American Journal of Water Resources, 2025, Vol. 13, No. 3, 69-76
Available online at http://pubs.sciepub.com/ajwr/13/3/1
Published by Science and Education Publishing
DOI:10.12691/ajwr-13-3-1

Monthly Water Level Forecasting of the Diani River Using a Hybrid ICEEMDAN-SE-ARIMA Model in Southern Guinea

Noukpo Médard Agbazo[1,*], Oumar Keita[1], Lonsenigbè Camara[1], Gideon Eustace Rwabona[2,3], Abdoulaye Sylla[1]

[1]Département d'Hydrologie, Université de N'Zérékoré, BP 50, N'Zérékoré, Guinée
[2]School of Computational and Communication Science and Engineering, Nelson Mandela African Institution of Science and Technology, Arusha, Tanzania
[3]Department of Mathematics and Statistics, Mbeya University of Science and Technology, Mbeya, Tanzania
*Corresponding author: agbmednou@gmail.com

Received June 12, 2025; Revised July 14, 2025; Accepted July 21, 2025

Abstract Flood control and water resources management are two critical tasks for hydrologists, and both heavily depend on accurate river water level forecasting. However, due to the intrinsic characteristics of water level series, it is difficult to achieve good forecasting accuracy. In Guinea, the forecasting of water level by physical models, and mathematical or data-driven models remains scarce. This study aims to implement for the first time in Guinea, the autoregressive integrated moving average (ARIMA) model and propose the improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) coupled with sample entropy (SE) and combined with ARIMA model namely as ICEEMDAN-SE-ARIMA to forecast Diani River monthly water level in southern guinea. The water level data of Diani hydrological station from 2000 to 2022 were used, in which the water data from 2000 to 2019 were used to build the models, and the data from 2020 to 2022 were used for validation. Seven statistical indices like Pearson's coefficient, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Symmetric Mean Absolute Percentage Error (sMAPE), Nash-Sutcliffe coefficient (NSCE), BIAS and Kolmogorov-Smirnov coefficient (DKS) are adopted to measure and compare the performance of the single ARIMA and ICEEMDAN-SE-ARIMA hybrid models. The results indicate that: (1) during the study period, six pseudo-periodic functions and one nonlinear trend contribute differently to Diani water level series forecasting, indicating their complexity; (2) Compared to the single ARIMA model, the Pearson's coefficient, DKS, BIAS, NSCE, RMSE, MAE and SMAPE of ICEEMDAN-SE-ARIMA were improved by 84.52%, 84.70%, 80%, 84.52%, 86%, 91%, 93%, and 80%, respectively; (3) ICEEMDAN-SE-ARIMA model outperformed the single ARIMA model. However, it seems that ICEEMDAN-SE-ARIMA model could be improved by combining ICEEMDAN-SE by other data-driven models. These findings are essential to enhance water resources management and flood mitigation in Guinea, mainly under climate change.

Keywords: *Guinea, Diani River, water level, forecasting, ICEEMDAN, SE, ARIMA, hybrid-model*

Cite This Article: Noukpo Médard Agbazo, Oumar Keita, Lonsenigbe Camara, Gideon Eustace Rwabona, and Abdoulaye Sylla, "Monthly Water Level Forecasting of the Diani River Using a Hybrid ICEEMDAN-SE-ARIMA Model in Southern Guinea." *American Journal of Water Resources*, vol. 13, no. 3 (2025): 69-76. doi: 10.12691/ajwr-13-3-1.

1. Introduction

Accurate forecasting of river water is essential for effective flood mitigation and for understanding the dynamics of river regime changes evolution [1,2,3]. However, due to the inherent randomness and complexity of hydrological variables such as water level fluctuations, this task remains highly challenging [4,5,6,7]. In the existing literature, hydrological forecasting models are generally classified in two groups: physical models, and statistical or data driven models [8]. Physical models are based on deterministic equations, and require extensive input a great deal of data and parameter calibration [9], which are often lacking or incomplete in many developing regions. This is particularly true in Guinea, where hydrological datasets are typically scare or incomplete. As a result, the application of physical models remains limited. In contrast, statistical models and data driven models require fewer parameters and input variables, making them more practical in data-scare contexts.

In recent years, hydrological forecasting has increasingly relied on artificial intelligence (AI) algorithms [3,10,11,12,13,14]. Despite their potential, these models often struggle to achieve high accuracy when applied directly to raw hydrological time series due to the inherent nonlinearity, and sensitivity of such data to climatic and

anthropogenic influences [15,5,16,17,18]. To overcome these limitations, several studies recommend hybrid modeling approaches that combine AI techniques with signal decomposition methods [16,17,19,20]. Among these, the Ensemble Empirical Mode Decomposition (EEMD) and its enhanced variants such as the Complete EEMD with Adaptive Noise (CEEMDAN) are widely used. These decomposition techniques allow complex and noisy time series to be broken down into simpler, more stationary components without assuming a predefined trend structure. They effectively address problems like mode mixing and noise contamination, enabling better modeling and interpretation of hydrological data [21,22,23].

In Guinea, accurate forecasting of hydrological variables such as river water levels has the potential to support not only flood management and water resource planning, but also gap-filing procedures in incomplete datasets. However, the implementation of both physical and statistical models remains limited. Physical modeling is particularly impractical due (1) the lack of long-term hydrological records, (2) frequent missing values in most time series, and the hydrological variables time series in most of the country's rivers, and (3) a sparse monitoring network across the country. Moreover, despite the growing global adoption of hybrid decomposition-based models, no such study has yet been conducted in Guinea.

In this context, combining an advanced decomposition method like ICEEMDAN with a statistical model such as Autoregressive Integrated Moving Average (ARIMA) appears promising. ARIMA performs well on stationary and linear series, while ICEEMDAN excels at decomposing complex, nonlinear, and nonstationary data into manageable components. Thus, a hybrid ICEEMDAN–ARIMA model could offer a robust solution for forecasting river water levels in Guinea. This study aims to develop and assess a hybrid ICEEMDAN–ARIMA model for forecasting the monthly water levels of the Diani River, one of the largest rivers in southern Guinea. The next section presents the dataset and methodology. This is followed by a discussion of the results, and finally, the study's conclusions are outlined.

2. Materials and Methods

2.1. Study Area and Data Description

Diani River (Figure 1) is one of most important rivers in Guinea. It is located in the forest region, precisely in Macenta's prefecture, sub-prefecture of N'zébéla. Among the rivers of southern Guinea, the Diani River is the only one equipped with a gauging station, located at the Diani Hydrological Station, approximately 4 km from N'Zébéla, along the national road connecting Macenta to N'Zérékoré. The river originates in a classified forest area near Vassérédou, along the Milo River, and serves as a natural boundary between Guinea and Liberia for nearly 50 km. Upon entering Liberian territory near Banié (Yomou Prefecture), it becomes known as the Saint Paul River. The Diani River watershed, one of the three main watersheds of Guinea Forest, covers a surface area of 9,333 km2 and exhibits an average slope of 3.31m/km. Its strategic location and hydrological characteristics make it a critical component for regional water monitoring and trans boundary water management.

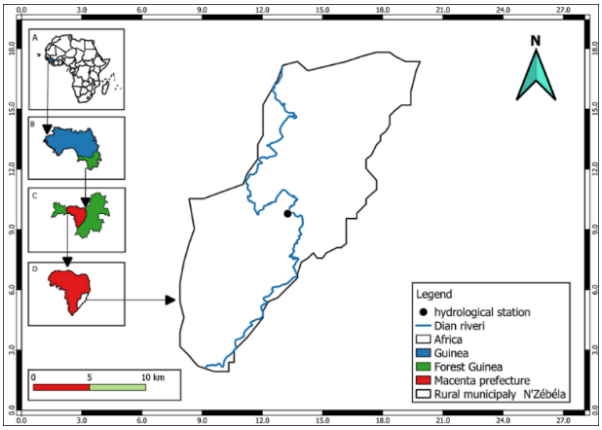

Figure 1. Geographic location (a) of Guinea within Africa, (b) of forested Guinea within Guinea, (c) Macenta within forested Guinea, (d) N'zébéla within Macenta, (e) Diani River within N'zébéla

From Figure 2, it can be seen that the water level time series from 2000 to 2020 at Diani hydrological station has certain inconsistent fluctuations, reflecting its strong variability.

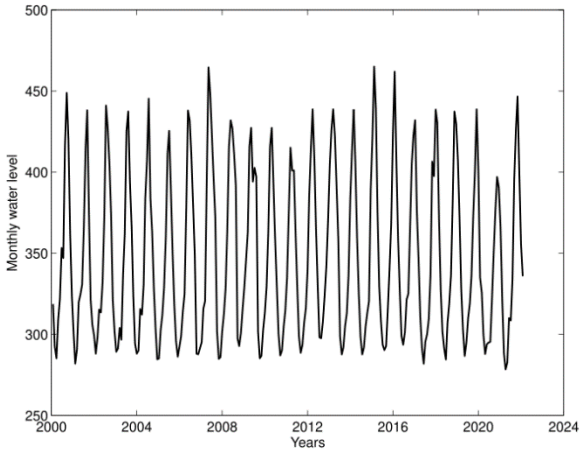

Figure 2. Temporal variation of monthly water levels measured at the Diani hydrological station from 2000-2022

2.2. Methods

2.2.1. Improved Complete EEMD with Adaptive Noise (ICEEMDAN) algorithm

Improved Complete Ensemble Empirical Mode Decomposition with adaptive noise (ICEEMDAN), which is an improved version of the Ensemble Empirical Mode Decomposition (EEMD), is adopted in this study to decompose complex water level series into several intrinsic mode functions (IMFs) and one nonlinear trend. Compared to the original water level series, IMF components are less nonstationary or stationary [24]. For a given observed series y(t), the specific steps of the ICEEMDAN algorithm are as follows [23]:

Step 1: Generate a new time series $x_i(t)$:

$$x_i(t) = y(t) + \beta_0 E_1(\eta_i(t)) \quad (1)$$

Where $\eta_i(t)$ is the ith zero-mean unit-variance white

Gaussian noise added; i is the number of times noise is added (i = 1, 2, …, m); β_0 ($\beta_0 > 0$) denotes the level of added white noise or the noise factor and the operator $E_k(\cdot)$ generates the kth EMD component of the white noise $\eta_i(t)$.

Step 2: Decompose fully $x_i(t)$ m times using the EMD algorithm to obtain the first residual and IMF of ICEEMDAN as follows:

$$R_1(t) = \frac{1}{m}\sum_{i=1}^{m} M\left(x_i(t)\right) \tag{2}$$

$$\widetilde{IMF}_1(t) = y(t) - R_1(t) \tag{3}$$

Where the operator M (·) represents the local mean of the envelope that satisfies the IMF sifting threshold same as EMD method

Step 3: Compute the second IMF and residual of ICEEMDAN as follows

$$\widetilde{IMF}_2(t) = R_1(t) - R_2(t) \tag{4}$$

$$R_2(t) = \frac{1}{m}\sum_{i=1}^{m} M\left(R_1(t) + \beta_1 E_2\left(\eta_i(t)\right)\right) \tag{5}$$

Step 4: Compute the kth IMF and residual of ICEEMDAN as follows:

$$\widetilde{IMF}_k(t) = R_{k-1}(t) - R_k(t) \tag{6}$$

$$R_k(t) = \frac{1}{m}\sum_{i=1}^{m} M\left(R_{k-1}(t) + \beta_{k-1} E_k\left(\eta_i(t)\right)\right) \tag{7}$$

Where $R_k(t)$ denotes the residual after the kth decomposition.

Step 5: Finally, go to Step 4 for the next k, until the stop criterion is reached (i.e., the residue does not have at least two extreme points, or its amplitude is less than the threshold).

