



Enhancing satellite clock bias prediction in BDS with LSTM-attention model

Chenglin Cai¹ · Mingyuan Liu¹ · Pinchun Li¹ · Zexian Li¹ · Kaihui Lv¹

Received: 11 September 2023 / Accepted: 5 March 2024 / Published online: 25 March 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

Satellite clock bias (SCB) is a critical factor influencing the accuracy of real-time precise point positioning. Nevertheless, the utilization of real-time service products, as supplied by the International GNSS Service, may be vulnerable to interruptions or network failures. In specific situations, users may encounter difficulties in obtaining accurate real-time corrections. Our research presents an enhanced predictive model for SCB using a long short-term memory (LSTM) neural network fused with a Self-Attention mechanism to address this challenge. This fusion enables the model to effectively balance global attention and localized feature capture, ultimately enhancing prediction accuracy and stability. We compared and analyzed our proposed model with convolutional neural network (CNN) and LSTM models. This analysis encompasses an assessment of the model's strengths and suitability for predicting SCB within the BeiDou navigation system, considering diverse satellites, orbits, and atomic clocks. Our results exhibit a substantial improvement in predictive accuracy through the LSTM-Attention model. There has been an improvement of 49.67 and 62.51% compared to the CNN and LSTM models in the 12-h prediction task. In the case of the 24-h prediction task, the improvements escalated to 68.41 and 71.16%, respectively.

Keywords SCB · BDS · Self-Attention · LSTM · CNN

Introduction

Precise point positioning (PPP) is a rapidly evolving technology, particularly RT-PPP, which has garnered significant attention (Malys and Jensen 1990). However, achieving widely applicable RT-PPP still faces limitations in the accuracy and real-time nature of satellite orbits and SCB products (El-Mowafy et al. 2017). The IGS has been providing RTS since 2013, including high-precision orbit and clock products (Wang et al. 2019b; Elsobeiey and Al-Harbi 2016). Compared with IGS rapid solutions, the accuracy of RTS clock bias products can reach 0.1–0.15 ns (Hadas and Bosy 2015). Nevertheless, practical applications may still encounter issues, such as interruptions in RTS products, errors and jumps in clock bias data, or data unavailability due to network failures (Zhang et al. 2019; Nie et al. 2017). To tackle these challenges, real-time high-precision positioning can be attained through the prediction of SCB during

interruptions in RTS products. This task entails predicting SCB for a future time interval based on the available clock bias data from the preceding period (Huang et al. 2014). Therefore, the establishment of a high-precision and short-term clock bias prediction model holds practical value.

In the realm of clock bias prediction, prior studies have frequently utilized methodologies like linear programming (LP) (Cernigliaro and Sesia 2012), quadratic polynomials (QP) (Huang et al. 2018), and gray models (GM) (Liang et al. 2015). However, it should be noted that the LP model does not account for the influence of clock drift on clock bias prediction. The QP model treats errors as noise that follows a normal distribution, leading to a reduction in prediction accuracy over time (Jonsson and Eklundh 2002). On the other hand, GM predictions rely on the assumption that the original function is smooth and exhibits exponential changes. This makes prediction accuracy sensitive to variations in function coefficients. Numerous studies have employed composite clock bias prediction models (Wang et al. 2017a, 2019a; Lu et al. 2018). Experimental findings consistently show that these composite models outperform single models in terms of prediction accuracy and stability.

✉ Mingyuan Liu
lmy_un@163.com

¹ College of Automation and Electronic Information, Xiangtan University, Xiangtan 411105, China

Owing to the vulnerability of satellite clocks to external environmental factors, the clock bias inevitably manifests in periodic and stochastic fluctuations (Qingsong et al. 2017). Traditional models have limitations in capturing the nonlinear characteristics of clock bias, which restrict the potential for further improving prediction accuracy. In contrast, neural networks are adept at addressing nonlinear challenges, surpassing the constraints of traditional models and enabling more precise predictions. Wavelet neural networks (WNN) (Wang et al. 2017b, 2021) and supervised learning long short-term memory (SL-LSTM) models (Huang et al. 2021) have been employed to predict SCB in the GPS satellite system, yielding promising results. The Transformer's encoder architecture was leveraged in previous research to model and predict GPS clock bias (Syam et al. 2023). In order to enhance the convergence speed and predictive accuracy of neural network models, there has been research focused on integrating optimization algorithms with neural networks. For instance, a clock bias prediction model based on mind evolution algorithm (MEA) optimization has been previously proposed to enhance the initial weights and thresholds during the training of a neural network using the back-propagation (BP) algorithm (Bai et al. 2023). An approach was introduced to integrate the particle swarm optimization (PSO) algorithm with a neural network model trained using the BP algorithm (Zhao et al. 2021). An enhanced neural BP network model, optimized through a combination of heterogeneous comprehensive learning and dynamic multi-swarm particle swarm optimizer (HPSO-BP), was introduced to address the potential issue of premature convergence associated with the PSO algorithm (Lv et al. 2022). Notably, the performance of this model surpasses that of conventional approaches.

In contrast to GPS satellites, BDS encompasses satellites of different types, utilizing diverse atomic clock technologies such as hydrogen maser clocks and rubidium clocks. This diversity results in the complexity and variability of SCB patterns. Presently, there exists a relatively limited body of research dedicated to the prediction of BeiDou SCB, presenting an opportunity for enhancing predictive accuracy. In previous research, LSTM was employed for the prediction of SCB with the third-generation BeiDou satellites (He et al. 2023). The results showed that the LSTM model performed better than the autoregressive integrated moving average (ARIMA) and QP model.

However, it should be noted that the LSTM model exhibits certain limitations when dealing with longer sequence lengths due to its inherently sequential nature. These limitations can potentially result in information loss and impede the effective capture of long-range dependencies (Huang et al. 2021). This research introduces an

LSTM model integrated with a Self-Attention mechanism (LSTM-Attention) to address this issue. Self-Attention is an attention mechanism relating different positions of a single sequence to compute a representation of the sequence. The Self-Attention mechanism allows the modelling of all dependencies without regard to their distance in the input or output sequences (Vaswani et al. 2023). Although the Self-Attention mechanism was initially designed to address issues in natural language processing (Lin et al. 2017; Cheng et al. 2016; Parikh et al. 2016), its uniqueness lies in its ability to consider global dependencies, thus helping overcome these limitations of LSTM models in long-time series. This research has developed an LSTM-Attention model for predicting BeiDou SCB. The SCB data are pre-processed using first-order difference and Euclidean norm normalization (L2 normalization) and subsequently utilized for modeling. The applicability of the model and methodology in SCB will be discussed with regard to various factors, including different satellites, orbits, and atomic clocks.

Methodology

In order to develop a model better suited for predicting the clock errors of BeiDou satellites, we integrated the LSTM model with a Self-Attention mechanism. In the subsequent sections, we will provide a detailed exposition of the characteristics of these two approaches while also delving into our enhancements to the LSTM model to address its limitations. Furthermore, we will introduce specific data pre-processing techniques aimed at further enhancing the model's performance.

Data pre-processing

The clock bias of the same satellite generally exhibits a linear trend. Deep learning networks tend to have difficulty in handling the original clock bias sequences with linear trends, as they are susceptible to the influence of trend components in the data. To remove this trend and simultaneously facilitate a better understanding of the nonlinear features within the SCB data, we apply a first-order difference to the original lock bias data used for model training. The following defines a set of n -dimensional SCB data sequences:

$$X = \{x_1, x_2, x_3, \dots, x_n\} \quad (1)$$

where $x_i, i = 1, 2, 3, \dots, n$ represents the SCB data for different epochs, totaling n epochs. By performing a first-order difference on the data from consecutive epochs, a new set of

SCB sequences can be obtained. This sequence is defined as follows:

$$X' = \{\Delta x_1, \Delta x_2, \Delta x_3, \dots, \Delta x_{n-1}\} \tag{2}$$

where $\Delta x_i = x_{i+1} - x_i$.

In order to eliminate the scale differences in the data after the first-order difference and enhance the stability and generalization capability of the model, it is necessary to normalize the sequence data using the L2 normalization. The L2 normalization formula is presented as follows:

$$\tilde{X} = \frac{X'}{\|X'\|_2} \tag{3}$$

where $\|X'\|_2$ is the L2 normalization of the first-order differenced sequence.

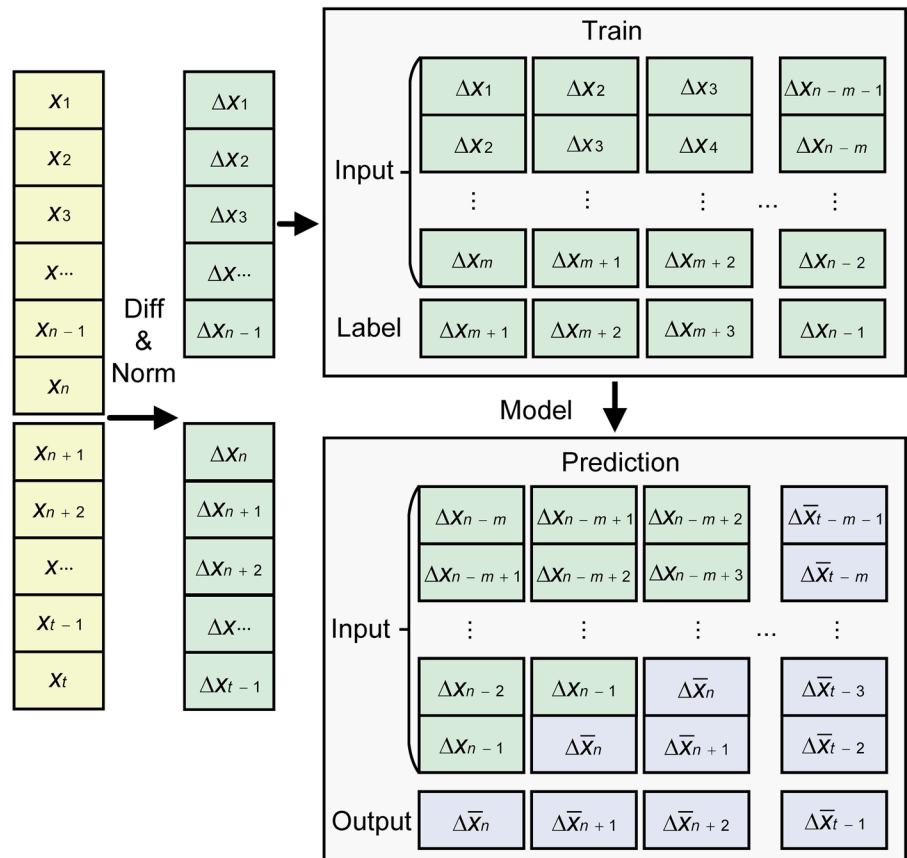
We have employed a sliding window approach for data processing to mitigate computational complexity and extend the predictive capabilities to longer target sequences. Here, we designate the window length as m and the normalized data can be structured into $(n - m - 1)$ distinct data groups as follows:

$$\begin{aligned} & \{\Delta x_1, \Delta x_2, \dots, \Delta x_m, \Delta x_{m+1}\} \\ & \{\Delta x_2, \Delta x_3, \dots, \Delta x_{m+1}, \Delta x_{m+2}\} \\ & \vdots \\ & \{\Delta x_{n-m-1}, \Delta x_{n-m}, \dots, \Delta x_{n-2}, \Delta x_{n-1}\} \end{aligned} \tag{4}$$

where $\Delta x_i, i = 1, 2, \dots, n - 1$ represents the SCB data that has undergone first-order difference and L2 normalization. For each data set, the first m data points serve as inputs to the model, while the last data point represents the predicted label value.

Figure 1 illustrates the specific data processing flow. During the training phase, we follow the training set described earlier and predict only one data point at a time. In the prediction phase, the first input data for the model consist of the preceding m known data points of the target data, denoted as $\{\Delta x_{n-m}, \Delta x_{n-m+1}, \dots, \Delta x_{n-2}, \Delta x_{n-1}\}$ in Fig. 1. This step yields the first predicted value. Subsequently, we remove the first data point Δx_{n-m} from the input sequence and simultaneously append the first predicted value $\Delta \bar{x}_n$ to the end of the data, serving as the input for the next prediction. Through this iterative process, we achieve the prediction of SCB data for 2 or 24 h, represented $\{\Delta \bar{x}_n, \Delta \bar{x}_{n+1}, \dots, \Delta \bar{x}_{t-2}, \Delta \bar{x}_{t-1}\}$ in Fig. 1. The output for each step is a vector containing the predicted value for each epoch. We only need to focus on the last value

Fig. 1 Data pre-processing of LSTM-Attention model. The figure illustrates the process of transforming the SCB data sequence into model inputs



in the vector. After undergoing a linear transformation, this value becomes a scalar, representing the final prediction of the model. The SCB data predictions for a time period are formed by multiple forecast results.

LSTM and Self-Attention

The recurrent neural network (RNN) is a neural network characterized by cyclic connections among nodes, utilizing its network structure to discover correlations in sequences. RNN is well suited for time series prediction and has found successful applications in various domains, including text, video, and speech processing. The LSTM represents a variant of RNN designed to address the issues of vanishing and exploding gradients during training (Yu et al. 2019). By introducing gate functions, LSTM endows the network with enhanced memory capabilities, thereby yielding improved results on longer time series. Figure 2 illustrates the cell structure of the LSTM neural network model. Each cell incorporates three essential gates: the forget gate, input gate, and output gate, as expressed by the following equations:

$$f_t = \text{sig}(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{5}$$

$$i_t = \text{sig}(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{6}$$

$$S_t = \text{tanh}(W_s \cdot [h_{t-1}, x_t] + b_s) \tag{7}$$

$$C_t = f_t \times C_{t-1} + i_t \times S_t \tag{8}$$

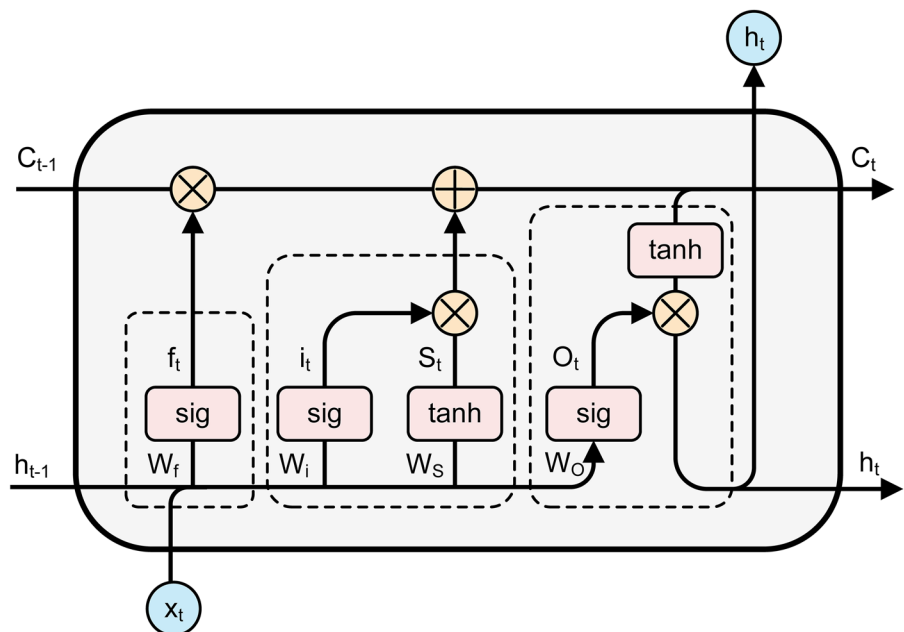
$$O_t = \text{sig}(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{9}$$

$$h_t = O_t \times \text{tanh}(C_t) \tag{10}$$

where h_t , C_t , and x_t , respectively, represent the hidden state, cell state, and cell input at the time step. W and b represent the weight matrix and bias of the current network layer, respectively. Equation (5) illustrates the forgetting process by applying a sigmoid mapping to $[h_{t-1}, x_t]$ to obtain a value between 0 and 1, which controls the extent of retention or forgetting. Equations (6) and (7) express the input process, where a sigmoid layer determines which information needs to be updated, and a hyperbolic tangent (Tanh) layer generates a candidate vector for updating the cell state. Equation (8) combines the forgotten and added information to update the current cell state. Finally, (9) and (10) demonstrate the output process, utilizing a sigmoid layer and a tanh layer to generate the hidden state.

The Self-Attention mechanism is an attention mechanism used for computing a representation of the sequence, effectively capturing correlations between different positions within a single sequence. It has been widely applied across various tasks, including natural language inference and abstractive summarization (Paulus et al. 2017; Vaswani et al. 2023). The core concept of the Self-Attention mechanism is that the representation at each position can be composed as a weighted sum of other positions in the sequence. Each position is assigned an attention weight in the Self-Attention mechanism, indicating its dependency on other positions in the sequence. The higher the dependency, the larger the corresponding attention weight. This capability enables the model to autonomously discern and

Fig. 2 A LSTM cell. Each cell of the LSTM model comprises three gates. From left to right, the three dashed boxes in the figure represent the forget, input, and output gates. Multiple LSTM cells are recurrently connected to form an LSTM neural network model



adapt to the varying degrees of association between positions, thereby facilitating the weighted aggregation of information from different positions.

As shown in Fig. 3, the Self-Attention mechanism involves the following steps: (1) For each position in the input sequence, generate query, key, and value vectors, which will be used to calculate the relevance weights between positions. (2) Calculate the dependencies between the query vector and all key vectors using dot products, then scale the computed dot products and normalize them to obtain attention weights. These attention weights indicate the degree of association between the current position and other positions, resembling a weight distribution. (3) Calculate the weighted sum of all value vectors based on the attention weights to obtain an aggregated representation. In summary, the expression of the Self-Attention mechanism can be formulated as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{11}$$

where Q , K , and V represent the query vector, key vector, and value vector, respectively. $\sqrt{d_k}$ stands for the input dimension of the key vectors. Here, it serves as a scaling factor to prevent the softmax function from being pushed

into a region with extremely small gradients when the dot products grow large.

Constructing the LSTM-attention model

LSTM is a type of recurrent neural network capable of effectively handling time dependencies, making it suited for processing time series data. In the field of time series data analysis and prediction, LSTM has been widely employed to capture time dependencies within data. However, LSTM also has its limitations. In the context of lengthy sequences, the model may encounter difficulties in effectively capturing intricate relationships between global and local contexts. Furthermore, the information flow within LSTM primarily relies on hidden states, occasionally resulting in constrained information dissemination.

To address these limitations and enhance model performance, we introduce the Self-Attention mechanism. When applied to sequence, Self-Attention assigns varying weights to different time steps, facilitating increased information exchange. This enables the model to better focus on the relationships between the global and local contexts, thereby improving its ability to capture long-term dependencies. We constructed an LSTM-Attention model for SCB prediction tasks by combining LSTM with the Self-Attention mechanism. Figure 4 depicts the core structure of our LSTM-Attention model, which comprises multiple data processing layers. Each layer includes an LSTM layer, a linear layer, and a Self-Attention layer. To maintain the model's expressive capacity, we utilize techniques such as residual connections and layer normalization within the Self-Attention layer.

The data processing layer comprises a stack of $N = 2$ identical layers. Within this layer, the first-differenced and normalized data undergo processing through an LSTM layer designed to capture long-term dependencies within the time series. The LSTM's function is to discern dynamic patterns and trends within the sequence, ultimately contributing to the improved accuracy of SCB prediction. Due to the sequential nature of LSTM, it processes the input differenced data based on the order of epochs. Each cell input processes the data for one epoch. The hidden state of each epoch represents the learned combination of both the current epoch's data features and historical data features. In theory, the hidden state can be understood as the predicted SCB result for the next epoch. The propagating cell state retains feature information about the historical SCB in each epoch. However, the cell state undergoes selective updates, deletions, or additions during the propagation process. This characteristic may lead to the loss of crucial portions of historical information, which is a drawback in

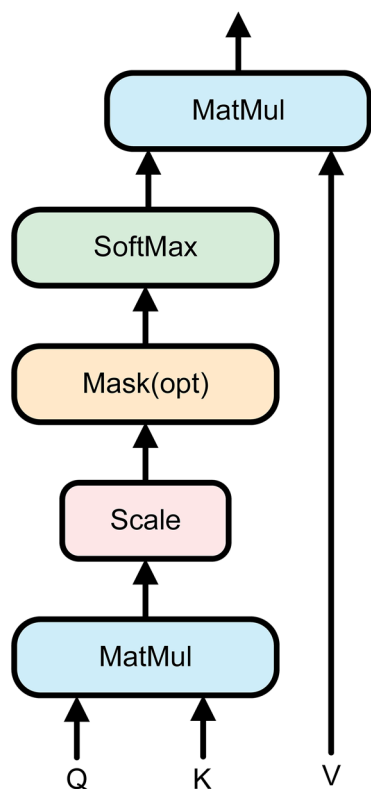
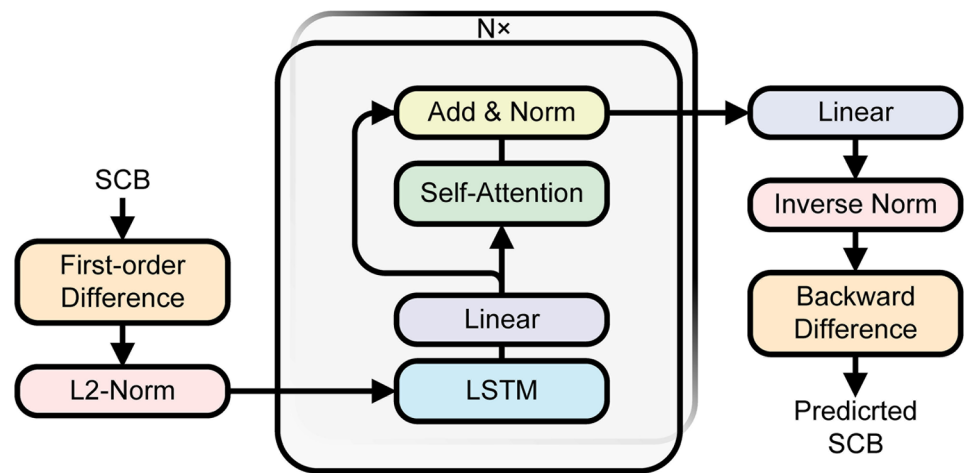


Fig. 3 Scaled Dot-Product attention. The computational process of the Self-Attention mechanism is demonstrated

Fig. 4 LSTM-Attention model. SCB prediction primarily consists of three stages: data preprocessing, model training and prediction, and the reconstruction of SCB sequences



traditional LSTM models. To address this issue, we have incorporated a Self-Attention mechanism. The outcomes from all time steps of the LSTM are subjected to a linear transformation. This transformation serves to simplify and align input dimensions for subsequent attention layers. Subsequently, these inputs are forwarded to the Self-Attention layer. The hidden states outputted at each epoch are linearly transformed with the query weight matrix, key weight matrix, and value weight matrix, resulting in corresponding query vectors, key vectors, and value vectors. These weight matrices are learned through training. Referring to (11), the Self-Attention mechanism calculates the dependencies between each hidden state and other hidden states using dot product operations. And apply a softmax function to obtain the weights on the values. Then, multiply these weights with the corresponding position's value vectors and sum them up, ultimately generating a weighted representation for that query. This approach allows the hidden states of each epoch to comprehensively consider all information, aiding in addressing the issue of information loss in the LSTM model. We employ residual connections and layer normalization after the Self-Attention layer. The purpose of this step is to prevent gradient vanishing. Specifically, the output of each sub-layer is given by $\text{LayerNorm}(x + \text{Attention}(x))$, where $\text{Attention}(x)$ represents the output of the attention layer. Following the propagation through multiple data layers, the ultimate SCB prediction is derived by applying a linear transformation, followed by inverse normalization and reverse differencing.

In the LSTM-Attention model, the LSTM layer effectively models dependencies within the time series, while the Self-Attention mechanism excels at capturing correlations across the sequence on a global scale. The model attains an exemplary equilibrium between capturing local and global

details by synergistically harnessing the capabilities of both the LSTM and Self-Attention layers.

Experiments and analysis

In order to comprehensively validate the performance of the LSTM-Attention model in predicting the complex characteristics of BeiDou satellite SCB, we carefully selected representative BeiDou satellites and conducted a thorough experimental analysis. Specifically, we focused on BeiDou second-generation satellites C06 and C14, as well as BeiDou third-generation satellites inclined geostationary orbit (IGSO) satellite C39, medium earth orbit (MEO) satellite C20 (Rubidium atomic clock), and MEO satellite C29 (Hydrogen atomic clock). Through a comprehensive performance evaluation of predictions involving different types of satellites, we gained a comprehensive understanding of the applicability of the LSTM-Attention model across various scenarios.

To ensure the quality and reliability of experimental data, we utilized the IGS post-processed SCB products provided by the NASA Crustal Dynamics Data Information System (CDDIS) as our data source. This data source is highly trusted for its authenticity, guaranteeing the credibility of our experimental results. In the experiments, the prediction target of our model was the SCB data on June 26, 2023. The data on the previous day (June 25, 2023) were used as the training. The data had a time interval of 30 s, covering a total of 5760 epochs.

We conducted training using the LSTM-Attention model and retained the final trained model. For subsequent predictions, we adopted a sliding window approach, leveraging the saved model to make consecutive predictions for multiple time intervals. Through a thorough analysis

and assessment of the model's predictive performance, we compared the actual post-satellite clock bias data with the model's predictions. In this evaluation, we utilized root mean square error (RMSE) and range error (RE) as metrics to gauge the predictive accuracy.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{pre,i} - x_i)^2} \tag{12}$$

$$d_{pre} = x_{pre,i} - x_i \tag{13}$$

$$RE = d_{premax} - d_{premin} \tag{14}$$

where $x_{pre,i}$ represents the predicted clock bias at the i -th epoch by the model, x_i denotes the actual clock bias at the i -th epoch as provided by the IGS, n represents the total number of epochs being predicted, and d_{pre} signifies the difference between the predicted clock bias and the actual clock bias.

Hyperparameter analysis

To comprehensively evaluate the influence of different window sizes on prediction accuracy, we selected specific BeiDou satellites for our preliminary experiments. These satellites included a BeiDou second-generation MEO satellite, C14. There were also two BeiDou third-generation satellites, C20 and C29, with distinct atomic clocks. Our experimental design encompassed different window sizes, including 30 epochs, 60 epochs, 120 epochs, and 240 epochs. Simultaneously, one full day of data consisting of 2880 epochs was employed for model training aimed at obtaining an optimized predictive model. In the experiments, we based our predictions on data from a selected day, using the data from the last time window as the input for the following day. The length of this time window precisely matched the size of the chosen moving window. Specifically, we focused on short-term (2-h) and medium-term (24-h) prediction scenarios to gain a comprehensive understanding of the model's predictive performance. To ensure the stability of our experiments,

Table 1 The RMSE over 2 h for different moving windows

PRN	Windows	RMSE in 2 h (ns)				Mean
C14	30	0.0665	0.0689	0.0685	0.0794	0.0708
	60	0.0715	0.0741	0.0847	0.0787	0.0773
	120	0.0799	0.3497	0.1504	0.1049	0.1712
	240	0.1823	0.2167	0.1364	0.1208	0.1641
C20	30	0.3237	0.2899	0.2778	0.2755	0.2917
	60	0.2333	0.2069	0.1796	0.2118	0.2079
	120	0.1448	0.1917	0.1779	0.1161	0.1576
	240	0.1040	0.0670	0.0809	0.0713	0.0808
C29	30	0.1202	0.1427	0.1047	0.1160	0.1209
	60	0.0951	0.0923	0.0798	0.0918	0.0898
	120	0.0723	0.0745	0.0786	0.0722	0.0744
	240	0.0733	0.0728	0.0731	0.0734	0.0732

Table 2 The RMSE over 24 h for different moving windows

PRN	Windows	RMSE in 24 h (ns)				Mean
C14	30	2.3145	2.4747	2.4575	1.7785	2.2563
	60	2.5646	1.8932	1.6819	2.7426	2.2206
	120	1.7696	2.1287	0.9253	1.3877	1.5528
	240	0.7774	0.8455	1.0413	1.1992	0.9659
C20	30	3.0920	2.6815	2.5347	2.5064	2.7037
	60	1.9906	1.6675	1.3311	1.7279	1.6793
	120	0.9013	1.4806	1.3096	0.5612	1.0632
	240	0.4403	0.6286	0.3888	0.5141	0.4930
C29	30	1.2746	1.3603	0.7988	0.9757	1.1024
	60	0.8688	0.8155	0.5393	0.8060	0.7574
	120	0.2219	0.3646	0.2094	0.2354	0.2578
	240	0.1194	0.1121	0.1164	0.1220	0.1175

we conducted four independent training runs and integrated the results of each run, as shown in Tables 1 and 2.

In the context of 24-h prediction results, we observed a consistent improvement in the predictive accuracy of the LSTM-Attention model as the size of the moving window increased. This trend was observed for both BeiDou second-generation satellite (C14) and BeiDou third-generation satellites (C20 and C29). This result implies that as the observation time increases, the model is capable of enhancing prediction accuracy by better capturing long-term dependencies and regularities within the time series data. In the case of a 2-h prediction, BeiDou second-generation satellites and BeiDou third-generation satellites exhibit distinct variation trends. For BeiDou second-generation satellites, with an increase in the moving window size, the model's prediction accuracy shows a slight decrease. However, the overall change is not substantial. This may imply that, for this type of satellite, short-term temporal patterns are more influenced by local features within the window, and longer moving windows may result in information confusion. Conversely, short-term predictions for BeiDou third-generation satellites demonstrate a different trend. As the moving window size increases, the model's prediction accuracy gradually improves. This indicates that information over a longer time may contribute to better predicting the clock bias of this type of satellite.

To assess the impact of training dataset size on model performance, we trained our models with different sizes of the training datasets. Various window sizes were also considered during modelling. In this experiment, the model was tasked with predicting SCB data for June 26, 2023. We performed four independent training runs and averaged the results, as detailed in Table 3.

From Table 3, it can be observed that with the increase in the training dataset size, there is no significant improvement in the forecast results. Overall, the prediction accuracy remains at a similar level. Based on experimental observations, we speculate that this may be due to the following reasons: Firstly, the closer the SCB data points are in proximity,

the stronger their feature correlation becomes. Secondly, although a more extensive training dataset can introduce more features, it may also introduce more noise, potentially impacting the model's performance. We observed that using a larger moving window improves predictions, which is consistent with prior results. Nevertheless, as the size of the moving window increases, the demand for computing resources and training time also grows. Therefore, considering prediction accuracy and computational resources, we opted for a sliding window size of 240 epochs and used one day's worth of data as the training dataset in our subsequent experiments. Prior to conducting these experiments, we also finalized the remaining hyperparameters for the LSTM-Attention model, which are detailed in Table 4.

Model performance analysis

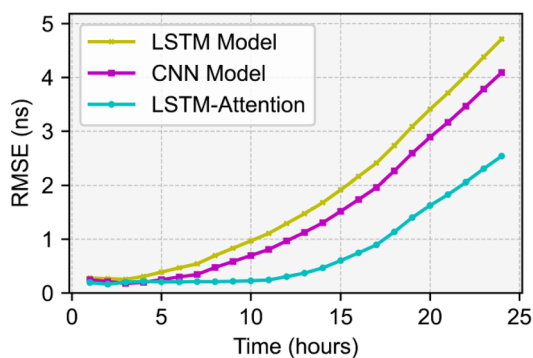
In order to comprehensively assess the predictive performance of the LSTM-Attention model, we selected samples from five different types of satellites for experimentation. The model construction was based on 24 h of data, and predictions were made for SCB data for the following 12 and 24 h. To ensure result stability, we conducted five independent model constructions and predictions and then took their average as the outcome. Currently, there have been studies applying CNN models to other time series predictions (Sayeed et al. 2021). Its multiple convolutional layers can progressively learn more abstract and advanced features within time series data. Given this advantage, we have decided to introduce CNN into the task of SCB prediction to evaluate its effectiveness in this specific domain and compare it with the LSTM-Attention model. Through this design, we aim to comprehensively understand the performance of different models in SCB prediction, providing a more holistic perspective for research. The predictive results of the LSTM-Attention model were compared with the performance of the CNN and LSTM models. In Fig. 5, we present RMSE of the predictions made by these three models for five different satellites, illustrating how their predictive accuracy varies with the prediction period.

Table 3 The RMSE over 24 h for different training datasets

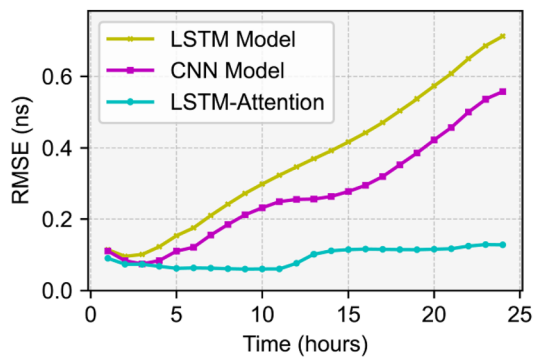
PRN	Windows	1 day data (ns)	2 days data (ns)	3 days data (ns)
C14	120	1.5528	1.5599	1.6230
	240	0.9659	1.0446	0.8083
	300	0.7874	0.8460	0.7912
C20	120	1.0632	0.6360	0.4159
	240	0.4930	0.4504	0.3849
	300	0.3633	0.3760	0.3841
C29	120	0.2578	0.2676	0.2338
	240	0.1175	0.1366	0.1136
	300	0.1178	0.1177	0.1303

Table 4 List of the LSTM-Attention model parameters

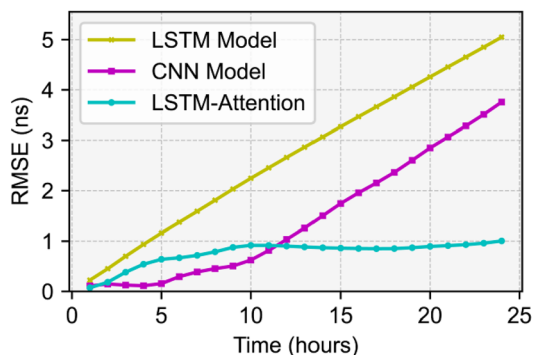
No	Parameter	Value
1	Loss function	MSE
2	Optimizer	Adam
3	Moving window size	240
4	Patience of early stopping	50
5	Learning rate	0.01
6	Batch size	1200
7	Number of data processing layers	2



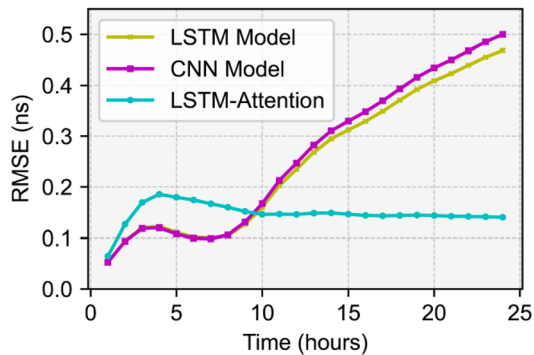
(a) RMSE for C06



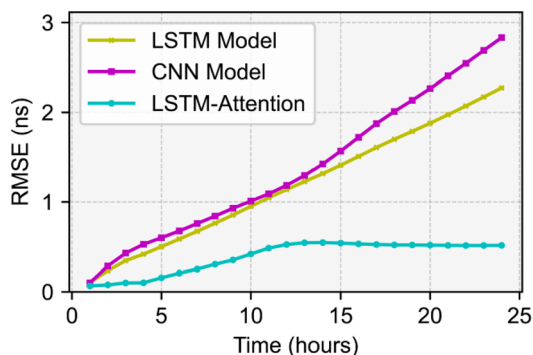
(d) RMSE for C29



(b) RMSE for C14



(e) RMSE for C39



(c) RMSE for C20

Fig. 5 The variations in prediction errors of three models with respect to the prediction period are examined. The top two figures represent BeiDou second-generation satellites (C06 and C14), while

the bottom three figures pertain to BeiDou third-generation satellites (C20, C29, and C39)

From Fig. 5, we can observe that in long-term prediction tasks, the prediction accuracy of the LSTM-Attention model surpasses that of the CNN and LSTM models. As the prediction horizon extends, the LSTM-Attention model exhibits relatively minor variations in prediction accuracy. This indicates higher prediction stability compared to the other two models. We have conducted specific performance

comparisons of these three models in 12- and 24-h prediction tasks, and the detailed results are presented in Tables 5 and 6, respectively.

Upon comparing the results presented in Tables 5 and 6, we can conclude that the LSTM-Attention model demonstrates superior predictive performance in both the 12- and 24-h prediction cycles. In the 12-h prediction task, the

Table 5 The accuracy statistics for 12-h forecasts of three models

PRN	CNN		LSTM		LSTM-Attention	
	RMSE (ns)	RE (ns)	RMSE (ns)	RE (ns)	RMSE (ns)	RE (ns)
C06	0.9685	2.4432	1.2921	2.9956	0.3034	1.3480
C14	1.0341	2.6715	2.6614	4.5524	0.9024	1.5135
C20	1.1856	2.1602	1.1387	2.1441	0.5254	1.1213
C29	0.2551	0.4730	0.3465	0.6784	0.0757	0.4413
C39	0.2469	0.7055	0.2349	0.6833	0.1458	0.4410

Table 6 The accuracy statistics for 24-h forecasts of three models

PRN	CNN		LSTM		LSTM-Attention	
	RMSE (ns)	RE (ns)	RMSE (ns)	RE (ns)	RMSE (ns)	RE (ns)
C06	4.0909	9.3427	4.7133	10.4409	2.5379	6.7741
C14	3.7574	8.0272	5.0492	9.0193	1.0051	3.2241
C20	2.8320	5.2938	2.2736	4.3992	0.5155	1.3186
C29	0.5572	1.1748	0.7128	1.4344	0.1276	0.5904
C39	0.5000	1.0211	0.4687	0.9724	0.1402	0.4834

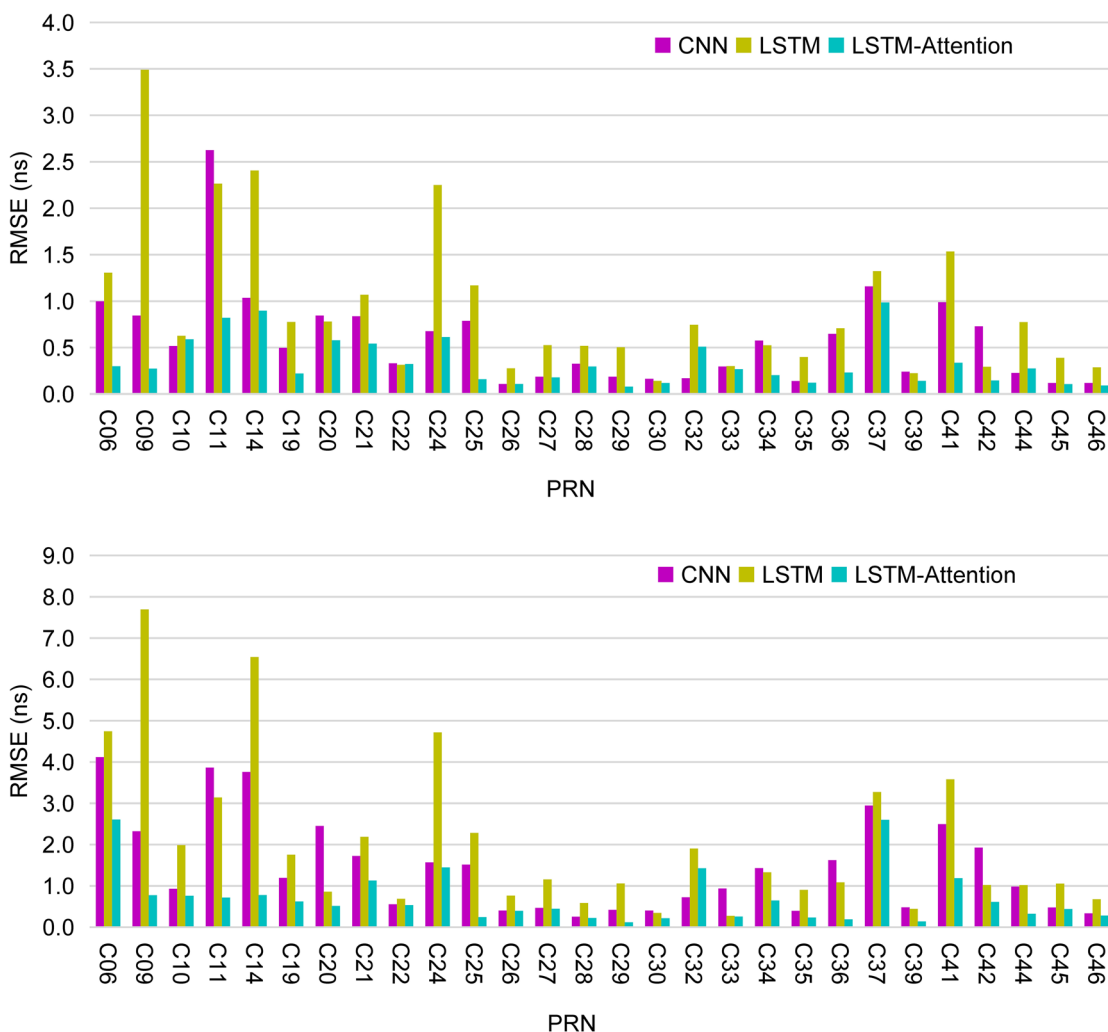


Fig. 6 RMSE histogram. The RMSE of the predictions for all BeiDou satellites by the three models over forecasting periods of 12 h (top) and 24 h (bottom)

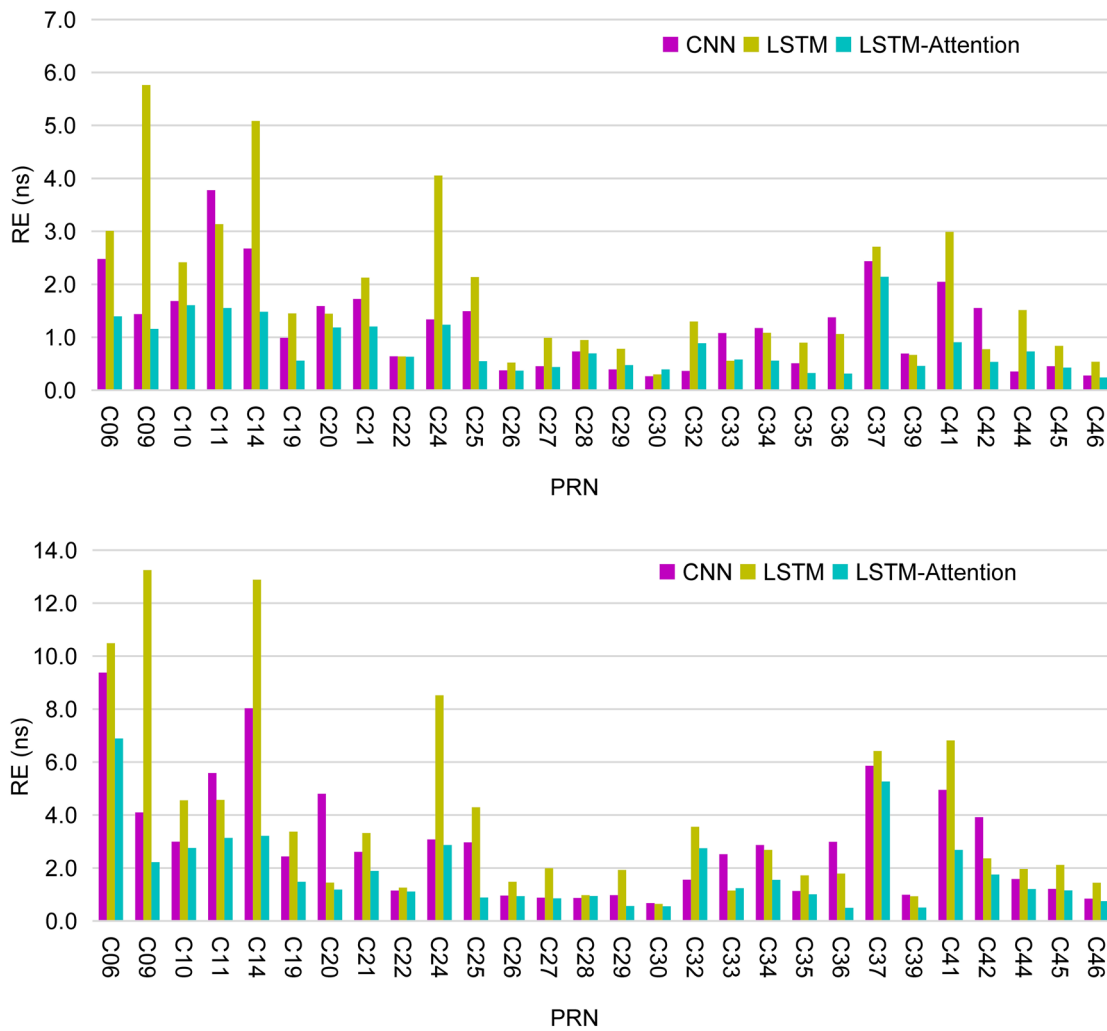


Fig. 7 RE histogram. The RE of the predictions for all BeiDou satellites by the three models over forecasting periods of 12 h (top) and 24 h (bottom)

LSTM-Attention model exhibits an accuracy improvement of 49.67 and 62.51% when compared to the CNN and LSTM models, respectively. In the 24-h prediction task, these improvements increase to 68.41 and 71.16%, respectively. Over longer prediction cycles, for both BeiDou second-generation and BeiDou third-generation satellites, the LSTM-Attention model yields smaller RMSE and RE in its predictions. In addition, through comparisons, it can be observed that the LSTM-Attention model exhibits higher prediction accuracy for BeiDou third-generation satellites compared to BeiDou second-generation satellites. For satellites in different orbits, such as IGSO satellite C39 and MEO satellites C20 and C29, the LSTM-Attention model consistently demonstrates outstanding predictive performance. Furthermore,

for satellites equipped with rubidium and hydrogen atomic clocks (C20 and C29), the LSTM-Attention model consistently yields lower RMSE values compared to the other two models. This highlights its superiority in SCB prediction across different clock types.

To comprehensively evaluate the predictive capabilities of the LSTM-Attention model, we extended our experiments to include a broader spectrum of BeiDou satellites. Validation was conducted for all BeiDou satellites available in the SCB files. Utilizing 24-h datasets for each satellite, we constructed LSTM-Attention models and conducted predictions for both the subsequent 12- and 24-h SCB data.

During the experimental process, we conducted comparisons between the predictions generated by the

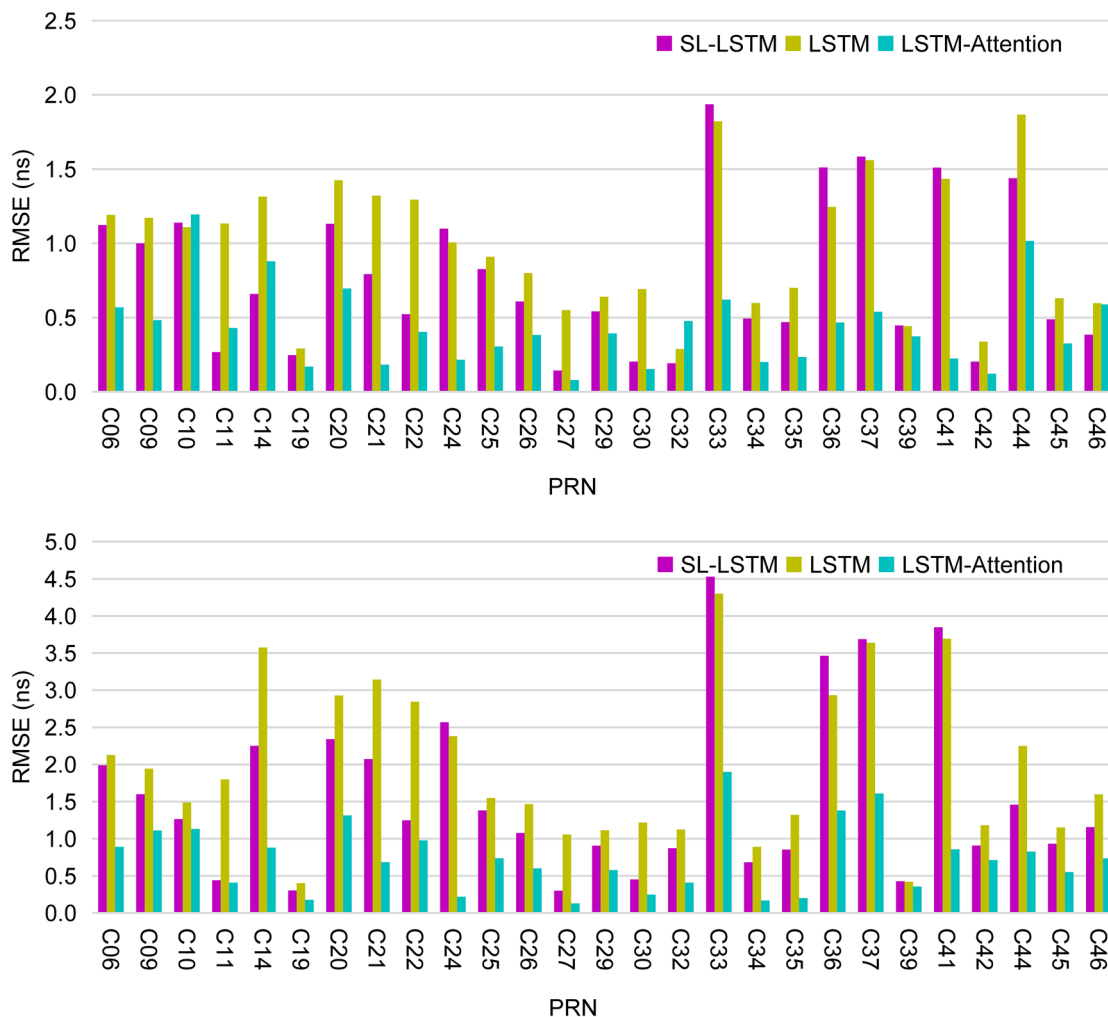


Fig. 8 RMSE histogram on November 1, 2021. The RMSE of the predictions for all BeiDou satellites by the three models over forecasting periods of 12 h (top) and 24 h (bottom)

LSTM-Attention model and those of two additional models. Figures 6 and 7, respectively, present these comparative results. Figure 6 illustrates the RMSE of predictions made by different models within 12- and 24-h forecasting intervals. From the figure, it can be observed that the LSTM-Attention model consistently achieves the lowest RMSE in most cases, indicating its superior accuracy in SCB prediction. To compare the RE of different models during the prediction process, Fig. 7 illustrates the RE for the three models. By comparing the RE, we gain a more comprehensive understanding of the performance of the LSTM-Attention model at different prediction time scales. The results from Fig. 7 clearly demonstrate that the LSTM-Attention model also exhibits certain advantages in terms of RE. This reaffirms its outstanding performance in prediction accuracy.

To reduce the randomness of experimental results, we replicated the above experiments using SCB data from

two additional days (October 31, 2021, and November 1, 2021). We used SCB data from the first day as the training set and data from the second day as the target for prediction. We compared the RMSE and RE of our experimental results with the LSTM model and the SL-LSTM model. The specific comparative results are shown in Figs. 8 and 9.

Figure 8 shows the results of the RMSE comparison. The LSTM-Attention model achieved better performance across all BDS satellites than other methods. Within a 12-h prediction time, the RMSE of the LSTM-Attention model is superior to the LSTM and SL-LSTM models for most satellites. In a few cases, the RMSE of the three models is at the same level. However, within a 24-h forecast time, the LSTM-Attention model demonstrates superior predictive performance. The forecasted RMSE for most satellites remains below 1 ns. From the experimental results shown in Fig. 9, we can observe that the predicted RE of the LSTM-Attention

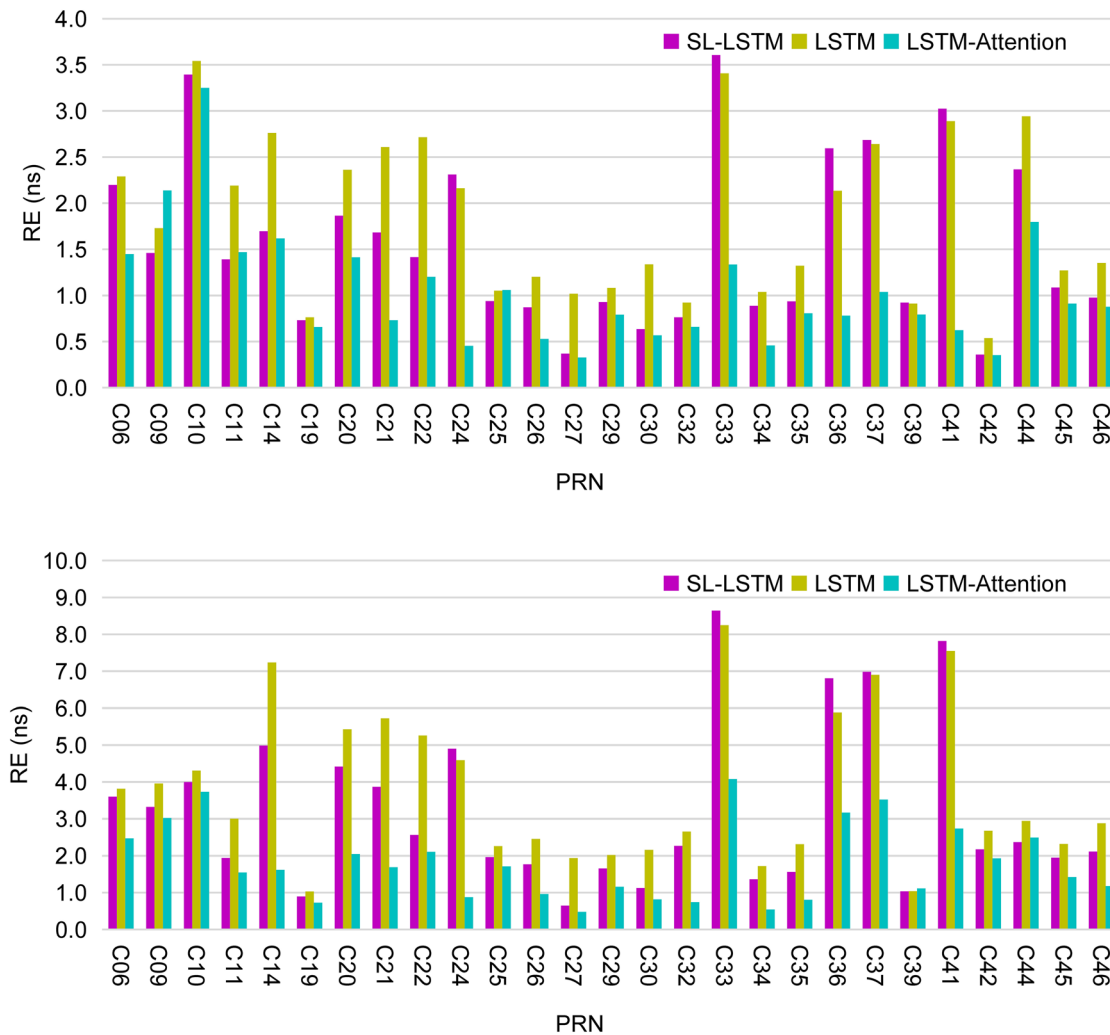


Fig. 9 RE histogram on November 1, 2021. The RE of the predictions for all BeiDou satellites by the three models over forecasting periods of 12 h (top) and 24 h (bottom)

model mostly remains below 2 ns within a 24-h prediction timeframe. Based on these experimental findings, it becomes evident that the LSTM-Attention model exhibits superior applicability and stability.

In summary, based on experimental validation involving multiple BeiDou satellites, we conclude that the LSTM-Attention model excels in the task of SCB prediction. Whether in a 12- or 24-h prediction time frame, its predictive performance exhibits notable improvements compared to the other two models.

Conclusions

This research leverages the LSTM-Attention model to forecast SCB, demonstrating its practical application in this context. We conducted a comprehensive evaluation of the

model's performance in predicting the intricate SCB of BeiDou in-orbit satellites through a series of experiments and analyses.

In our experimental investigations, we conducted comprehensive experiments encompassing multiple representative BeiDou satellites. These experiments entailed a comprehensive exploration of clock bias predictions, considering various satellite and atomic clock types. Through experiments with varying window sizes, we observed that the LSTM-Attention model tends to improve prediction accuracy as the window size increases in most cases. In short-term and medium-term prediction tasks, different satellite types exhibited distinct trends. These trends indicate variations in temporal characteristics among different satellite types in response to window size adjustments. Furthermore, we conducted a comprehensive performance comparison between the LSTM-Attention model

and CNN and LSTM models. The results demonstrate that over longer prediction horizons, the LSTM-Attention model exhibits superior predictive performance across different satellites and higher prediction stability. Notably, the LSTM-Attention model demonstrates a pronounced advantage in SCB prediction, particularly for satellites equipped with rubidium and hydrogen atomic clocks. Also, we investigated the predictive performance of the LSTM-Attention model across multiple BeiDou satellites. In comparison with other models, it exhibited lower RMSE in both the 12- and 24-h prediction tasks, concurrently demonstrating a certain advantage in RE. This validates the broad applicability of the LSTM-Attention model across various BeiDou satellites.

The average total time to process each satellite, i.e., train the model using 24 h of data and use the fitted model to predict SCB in the next 24 h, is about 5 min. The expediency and potential latency of SCB prediction depend on the hardware used and the amount of data. In our operating environment, equipped with a CPU I7 12700 K and a graphics card RTX3070, the prediction of SCB for the next 24 h takes approximately 5 min, which is considered acceptable.

However, this research has some limitations, such as the data coverage range and sample size. Subsequent research could expand the experimental datasets and explore the performance of the LSTM-Attention model with different satellite types and atomic clock types in more depth. This would further enhance its reliability and applicability in SCB prediction. Additionally, advanced model fusion strategies could be considered to further improve prediction accuracy and robustness.

Acknowledgements Not applicable

Authors' contributions CC and ML helped in conceptualization; CC and ML helped in methodology; ML worked in software; PL and ZL contributed to validation; ML and PL helped in data curation; ML and KL helped in the investigation; ML and ZL wrote the main manuscript text; CC, PL, and KL helped in writing—review and editing; CC and ML worked in project administration. All authors reviewed the manuscript.

Funding This work was supported by the National Key Research and Development Program of China (2020YFA0713501), the Hunan Provincial Innovation Foundation for Postgraduate under Grant (CX20220551), and the Xiangtan University Innovation Foundation for Postgraduate under Grant (XDCX2022Y084).

Data availability The experimental data in the manuscript are all public data and can be downloaded from <https://cdsis.nasa.gov/archive>.

Declarations

Competing interests The authors declare no competing interests.

Ethics approval Not applicable.

Consent for publication All authors gave their consent for the publication of this article.

References

- Bai H, Cao Q, An S (2023) Mind evolutionary algorithm optimization in the prediction of satellite clock bias using the back propagation neural network. *Sci Rep* 13(1):2095. <https://doi.org/10.1038/s41598-023-28855-y>
- Cernigliaro A, Sesia I (2012) INRIM tool for satellite clock characterization: frequency drift estimation and removal. *Mapan* 27(1):41–48. <https://doi.org/10.1007/s12647-012-0001-5>
- Cheng J, Dong L, Lapata M (2016) Long short-term memory-networks for machine reading. *arXiv preprint arXiv:1601.06733*. <https://doi.org/10.48550/arXiv.1601.06733>
- El-Mowafy A, Deo M, Kubo N (2017) Maintaining real-time precise point positioning during outages of orbit and clock corrections. *GPS Solutions* 21(3):937–947. <https://doi.org/10.1007/s10291-016-0583-4>
- Elseibeey M, Al-Harbi S (2016) Performance of real-time precise point positioning using IGS real-time service. *GPS Solutions* 20(3):565–571. <https://doi.org/10.1007/s10291-015-0467-z>
- Hadas T, Bosy J (2015) IGS RTS precise orbits and clocks verification and quality degradation over time. *GPS Solutions* 19(1):93–105. <https://doi.org/10.1007/s10291-014-0369-5>
- He S, Liu J, Zhu X, Dai Z, Li D (2023) Research on modeling and predicting of BDS-3 satellite clock bias using the LSTM neural network model. *GPS Solutions* 27(3):108. <https://doi.org/10.1007/s10291-023-01451-3>
- Huang GW, Zhang Q, Xu GC (2014) Real-time clock offset prediction with an improved model. *GPS Solutions* 18(1):95–104. <https://doi.org/10.1007/s10291-013-0313-0>
- Huang G, Cui B, Zhang Q, Fu W, Li P (2018) An improved predicted model for BDS ultra-rapid satellite clock offsets. *Remote Sensing* 10(1):60. <https://doi.org/10.3390/rs10010060>
- Huang B, Ji Z, Zhai R, Xiao C, Yang F, Yang B, Wang Y (2021) Clock bias prediction algorithm for navigation satellites based on a supervised learning long short-term memory neural network. *GPS Solutions* 25(2):80. <https://doi.org/10.1007/s10291-021-01115-0>
- Jonsson P, Eklundh L (2002) Seasonality extraction by function fitting to time-series of satellite sensor data. *IEEE Trans Geosci Remote Sens* 40(8):1824–1832. <https://doi.org/10.1109/TGRS.2002.802519>
- Liang YJ, Ren C, Yang XF, Pang GF, Lan L (2015) Grey model based on first difference in the application of the satellite clock bias prediction. *Acta Astronom Sinica* 56(3):264–277
- Lin Z, Feng M, dos Santos CN, Yu M, Xiang B, Zhou B, Bengio Y (2017) A structured Self-Attentive sentence embedding. *arXiv preprint arXiv:1703.03130*. <https://doi.org/10.48550/arXiv.1703.03130>
- Lu J, Zhang C, Zheng Y, Wang R (2018) Fusion-based satellite clock bias prediction considering characteristics and fitted residue. *J Navig* 71(4):955–970. <https://doi.org/10.1017/S0373463317001035>
- Lv D, Liu G, Ou J, