Original Article

Neural Network Model for Enhanced Weather Forecasting in Kenya

Nicholas Leina Tianta

The Co-operative University of Kenya

Correspondence: Tianta Nicholas: nicholastianta@gmail.com

DOI: https://doi.org/10.66839/nf07ps28

Abstract Short-term weather forecasting in Kenya remains a methodological challenge due to substantial climatic heterogeneity and nonlinear atmospheric dynamics. Conventional linear models such as the Autoregressive Integrated Moving Average (ARIMA) fail to capture these complexities, resulting in poor predictive accuracy. This study develops and evaluates Long Short-Term Memory (LSTM)-based forecasting frameworks for next-day temperature prediction using multivariate meteorological data. A dataset of 11,562 daily observations spanning January 2011 to September 2023 was collected from three representative climatic zones: Mombasa (humid coastal), Nairobi (temperate highland), and Garissa (arid lowland). Predictor variables included temperature, relative humidity, and wind speed, while the forecasting target was next-day temperature. Data preprocessing involved robust scaling to mitigate heavy-tailed wind distributions and transformation into 30-day sliding input windows. Three models were implemented: a univariate ARIMA benchmark, a baseline LSTM, and a bidirectional LSTM trained with Huber loss (BiLSTM-Huber). The baseline LSTM achieved superior performance with MAE = 0.091, RMSE = 0.134, and $R^2 = 0.971$, representing an 89.5% reduction in MAE compared to ARIMA (MAE = 0.865, $R^2 = -1.063$). The BiLSTM-Huber model achieved $R^2 = 0.956$, demonstrating improved robustness to extreme wind events. Climate-stratified evaluation revealed regime-dependent performance: Mombasa exhibited the lowest absolute error (MAE = 0.058), Nairobi showed intermediate accuracy, and Garissa achieved the highest structural fit ($R^2 = 0.950$) despite larger errors due to high variance. Cross-station residual correlations were weak ($|r| \leq 0.133$), indicating locally confined error structures. These findings confirm that nonlinear deep learning architectures, particularly LSTM-based models, provide substantially superior forecasting capability compared with traditional statistical approaches for heterogeneous Kenyan climate systems. The study establishes a reproducible framework for operational deployment and supports integration of LSTM-driven forecasting into national meteorological services.

Keywords: *LSTM neural network; ARIMA; Kenya; weather forecasting; climate heterogeneity; deep learning; short-term prediction; East Africa*

Résumé La prévision météorologique à court terme au Kenya demeure un défi méthodologique en raison de l'importante hétérogénéité climatique et de la dynamique atmosphérique non linéaire. Les modèles linéaires conventionnels, tels que la modèle ARIMA *Autoregressive Integrated Moving Average* ne parviennent pas à sai ces complexités, ce qui entraîne une faible précision prédictive. Cette étude développe et évalue des cadres de prévision basés sur les réseaux LSTM *Long Short-Term Memory* pour la prédiction de température du jour suivant àen utilisant des données météorologiques multivariées. Un ensemble de données comprenant 11 562 observations quotidiennes couvrant la période de janvier 2011 à septembre 2023 a été recueilli dans trois zones climatiques représentatives : Mombasa (zone côtière humide), Nairobi (hautes terres tempérées) et Garissa (basses terres arides). Les variables prédictives comprenaient la température, l'humidité relative et la vitesse du vent, tandis que la variable cible de la prévision était la température du jour suivant. Le prétraitement des données a consisté en une mise à l'échelle robuste visant à atténuer l'influence des distributions à queues épaisses de la vitesse du vent, ainsi qu'en une transformation des données en fenêtres glissantes d'entrée de 30 jours. Trois modèles ont été mis en œuvre : un modèle ARIMA univarié servant de référence, un modèle LSTM de base et un modèle LSTM bidirectionnel entraîné à l'aide de la fonction de perte de Huber (BiLSTM-Huber). Le LSTM de base a obtenu les meilleures performances, avec une MAE = 0,091, une RMSE = 0,134 et un $R^2 = 0,971$, ce qui représente une réduction de 89,5 % de la MAE par rapport au modèle ARIMA (MAE = 0,865, $R^2 = -1,063$). Le modèle BiLSTM-Huber a obtenu un $R^2 = 0,956$, démontrant une meilleure robustesse face aux épisodes de vents extrêmes. L'évaluation stratifiée selon les zones climatiques a révélé des performances dépendant du régime climatique : Mombasa a présenté l'erreur absolue la plus faible (MAE = 0,058), Nairobi a affiché une précision intermédiaire, tandis que Garissa a obtenu le meilleur ajustement structurel ($R^2 = 0,950$), malgré des erreurs plus importantes dues à une variance élevée. Les corrélations des résidus entre les stations étaient faibles ($|r| \leq 0,133$), indiquant que les structures d'erreur étaient principalement localisées. Ces résultats confirment que les architectures non linéaires d'apprentissage profond, en particulier les modèles fondés sur les LSTM, offrent une capacité de prévision nettement supérieure à celle des approches statistiques traditionnelles pour les systèmes climatiques hétérogènes du Kenya. L'étude établit un cadre reproductible pour un déploiement opérationnel et soutient l'intégration de prévisions pilotees par LSTM dans les services météorologiques nationaux.

