# A Comparative Analysis of Random Forest and Gradient Boosting Regression Techniques in Google Colab for Air Temperature Prediction in the Greater Accra Region, Ghana

M. S. Peprah[1]*, A. Odoom[1], E. K. Larbi[2], M. A. Mwin[1]
[1]School of Mines and Built Environment, University of Energy and Natural Resources, Sunyani, Ghana
[2]Geo-Informatics Division, Building and Road Research Institute (CSIR-BRRI), Kumasi, Ghana
*Corresponding Author: mspeprah91@gmail.com

(Received on 29 May 2025; In final form on 05 April 2026)

***DOI:*** *https://doi.org/10.58825/jog.2026.20.1.259*

**Abstract:** Accurate temperature prediction is essential for climate adaptation, environmental monitoring, and sustainable urban planning. This study evaluates the performance of two machine learning techniques Random Forest (RF) and Gradient Boosting Regression (GBR) for predicting near-surface air temperature in the Greater Accra Region of Ghana. Daily temperature observations obtained from the Ghana Meteorological Agency covering the period 1960–2018 were used for model development. The average daily air temperature was computed from minimum and maximum temperature observations. The predictive performance of the models was compared with a classical statistical time-series model, Autoregressive Integrated Moving Average (ARIMA). Model evaluation was performed using five-fold cross-validation to improve the robustness of the results. Performance metrics included Mean Squared Error (MSE) and the coefficient of determination ($R^2$). The results show that the Random Forest model achieved the highest predictive accuracy with MSE = 0.0010 °C and $R^2 = 0.9996$, while the Gradient Boosting Regression model produced MSE = 0.0015 °C and $R^2 = 0.9994$. The ARIMA model showed significantly lower performance with MSE ≈ 0.598 °C and $R^2 \approx 0.30$. The high predictive performance of the machine learning models is partly attributed to the deterministic relationship between the input variables and the computed target temperature. The study demonstrates the potential of machine learning approaches for climate-related prediction tasks and provides insights for environmental planning and climate resilience strategies in rapidly urbanizing regions such as Greater Accra.

**Keywords:** Air temperature prediction, Random forest, Gradient boosting regression, ARIMA, Machine learning

## 1. Introduction

Climate change has intensified global concern about rising temperatures and increasing climate variability. Accurate prediction of temperature trends is essential for agriculture, water resource management, urban planning, and environmental sustainability (Gupta et al. 2022; Shams et al. 2023). In rapidly urbanizing regions such as the Greater Accra Region of Ghana, rising temperatures are closely linked to urban expansion, land-use change, and population growth.

Traditional temperature forecasting approaches rely on statistical models such as linear regression and autoregressive time-series models. Among these, the Autoregressive Integrated Moving Average (ARIMA) model has been widely used for climate forecasting due to its ability to capture temporal dependencies within time-series data (Romilly, 2005; Box and Jenkins., 1976). However, these models often struggle to represent nonlinear relationships that frequently occur in climatic datasets (Shams et al. 2023).

Recent advances in machine learning have introduced powerful techniques capable of modelling nonlinear interactions within environmental datasets. Algorithms such as Random Forest and Gradient Boosting Regression have demonstrated strong predictive performance in environmental modelling due to their ability to capture nonlinear relationships and interactions between variables (Breiman, 2001; Zhou et al. 2023). Random Forest is an ensemble learning algorithm that constructs multiple decision trees using bootstrap sampling and aggregates their predictions to improve predictive accuracy and reduce overfitting (Breiman, 2001). Gradient Boosting Regression builds predictive models sequentially by correcting the errors made by previous models in the ensemble (Sibindi et al. 2023). Boosting-based algorithms are particularly effective in environmental modelling because they can capture nonlinear interactions between climate variables and surface processes. Recent studies have demonstrated the robustness of boosting models in modelling temperature-related environmental processes such as evapotranspiration and surface energy dynamics (Kumar et al. 2024; Heramb et al. 2023; Wang et al. 2025; Shu et al. 2026).