Therefore, the observed series y(t) can be written as:

$$y(t) = \sum_{k}^{n} \widetilde{IMF}_k(t) + R(t) \tag{8}$$

R (t) represents the nonlinear trend of the observed series y(t). In this study, m=1000, and $\beta_k = 0.2$.

2.2.2. Autoregressive Integrated Moving Average (ARIMA)

ARIMA (p, d, q) model [25] consists of three parts: p-autoregressive (AR) model, I(d) differential order and q-moving average (MA) model. For a given time series X_t(t = 1, 2, 3,⋯, N), the structure of ARIMA (p, d, q) model is written as follows [26]:

$$\begin{cases} \Phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0, \mathrm{var}(\varepsilon_t) = \sigma^2, E(\varepsilon_t, \varepsilon_s) = 0, s \neq t \\ E(x_s, \varepsilon_t) = 0 \forall s < t \end{cases} \tag{9}$$

Where $\nabla^d = (1-B)^d$, B is the delay operator; $\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\Theta(B) = 1 - \theta_1 B - \cdots - \theta_p B^P$ are a polynomial of all auto-regression coefficients and of all moving average coefficients in the ARIMA model, respectively; $q \geq 0$, $p \geq 0$; ε_t represents white noise subject to $N\left(0, \sigma^2\right)$; $E\left(\varepsilon_t, \varepsilon_s\right)$ denotes the correlation between factors before and after the ε_t sequence.

2.2.3. Sample Entropy

Sample entropy (SE) algorithm developed by [27] is adopted to distinguish the complexity of each IMF and reconstruct IMF with similar entropy values. As in [28], the reconstruction of IMFs is done as follows: the IMFs with entropy > 0.35 are summed and named component1; the IMF with entropy between 0.005 and 0.008 are summed and named component2 and the IMF with entropy < 0.005 are summed and named component3. The nonlinear trend obtained from ICEEMDAN is not concerned by the reconstruction.

For a given time series X_t(t = 1, 2, 3,⋯, N), and a real number r > 0. In this study r is the standard deviation of the original series. The computation steps of SE are as follows:

Step 1: compute the dimension between the vector Xi and Xj is defined as:

$$\begin{aligned} &d\left[X_m(i), X_m(j)\right] = \\ &\max_{k=1,2,\cdots,m}\left(\left|X(i+k-1) - X(j+k-1)\right|\right) \end{aligned} \tag{10}$$

Step 2: Compute the average value $B^m(r)$

$$B^m(r) = \frac{\sum_{j=1}^{N-m}\left(\frac{\sum_{j\neq 1, j\neq i}^{N-m}\left(nbt\left(d\left[X_m(i), X_m(j)\right] \leq r\right)\right)}{N-m-1}\right)}{N-m} \tag{11}$$

Step 3: Compute the average value $A^m(r)$

$$A^m(r) = \frac{\sum_{j=1}^{N-m}\left(\frac{\sum_{j\neq 1, j\neq i}^{N-m}\left(nbt\left(d\left[X_{m+1}(i), X_{m+1}(j)\right] \leq r\right)\right)}{N-m-1}\right)}{N-m} \tag{12}$$

Where nbt represents the number of times that the condition is verified.

Step 4: Compute SE as follows

$$SE(m,r) = \lim_{N\to\infty}\left[-\ln\left(\frac{A^m(r)}{B^m(r)}\right)\right] \tag{13}$$

Where m indicates the embedding dimension, and r represents the conditional threshold.

2.2.4. Combined model based on ICEEMDAN-SE-ARIMA

Figure 3 shows the flowchart of the combined ICEEMDAN-SE-ARIMA forecasting model. In the

forecasting process, data of 2000-2019 period were used as training sets and those of 2020-2022 as test sets.

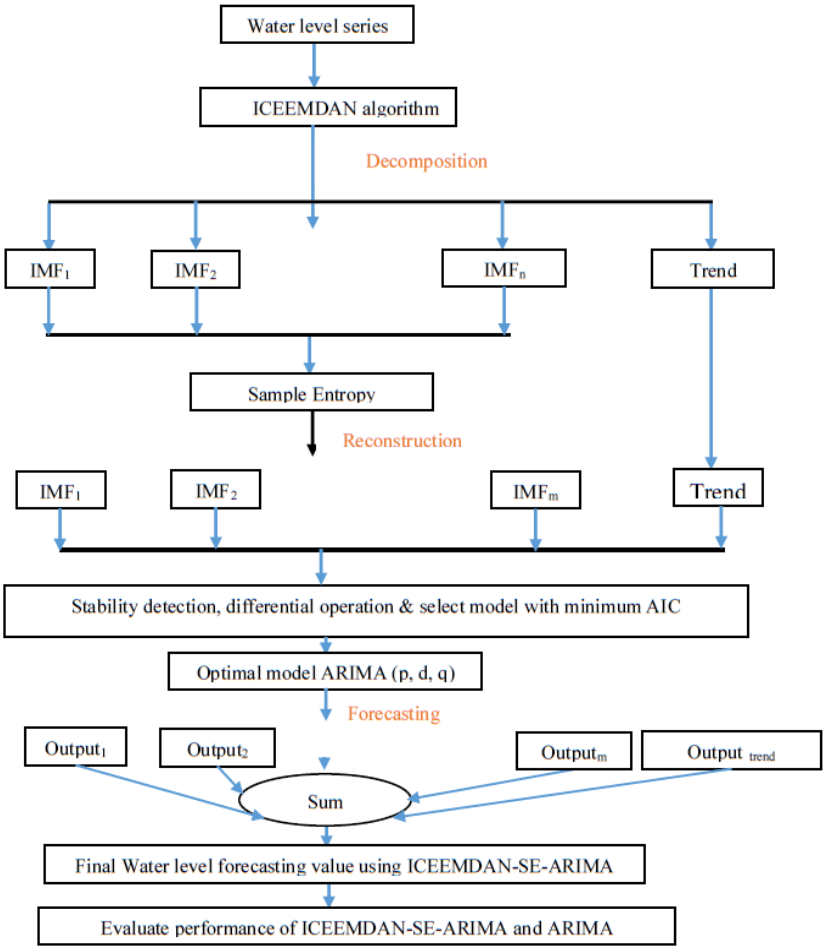

Figure 3. Flowchart of the combined ICEEMDAN-SE-ARIMA forecasting model and this study