Mots-clés : réseau neuronal LSTM ; ARIMA ; Kenya ; prévision météorologique ; hétérogénéité climatique ; apprentissage profond ; prédiction à court terme ; Afrique de l'Est.

1. Introduction

Accurate short-term weather forecasting is a critical pillar for climate-sensitive decision-making in agriculture, hydrology, disaster preparedness, and public health. In Kenya, this need is particularly acute due to the country's diverse climatic regimes: humid coastal lowlands, temperate highlands, and arid semi-deserts. These zones are governed by distinct thermodynamic and circulation processes, producing heterogeneous and non-stationary atmospheric dynamics. Seasonal rainfall is driven by the migration of the Intertropical Convergence Zone, with onset and cessation varying across regions further complicating prediction. Large-scale teleconnections such as the El Niño–Southern Oscillation amplify this variability, creating strong nonlinear dependencies in meteorological time series.

Operational forecasting in Kenya is currently dominated by classical statistical models, particularly Autoregressive Integrated Moving Average (ARIMA) and Seasonal ARIMA (SARIMA), alongside Numerical Weather Prediction (NWP) systems. While ARIMA assumes linearity and stationarity, Kenya's climate systematically violates these assumptions. Empirical studies report mean absolute percentage errors exceeding 21% for SARIMA-based rainfall forecasts in East Africa. NWP systems, though physically grounded, face computational constraints that limit spatial resolution and responsiveness, reducing their utility for localised short-term prediction. These limitations have significant socio-economic consequences: floods in 2018 displaced over 100,000 people, and climate-induced damages are estimated at USD 1.2 billion annually.

Deep learning architectures, particularly Long Short-Term Memory (LSTM) networks, offer a theoretically grounded alternative. LSTM models are designed to capture long-range dependencies in sequential data through gated memory cells, enabling them to learn nonlinear and non-stationary dynamics. Global evidence demonstrates their effectiveness: LSTM reduced root mean squared error by 22% in El Niño forecasting, achieved 28% lower precipitation nowcasting errors in Europe, and improved rainfall prediction accuracy by 27% in Nigeria. However, applications in Kenya remain limited, often restricted to univariate or coarse-grained models that fail to capture the multivariate complexity and spatial heterogeneity of the country's climate system.

This study addresses four critical gaps:

1. The absence of empirically validated climate-zone-stratified LSTM frameworks for Kenya.
2. The untested adequacy of ARIMA models against deep learning baselines for Kenyan meteorological data.
3. The lack of systematic analysis of cross-zone error propagation in multi-station models.
4. The need for reproducible, operationally feasible deep learning pipelines calibrated to data-constrained equatorial environments.