Despite these advances, comparisons between machine learning techniques and classical statistical models remain important for understanding their relative strengths and limitations. Such comparisons help determine the suitability of different modelling approaches for climate prediction tasks. This study therefore aims to compare the predictive performance of Random Forest, Gradient Boosting Regression, and ARIMA models for forecasting air temperature variations in the Greater Accra Region using long-term meteorological observations.

## 2. Study Area

The study focuses on the Greater Accra Region of Ghana (Figure 1), which is the political and economic hub of the country. The region lies between longitude 0°03′W and 0°25′W and latitude 5°30′N and 5°53′N, covering an area of approximately 3,533 km² (Frimpong and Asonomaso-Kwabre, 2022; Acheampong, 2021).

 

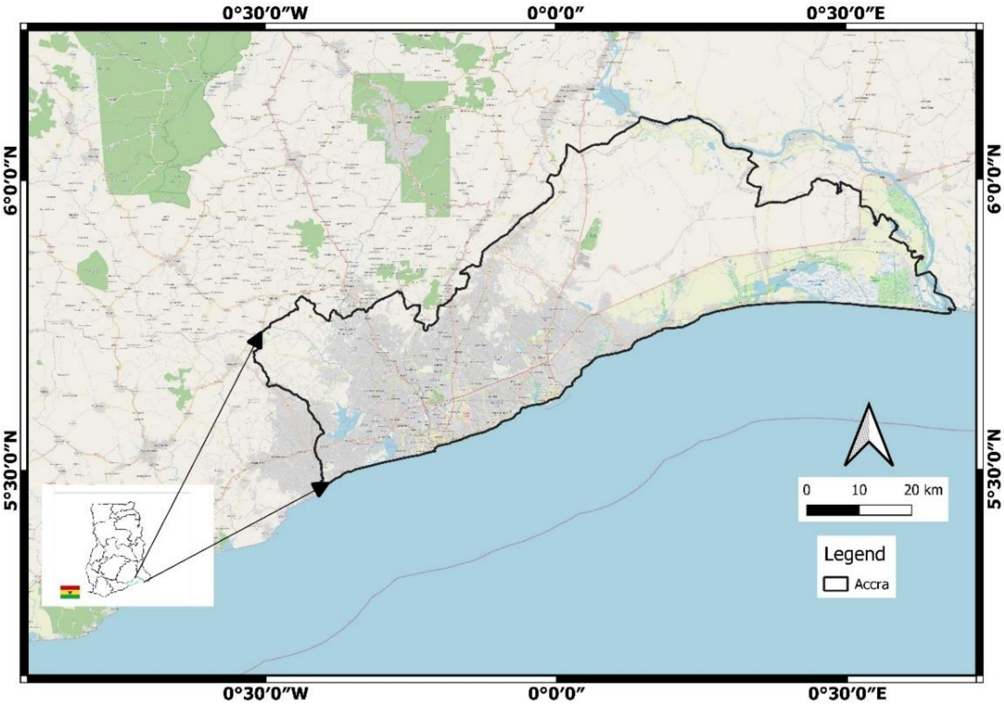


**Figure 1: Study Area Map**

The region is bordered by the Gulf of Guinea to the south, the Eastern Region to the north, and the Central Region to the west. Greater Accra is characterized by coastal plains and low-lying terrain that are highly influenced by coastal climatic conditions (Darko et al. 2022; Appeaning, 2015). The climate of the region is classified as tropical savannah, characterized by two rainfall seasons and a prolonged dry period influenced by Harmattan winds. The major rainy season occurs between May and June, while the minor rainy season occurs between October and November. Average annual rainfall ranges between 740 mm and 1200 mm (Bessah et al. 2022; Adu-Prah et al. 2019). Temperatures in the region remain relatively high throughout the year, with average monthly temperatures ranging between 25 °C and 29 °C. Rapid urbanization and land-use change have significantly influenced the thermal environment of the region, contributing to localized warming effects often associated with urban heat island dynamics (Waylen et al. 2014; Padi et al. 2022; Wemegah et al 2020).