2.2.5. Performance Evaluation Criteria

Seven statistical metrics: (a) Root Mean Square Error (RMSE), (b) Mean Absolute Error (MAE), (c) Symmetric Mean Absolute Percentage Error (sMAPE), (d) Nash–Sutcliffe coefficient (NSCE), (e) BIAS, (f) Pearson's coefficient and (g) the Kolmogorov-Smirnov coefficient (DKS) are used to measure and compare the forecasting accuracy. The range of RMSE, MAE and SMAPE is (0, +∞) and the smaller the RMSE, MAE, SMAPE value is, the higher the accuracy of the forecasting model. The perfect value of RMSE, MAE and SMAPE is 0, suggesting that the forecasting model is perfect [29,30,31]. If the value of SMAPE is greater than 100%, this means that the forecasting model is very poor. The range of NSCE is (-∞, 1) and the perfect value of NSCE is 1, suggesting perfect agreement between the forecasted results and the observed data [4]. The range of BIAS is (-∞, +∞) and its perfect value is 0. A positive value suggests an overall high-water level, while a negative value suggests a low water level [4]. DKS evaluates similarity between the distribution of forecasted and observed water level. The DKS varies from 0 (for identical distributions) to 1 (for distributions that do not even overlap) [32].

RMSE, MAE, SMAPE, NSCE and BIAS formulas are as follows:

$$RMSE = \left(\frac{1}{N} \sum_{i=1}^{n} \left(W_{\mathrm{mod}}(i) - W_{obs}(i) \right)^2 \right)^{1/2} \quad (14)$$

$$MAE = \frac{1}{N} \sum_{i=1}^{n} \left| W_{\mathrm{mod}}(i) - W_{obs}(i) \right| \quad (15)$$

$$SMAPE = \frac{1}{N} \sum_{i=1}^{n} 200 * \left(\frac{\left| W_{obs}(i) - W_{\mathrm{mod}}(i) \right|}{W_{obs}(i) + W_{\mathrm{mod}}(i)} \right) \quad (16)$$

$$NSCE = 1 - \frac{\sum_{i=1}^{n} \left(W_{obs}(i) - W_{\mathrm{mod}}(i) \right)^2}{\sum_{i=1}^{n} \left(W_{obs}(i) - \bar{W}_{obs}(i) \right)^2} \quad (17)$$

$$BIAS = \frac{\left(W_{\mathrm{mod}}(i) - W_{obs}(i) \right)}{\sum_{i=1}^{n} W_{obs}(i)} \quad (18)$$

Where W_{mod} and W_{obs} are respectively are the forecasted and observed of water level at time i, respectively; N is the length.

DKS is computed as follows [32]:

$$DKS = \sup_x \left| F_a(x) - F_b(x) \right| \quad (19)$$

Where $F_a(x)$ and $F_b(x)$ are cumulative distribution functions of the sample A et B, respectively, while, $\sup_x$

stands for supremum of the set of distances, which means the largest value over all possible values of x.

3. Results

As illustrated in Figure 4, six intrinsic mode functions (IMFs) and one nonlinear trend characterized the multi-scale variability of water level series measured during 2000-2022 period at Diani. Thus, six pseudo-periodic functions and one nonlinear trend contribute differently to Diani water level series forecasting. Sample entropy values of each IMF (not shown) indicate that the IMFs can be grouped in three components: component$_1$ (IMF_1+IMF_2), component$_2$ (IMF_3+IMF_4) and component$_3$ (IMF_5+IMF_6).

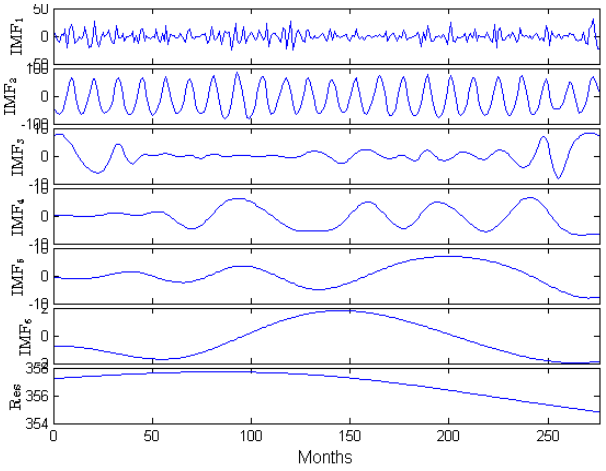

Figure 4. ICEEMDAN decomposition of monthly water level from 2000 to 2022 at Diani hydrological station

Figure 5 presents the forecasting results when the single ARIMA model is used. The forecasting results is compared with the test sets (Figure 5b). It can be clearly observed from Figure 5b that qualitatively, ARIMA model forecasts well the shape and trend of test sets. Thus, the forecasted data with the single ARIMA model seem to be in sync with the test sets. However, quantitatively ARIMA model either underestimates or overestimates the test sets.