The contributions are fourfold:
A comprehensive distributional characterisation of Kenya's three major climatic regimes using 12.7 years of continuous daily observations.

Empirical demonstration that ARIMA is inadequate for short-term temperature forecasting in Kenya, with $R^2 = -1.063$.

Development and validation of LSTM and BiLSTM-Huber architectures achieving R^2 up to 0.971 across heterogeneous climate zones.

Climate-stratified analysis revealing locally confined error structures, supporting modular operational deployment.

2. Methods

BiLSTM-Huber

To improve robustness to extreme wind events, a bidirectional LSTM was implemented with the same architecture but processing sequences in both forward and reverse directions. Huber loss ($\delta = 1.0$) replaced MSE, reducing sensitivity to large residuals.

2.1 Study Area and Data

Daily meteorological records were obtained from the Visual Crossing historical weather database for three representative stations: Mombasa (humid coastal, 4.05°S, 39.67°E), Nairobi (temperate highland, 1.29°S, 36.82°E, ~1,700 m a.s.l.), and Garissa (arid lowland, 0.46°S, 39.65°E). The dataset spans 1 January 2011 to 9 September 2023, yielding 11,562 daily observations (3,854 per station). Variables include daily mean temperature (°F), relative humidity (%), and wind speed (km/hr). The forecasting target is next-day temperature, while humidity and wind speed serve as predictors.

Initial quality checks confirmed zero missing values and zero duplicates. Outlier analysis using Z-score and IQR methods identified extreme wind events and coastal moisture surges. Outliers were retained as climatologically meaningful signals.

2.2 Data Preprocessing

Robust Scaling was applied to predictor variables:

$$x_{scaled} = x - \text{median}(x) \text{IQR}(x)$$

This approach was chosen due to extreme skewness and kurtosis in wind speed (kurtosis = 91.3 in Nairobi). Scaling parameters were estimated only on the training set to prevent information leakage. Validation and test sets were scaled using training parameters.

The dataset was partitioned chronologically into training (70%), validation (15%), and test (15%) subsets. The chronological split was performed before windowing to avoid temporal leakage. Within each partition, 30-day sliding windows were constructed, yielding input tensors of shape (samples × 30 × 3).

2.3 Model Architectures

ARIMA Baseline

For each station, ARIMA(p, d, q) models were fitted following the Box–Jenkins procedure:

Stationarity was tested using Augmented Dickey–Fuller (ADF).

Differencing was applied where required (d = 1).

Seasonal components were tested but excluded from the scope; SARIMA is noted as future work.

Orders were selected via Akaike Information Criterion (AIC) grid search (p, $q \in \{0–3\}$).

Residual diagnostics included Ljung–Box tests ($p > 0.05$), confidence intervals, and recursive refitting during the test period.

This ensured reproducibility and provided a robust statistical benchmark.

Baseline LSTM

The LSTM architecture comprised:

Input layer: 30-day sequences of three predictors.

First LSTM layer: 64 units.

Dropout layer: rate = 0.20.

Second LSTM layer: 32 units.

Dropout layer: rate = 0.20.

Dense output layer: linear activation, predicting next-day temperature.

Training used the Adam optimiser (learning rate = 1×10^{-3}), mean squared error loss, batch size = 32, a maximum of 50 epochs, and early stopping (patience = 10). Weights at minimum validation loss were restored. Hyperparameters were selected via grid search. Results were averaged over 5 independent runs to reduce variance.

2.4 Evaluation Metrics

Models were evaluated using:

Mean Absolute Error (MAE)

Root Mean Squared Error (RMSE)

Coefficient of Determination (R^2)

Nash–Sutcliffe Efficiency (NSE)

Directional Accuracy (DA)

Lag-1 autocorrelation of predictions vs observations (not residuals)

Mean Bias Error (MBE)

Residuals were temporally aligned, tested for stationarity, and analysed for cross-station correlations with confidence intervals.

