

Deep Learning-Based Sea Level Forecasting Using Informer in Cilacap, Indonesia

Reva Rivandi Salim¹, Didit Adytia^{2*}

¹*School of Computing, Telkom University, Bandung, Indonesia*

²*Center of Excellence of Sustainable Energy and Climate Change (SECC), Telkom University, Bandung, Indonesia*

*corr-author: adytia@telkomuniversity.ac.id

Abstract - Sea level forecasting is very important for coastal risk management and operational planning, especially in regions vulnerable to frequent tidal flooding events. Tidal Harmonic Analysis (THA) and other traditional methods can effectively reconstruct tidal components but typically overlook non-tidal influences such as meteorological variability and ocean swell. This study mitigates these limitations by proposing the Informer model, a Transformer-based deep learning architecture for long-range sequence forecasting, to predict sea levels using 11 months of hourly observational data (December 2023 – October 2024) from Cilacap, a tropical coastal region in Indonesia. A new preprocessing pipeline is introduced, integrating THA-based tidal reconstruction with interpolation techniques to handle missing data. Forecasting performance is evaluated across multiple prediction horizons (1, 3, 5, 7, and 14 days) and compared against XGBoost, LSTM, and the standard Transformer. The results show that Informer does better than the other models, especially over longer horizons. It has the lowest RMSE (0.091), the lowest MAPE (2.14%), and the highest correlation coefficient (0.98) on the 14-day forecast. In this study, we focused on the Informer's capability for long horizon from sea level data for providing a reliable solution for sea level prediction. This results show that the model is applicable for integration into early warning systems.

Keywords: sea level forecasting, informer, deep learning, time series prediction

I. INTRODUCTION

Reliable sea level forecasting is crucial for coastal planning, disaster mitigation, and adaptation to climate change [1]. Rising sea levels are endangering coastal areas worldwide, causing phenomena such as tidal flooding, shoreline erosion, saltwater intrusion, and infrastructure destruction [2-3]. These phenomena, driven by complex interactions among tides, ocean currents, atmospheric pressure, and large-scale climate oscillations, pose significant risks to populations and ecosystems in low-lying coastal zones [4]. A reliable predictive model can help most people prevent or warn

them for early warning systems, design resilient infrastructure, and manage coastal resources more effectively [5].

In practice, acquiring consistent and high-quality sea level observations remains problematic, especially in developing regions such as Indonesia. Sensor malfunctions, transmission errors, and logistical limitations frequently introduce gaps and inaccuracies into observational datasets. To fill this missing gap, Tidal Harmonic Analysis (THA) has traditionally been used to reconstruct and interpolate missing portions of sea level data. Although THA effectively captures tidal periodicity, this method is inherently limited in representing non-tidal influences such as meteorological events, ocean waves, and seasonal climate variability [6-7].

Many significant breakthroughs in machine learning and deep learning have come out in recent years. These breakthroughs have led to some promising new ways to estimate sea level. It is hard for traditional statistical methods like ARIMA and Kalman filters to model nonlinear patterns and long-term relationships that are naturally present in oceanographic data [1,8]. Consequently, approaches based on Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and hybrid models combining Convolutional Neural Networks (CNN) and LSTM have emerged, demonstrating better short-term prediction accuracy compared to traditional models [9,-10]. However, even though there are many issues such as gradient vanishing, error accumulation, and computing inefficiency, these issues usually arise when using long-term predictions [11].

Recent research in Southeast Asia and tropical coastal regions has tested various advanced models, such as Temporal Convolutional Networks (TCN) and combinations with attention mechanism, to improve prediction accuracy. For example, Ref. [12] applied TCN in 14-day water level forecasting in Surabaya, Indonesia. The model demonstrated consistent performance, although it required a large amount of multivariate data.

On the other hand, Puspita Sari and Adytia [13] used a LSTM with an attention mechanism to predict sea level height in Jakarta. The results were moderate but limited by the short length of the input data and the simplification of the tidal system. Although offering positive prospects, these studies still rely on short-term memory models, are highly location-specific, and have not been extensively tested under complex oceanographic conditions, such as those caused by the Indian Ocean Dipole (IOD) or monsoon variability. Additionally, there are few studies that systematically examining long-term predictions using univariate data in dynamic and exposed coastal area. These conditions highlight the importance of utilizing more scalable architectures, such as Informer, which is designed to capture long-term temporal relationships while functioning reliably even when observational data are sparse, noisy, and univariate [14].

Recent research investigates Transformer-based deep learning architectures that utilize a self-attention method to address restrictions by effectively capturing long-range temporal relationships. The Informer model stands out by introducing ProbSparse self-attention and generative decoders designed specifically to handle large time-series data efficiently [14]. ProbSparse computes attention only for the most 'informative' queries, cutting complexity from $O(L^2)$ toward $O(L \log L)$ while preserving long-range dependencies. The generative decoder then produces the entire forecast horizon in one pass (non-autoregressive), speeding up inference and reducing error accumulation in multi-step prediction. Despite Informer's demonstrated success in forecasting applications, its performance and applicability for sea level prediction, especially in complex tropical coastal environments with intense climatic variability, remain underexplored. The architecture allows the model to selectively focus on informative past values, making it particularly effective for forecasting complex geophysical signals such as sea level, where both periodic and non-periodic patterns must be accurately modeled [15].

This study closes the research gap by carrying out a thorough comparative analysis of the Informer model and well-established forecasting methods XGBoost, LSTM, and the standard Transformer model, using hourly sea level data from Cilacap in Indonesia. Cilacap was chosen because Cilacap is frequently affected by disasters such as coastal flooding events and exposure to dynamic oceanographic conditions influenced by the Indian Ocean Dipole (IOD), monsoonal patterns, and regional swell activities. To our knowledge, this study is the first application of the Informer model for sea level

forecasting in Cilacap, Indonesia. Although previous study has explored the Informer model in other coastal environments, none have evaluated its performance in dynamically exposed tropical region such as Cilacap using hourly sea level data over a long forecasting horizon.