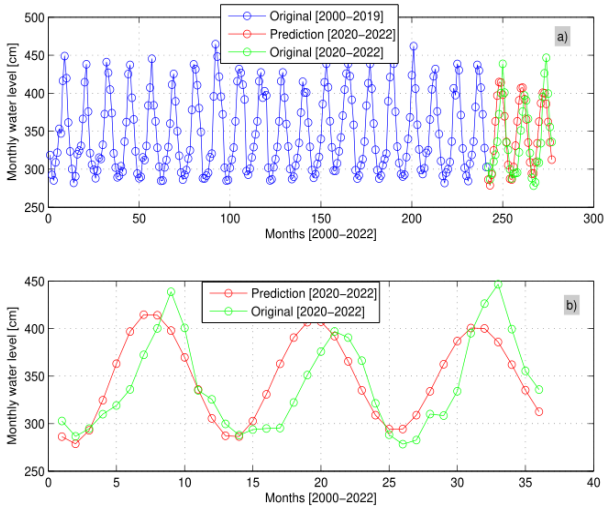

Figure 5. (a) Forecasting results of the single ARIMA model. (b) Comparison of test sets and forecasted values

Comparison of component1(IMF_1+IMF_2), component2 (IMF_3+IMF_4) and component3 (IMF_5+IMF_6) with their forecasted values by using ICEEEMDAN-SE-ARIMA model is depicted in Figures 6, 7, 8 and 9 respectively. In these figures the red line shows the forecasted records, while the green line is the original records. From these figures, it can be noted that the forecasted time series of component1, component2, component3 and nonlinear trend are in sync and in high agreement (except for component1) with their corresponding original time series. Moreover, the red line is generally close to the green line, suggesting that the forecasting of component1, component2, component3 and the nonlinear trend by ICEEMDAN-SE-ARIMA accord with the observation. Thus, it seems that the ICEEMDAN-SE-ARIMA model forecasts the components and the intrinsic nonlinear trend of water level at Diani hydrological station with small forecasting errors.

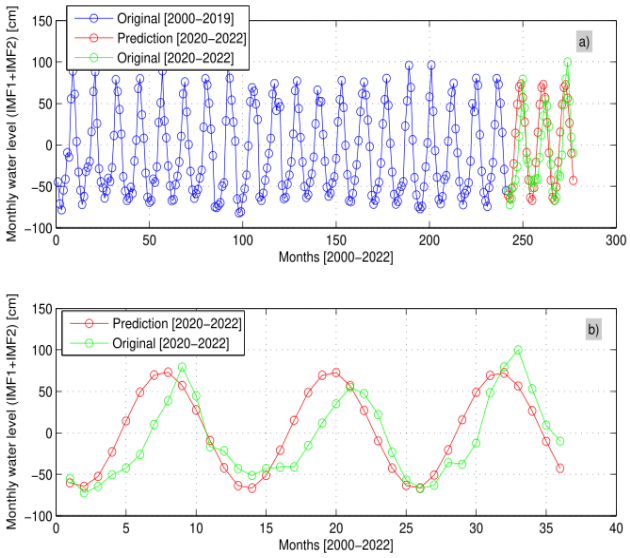

Figure 6. (a) Forecasting results of ICEEMDAN-SE-ARIMA model. (b) Comparison of component1 (IMF1+IMF2) with its forecasted values using ICEEMDAN-SE-ARIMA model

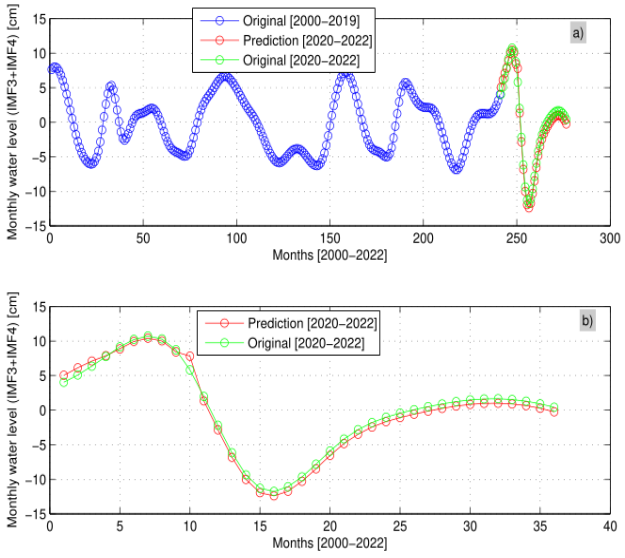

Figure 7. (a) Forecasting results of ICEEMDAN-SE-ARIMA model. (b) Comparison of component2 (IMF3+IMF4) with the forecasted values using ICEEMDAN-SE-ARIMA model

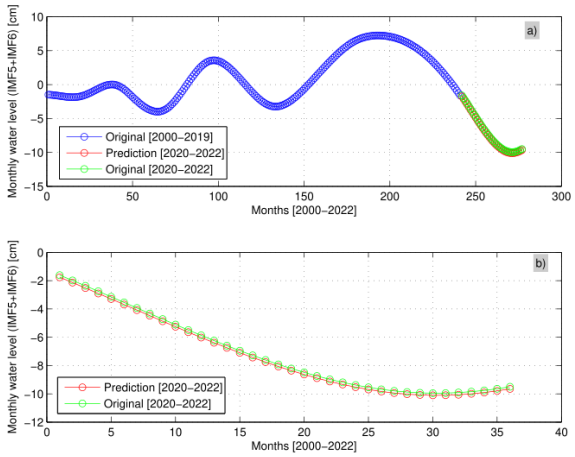

Figure 8. (a) Forecasting results of ICEEMDAN-SE-ARIMA model. (b) Comparison of component3 (IMF5+IMF6) with its forecasted values using ICEEMDAN-SE-ARIMA model

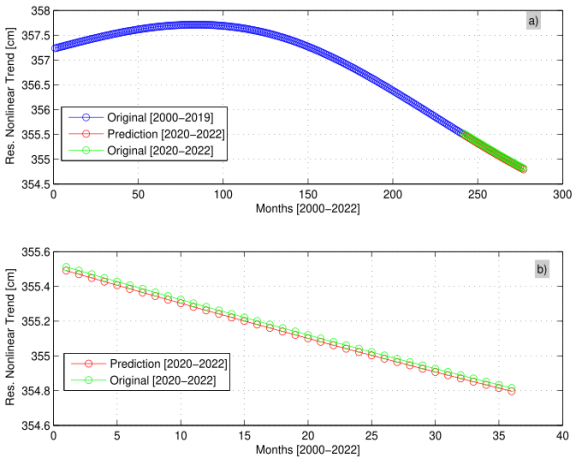

Figure 9. (a) Forecasting results of ICEEMDAN-SE-ARIMA model. (b) Comparison of nonlinear trend of the observed series with its forecasted values using ICEEMDAN-SE-ARIMA model

Figure 10 displays the boxplot of the forecasting error obtained from ICEEMDAN-SE-ARIMA model and single ARIMA model. It can be seen from this figure that the single ARIMA model has a large forecasting error compared to ICEEMDAN-SE-ARIMA model. Thus, the forecasting error of the single ARIMA model is larger than ICEEMDAN-SE-ARIMA model ones.