## 3. Materials and Methods

### 3.1 Data Source

Daily temperature data were obtained from the Ghana Meteorological Agency for the period 1960–2018. The dataset includes:

- Minimum air temperature (Tmin)
- Maximum air temperature (Tmax)

The average daily air temperature was calculated according to Equation 1 as:

$$T_{Avg} = \sum_{j=1}^{i} \frac{T_{\min} + T_{\max}}{2} \quad (1)$$

Where $T_{min}$ and $T_{max}$ are the daily minimum and maximum time series data, i is an integer ranging from 1 to n.

### 3.2 Data Preprocessing

Prior to model development, the dataset was preprocessed to ensure consistency and reliability. Data cleaning procedures included removal of missing values and normalization of temperature variables to improve model stability during training.

### 3.3 Random Forest Model

Random Forest is an ensemble machine learning algorithm that constructs multiple decision trees and aggregates their predictions to produce the final model output. The method reduces overfitting and improves predictive accuracy through bootstrap aggregation (Breiman, 2001). Random Forest models are widely used in environmental modelling due to their ability to capture nonlinear relationships between predictors and response variables (Li et al. 2022)

### 3.4 Gradient Boosting Regression

Gradient Boosting Regression is another ensemble learning technique that sequentially builds decision trees to minimize prediction errors. Each new tree is trained to correct the residual errors of the previous tree (Sibindi et al, 2023). Boosting-based algorithms are well suited for climate modelling because they effectively capture nonlinear interactions between variables and can handle complex environmental datasets (Kumar et al. 2024; Heramb et al. 2023; Zounemat-Kermani and Kheimi, 2025; Zhang et al. 2025).

### 3.5 ARIMA Model

The ARIMA model is a classical statistical method for time-series forecasting that combines autoregressive (AR), differencing (I), and moving average (MA) components (Box and Jenkins, 976). In this study, an ARIMA (0,1,1) model was selected after testing for stationarity and performing first-order differencing on the temperature time series.

### 3.6 Model Evaluation

To ensure reliable model evaluation, five-fold cross-validation was applied. This approach reduces bias associated with a single train-test split and improves the robustness of the model performance assessment.

Model performance was evaluated using:

- Mean Squared Error (MSE)
- Coefficient of Determination ($R^2$)

## 4. Results and Discussion

### 4.1 Temperature Time-Series Trends

The time-series analysis of average air temperature in the Greater Accra Region indicates fluctuations in temperature values over the study period with gradual warming trends in recent decades. Figure 2 is the time series of average air temperature.

### 4.2 Model Performance Comparison

The Random Forest model achieved the highest predictive accuracy with MSE = 0.0010 °C and $R^2 = 0.9996$, followed by the Gradient Boosting Regression model with MSE = 0.0015 °C and $R^2 = 0.9994$. In contrast, the ARIMA model showed considerably lower predictive accuracy. Figures 3, Figure 4 and Figure 5 are the observed and predicted temperature by each model and Table 1 is the presents the model evaluation results of the RF, GBR, and ARIMA models.