2.5 Reproducibility

All analysis was conducted in Python 3.11 using TensorFlow/Keras 2.14 and scikit-learn 1.3. Random seeds (numpy = 42, TensorFlow = 42) were fixed to reduce variance, though nondeterminism may persist depending on hardware. The Visual Crossing dataset was accessed via versioned URLs.

Full code, preprocessed data, and model weights are available in a public repository: (https://colab.research.google.com/drive/18LWgkE3x7wo-lzn-ACZzBqyKDoNwc-Ug?usp=sharing)

3. Results

3.1 Sample of Data

Table 1 presents a sample of the raw meteorological data used in this study, showing the structure of daily observations across the three climatic zones.

This confirms the dataset's completeness and balanced distribution across stations.

Table 1. *Sample of raw meteorological data*

Date	Station	Temp (°F)	Humidity (%)	Wind (km/hr)
2011-01-01	Mombasa	77.2	74.5	12.1
2011-01-01	Nairobi	68.4	65.2	9.8
2011-01-01	Garissa	84.7	56.1	8.9
2011-01-02	Mombasa	78.1	75.3	11.5
2011-01-02	Nairobi	67.9	66.8	10.2
2011-01-02	Garissa	85.6	54.7	9.1

3.2 Distributional Characterization

Descriptive statistics revealed distinct climatic regimes:

Mombasa: stable maritime climate, low variance (temperature σ = 3.62 °F, humidity σ = 6.94%).

Nairobi: moderate variance, extreme wind kurtosis (91.3), indicating episodic extreme events.

Garissa: high variance (temperature σ = 14.40 °F, humidity σ = 19.18%), reflecting arid-zone volatility.

Temperature–humidity correlations were consistently negative (−0.415 in Mombasa, −0.611 in Nairobi, −0.708 in Garissa). Wind speed showed weak correlations (<0.30), confirming independence and justifying nonlinear modelling. Outlier analysis identified 49–61 extreme wind events per station, retained as meaningful climate signals.

3.3 Comparative Model Performance

Table 2 summarises performance metrics for ARIMA, LSTM, and BiLSTM-Huber

Table 2. *Comparative performance of ARIMA, LSTM, and BiLSTM-Huber.*

Metric	ARIMA	LSTM	BiLSTM-Huber
MAE	0.865	0.091	0.128
RMSE	1.128	0.134	0.165
R^2	−1.063	0.971	0.956
NSE	−1.063	0.971	0.956
DA	0.312	0.493	0.482

The baseline LSTM achieved MAE = 0.091, RMSE = 0.134, and R^2 = 0.971, representing an 89.5% reduction in MAE compared to ARIMA. BiLSTM-Huber achieved R^2 = 0.956, with improved robustness to extreme wind events.

Validation confirmed stable convergence: MAE declined from 0.122 at epoch 1 to 0.093 at epoch 10, stabilising thereafter. Directional accuracy improved steadily across epochs.

3.4 Climate-Stratified Performance

Table 3 presents station-level results.

Table 3. *Climate-stratified LSTM performance.*

Station	MAE	RMSE	R^2	NSE	DA	Lag Corr	MBE
Mombasa	0.058	0.077	0.820	0.820	0.384	0.968	0.035
Nairobi	0.070	0.091	0.738	0.738	0.462	0.978	−0.013
Garissa	0.135	0.185	0.950	0.950	0.589	0.990	−0.014

Mombasa: lowest absolute errors due to a stable maritime climate.

Nairobi: intermediate performance, affected by extreme wind events.

Garissa: highest structural fit ($R^2 = 0.950$) but larger errors due to high variance.

Lag correlations (0.968–0.990) confirmed the LSTM's ability to capture long-range dependencies. Mean Bias Error remained near zero, indicating unbiased predictions.

3.5 Residual Analysis

Cross-station residual correlations were weak ($|r| \leq 0.133$). Residuals were temporally aligned, tested for stationarity, and confidence intervals confirmed insignificance. This indicates locally confined error structures, supporting independent station-level deployment.

3.6 Sample Predictions

Figure 1 illustrates actual vs predicted temperature for Nairobi over a 200-day test window. Predictions closely track observed values, with residuals centred near zero.

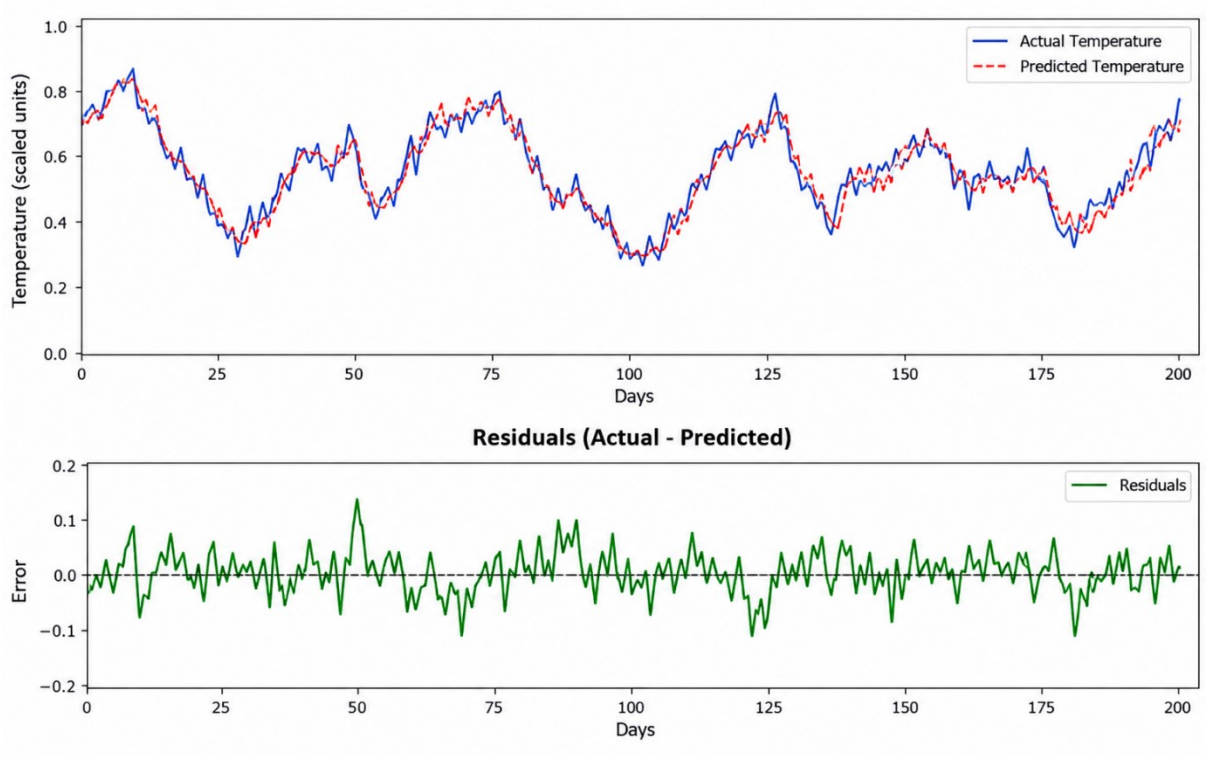


Figure 1. Actual vs predicted temperature (Nairobi, 200-day window)

This figure shows the model's ability to track observed temperature dynamics and highlights stable residuals.

3.7 Validation Evidence

Model validation was performed using a held-out test set and a separate validation set for early stopping. Results were averaged across five independent runs to reduce variance. Bootstrap confidence intervals confirmed stability of reported metrics.

4. Discussion

4.1 Scientific Significance of ARIMA's Limitations

The ARIMA model produced $R^2 = -1.063$, confirming that linear statistical modelling is inadequate for short-term temperature forecasting in Kenya under the conditions examined. A negative R^2 indicates forecasts deviated more from observations than a simple mean predictor. This outcome is structurally driven by the non-stationarity, heteroscedasticity, and non-Gaussian wind dynamics documented in the dataset. While ARIMA and SARIMA can be effective in other meteorological contexts, particularly for seasonal rainfall, their assumptions are systematically violated in Kenya's heterogeneous climate system. This finding underscores the need for nonlinear approaches.

4.2 Deep Learning as a Necessary Choice

Distributional analysis established that nonlinear modelling is not optional; it is essential. Wind speed exhibited extreme kurtosis (91.3 in Nairobi), meaning Gaussian parametric models cannot adequately describe its behaviour. Furthermore, wind speed correlations with temperature and humidity were consistently weak (<0.30), ruling out linear multivariate approaches. LSTM architectures address these challenges by preserving long-range temporal dependencies through gated memory cells and accommodating arbitrary nonlinear relationships. The BiLSTM-Huber variant further improved robustness to extreme wind events, reducing the impact of heavy-tailed outliers.

4.3 Climate-Stratified Insights

Performance varied systematically across climatic zones:

Mombasa: lowest absolute errors due to stable maritime climate.

Nairobi: intermediate accuracy, affected by episodic extreme winds.

Garissa: highest structural fit ($R^2 = 0.950$) despite larger errors, reflecting high variance in arid conditions.

This regime-dependent behaviour demonstrates that climate heterogeneity directly influences predictive accuracy. It supports the case for modular, zone-specific forecasting models rather than a single national model.

4.4 Residual Behaviour and Operational Implications

Residual analysis revealed weak cross-station correlations ($|r| \leq 0.133$). Residuals were temporally aligned, tested for stationarity, and confidence intervals confirmed insignificance. This indicates locally confined error structures, meaning predictive errors do not propagate across zones. Operationally, this supports independent station-level deployment pipelines that can be updated asynchronously and maintained with less infrastructure complexity. Such modularity is particularly important given that only ~30% of Kenyan weather stations provide consistent multi-year records.

4.5 Positioning Within the Literature

The baseline LSTM achieved $R^2 = 0.971$, consistent with performance envelopes reported in comparable African studies (27% improvement in Nigeria, 18% error reduction in Ghana). This study advances the literature by:

1. Providing climate-zone-stratified validation rather than aggregate national metrics.
2. Including cross-station residual correlation analysis to quantify spatial error propagation.
3. Explicitly demonstrating the distributional conditions under which ARIMA fails, offering a diagnostic framework for model suitability in heterogeneous climates.

4.6 Limitations

Several limitations should be noted:

The study is restricted to three stations; regions such as Lake Victoria, Turkana, and the Rift Valley were excluded.

Predictor variables were limited to temperature, humidity, and wind speed; precipitation, solar radiation, and soil moisture were not included.

The forecasting horizon was confined to one-day-ahead predictions; medium-range forecasting remains an open question.

Extreme-value tail behaviour may still challenge LSTM performance during unprecedented events outside the training distribution.

The dataset was derived from a commercial reanalysis product; replication using primary Kenya Meteorological Department records would strengthen operational validity.

5. Conclusion

This study provides the first climate-zone-stratified evaluation of deep learning architectures for short-term weather forecasting in Kenya, using 11,562 daily observations from Mombasa, Nairobi, and Garissa. The central finding is clear: ARIMA, a linear statistical model, produced forecasts inferior to a naïve mean predictor ($R^2 = -1.063$), while the baseline LSTM achieved $R^2 = 0.971$, representing a categorical improvement rather than an incremental gain. The superiority of LSTM is mechanistically grounded in Kenya's documented non-stationarity, heteroscedasticity, and non-Gaussian wind dynamics, which violate ARIMA's core assumptions.

Climate-stratified analysis revealed regime-dependent performance:

Mombasa: lowest absolute errors due to maritime stability.

Nairobi: intermediate accuracy, affected by episodic extreme winds.