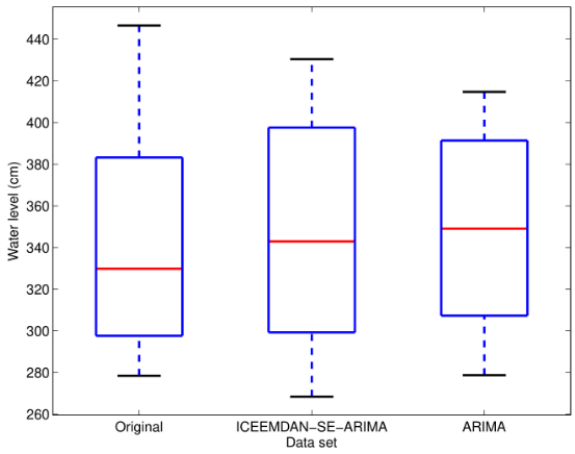

Figure 10. Box plot of the forecasting error obtained with ICEEMDAN-SE-ARIMA model and the single ARIMA model

Table 1 presents the Pearson coefficient and the associated lower and upper bound of 95% confidence interval of the monthly water level forecasting by ICEEMDAN-SE-ARIMA model and the single ARIMA model. It can be clearly observed from Table 1 that the Pearson coefficient of the both models vary from 0.70 to 0.85 with positive lower and upper bound of 95% confidence interval, indicating that the correlation between observed and forecasted data are statistically significant at the 95% confidence level, regardless of the models. These findings suggest that ICEEMDAN-SE-ARIMA and the single ARIMA model present a good capacity to replicate Diani River water level patterns. However, ICEEMDAN-SE-ARIMA model presents the highest and strong Pearson's coefficient suggesting that ICEEMDAN-SE-ARIMA model explains much better a greater percentage of the variability in the observed water levels than the single ARIMA and presents a better fit between forecasting and observation. Thus, it seems that the forecasting performance of ICEEMDAN-SE-ARIMA model is better than the single ARIMA ones.

Table 1. Pearson correlation

Model	Pearson coefficient	Lower bound of 95% confidence interval	Upper bound of 95% confidence interval
ICEEMDAN-SE-ARIMA	0.84	0.62	0.89
ARIMA	0.71	0.54	0.85

Table 2 summarizes the evaluation indexes for forecasting performance of the monthly water level forecasting by ICEEMDAN-SE-ARIMA model and the single ARIMA model.

Table 2. Evaluation Indexes for Forecasting Performance

Model	RMSE	MAE	SMAPE	NSCE	BIAS	DKS
ICEEMDAN-SE-ARIMA	31.24	22.50	7.33	0.57	0.016	0.11
ARIMA	34.08	27.83	7.89	0.49	0.020	0.17

It can be seen from Table 2:

(1). ICEEMDAN-SE-ARIMA model has lower DKS value than the single ARIMA model. Thus, based on the fact DKS evaluates similarity between the distribution of forecasted and observed water level, it is clear that compared to the single ARIMA model, the distribution of forecasted by ICEEMDAN-SE-ARIMA is more identical to observed ones. Therefore, ICEEMDAN-SE-ARIMA model represents better the distribution of water level at Diani than the single ARIMA model. However, the DKS value of ICEEMDAN-SE-ARIMA (0.11) is not very close to zero, suggesting that this model could be more improved.

(2). The BIAS values for ICEEMDAN-SE-ARIMA and the single ARIMA models are 0.016 and 0.020, respectively. Thus, positive BIAS values are obtained for the both models, suggesting an overall high water level compared to the observed data. However, the BIAS value of the ICEEMDAN-SE-ARIMA model is below 0.2 and reduced by 80% compared to the single ARIMA model. Therefore,

ICEEMDAN-SE-ARIMA model achieves the highest accuracy in terms of BIAS.

(3). The NSCE values for the single ARIMA model are positive and below 0.5, while it is positive and above 0.55 for the ICEEMDAN-SE-ARIMA model. Furthermore, compared to the standalone ARIMA model, the NSE is improved by 86%. These results suggest good agreement between the forecasted results and the observed data for ICEEMDAN-SE-ARIMA model compared to the single ARIMA model. Thus, the ICEEMDAN-SE-ARIMA model performs best in terms of NSCE.

(4). Compared with the single ARIMA model, RMSE, MAE and SMAPE of ICEEMDAN-SE-ARIMA model are lower than that ARIMA model. Moreover, compared to the standalone ARIMA model, the RMSE, MAE and SMAPE of ICEEMDAN-SE-ARIMA model is reduced by 91%, 93% and 80% respectively. These findings suggest that ICEEMDAN-SE-ARIMA model has lower prediction error metrics and the highest prediction accuracy compared to the single ARIMA model.

4. Discussion

Overall, the findings demonstrate that the ICEEMDAN-SE-ARIMA model consistently outperforms the ARIMA model in water level forecasting. However, it seems that replacing ARIMA by others artificial intelligence (AI) algorithms or data-driven model could improve the forecasting. Thus, our findings demonstrate and confirm that coupling artificial intelligence (AI) algorithms or data-driven models with data decomposition techniques enhances the accuracy of water level forecasting compared to using AI algorithm or data-driven model alone. Although our findings could not be compared to others studies, as research on forecasting of hydrological parameters remains scarce in Guinea and has received limited attention in West African. However, the results may be qualitatively compared with previous studies conducted in other parts of the world. Our findings related to the impact of data decomposition technique on the improvement of forecasting are consistent with the results reported on hydrological variables by [16,17,10,33,34] [35,36,37,38,39,30,15].

