forests performed the best out of the four models that were used. LSTM Sequence to Sequence Model with one encoder layer and one decoder layer shows embarrassing results predicting SST values for the future with high-accuracy and robustness in analyzing patterns of SST over the years. For simulation of SST temporal variation, the performance of LSTM and RF models is higher in Antalya and Adana than other ML techniques in respectively.

Forecasting Precipitation was more complicated with the models we used, especially with tuning the models that we used. Precipitation of last year seems to have no or weak coefficient correlation with the precipitation of the next years, which made the models used for SST unable to fit the Precipitation forecasts. For short-term forecasting the performance of the models is better. The forthcoming study would be related to the application of more efficient hybrid methods by using wavelet transformations on the future construction of data.

ACKNOWLEDGEMENTS

This study is supported by the North Atlantic Treaty Organization (NATO), the Science for Peace and Security (SPS) Multi-Year Project number G5970 named Cube4EnvSec: “Big Earth Data-cube Analytics for Transnational Security and Environment Protection.

REFERENCES

AIRS/Aqua L3 Daily Standard Physical Retrieval (AIRS-only) 1 degree x 1 degree V006 10.5067/Aqua/AIRS/DATA303.

Aqua/AIRS L3 Daily Standard Physical Retrieval (AIRS-only) 1 degree x 1 degree V7.0 (AIRS3STD) 10.5067/UO3Q64CTTS1U, (accessed date 1-20 August, 2024)

AIRS/Aqua L3 Daily Standard Physical Retrieval (AIRS+AMSU) 1 degree x 1 degree V006 10.5067/Aqua/AIRS/DATA301

Cramer, S., M. Kampouridi, A. A. Freitas, A. K. Alexandridis, (2017). An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives, Expert Systems with, Applications, Volume 85, 1 November 2017, Pages 169-181.

Das, S., R. Chakraborty and A. Maitra, (2017): A random forest algorithm for now-casting of intense precipitation events, Advances in Space Research, Volume 60, Issue 6, 15 September 2017, Pages 1271-1282

GLDAS Catchment Land Surface Model L4 daily 0.25 x 0.25 degree GRACE-DA1 V2.2 10.5067/TXBMLX370XX8

GLDAS Noah Land Surface Model L4 3 hourly 0.25 x 0.25 degree V2.1 (GLDAS_NOAH025_3H) 10.5067/E7TYRXPJKWOQ

Han J. and M. Kanber, (2006): Data Mining Concepts and Techniques, Elsevier, Morgan-Kaufmann Publisher. pp. 743, CA, USA.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York, Springer.

Platnick, S., P. Hubanks, K. Meyer, and M. D. King, 2015: MODIS Atmosphere L3 Monthly Product (08_L3). NASA MODIS Adaptive Processing System, Goddard Space Flight Center

Shalev-Shwartz, S., Ben-David, S. (2014). Understanding Machine Learning - From Theory to Algorithms. Cambridge University Press. ISBN: 978-1-10-705713-5

Yang, N., Wang, C., & Li, X. (2024). Improving tropical cyclone precipitation forecasting with deep learning and satellite image sequencing. Journal of Geophysical Research: Machine Learning and Computation, 1, e2024JH000175. Https:// doi.org/10.1029/2024JH000175.

Web sites:

Web-1: https://marine.copernicus.eu/ (accessed date, 1 Julay-20 August).