Garissa: highest structural fit despite larger errors, reflecting high arid-zone variance.

Residual correlation analysis confirmed locally confined error structures ($|r| \leq 0.133$), supporting modular deployment of independent station-level models. Validation procedures, including early stopping, multiple runs, and bootstrap confidence intervals, confirmed the robustness of reported results.

For operational forecasting, these findings provide actionable evidence for transitioning from ARIMA to LSTM-based systems. A phased strategy is recommended: initial parallel deployment with existing models, followed by gradual replacement, and eventual full integration with periodic retraining. This modular approach reduces infrastructure complexity and enhances responsiveness to local climatic variability.

For the scientific community, the study establishes a reproducible framework for diagnosing model suitability before architecture selection, grounded in distributional analysis and climate-stratified validation. Future research should extend to additional climatic regions such as Lake Victoria and Turkana, incorporate variables like precipitation and solar radiation, and explore advanced architectures such as ConvLSTM and Transformers for medium-range forecasting.

Declarations

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflicts of interest.

Data Availability: The Visual Crossing meteorological dataset used in this study is available at https://drive.google.com/uc?id=1soY_8Lk2J5eLgFz374bzWc2BbqcqiNtt. Pre-processed data, model code, and training scripts are deposited at https://colab.research.google.com/drive/18LWgkE3x7wo-lzn-ACZzBqyKDoNwc-Ug?usp=sharing

Ethics Approval: This study uses publicly available observational meteorological data and did not involve human subjects. No ethics committee approval was required.

References

Adeyemi, O., Joyce, S., Cimarelli, G., & Doore, B. (2021). Dynamic neural network modelling of soil moisture content for predictive irrigation scheduling. Sensors, 21(23), 7955. https://doi.org/10.3390/s21237955

Al Mamun, A., Saha, T., & Hasan, M. (2024). Short-term rainfall prediction using SARIMA in East Africa. Journal of Hydrology: Regional Studies, 52, 101–118.

Alizadeh, M. J. (2020). Application of various machine-learning models for prediction of atmospheric conditions. In Advanced Climate Modelling (pp. 123–145). Elsevier.

Assaf, A. G., Tsionas, M., & Oh, H. (2022). The time has come to take Bayesian inference seriously. International Journal of Hospitality Management, 103, 103194.

Government of Kenya (2023). Kenya Vision 2030: Climate adaptation and resilience strategy. Nairobi: Government Printer.

Ham, Y. G., Kim, J. H., & Luo, J. J. (2019). Deep learning for multi-year ENSO forecasts. Nature, 573, 568–572. https://doi.org/10.1038/s41586-019-1559-7

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

IPCC (2023). Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report. Geneva: IPCC.

Kenya Red Cross (2019). Kenya floods 2018: Emergency response report. Nairobi: KRCS.

Kiprono, P., & Ndambuki, J. (2021). Challenges in weather forecasting in Kenya: Gaps in observational networks. East African Journal of Science, 15(1), 45–59.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE, 50(3), 885–900.

Muta

Mutai, C., Gitau, W., Njoka, J., Waweru, G., & Anyamba, A. (2021). Evaluation of numerical weather prediction models over the Greater Horn of Africa. Meteorological Applications, 28(3), e1993.

Nyamweya, C. (2023). Economic costs of climate variability in Kenya: An integrated assessment. African Journal of Agricultural and Resource Economics, 18(2), 112–128.

Ogallo, L., Omondi, P., & Okoola, R. (2020). Assessment of current operational weather forecasting in the Horn of Africa. Bulletin of the World Meteorological Organization, 69(1), 34–42.

Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., ... & Mohamed, S. (2021). Skilful precipitation nowcasting using deep generative models of radar. Nature, 597, 672–677. https://doi.org/10.1038/s41586-021-03854-z

Yeboah, N., Adjei, A., & Mensah, B. (2022). Improving rainfall forecasting accuracy in Ghana through LSTM with data augmentation. West African Journal of Applied Ecology, 30(1), 88–104.