The primary contributions of this research include (1) applying and evaluating the Informer model; (2) providing a robust comparative analysis between Informer and several state-of-the-art baseline methods; and (3) proposing an effective preprocessing approach combining THA with interpolation to handle missing observational data robustly. Results from this study offer critical insights for developing effective coastal flood forecasting systems and inform practical decision-making for coastal hazard mitigation and adaptive management in tropical regions.

II. METHOD

The objective of this study is to utilize the Informer architecture combine with and Tidal Harmonic Analysis (THA) for data pre-processing. The study area is located in Cilacap. Cilacap is a coastal city in Central Java, Indonesia that is frequently affected by tidal flooding (*banjir rob*) and seasonal variations.

The research process begins with the collection of sea level data from Intergovernmental Oceanographic Commission (ICO). The actual data from IOC was one-minute intervals, the data is then subjected to a pre-processing step, which includes resampling minute intervals to hourly intervals and handling missing values with THA. THA is subsequently applied to reconstruct tidal signals, separating periodic tidal components from residual influences, and filling data gaps.

After pre-processing, the forecasting model is constructed using Informer architecture, which is especially well-suited for capturing and studying long-term dependencies in time series data [16]. Besides that, we used three baseline models, the models are: XGBoost, LSTM, and standard Transformer. This model was trained and tested on a dataset that had been processed using THA and interpolation beforehand, and its performance in forecasting was evaluated using three evaluation metrics: Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Correlation Coefficient (CC) [17-18]. These metrics are used to measure the accuracy of predictions and the robustness of the linear correlation between actual and predicted sea level values.

To compare the performance of the Informer model, a comparison was made using baseline models. These models were selected based on their frequent use in time

series forecasting tasks and serve as benchmarks for evaluating the added value of the Informer architecture [19]. This forecasting task in all these models were performed using multiple prediction horizons, specifically 1 day, 3 days, 5 days, 7 days, and 14 days ahead, to assess the model’s ability to generalize across short to long-range temporal scales. The complete methodology is illustrated in Fig. 1, which outlines the sequence from data collection and preprocessing to model training, validation, comparison, and evaluation.

Each of the following subsections describes the components of this methodology in greater detail, beginning with the source and characteristics of the sea level dataset.

A. Sea Level Data

Hourly sea-level observations utilized in this study were collected from a buoy station maintained by the Intergovernmental Oceanographic Commission (IOC) near Cilacap, Indonesia (latitude -7.75° , longitude 109.00°). Cilacap is strategically chosen due to its vulnerability to tidal flooding events known locally as *banjir rob* and complex oceanographic influences from the Indian Ocean Dipole (IOD), seasonal monsoon winds, and regional wave-swell dynamics [20].

The dataset spans from December 11, 2023, to October 28, 2024, providing sufficient temporal coverage for capturing diverse tidal cycles and seasonal climatic variations critical for forecasting. Such temporal diversity ensures model robustness and generalizability. The location of the buoy, including its regional and local spatial context along the southern coast of Java, is illustrated in Fig. 2.

B. Layout

Raw sea level data were originally sampled at one-minute intervals. To reduce computational complexity and data dimensionality, we performed temporal resampling by aggregating to hourly intervals. This aggregation utilized hourly mean values, calculated in Equation (1) where \bar{x}_h represents the hourly averaged sea level, x_i indicates minute-level observations, and n is the total number of observations per hour.

$$\bar{x}_h = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

After resampling from minute intervals to hourly intervals, gaps caused by sensor errors or transmission disturbances remained in dataset. To address this issue, we applied Tidal Harmonic Analysis (THA), a reliable technique that reconstructs missing tidal signals by decomposing the data into sinusoidal constituents using the cosine function [7]. The THA equation for reconstruction is defined by:

$$y(t) = \sum_{i=1}^m A_i \cos(\omega_i t + \phi_i) \quad (2)$$

where $y(t)$ in Equation (2), denotes reconstructed tidal height, A_i , ω_i , and ϕ_i represents the amplitude angular frequency, and phase of each tidal constituent, respectively. This preprocessing step substantially improved data completeness and consistency, essential for effective model training.

This harmonic synthesis allows for accurate reconstruction of the tidal component during periods of missing data, ensuring that the final dataset retains the essential periodicity to tidal forecasting. The unprocessed hourly time series, with missing data segments highlighted, is shown in Fig. 3a After applying THA, the resulting continuous and smoothed time series is illustrated in Fig. 3b, clearly showing improvements in completeness and coherence.

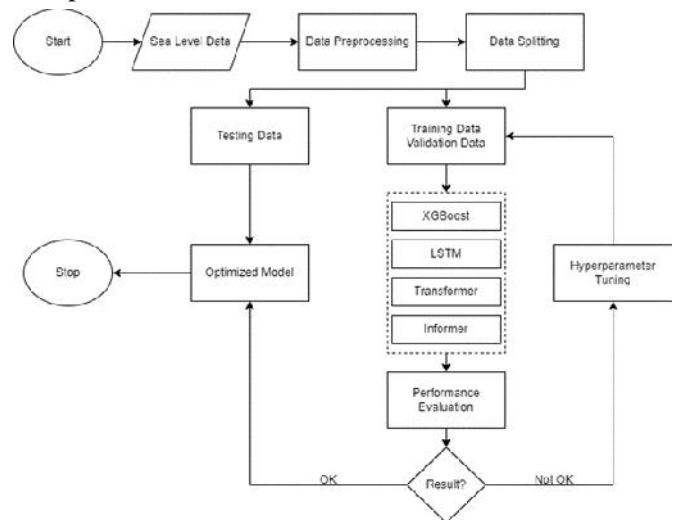
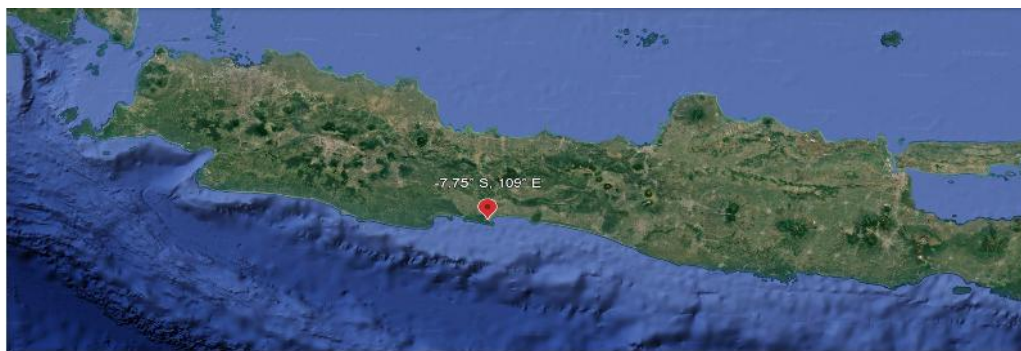
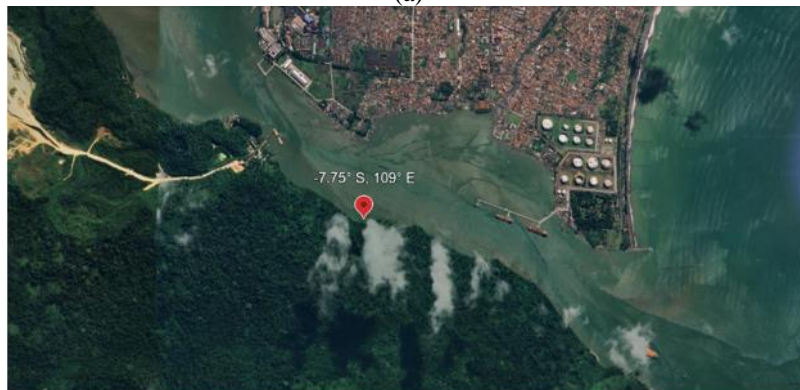


Fig. 1 Methodological flowchart

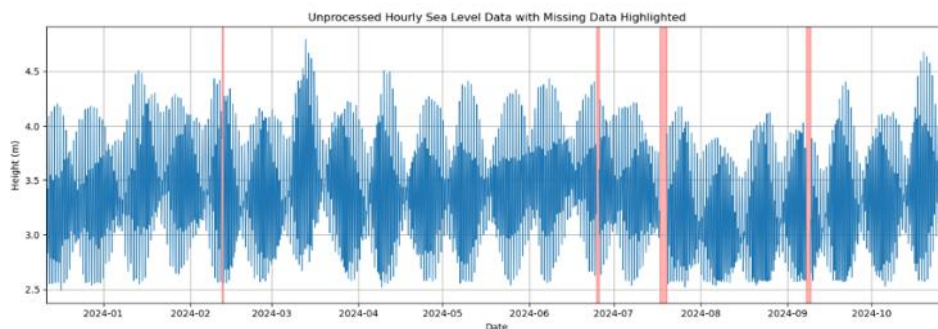


(a)

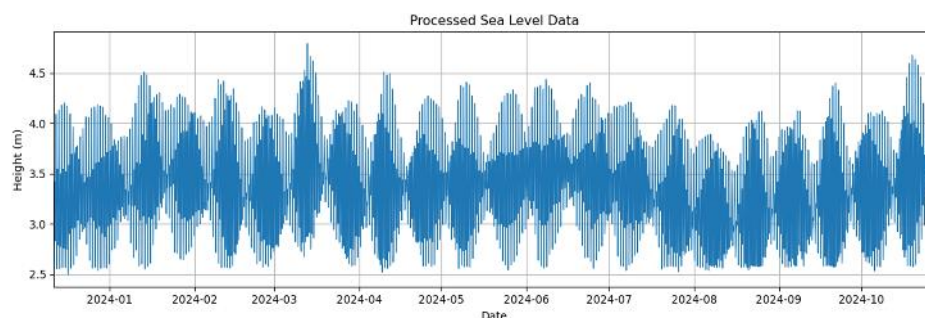


(b)

Fig. 2 Location of the tide gauge station near Cilacap, Central Java, Indonesia: (a) regional map of Java; (b) zoomed-in view



(a)



(b)

Fig. 3 Hourly sea level data at Cilacap: (a) unprocessed data hourly sea level data. Missing timestamps are highlighted in red; 1.76% of the 7771 hourly records were missing (135 timestamps); (b) processed hourly sea level data after gap filling with THA (7711 records)

To quantify the quality of THA reconstruction, we ran a mask-and-reconstruct validation, we randomly withheld 10% of observed timestamps and reconstructed them with the same THA settings used for real gaps. Comparing the reconstructions to the ground truth yielded $RMSE = 0.0703$, $MAPE = 1.71\%$, and correlation coefficient = 0.99. These results indicate that THA introduces only small errors relative to the observed variability while restoring continuity for downstream modelling.

Once resampled and gap-filled, the dataset is split into three subsets for model training and evaluation using a chronological approach. The first 70% (5398 records) of the data is allocated for training, allowing the model to learn underlying tidal and seasonal patterns. The next 15% (1157 records) is used as a validation set for hyperparameter tuning and early stopping, while the final 15% (1156 records) is reserved as an unseen test set for model evaluation. This chronological approach ensured the preservation of temporal causality and prevented data leakage, a critical aspect for reliable forecasting performance evaluation. The result of this splitting strategy is visualized in Fig. 4 which shows the distribution of training (blue), validation (orange), and testing (green) segments across the 11-month dataset.

To provide sufficient temporal variation and learning depth, a continuous 11-month period of hourly sea level data is selected for use. This window captures various tidal cycles, seasonal trends, and transient anomalies, offering a rich and representative temporal dataset for training a long sequence forecasting model.

C. Model Hyperparameters and Training Setup

All models were trained to minimize MSE. The data were split chronologically into training, validation, and test sets as described previously. Unless noted otherwise, optimization used Adam with a maximum of 100 epochs and early stopping based on validation RMSE with patience of 10, restoring the best checkpoint. Each configuration was run 10 times with distinct random seeds to support subsequent statistical analysis. Complete hyperparameters appear in Table I.

We first consider XGBoost, adopting direct multi-horizon strategy in which a separate regressor is fit for each prediction step. This design avoids error accumulation from recursive forecasting and aligns the learning objective with the evaluation horizons. The same input window and calendar features used for the neural baselines are applied to ensure a fair comparison. Our second backbone is LSTM, which models sequential dependencies through gated recurrent dynamics. This architecture is well suited to capturing short- and mid-

range temporal patterns while offering a strong, widely used baseline in time series forecasting. Third, we include a standard Transformer to capture long-range dependencies via self-attention. Finally, we evaluate the Informer architecture. Informer combines ProbSparse self-attention with time-feature embeddings to scale attention to long sequences efficiently while retaining expressiveness. This makes it a natural candidate for multi-horizon forecasting where long historical context can be informative. Experiments were executed on a single workstation with NVIDIA GeForce GTX 1660 Super VRAM 6 GB, Intel Core i5-12400F, and 32GB RAM.

D. Transformer

The Transformer architecture was first published by Vaswani et al. (2017) in their paper titled “Attention Is All You Need” as an alternative to recurrent neural networks (RNNs) and convolutional models in sequence-to-sequence tasks. The innovation of the Transformer lies in its use of a self-attention mechanism to model dependencies between elements in a sequence, allowing the model to process inputs in parallel rather than sequentially [21].

One of the main advantages of Transformers is its ability to effectively model long-range dependencies, a task that becomes increasingly difficult for traditional neural network based models as the length of the sequence increases. This makes Transformer particularly well-suited for tasks where contextual relationships are not limited to local proximity. Examples of the use of Transformer are language translation, document modelling, and time series forecasting.

The architecture follows an encoder-decoder framework, in which both components are built from several repeated layers. Within each encoder layer, a multi-head self-attention module is paired with a position-wise feed-forward network, and both are integrated with residual connections and layer normalization. The decoder includes an additional masked self-attention layer to maintain autoregressive prediction properties. The encoder processes the input sequence into abstract representations, while the decoder generates output predictions step-by-step based on these representations and previously generated outputs.

As illustrated in Fig. 5, the standard Transformer architecture consists of an encoder-decoder framework with self-attention mechanism. The figure shows the flow of information between the encoder and decoder blocks, emphasizing how each input token is connected through multi-head attention and processed in parallel. Each layer incorporates positional encoding to maintain the sequential nature of the data despite the absence of recurrence.

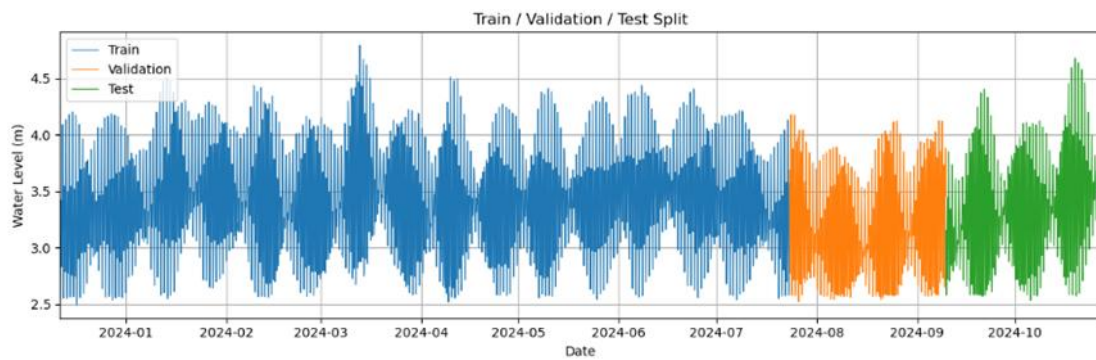


Fig. 4 Data splitting

At the heart of the Transformer lies the scaled dot-product attention mechanism in Equation (3), which enables the model to determine the relative importance of different elements within an input sequence. The attention mechanism takes three matrices as input: queries (Q), keys (K), and values (V). The attention score is computed by taking the dot product between Q and K, scaling it by the square root of the key dimension, applying a softmax function to obtain normalized weights, and then using those weights to compute a weighted sum of the values. The equation is formally expressed by the following operation:

$$At(Q, K, V) = sc \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (3)$$

Although Transformer have many advantages, the Transformer architecture has several limitations when used for long sequential time series forecasting. One of the main weakness are the quadratics time and the memory complexity, as attention operations must calculate interactions between all pairs of time steps sequentially. This results in low computational efficiency and high memory usage especially for datasets with thousands sequences [22]. Moreover, the model may not inherently encode the temporal continuity or physical dynamics often present in time-series data, which can affect generalization and stability when forecasting far into the future.

Several developments of the Transformer have been proposed to address these limitations. One such innovation is the Informer model, which introduces ProbSparse self-attention mechanism and a generative decoder specifically designed for efficient long-term sequence forecasting. The Informer model remains capable of modeling global dependencies and autoregressive mechanism while reducing computational costs, making it more suitable for tasks such as predicting data over long periods of time.

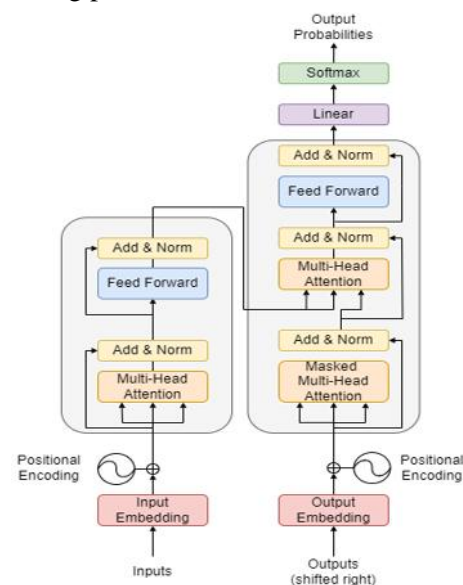


Fig. 5 Transformer architecture

TABLE I
HYPERPARAMETER SETTINGS FOR EACH MODELS

Model	Batch	Epoch	Hidden size	n_heads/dim_ff	Layers (enc/dec)	Dropout
Informer	32	100	512	8/-	2/1	0.05
Transformer	64	100	128	4/128	2/1	0.1
LSTM	64	100	128		2 (stacked with activation = 'tanh')	0.1
	n_estimators	subsample	colsample_bytree			
XGBoost	100	0.8	0.8			

The most notable innovation from Informer is the ProbSparse self-attention, a sparsity-aware attention mechanism designed to reduce the high computational complexity of standard self-attention. In traditional Transformers, attention is calculated between all query-key pairs, resulting in a quadratic complexity of $O(L^2)$ for a sequence of length L . In contrast, Informer performs attention computation only on a subset of the most “informative” queries, based on a sparsity measurement. This mechanism significantly lowers the complexity to approximately $O(L \log L)$, enabling the model to scale efficiently with longer sequences, which is crucial for forecasting problems such as sea level prediction [14,15].

ProbSparse self-attention computes attention exclusively between dominant queries and their relevant keys, rather than across the entire sequence [14]. Based on the proposed sparsity measurement, the mechanism allows each key to attend only to the top- u most informative queries [14]. This ProbSparse self-attention focuses calculations on the most relevant interactions while ignoring the least relevant ones. This reduces redundancy without sacrificing prediction accuracy. This sparsity-aware selection process enables Informer to effectively modeling global dependencies while operating with much lower memory and runtime requirements compared to standard full-attention approaches especially Transformer.

The Informer incorporates a self-attention distilling mechanism that hierarchically reduces the length of the output sequence across layers after ProbSparse self-attention by filtering out redundant temporal patterns [14]. This distillation method allows the model to reduce the least relevant data, concentrate on predominant, and significant patterns in the data, hence improving representation quality and computing efficiency. Moreover, Informer employs a generative decoder that forecasts the complete output sequence in one and not sequentially [14].

As shown in Fig. 6, the Informer architecture employs ProbSparse self-attention and a generative decoder for long-horizon forecasting. The encoder consists of stacked layers of ProbSparse self-attention followed by distilling layers, which progressively compress the temporal sequence while preserving essential features. The decoder includes masked attention layers and cross-attention with encoder outputs, followed by a linear projection layer to produce the final forecasted sequence.

Informer retains the core attention operation used in the original Transformer. The attention mechanism is defined as Equation 3. Where Q , K , and V are the query, key, and value matrices, and d_k is the dimension of the key vectors. The softmax function ensures that attention weights are normalized, allowing each output token to be computed as a weighted sum of all values in the sequence.

By combining computational efficiency through sparse attention, pattern compression via distillation, and faster inference with a generative decoder, Informer effectively addresses the major shortcomings of standard Transformers in time-series forecasting. These design enhancements make it highly suitable for long-horizon sea level prediction, where both accuracy and scalability are critical.

E. Evaluation Metrics

To quantitatively assess the performance of the forecasting models, this study adopts three widely used evaluation metrics: Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the Correlation Coefficient (CC). These metrics are selected to capture different aspects of prediction accuracy, including absolute error, percentage deviation, and the linear relationship between predicted and observed values.

RMSE measures the average magnitude of the error by taking the square root of the mean squared differences between predicted and actual values. RMSE penalizes large errors more severely, making it sensitive to extreme deviations as it shown in Equation (4).

$$R = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (4)$$

where y_i and \hat{y}_i are the observed and predicted sea level values at time i , and N is the total number of observations.

MAPE provides a normalized measure of prediction accuracy by expressing the absolute error as a percentage of the actual values. This metric is particularly useful for interpreting the relative size of prediction errors and it is calculated as Equation (5).

$$M = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (5)$$

However, MAPE can be unstable when actual values are close to zero. To minimize distortion, extreme low values in the denominator are carefully managed during implementation.

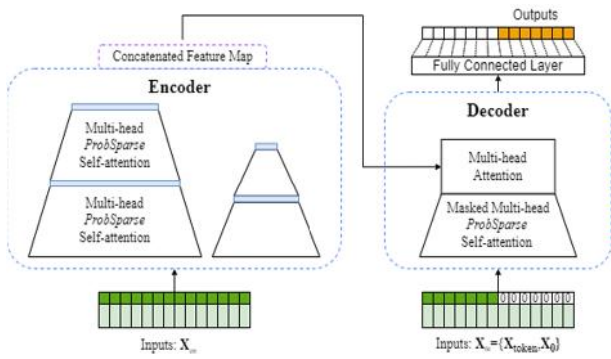


Fig. 6 Informer architecture

CC evaluates the strength and direction of the linear relationship between the predicted and actual values. A higher CC indicates that the predicted trend closely follows the actual trend, even if some absolute errors are present. The coefficient is computed using Equation (6), where \bar{y} and \hat{y} are the mean values of the actual and predicted sequences, respectively.

$$C = \frac{\sum_{i=1}^N (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^N (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^N (\hat{y}_i - \bar{\hat{y}})^2}} \quad (6)$$

Together, these three metrics provide a robust and comprehensive evaluation framework. RMSE highlights the presence of large outliers, MAPE offers interpretability in percentage terms, and CC reflects the overall similarity in pattern and trend between the observed and predicted series. This combination enables a balanced assessment of model performance across both accuracy and structural consistency.

III. RESULT AND DISCUSSION

This study evaluates the predictive performance of four models: XGBoost, LSTM, Transformer, and Informer. The primary forecasting model is Informer, while the other three serve as baseline models. These models are trained using hourly sea level data and tested on multi-step predictions over five forecast horizons: 1, 3, 5, 7, and 14 days ahead. The accuracy of the models is assessed using RMSE, MAPE, and CC.

Table II summarizes forecasting performance across horizons. Informer achieves the lowest RMSE and MAPE values and maintains high CC values across all forecast horizons, particularly for longer prediction windows. For instance, Informer records an RMSE of 0.058 and MAPE of 1.44% on the 1-day forecast, while still maintaining strong performance at 14 days with RMSE 0.091 and MAPE 2.14%.

In comparison, LSTM demonstrates competitive short-term accuracy but declines more rapidly on longer

forecasts, with RMSE reaching 0.158 and MAPE 3.68% at 14 days. Transformer performs well in shorter horizons but becomes less consistent on extended forecasts. XGBoost shows solid results for 1 to 3-day horizons but fails to capture long-term patterns due to its non-sequential nature [18,23]. In contrast, baseline models like LSTM and Transformer demonstrate declining accuracy as forecast horizons extend, indicating Informer’s strength in modeling long-term temporal dependencies.

Fig. 7 compares forecasts over 14 days. Informer closely tracks both tidal phase and amplitude, whereas Transformer deviates after day 10, XGBoost fails to preserve periodicity, and LSTM smooths rapid fluctuations. These results highlight Informer’s architectural advantage in preserving long-range patterns.

Findings indicate that the Informer model delivers the most consistent and precise forecasts, showing clear advantages when applied to extended horizons. Its architecture, which features ProbSparse self-attention and attention distillation, enables it to maintain consistent performance by focusing on dominant temporal patterns and filtering noise. This makes it well-suited for operational sea-level forecasting in coastal management and early warning systems.

The experimental results demonstrate that the Informer model consistently outperforms traditional and deep learning baseline methods, namely XGBoost, LSTM, and standard Transformer, across all tested prediction horizons (1, 3, 5, 7, and 14 days). The Informer achieves the low RMSE (0.091) and high correlation coefficient (0.98) on the 14-day forecast, affirming its ability to model long-range dependencies in sea level time series.

Using the 10 paired runs at the primary horizon (14-day), Wilcoxon signed-rank tests showed that Informer achieves significantly lower RMSE than XGBoost (median RMSE = -0.0357; p_adj = 0.00293), LSTM (median RMSE = -0.0608; p_adj = 0.00195), and Transformer (median RMSE = -0.0521; p_adj = 0.00098). These results statistically substantiate the performance gaps summarized in Table II.

Informer’s superior performance reflects its efficient design, which scales attention to long sequences while avoiding vanishing-gradient issues common in recurrent models. In addition to the architectural advantages, we also note that model performance is not uniform across all conditions. Informer maintained stable accuracy under most tidal regimes, but relative errors tended to increase during periods with stronger non-tidal influences, such as during the northwest monsoon season when residuals are larger. This suggests that while

ProbSparse self-attention effectively emphasizes the dominant tidal periodicities (semidiurnal and diurnal), the model is less resilient to extreme meteorological disturbances. Therefore, Informer’s advantage appears strongest when tidal signals dominate, whereas during monsoon-driven events its accuracy converges closer to that of the baseline models. This observation highlights both the strength of Informer in capturing periodic structure and its limitation under conditions where exogenous forcing play a greater role.

While Informer demonstrated superior performance, several limitations remain. The current study uses only univariate sea level data; integrating additional meteorological and oceanographic parameters such as wind speed and pressure might further improve predictive accuracy. Future work should explore multivariate forecasting scenarios and assess Informer’s

transferability to other region with distinct oceanographic characteristics. Additionally, interpretability analysis of attention patterns could provide deeper insights into the model’s internal decision-making process.

To provide further context, several recent research have explored alternative models for sea level and water level forecasts. For instance, Puspita Sari and Adytia [13] implemented an LSTM with attention mechanism on Jakarta’s sea level data and reported an RMSE of 0.14 and a correlation coefficient of 0.93 on the 14-day horizon. Although their RMSE is comparable, the geographical environment differs substantially. Jakarta Bay is relatively confined and exhibits simpler tidal dynamics compared to Cilacap, which is open to the Indian Ocean and vulnerable to monsoonal effects, IOD, and swell.

TABLE II
FORECASTING RESULTS OF MODELS ACROSS MULTIPLE PREDICTION HORIZONS

Prediction (Days)	XGBoost			LSTM			Transformer			Informer		
	RMSE	MAPE	CC	RMSE	MAPE	CC	RMSE	MAPE	CC	RMSE	MAPE	CC
1	0.063	1.49	0.99	0.071	1.71	0.98	0.059	1.39	0.99	0.058	1.44	0.99
3	0.088	2.04	0.98	0.082	1.90	0.98	0.074	1.73	0.98	0.067	1.61	0.99
5	0.106	2.44	0.97	0.139	3.13	0.95	0.109	2.47	0.97	0.054	1.26	0.99
7	0.117	2.68	0.96	0.148	3.38	0.94	0.122	2.82	0.97	0.068	1.63	0.99
14	0.122	2.88	0.96	0.158	3.68	0.93	0.115	2.72	0.97	0.091	2.14	0.98

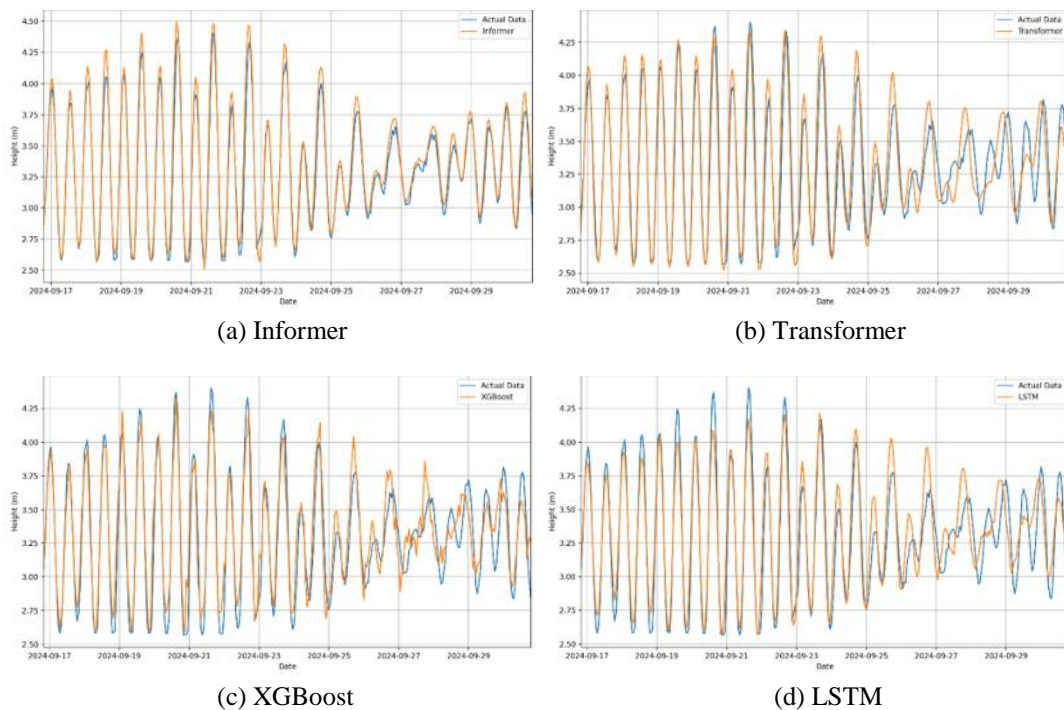


Fig. 7 Comparison models for 14-day forecast vs actual sea level data

IV. CONCLUSION

This study proposes a deep learning approach with Informer as its main model for sea level forecasting in Cilacap, Indonesia. This study applies this approach to hourly sea level data over 11 months from Cilacap. The data pre-processing process includes interpolation data to hourly intervals and filling missing gap using Tidal Harmonic Analysis (THA). Based on the result, THA performs well to filling the missing gaps in data. Across all forecasting horizons from 1 to 14 days, Informer outperformed XGBoost, LSTM, and Transformer in terms of Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and correlation coefficient. From the metrics evaluation, Informer demonstrated consistent metric evaluation results, and at longer horizons, Informer achieved an RMSE of 0.091, MAPE of 2.14, and CC of 0.98, which clearly outperforming all benchmark models. To validate more, we tested using paired Wilcoxon test with Holm correction with RMSE as the main metric using 14 days as the forecast horizon. The results of metric evaluation and Wilcoxon paired test prove the Informer's ability to model and predict complex temporal patterns in tropical coastal environments especially in Cilacap. Future research could incorporate multivariate inputs, such as meteorological or oceanographic features, investigate the model's adaptability to other coastal regions with different physical dynamics, and forecast at the longer horizons.

ACKNOWLEDGEMENT

This research was funded by Telkom University, Indonesia, under the International Research Scheme with contract number No. 380/LIT06/PPM-LIT/2025.

REFERENCES

- [1] S. Tumse and U. Alcansoy, "Statistical and Deep Learning Approaches in Estimating Present and Future Global Mean Sea Level Rise," *Natural Hazards*, 2025, doi: 10.1007/s11069-025-07203-5.
- [2] Kiera L. O'Donnell, Emily S. Bernhardt, Xi Yang, Ryan E. Emanuel, Marcelo Ardón, Manuel T. Lerdau, Alex K. Manda, Anna E. Braswell, Todd K. BenDor, Eric C. Edwards, Elizabeth Frankenberg, Ashley M. Helton, John S. Kominoski, Amy E. Lesen, Lindsay Naylor, Greg Noe, Kate L. Tully, Elliott White, and Justin P. Wright, "Saltwater Intrusion and Sea Level Rise Threatens U.S. Rural Coastal Landscapes and Communities," Mar. 2024, *Elsevier Ltd.* doi: 10.1016/j.ancene.2024.100427.
- [3] A. Nurlatifah, Martono, I. Susanti, and M. Suhermat, "Variability and Trend of Sea Level in Southern Waters of Java, Indonesia," *Journal of Southern Hemisphere Earth Systems Science*, vol. 71, no. 3, pp. 272–283, Dec. 2021, doi: 10.1071/ES21004.
- [4] Y. Li, J. Feng, X. Yang, S. Zhang, G. Chao, L. Zhao, and H. Fu, "Analysis of Sea Level Variability and Its Contributions in the Bohai, Yellow Sea, and East China Sea," *Frontiers in Marine Science*, vol. 11, 2024, doi: 10.3389/fmars.2024.1381187.
- [5] N. Alenezi, A. Alsulaili, and M. Alkhalidi, "Prediction of Sea Level in the Arabian Gulf Using Artificial Neural Networks," *Journal of Marine Science and Engineering*, vol. 11, no. 11, Nov. 2023, doi: 10.3390/jmse11112052.
- [6] A. W. Ramadhan, D. Adytia, D. Saepudin, S. Husrin, and A. Adiwijaya, "Forecasting of Sea Level Time Series using RNN and LSTM Case Study in Sunda Strait," *Lontar Komputer: Jurnal Ilmiah Teknologi Informasi*, vol. 12, no. 3, p. 130, Oct. 2021, doi: 10.24843/lkjiti.2021.v12.i03.p01.
- [7] E. J. Lee, K. Kim, and J. H. Park, "Reconstruction of long-term sea-level data gaps of tide gauge records using a neural network operator," *Frontiers in Marine Science*, vol. 9, Oct. 2022, doi: 10.3389/fmars.2022.1037697.
- [8] E. Forootan, S. Farzaneh, K. Naderi, and J. P. Cederholm, "Analyzing GNSS Measurements to Detect and Predict Bridge Movements Using the Kalman Filter (KF) and Neural Network (NN) Techniques," *Geomatics*, vol. 1, no. 1, pp. 65–80, Feb. 2021, doi: 10.3390/geomatics1010006.
- [9] M. Cho, C. Kim, K. Jung, and H. Jung, "Water Level Prediction Model Applying a Long Short-Term Memory (LSTM)-Gated Recurrent Unit (GRU) Method for Flood Prediction," *Water (Switzerland)*, vol. 14, no. 14, Jul. 2022, doi: 10.3390/w14142221.
- [10] Cornelius Stephanus Alfredo and D. A. Adytia, "Time Series Forecasting of Significant Wave Height using GRU, CNN-GRU, and LSTM," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 5, pp. 776–781, Oct. 2022, doi: 10.29207/resti.v6i5.4160.
- [11] M. Waqas and U. W. Humphries, "A Critical Review of RNN and LSTM Variants in Hydrological Time Series Predictions," *MethodsX*, vol. 13, p. 102946, Dec. 2024, doi: 10.1016/j.mex.2024.102946.
- [12] D. Saepudin, Egi Shidqi Rabbani, Dio Navialdy, and Didit Adytia, "Water Level Time Series Forecasting Using TCN Study Case in Surabaya," *Jurnal Online Informatika*, vol. 9, no. 1, pp. 61–69, Apr. 2024, doi: 10.15575/join.v9i1.1312.
- [13] A. Puspita Sari and D. Adytia, *Sea Level Time Series Forecasting by using LSTM with Attention Mechanism, A Case Study in Jakarta, Indonesia*. IEEE, 2024. doi: 10.1109/ICoDSA62899.2024.10652146.

- [14] H. Zhou, "Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting," 2021. doi: <https://doi.org/10.48550/arXiv.2012.07436>.
- [15] P. Lu, Y. He, W. Li, Y. Chen, R. Kong, and T. Wang, "An Informer-based multi-scale model that fuses memory factors and wavelet denoising for tidal prediction," *Electronic Research Archive*, vol. 33, no. 2, pp. 697–724, 2025, doi: 10.3934/era.2025032.
- [16] N. Tepetidis, D. Koutsoyiannis, T. Iliopoulou, and P. Dimitriadis, "Investigating the Performance of the Informer Model for Streamflow Forecasting," *Water (Switzerland)*, vol. 16, no. 20, Oct. 2024, doi: 10.3390/w16202882.
- [17] Zhang, T., Lin, P., Liu, H., Wang, P., Wang, Y., Zheng, W., Yu, Z., Jiang, J., Li, Y., & He, H. (2025). A New Transformer Network for Short-Term Global Sea Surface Temperature Forecasting: Importance of Eddies. *Remote Sensing*, 17(9), 1507. <https://doi.org/10.3390/rs17091507>
- [18] A. Frifra, M. Maanan, M. Maanan, and H. Rhinane, "Harnessing LSTM and XGBoost algorithms for storm prediction," *Scientific Reports*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-62182-0.
- [19] C. Chen and J. Dong, "Deep Learning Approaches For Time Series Prediction in Climate Resilience Applications," *Frontiers in Environmental Science* vol. 13, 2025, doi: 10.3389/fenvs.2025.1574981.
- [20] M. Vicens-Miquel, P. E. Tissot, and F. A. Medrano, "Exploring Deep Learning Methods for Short-Term Tide Gauge Water Level Predictions," *Water (Switzerland)*, vol. 16, no. 20, Oct. 2024, doi: 10.3390/w16202886.
- [21] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need," Jun. 2017, [Online]. Available: <http://arxiv.org/abs/1706.03762>
- [22] I. Aguilera-Martos, A. Herrera-Poyatos, J. Luengo, and F. Herrera, "Local Attention Mechanism: Boosting the Transformer Architecture for Long-Sequence Time Series Forecasting," Oct. 2024, doi: <https://doi.org/10.48550/arXiv.2410.03805>.
- [23] M. N. Adli Zakaria, A. N. Ahmed, M. A. Malek, A. H. Birima, Md. M. Hayet Khan, M. Sherif, and A. Elshafie, "Exploring Machine Learning Algorithms For Accurate Water Level Forecasting in Muda River, Malaysia," *Heliyon*, vol. 9, no. 7, Jul. 2023, doi: 10.1016/j.heliyon.2023.e17689.
- [24] N. Raj, Z. Gharineiat, A. A. M. Ahmed, and Y. Stepanyants, "Assessment and Prediction of Sea Level Trend in the South Pacific Region," *Remote Sensing (Basel)*, vol. 14, no. 4, Feb. 2022, doi: 10.3390/rs14040986.