5. Conclusions

This study aims to implement the autoregressive integrated moving average (ARIMA) model and to develop and assess a hybrid ICEEMDAN-SE-ARIMA model, which combines the improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) coupled with sample entropy (SE) and ARIMA to forecast the monthly water levels of the Diani River, one of the largest rivers in southern Guinea. The monthly historical water level data of Diani hydrological station from 2000 to 2022 were used, in which the water level data from 2000-2019 period were used to build the models, and the data from 2020 to 2022 were used as validation set. Seven statistical indices like Pearson's coefficient; Root Mean Square Error (RMSE); Mean Absolute Error (MAE); Symmetric Mean Absolute Percentage Error (sMAPE); Nash-Sutcliffe coefficient (NSCE); BIAS and Kolmogorov-Smirnov coefficient (DKS) are adopted to measure and compare the performance of the single ARIMA and ICEEMDAN-SE-ARIMA hybrid models. The following conclusions are drawn.

(1). Six pseudo-periodic functions and one nonlinear trend characterize Diani water level series variation. This indicates that the studied water level series show strong nonstationary and nonlinear characteristics, which contribute differently to Diani water level series forecasting;

(2). Compared to the single ARIMA model the:

(i). Pearson's coefficient of ICEEMDAN-SE-ARIMA were improved by 84.52%, indicating its superiority in identifying and recreating the dynamics and patterns of validation data.

(ii). DKS of ICEEMDAN-SE-ARIMA were improved by 84.70%, indicating its superiority in reducing the largest difference between the empirical cumulative density function of the forecasted and observed data.

(iii). NSCE of ICEEMDAN-SE-ARIMA were improved by 84.52%, suggesting its better performance.

(iv). RMSE, MAE and SMAPE of ICEEMDAN-SE-ARIMA were improved by 91%, 93%, and 80%, respectively, indicating its high forecasting precision.

(3). DKS, BIAS, NSCE, RMSE, MAE and SMAPE (Pearson's coefficient) of ICEEMDAN-SE-ARIMA model are smaller (greater) than those of the single ARIMA model. Therefore, ICEEMDAN-SE-ARIMA gives a better performance and high forecasting precision than the single ARIMA model. Thus, the decomposition of the original water level series enhances their forecasting at Diani station in southern Guinea. However, the ICEEMDAN-SE-ARIMA model proposed in this study need to be improved because is BIAIS is positive. Our future research will focused on the improving of this model.

ACKNOWLEDGEMENTS

We acknowledge the Guinean National Agency for Hydraulic for providing the dataset used in this study. We also thank the editor and anonymous reviewers for their constructive comments.

Funding

No funding received for doing this study

Ethical Compliance

Compliance with ethical standards

Conflicts of Interest

The authors declare no conflict of interest

References

[1] Hu, Q. F.; Cao, S. Y.; Yang, H. B.; Wang, Y. T.; Li, L. J.; Wang, L. H. Daily runoff predication using LSTM at the Ankang Station, Hanjing River. Progress in Geography 2020, 39 (04), 636–642.

[2] Yu, H.; Yang, Q. Applying Machine Learning Methods to Improve Rainfall–Runoff Modeling in Subtropical River Basins. Water 2024, 16, 2199.

[3] Zhang, X.; Wang, R.; Wang, W.; Zheng, Q.; Ma, R.; Tang, R.; Wang, Y. Runoff prediction using combined machine learning models and signal decomposition. Journal of Water and Climate Change 2025, 16 (1): 230–247.

[4] Peng, T.; Zhou, J.; Zhang, C.; Fu,W. Streamflow forecasting using empirical wavelet transform and artificial neural networks, Water 2017, 9 (6), 406–406.

[5] Ali, M.; Khan, A.; Rehman, N.U. Hybrid multiscale wind speed forecasting based on variational mode decomposition. Int Trans Electr Energy Sys 2018, 28, 2466.

[6] Jiao, X.; He, Z. A novel coupled rainfall prediction model based on stepwise decomposition technique. Sci Rep 2024, 14 10853.

[7] Ling, M.; Xiao, L.Y.; Zhao, J.; Wang, P.; Wang, Y.; Xiang, K. Daily precipitation prediction based on SVM-CEEMDAN-BiLSTM Mode. Pearl River 2023, 44(09) 61-68.

[8] Zhang, X.; Zhang, Q.; Zhang, G.; Nie, Z.; Gui, Z. A Hybrid Model for Annual Runoff Time Series Forecasting Using Elman Neural Network with Ensemble Empirical Mode Decomposition. Water 2018, 10, 416.

[9] Devia, G.K.; Ganasri, B.P.; Dwarakish, G.S. A review on hydrological models. Aquat. Procedia 2015, 4, 1001–1007. [CrossRef].

[10] Ali, M.; Prasad, R.; Xiang, Y.; Yaseen, Z.M. Complete ensemble empirical mode decomposition hybridized with random forest and kernel ridge regression model for monthly rainfall forecasts. Journal of Hydrology 2020, 584, 124647, ISSN 0022-1694.

[11] Wani, O.A.; Mahdi, S.S.; Yeasin, M.; Kumar, S.S.; Gagnon, A.S.; Danish, F.; Al-Ansari, N.; El-Hendawy, S.; Mattar, M.A. Predicting rainfall using machine learning, deep learning, and time series models across an altitudinal gradient in the North-Western Himalayas. Sci Rep. 2024, 14(1): 27876.

[12] Wang, C.; Jia, Z.Y.; Yin, Z.H.; Liu, F.; Lu, G.P. Improving the accuracy of subseasonal forecasting of China precipitation with a machine learning approach. Front. Earth Sci 2021.

[13] Huang, C.; Li, Q.P.; Xie, Y.J.; Peng, J. Prediction of summer precipitation in Hunan based on machine learning. Trans. Atmos. Sci 2022, 45(2) 191-202.

[14] Kumar, V.; Kedam, N.; Sharma, K.V.; Mehta, D.J.; Caloiero, T. Advanced Machine Learning Techniques to Improve Hydrological Prediction: A Comparative Analysis of Streamflow Prediction Models. Water 2023, 15, 2572.

[15] Zhang, X.; Cheng, W.; Zhang, Y.; Zhua, J.; Rena, H. A combined model based on secondary decomposition and the optimized support vector machine algorithm for regional rainfall forecasting. Journal of Water and Climate Change 2025, 16 (2) 474.

[16] Adarsh, S.; Reddy, M.J. Multiscale characterization and prediction of monsoon rainfall in India using Hilbert–Huang transform and time-dependent intrinsic correlation analysis. Meteorol Atmos Phys 2018, 130 667-688.

[17] Adarsh, S.; Janga Reddy, M. Multi-scale Spectral Analysis in Hydrology: From Theory to Practice (1st ed.) CRC Press 2021.

[18] Poongadan, S.; Lineesh, M.C. Non-linear Time Series Prediction using Improved CEEMDAN, SVD and LSTM. Neural Process Lett 2024, 56 164.

[19] Khan, M.M.H.; Muhammad, N.S.; El-Shafie, A. Wavelet based hybrid ANN-ARIMA models for meteorological drought forecasting. J. Hydrol. 2020, 590 125380.

[20] Danandeh, M.A. Seasonal rainfall hindcasting using ensemble multi-stage genetic programming. Theor. Appl. Climatol 2021, 143 1 461-472.

[21] Wu, Z.; Huang, N.E. Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method. Adv Adapt Data Anal 2009, 1 1-41.

[22] Torres, M.E.; Colominas, M.A.; Schlotthauer, G.; Flandrin, P. A complete ensemble empirical mode decomposition with adaptive noise. In 2011 IEEE International Conference on Acoustics, Speech, and Signal Processing 2011, 4144-4147.

[23] Colominas, M.A.; Schlotthauer, G.;Torres, M. Improved complete ensemble EMD: A suitable tool for biomedical signal processing. Biomed. Signal Process Control 2014, 14 19-29. [CrossRef].

[24] Zhang, X.; Zhao, D.; Wang, T.; Wu, X.; Duan, B. A novel rainfall prediction model based on CEEMDAN-PSO-ELM coupled model. Water Supply 2022, 22, 4, 4531.

[25] Box, G. E. P.; Jenkins, G. M. Some comments on a paper by Chatfield and Prothero and on a review by Kendall. Journal of the Royal Statistical Society Series A (General) 1973, 136 (3), 337–352.

[26] Liu, X; Zhang, Y; Zhang, Q; Comparison of EEMD-ARIMA, EEMD-BP and EEMD-SVM algorithms for predicting the hourly urban water consumption. Journal of Hydroinformatics 2022, 24 (3): 535–558.

[27] Richman, J.S.; Lake, D.E.; Moorman, J.R. Sample Entropy. Methods Enzymol. 2004, 384, 172–184.

[28] Hu, T.; Zhou, M.; Bian, K.; Lai, W.; Zhu, Z. Short-Term Load Probabilistic Forecasting Based on Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise Reconstruction and Salp Swarm Algorithm. Energies 2022, 15, 147.

[29] Jierula, A.;Wang, S.; OH, T.-M.;Wang, P. Study on Accuracy Metrics for Evaluating the Predictions of Damage Locations in Deep Piles Using Artificial Neural Networks with Acoustic Emission Data. Appl. Sci. 2021, 11, 2314.

[30] Wei, X.; Chen, M.; Zhou, Y. et al. Research on optimal selection of runoff prediction models based on coupled machine learning methods. Sci Rep 2024, 14, 32008.

[31] Li, C.; Zhu, L.; He, Z.; Gao, H.; Yang, Y.; Yao, D.; Qu, X. Runoff Prediction Method Based on Adaptive Elman Neural Network. Water 2019, 11(6), 1113.

[32] Wilks, D.S. Statistical methods in the atmospheric sciences. In: International Geophysics Series 59 2nd edition. Burlington MA: Elsevier Academic Press 2006, 627.

[33] Johny, K.; Pai, M.L.; Adarsh, S. Adaptive EEMD-ANN hybrid model for Indian summer monsoon rainfall forecasting. Theor Appl Climatol 2020, 141 1-17.

[34] Tan, Q.F.; Lei, X.H.; Wang, X.; Wang, H.; Wen, X.; Ji Y.; Kang A.Q. An adaptive middle and long-term runoff forecast model using EEMD-ANN hybrid approach. J Hydrol, 2018, 567 767-780.

[35] Yang, Q.; Qin, L.; Gao, P.; Zhang, R. BAnnual precipitation prediction in the economic zone of the northern slope of Tianshan Mountain based on EEMD-LSTM model. Arid Zone Research 2021, 38 (05), 1235–1243.

[36] Wang, W.C.; Chau, K. W.; Xu, D. M.; Chen, X. Y. Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. Water Resources Management 2015, 29, 2655–2675.

[37] Wang, Z.-Y.; Qiu, J.; Li, F.-F. Hybrid Models Combining EMD/EEMD and ARIMA for Long-Term Streamflow Forecasting. Water 2018, 10, 853.

[38] Rezaiy, R.;Shabri, A. Enhancing drought prediction precision with EEMD-ARIMA modeling based on standardized precipitation index. Water Sci Technol. 2024, 89(3): 745-770. (2024) 89 (3): 745–770.

[39] Liao S.; Wang H.; Liu B.; Ma X.; Zhou B.; Su H. Runoff Forecast Model Based on an EEMD-ANN and Meteorological Factors Using a Multicore Parallel Algorithm," Water Resources Management: An International Journal, Published for the European Water Resources Association (EWRA), Springer; European Water Resources Association (EWRA) 2023, 37(4), 1539-1555.