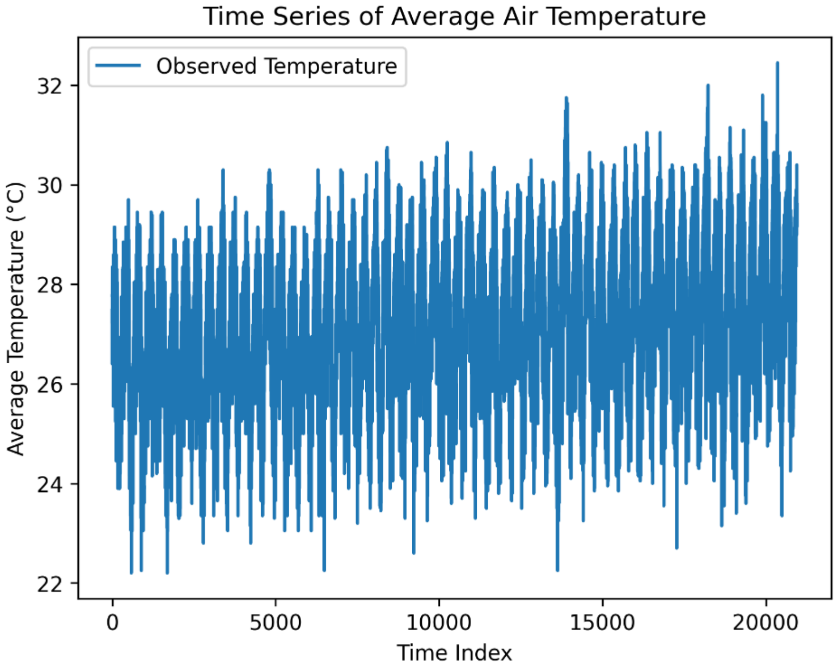


**Figure 2: Time Series of Average Air Temperature**

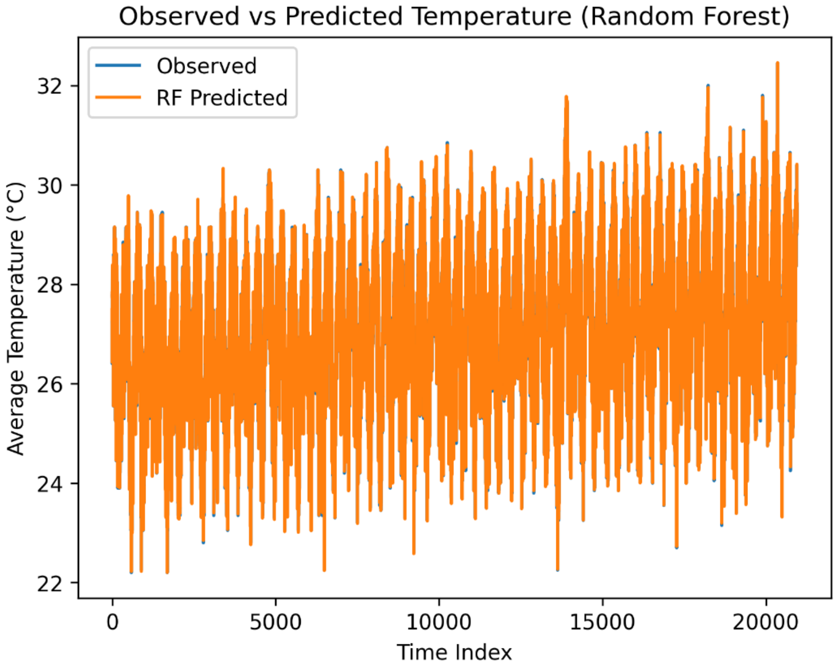


**Figure 3: Observed vs Predicted Temperature by the Random Forest Model**

**Table 1: Model Evaluation of the Models (units in Degree Celsius)**

| PCI | $MSE$ | $R^2$ |
|---|---|---|
| RF | 0.001022 | 0.9996 |
| GBR | 0.001521 | 0.9994 |
| ARIMA | 0.598179 | 0.3000 |

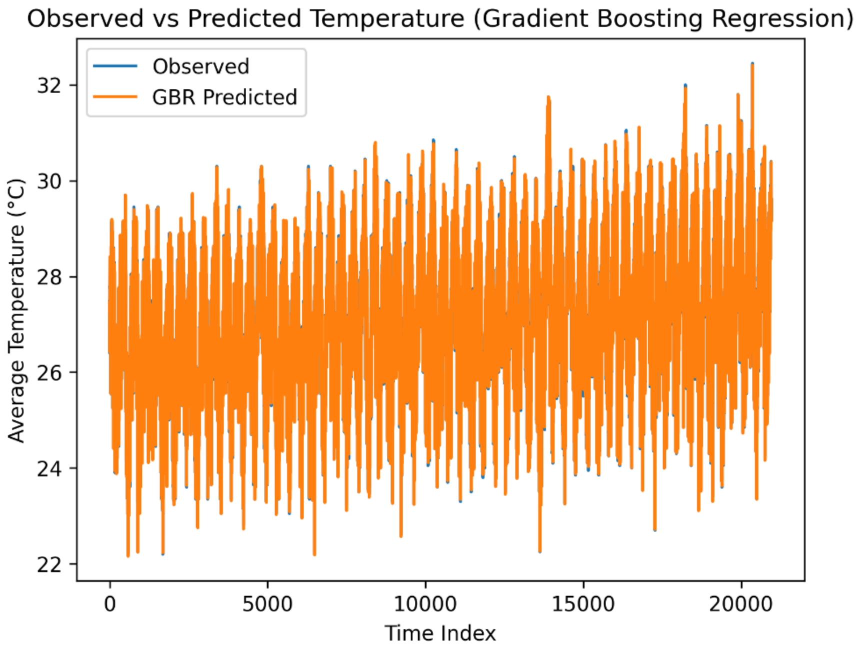


**Figure 4: Observed vs Predicted Temperature by the Gradient Boosting Regression Model**

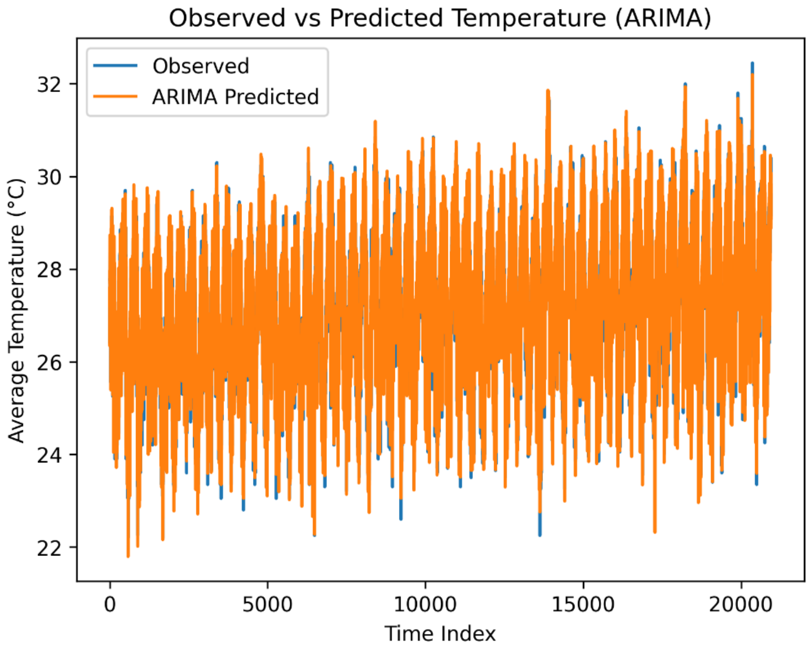


**Figure 5: Observed vs Predicted Temperature by the ARIMA Model**

### 4.3 Interpretation of Model Accuracy

The extremely high predictive accuracy observed in the machine learning models is largely due to the deterministic relationship between the predictor variables and the computed target variable. Because the average temperature is derived directly from minimum and maximum temperature values, the models effectively learn a direct mathematical relationship rather than complex physical temperature dynamics. Similar observations have been reported in recent studies where high model accuracy may occur when predictors are directly related to the target variable (Kumar et al. 2025; Fisher et al. 2019; Ortiz-Garcia et al. 2014).

### 4.4 Comparison with Previous Studies

Previous studies have demonstrated the effectiveness of machine learning models such as Random Forest and Gradient Boosting Regression in environmental prediction tasks (Niu et al. 2019; Kumar et al. 2024). Recent research also highlights the importance of integrating machine learning models with physically meaningful predictors to improve interpretability and reliability in environmental modelling (Holzman et al. 2025).

## 5. Conclusions

This study compared the performance of Random Forest, Gradient Boosting Regression, and ARIMA models for predicting air temperature in the Greater Accra Region of Ghana using long-term meteorological data. The machine learning models demonstrated significantly higher predictive accuracy than the classical ARIMA model. However, the extremely high accuracy observed in the models is partly due to the deterministic relationship between the predictor variables and the computed target variable. Future research should incorporate additional environmental predictors such as humidity, solar radiation, vegetation indices, and land-use characteristics to improve the physical interpretability of temperature prediction models. The findings demonstrate the potential of machine learning approaches for climate prediction and environmental monitoring in rapidly urbanizing regions.

**Acknowledgments**

The Authors appreciate the anonymous reviewers' constructive criticism, time, and efforts in helping to make this paper better.

**Conflict of Interest**

The Authors declare that there is no potential conflict of interest in this work. The work does not infringe any copyright, proprietary right, or any other right of any third parties, and the Authors are the sole owners of the work.

**References**

Acheampong, R. A. 2021. "Accra: City Scoping Study." *African Cities Research Consortium*:1–11.

Adu-Prah, S., S. Appiah-Opoku, and D. Aboagye. 2019. "Spatiotemporal Evidence of Recent Climate Variability in Ghana." *African Geographical Review* 38(2):172–190.

Appeaning Addo, K. 2015. "Monitoring Sea Level Rise-Induced Hazards along the Coast of Accra in Ghana." *Natural Hazards* 78(2):1293–1307.

Bessah, E., W. Amponsah, S. O. Ansah, A. Afrifa, B. Yahaya, C. S. Wemegah, et al. 2022. "Climatic Zoning of Ghana Using Selected Meteorological Variables for the Period 1976–2018." *Meteorological Applications* 29(1):e2049.

Box, G. E. P., and G. M. Jenkins. 1976. *Time Series Analysis: Forecasting and Control.* Boca Raton, FL: Holden-Day.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45(1):5–32.

Darko, G., S. Bi, I. Sarfo, S. O. Y. Amankwah, F. E. Azeez, E. Yeboah, et al. 2022. "Impacts of Climate Hazards on Coastal Livelihoods in Ghana: The Case of Ningo-Prampram in the Greater Accra Region." *Environment, Development and Sustainability* 24(1):1445–1474.

Fisher, A., C. Rudin, and F. Dominici. 2019. "All Models Are Wrong, but Many Are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously." *Journal of Machine Learning Research* 20(177):1–81.

Frimpong, B. F., and A. R. G. Asonomaso-Kwabre. 2022. *Analysis of the Relationship between Spatial Urban Expansion and Temperature Utilizing Remote Sensing and GIS Techniques in the Accra and Kumasi Metropolises in Ghana.* Doctoral dissertation, Brandenburgische Technische Universität Cottbus-Senftenberg.

Gupta, I., H. Mittal, D. Rikhari, and A. K. Singh. 2022. "MLRM: A Multiple Linear Regression Based Model for Average Temperature Prediction of a Day." *arXiv Preprint* arXiv:2203.05835.

Heramb, P., K. V. Ramana Rao, A. Subeesh, and A. Srivastava. 2023. "Predictive Modelling of Reference Evapotranspiration Using Machine Learning Models Coupled with Grey Wolf Optimizer." *Water* 15(5):856.

Holzman, M., A. Srivastava, R. Rivas, and A. Huete. 2025. "Evaluating the Relationship between Vegetation Status and Soil Moisture in Semi-Arid Woodlands, Central Australia, Using Daily Thermal, Vegetation Index, and Reflectance Data." *Remote Sensing* 17(4).

Kumar, M., Y. Agrawal, S. Adamala, Pushpanjali, A. V. M. Subbarao, V. K. Singh, and A. Srivastava. 2024. "Generalization Ability of Bagging and Boosting Type Deep Learning Models in Evapotranspiration Estimation." *Water* 16(16):2233.

Kumar, U., Rashmi, A. Srivastava, H. V. Parmar, H. H. Mashru, P. A. Pandya, et al. 2025. "Modelling Soil Temperature at Multiple Depths in Saurashtra Region (Junagadh) of Gujarat Using Machine Learning and Shapely Approach." *Discover Applied Sciences* 7(9):981.

Li, Z., S. Bi, S. Hao, and Y. Cui. 2022. "Aboveground Biomass Estimation in Forests with Random Forest and Monte Carlo-Based Uncertainty Analysis." *Ecological Indicators* 142:109246.

Niu, L., R. Tang, Y. Jiang, and X. Zhou. 2019. "Spatio-Temporal Patterns and Drivers of the Surface Urban Heat Island in 36 Major Cities in China: A Comparison of Two Different Methods for Delineating Rural Areas." *Sustainability* 12:478.

Ortiz-García, E. G., S. Salcedo-Sanz, and C. Casanova-Mateo. 2014. "Accurate Precipitation Prediction with Support Vector Classifiers: A Study Including Novel Predictive Variables and Observational Data." *Atmospheric Research* 139:128–136.

Padi, M., B. A. K. Foli, E. S. Nyadjro, K. Owusu, and G. Wiafe. 2022. "Extreme Rainfall Events over Accra, Ghana, in Recent Years." *Remote Sensing in Earth Systems Sciences* 5(1):71–82.

Romilly, P. 2005. "Time Series Modelling of Global Mean Temperature for Managerial Decision-Making." *Journal of Environmental Management* 76:61–70.

Shams, M. Y., Z. Tarek, A. M. Elshewey, M. Hany, A. Darwish, and A. E. Hassanien. 2023. "A Machine Learning-Based Model for Predicting Temperature under the Effect of Climate Change." Pp. 61–81 in *The Power of Data: Driving Climate Change with Data Science and Artificial Intelligence Innovations*, edited by A. E. Hassanien and A. Darwish. Studies in Big Data, Vol. 118. https://doi.org/10.1007/978-3-031-22456-0-4

Shu, Y., K. Zhang, H. Duan, and L. Wu. 2026. "Simulation Method for Considering the Lag of Turbulent Evapotranspiration." *Ecohydrology* 19(1):e70184.

Sibindi, R., R. W. Mwangi, and A. G. Waititu. 2023. "A Boosting Ensemble Learning-Based Hybrid Light Gradient Boosting Machine and Extreme Gradient Boosting Model for Predicting House Prices." *Engineering Reports* 5(4):e12599.

Wang, L., J. Kong, Q. Zhang, and Y. Zhong. 2025. "A Dual-Source Energy Balance Model-Assisted Ensemble Learning Approach for Estimating Latent Heat Flux." *Theoretical and Applied Climatology* 156(7):399.

Waylen, P., and K. Owusu. 2014. "Changes in Expectations and Extremes in the Rainfall Climatology of Accra, Ghana, 1895–2005." *Applied Geography* 52:99–109.

Wemegah, C. S., E. I. Yamba, J. N. Aryee, F. Sam, and L. K. Amekudzi. 2020. "Assessment of Urban Heat Island Warming in the Greater Accra Region." *Scientific African* 8:e00426.

Zhang, B., H. Yu, Z. Hu, P. Yue, Z. Tang, H. Luo, et al. 2025. "A Machine Learning-Based Observational Constraint Correction Method for Seasonal Precipitation Prediction." *Advances in Atmospheric Sciences* 42(1):36–52.

Zhou, Z., C. Qiu, and Y. Zhang. 2023. "A Comparative Analysis of Linear Regression, Neural Networks and Random Forest Regression for Predicting Air Ozone Employing Soft Sensor Models." *Scientific Reports* 13(1):22420.

Zounemat-Kermani, M., and M. Kheimi. 2025. "Ensemble Boosting Methods for Surface Water Quality Modeling: A Review." *Archives of Computational Methods in Engineering*:1–38.