ELSEVIER Scopus doi

Journal of Vibration Engineering

ISSN:1004-4523

Registered

- SCOPUS
- DIGITAL OBJECT IDENTIFIER (DOI)
- GOOGLE SCHOLAR
- IMPACT FACTOR 6.1

Our Website
www.jove.science

Data-Driven Modelling of Geoclimatic Factor in a Coastal Tropical Environment Using ERA5 Reanalysis

Y. B. Lawal[1*], P. A. Owolawi[2], C. Tu[3], E. Van Wyk[4]
[1,2,3]Department of Computer Systems Engineering, Tshwane University of Technology, South Africa
[4]Faculty of Information and Communications Technology, Tshwane University of Technology

Abstract

Accurate characterization of the geoclimatic factor is essential for reliable terrestrial radio link design in coastal tropical environments, where strong moisture variability often induces anomalous propagation conditions. This study presents a data-driven framework for predicting the geoclimatic factor over Toru-Orua, a Niger Delta coastal location, using ERA5 reanalysis data and machine learning techniques. Monthly meteorological variables spanning 2015–2025, including surface temperature, dew point temperature, relative humidity, vapour pressure, and wind speed, were used as input predictors. Multiple models were evaluated, comprising Random Forest (RF), Extra Trees (ET), and Gradient Boosting Regression (GBR). The results indicate that Gradient Boosting Regression substantially outperform other tree-based ensemble models, achieving the highest predictive accuracy with MAE, RMSE, and coefficient of determination (R^2) of 5×10^{-6}, 10^{-6} and 0.76 respectively. The winner model performance closely reproduces observed seasonal variability of geoclimatic. Feature importance analysis reveals that moisture-related parameters, particularly dew point temperature, dominate geoclimatic factor variability, underscoring the critical role of atmospheric water vapour in coastal radioclimatic processes. The findings are consistent with previous studies in southern Nigeria and demonstrate the suitability of ERA5-driven machine learning models for improved fade margin estimation, interference mitigation, and robust planning of terrestrial line-of-sight radio communication systems in humid tropical regions.

Keyword: Geoclimatic Factor, Refractivity gradient, ERA5 reanalysis data, Machine Learning, Terrestrial radio link.

1.0 Introduction

The geoclimatic factor is a key secondary radioclimatic parameter that encapsulates the combined influence of atmospheric moisture, temperature structure, and refractivity gradients on radio wave propagation. It is widely employed in terrestrial and Earth–space propagation models to characterize climatic severity, particularly in tropical and coastal regions where anomalous propagation and rain-induced impairments are prevalent (Ajayi & Barbaliscia, 1990; ITU-R P.530-17). Accurate estimation of the geoclimatic factor is therefore essential for reliable fade margin determination, interference mitigation, and robust design of terrestrial microwave links. In humid tropical environments, such as the Niger Delta region of Nigeria, atmospheric conditions are highly variable in time and space. Persistent high humidity, strong boundary-layer stratification, and pronounced seasonal transitions associated with the West African monsoon system frequently lead to super-refraction and ducting, extending radio horizons beyond standard assumptions (Willoughby, 2002; Adeyemi & Emmanuel, 2011). Conventional approaches to estimating the geoclimatic factor are largely empirical, often based on radiosonde observations or long-term climatological averages. While these methods have provided valuable insights, they are limited by sparse spatial coverage, data gaps, and an inability to capture nonlinear interactions among multiple meteorological drivers, especially at high temporal resolution.

Recent advances in atmospheric reanalysis products and machine learning (ML) techniques provide an opportunity to overcome these limitations. The ERA5 reanalysis, produced by the Copernicus Climate Change Service, offers globally consistent, high-resolution meteorological data with demonstrated reliability for radioclimatic studies (Hersbach et al., 2020). At the same time, ML models such as Random Forest, Extreme Gradient Boosting, and Gaussian Process Regression have shown strong capability in learning complex nonlinear relationships between atmospheric variables and propagation-relevant parameters (Salonen & Uppala, 1991; Abimbola et al., 2021). Despite these advances, the application of ML to the prediction of geoclimatic factor remains limited, particularly for coastal tropical environments in West Africa.

This study focuses on Toru-Orua, the host community of the University of Africa, located in Bayelsa State within the core Niger Delta region of southern Nigeria. Toru-Orua lies in a low-lying coastal plain characterized by extensive wetlands, proximity to the Atlantic Ocean, and a tropical monsoon climate with a prolonged wet season (April–October) and a short dry season (November–March). These geographic and climatic attributes make the area especially susceptible to strong moisture gradients, frequent anomalous propagation, and high rain attenuation risk, thereby posing significant challenges to terrestrial wireless communication systems. The motivation for this research is threefold. First, there is a clear need for site-specific and data-driven models that can reliably predict geoclimatic factor under the highly variable conditions of the Niger Delta. Second, existing studies over southern Nigeria have largely relied on traditional statistical or empirical methods, with limited exploitation of modern ML techniques and high-resolution reanalysis data. Third, accurate prediction of geoclimatic factor using readily available surface meteorological variables would provide a practical tool for network planners and engineers, reducing reliance on scarce upper-air measurements. Accordingly, this study aims to develop and evaluate a machine learning framework for predicting the geoclimatic factor over Toru-Orua using ERA5 reanalysis data spanning 2017–2025. By leveraging multiple ML models and robust statistical evaluation metrics, the work seeks to enhance understanding of the dominant atmospheric controls on geoclimatic factor and to provide actionable insights for reliable terrestrial radio link design in coastal tropical environments.

1.1 Theoretical Background of the applied Machine Learning Models

The prediction of the geoclimatic factor constitutes a supervised regression problem in which a nonlinear functional relationship is established between a set of surface meteorological predictors and a continuous target variable. Let the input feature vector be x as shown in equation (1).

$$x = [T, T_d, e, RH, WS] \tag{1}$$

where T denotes surface air temperature, T_d is dew point temperature, e represents water vapour pressure, RH is relative humidity, and WS is wind speed. The objective of the learning process is to approximate a function $f(\cdot)$ such that

$$\widehat{K} = f(x) \tag{2}$$

where $\widehat{K}$ is the predicted geoclimatic factor. The theoretical background of the applied machine learning models with reference to equations (1) and (2) are described briefly in the subsequent paragraph.

The Random Forest (RF) model is an ensemble learning method that constructs multiple decision trees using bootstrap samples of the training dataset and random subsets of predictors at each node. The RF prediction is obtained by averaging the outputs of all individual trees and is given by equation (3).

$$\widehat{K}_{RF} = \frac{1}{M}\sum_{m=1}^{M} f_m(x) \tag{3}$$

where $f_m(x)$ denotes the prediction from the m-th decision tree and M is the total number of trees. This ensemble averaging mechanism reduces variance and enhances robustness, making RF particularly suitable for noisy atmospheric datasets.

Extra Trees (ET) follows a similar ensemble structure to random forest but introduces stronger randomization by selecting split thresholds randomly rather than optimizing them. Using the entire training set instead of bootstrap samples, the ET prediction is expressed as

$$\widehat{K}_{ET} = \frac{1}{M}\sum_{m=1}^{M} \tilde{f}_m(x) \tag{4}$$

where $\tilde{f}_m(x)$ represents an extremely randomized decision tree. This increased randomness often leads to improved generalization for relatively small climatological datasets.

Gradient Boosting Regression (GBR) is a sequential ensemble technique in which models are trained iteratively to correct the residuals of preceding models. At iteration *k*, the prediction is updated as

$$\widehat{K}_{GPR} = \widehat{K}^{(k-1)} + \eta h_k(x) \tag{5}$$

where h_k(·) is the newly fitted regression tree and η is the learning rate controlling the contribution of each tree. The model minimizes a specified loss function through gradient descent, enabling it to capture subtle nonlinear dependencies associated with moisture-driven refractivity processes.

These models provide a balanced framework combining interpretability, nonlinear learning capability, and uncertainty quantification for modelling geoclimatic factor variability in coastal tropical environments.

2.0 Methodology

2.1 Data Description

This study employs ERA5 reanalysis data produced by the Copernicus Climate Change Service (C3S) to develop a machine learning–based prediction model for the geoclimatic factor over Toru-Orua, Bayelsa State, Nigeria. ERA5 provides globally consistent, high-quality atmospheric data generated through the assimilation of in situ observations and satellite measurements using advanced numerical weather prediction models, and has been widely validated for radioclimatic and propagation studies (Hersbach et al., 2020). Hourly ERA5 data covering the period 2017–2025 were extracted for the study location. The predictor variables selected are surface air temperature, dew-point temperature, relative humidity, water vapour pressure, and wind speed, as these parameters directly influence atmospheric moisture content, boundary-layer stability, and refractivity structure, which are fundamental drivers of geoclimatic variability (Adeyemi & Emmanuel, 2011; ITU-R P.453-14). Water vapour pressure was derived from temperature and dew-point temperature using standard thermodynamic relationships consistent with ITU-R recommendations. The target variable, the geoclimatic factor, was derived from refractivity gradients following established formulations adopted in tropical propagation studies and ITU-R–based models. Monthly refractivity gradients at 1 km altitude were used to compute the geoclimatic factor, ensuring temporal smoothing and reducing sensitivity to near-surface noise, while preserving climatological relevance.

All datasets were quality-checked for consistency, missing values, and outliers prior to modelling. The use of ERA5 enables continuous long-term analysis without the spatial and temporal limitations associated with radiosonde observations, making the dataset particularly suitable for data-driven modelling in data-sparse coastal environments such as the Niger Delta.

2.2 Data Processing and Model Training

The extracted ERA5 datasets were subjected to systematic pre-processing prior to model development. Quality control procedures were applied to remove incomplete records and ensure temporal consistency across all predictor variables. Water vapour pressure was computed from surface temperature and dew-point temperature using standard thermodynamic relationships, after which all input variables were normalised to eliminate scale-related bias during training. The complete dataset was randomly partitioned into training (70%), validation (15%), and testing (15%) subsets to enable robust model development and independent performance assessment. Random Forest, Extreme Gradient Boosting, and Gaussian Process Regression models were implemented using the respective equations and identical input features describes in the previous section to ensure fair comparison. Parameters for the models were optimised using the validation set, while overfitting was controlled through regularisation and ensemble averaging. The parameters used for training the models are presented in Table 1. Model performance was finally evaluated on the unseen test dataset using statistical and error-based metrics.

Table 1: Model Hyperparameters Used for Geoclimatic Factor Prediction

Model	Parameter	Value	Description
Random Forest (RF)	n_estimators	300	Number of trees
	max_depth	None	Maximum tree depth
	min_samples_split	2	Minimum samples to split a node
	min_samples_leaf	1	Minimum samples per leaf
	bootstrap	True	Bootstrap sampling enabled

	random_state	42	Reproducibility
Extra Trees (ET)	n_estimators	300	Number of trees
	max_depth	None	Maximum tree depth
	min_samples_split	2	Minimum samples to split a node
	min_samples_leaf	1	Minimum samples per leaf
	bootstrap	False	No bootstrap sampling
	random_state	42	Reproducibility
Gradient Boosting (GBR)	n_estimators	300	Number of boosting stages
	learning_rate	0.05	Learning rate
	max_depth	3	Depth of individual trees
	loss	Squared error	Loss function
	subsample	1.0	Fraction of samples used
	random_state	42	Reproducibility

3.0 RESULTS AND DISCUSSION

3.1 Model Performance Comparison and Statistical Analysis of Prediction

The parameters of each prediction model were tuned to obtained best results. Table 2 present a comparative evaluation of the machine-learning models used for monthly geoclimatic factor prediction over Toru-Orua for the period 2015–2025. Overall, all three tree-based models demonstrate strong predictive capability, with clear differences in accuracy and robustness.

Table 2: Comparison of performance metrics for all models

Models	MAE ($\times 10^{-6}$)	RMSE ($\times 10^{-5}$)	R^2
Random Forest	6	1.1	0.67
Extra Trees	5	1.1	0.71
Gradient Boosting Regression	5	0.1	0.76

The Gradient Boosting Regression (GBR) model exhibits the best performance, achieving the lowest error metrics (RMSE ≈ 1.0×10^{-6}, MAE ≈ 5×10^{-6}) and the highest coefficient of determination ($R^2 \approx 0.76$). This indicates that GBR captures both the magnitude and temporal variability of the geoclimatic factor more effectively than the other models. The superior performance of GBR can be attributed to its sequential error-correction mechanism, which is well suited for learning nonlinear atmospheric dependencies embedded in reanalysis data. Similar advantages of boosting techniques for atmospheric and radioclimatic prediction have been reported in recent studies over tropical and coastal regions (Sabiru et al., 2024; Bello et al., 2024). The Extra Trees model follows closely ($R^2 \approx 0.71$), benefiting from enhanced randomization that improves generalization while maintaining low bias. The Random Forest model also performs satisfactorily ($R^2 \approx 0.67$), confirming its reliability for geoclimatic modelling, as widely documented in Nigerian and West African propagation studies (Lawal & Omotoso, 2023; Abimbola et al., 2021).

Figure 1 further shows that all models reproduce the observed seasonal peaks and interannual fluctuations, particularly during high-variability periods. A critical inspection of the figure reveals that all the models follow the seasonal variability of geoclimatic factor over the study period. The comparatively low of RMSE and MAE and of determination coefficient of 0.76 reinforce the suitability and reliability of the GBR prediction model. Collectively, these results highlight the suitability of ensemble tree-based models for accurate, data-driven estimation of geoclimatic factors in coastal tropical environments.

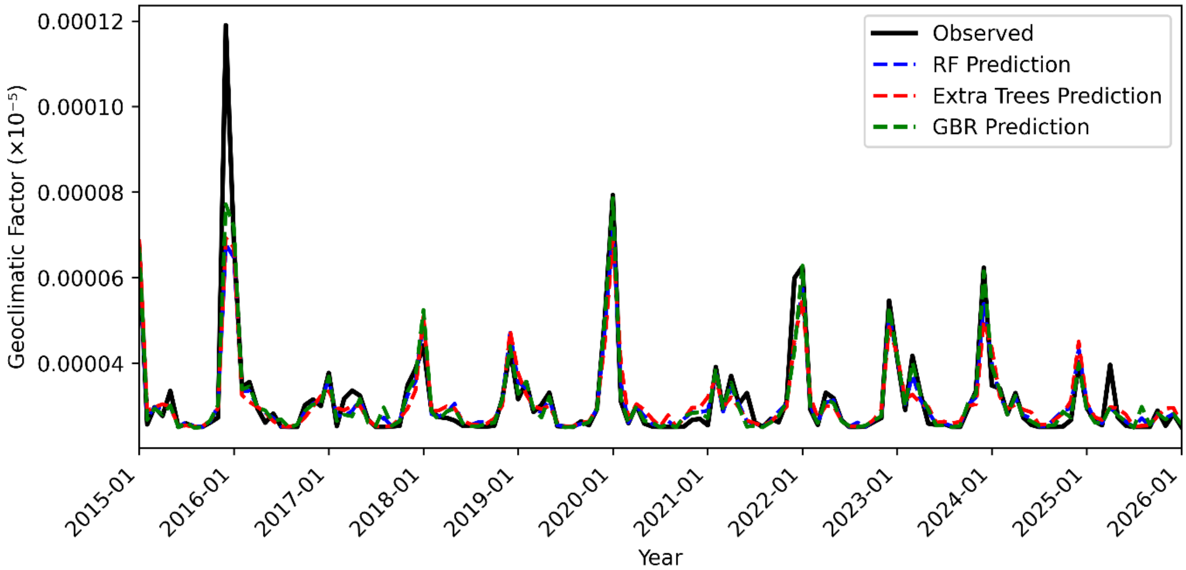

Figure 1: Times series plot of the actual and predicted Geoclimatic Factor

3.2 Feature Importance of the meteorological predictors

The contribution of each meteorological parameter used as input variables during the model training were investigated to determine their influence on geoclimatic prediction. This is a crucial as it reveals the main drivers of geoclimatic variability that plays significant role in radio wave propagation. Figure 2 presents the relative importance of the input predictors in the winner model (GBR) for geoclimatic factor prediction over Toru-Orua.

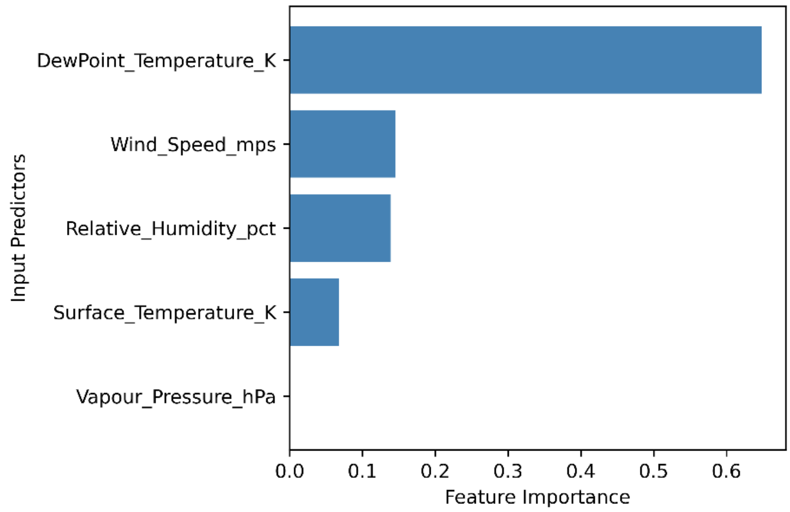

Figure 2: Feature Importance Ranking for Geoclimatic Factor Prediction

The results clearly indicate that dew point temperature is the dominant predictor, contributing approximately 65% of the total importance. This highlights the strong control of near-surface moisture content on refractivity gradients and, consequently, on the geoclimatic factor. Dew point temperature directly governs atmospheric water vapour concentration, which plays a critical role in modifying the vertical refractivity structure in humid coastal environments (Adeyemi & Emmanuel, 2011; Lawal & Omotoso, 2023). Wind speed and relative humidity follow with comparable contributions of about 14.5% and 13.9%, respectively. Wind speed influences turbulent mixing and vertical moisture redistribution within the lower troposphere, thereby modulating refractivity gradients, while relative humidity reflects moisture availability that enhances anomalous propagation conditions (Ajayi & Barbaliscia, 1990; Bello et al., 2024). Surface temperature exhibits a relatively smaller influence (~6.8%), suggesting that its impact on geoclimatic factor is largely indirect, mainly through its coupling with moisture-related parameters. Interestingly, vapour pressure shows negligible importance, likely due to strong collinearity with dew point temperature and relative humidity, which already encapsulate moisture variability in the model.

Overall, the feature ranking confirms that moisture-related parameters dominate geoclimatic factor variability in the Niger Delta coastal environment, reinforcing earlier climatological findings and validating the physical consistency of the machine learning framework adopted in this study (Etokebe et al., 2016; Sabiru et al., 2024).

3.3 Comparison with Previous Studies and Implications for Terrestrial Radio Link Design

The predicted geoclimatic factor over Toru-Orua exhibits magnitudes and temporal behaviour consistent with earlier empirical and climatological studies conducted in Nigeria's coastal and Niger Delta regions. Previous works over Lagos, Port Harcourt, Yenagoa, and Calabar have similarly reported elevated geoclimatic factors driven by persistent high moisture content, weak vertical mixing, and frequent super-refractive conditions (Ajayi & Barbaliscia, 1990;

Etokebe et al., 2016; Lawal & Omotoso, 2023; Bello et al., 2024). The close agreement between observed and machine-learning-predicted values in this study reinforces the reliability of data-driven approaches in capturing complex atmospheric influences that are often simplified in traditional climatological models.

Compared with conventional statistical estimation techniques, the machine learning framework demonstrates improved predictive accuracy and temporal adaptability, making it more suitable for modern radio network planning. Accurate prediction of the geoclimatic factor is crucial for realistic fade margin estimation, interference mitigation, and link availability assessment, particularly for line-of-sight terrestrial microwave systems operating in humid coastal environments. The methodology adopted here is transferable to other coastal tropical regions with similar climatic characteristics, supporting broader applications in regional radio propagation modelling.

Conclusion

This study developed a robust data-driven framework for predicting the geoclimatic factor over Toru-Orua, a coastal tropical environment in the Niger Delta, using ERA5 reanalysis data and machine learning techniques. The results demonstrate that tree-based ensemble models, particularly Gradient Boosting Regression, provide high predictive accuracy and strong agreement with observed monthly geoclimatic factor values. This confirms the suitability of machine learning approaches for capturing the nonlinear interactions between atmospheric variables that govern refractivity gradients in humid coastal regions. Feature importance analysis revealed that moisture-related parameters, especially dew point temperature, exert dominant control on geoclimatic factor variability, highlighting the critical role of atmospheric water vapour in anomalous propagation conditions. The findings are consistent with previous empirical studies conducted in southern Nigeria and reinforce the reliability of ERA5-driven modelling for radioclimatic assessment. From an engineering perspective, the improved prediction of geoclimatic factor has direct implications for accurate fade margin estimation, interference mitigation, and enhanced reliability of terrestrial line-of-sight radio links. Future research should consider incorporating longer climatological datasets, multi-level atmospheric profiles, and hybrid or deep-learning architectures to further improve model generalisation and operational applicability across diverse tropical coastal regions.

References

Abimbola, O. J., Afullo, T. J., & van Rensburg, L. J. (2021). Estimation of radio refractivity from satellite-derived data over West Africa. *Results in Physics, 26*, 104353. https://doi.org/10.1016/j.rinp.2021.104353

Adeyemi, O., & Emmanuel, O. (2011). Variability of radio refractivity and propagation conditions in Nigeria. *Journal of Atmospheric and Solar-Terrestrial Physics, 73*(2–3), 273–281. https://doi.org/10.1016/j.jastp.2010.10.006

Ajayi, G. O., & Barbaliscia, F. (1990). Radiowave propagation characteristics of the Nigerian troposphere. *IEEE Transactions on Antennas and Propagation, 38*(9), 1311–1319. https://doi.org/10.1109/8.59774

Bello, G., Akpootu, D. O., Zoramawa, A. B., & Yusuf, A. (2024). Performance analysis of tropospheric radio refractivity, refractivity gradient and effective Earth radius in the coastal zone of Nigeria. *UMYU Scientifica Journal, 3*(3), 251–266. https://doi.org/10.56919/usci.2433.028

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Etokebe, I. J., Uko, M. C., & Chinwe, I. U. (2016). Determination of refractivity gradient and geoclimatic factor using radiosonde data. *International Journal of Systems Science and Applied Mathematics, 1*(4), 76–82. https://doi.org/10.11648/j.ijssam.20160104.17

Lawal, Y. B., & Omotoso, E. T. (2023). Investigation of point refractivity gradient and geoclimatic factor in a coastal Nigerian environment. *Journal of the Nigerian Society of Physical Sciences, 5*, 1081–1091. https://doi.org/10.46481/jnsps.2023.1081

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. MIT Press.

Sabiru, A. Y., Akinbolati, A., Ikechiamaka, F. N., & Abe, B. T. (2024). Characterization of some secondary radio-climatic factors for reliable radio wave propagation and link design over Northwestern Nigeria. *Advances in Space Research, 74*(3), 890–898. https://doi.org/10.1016/j.asr.2024.04.038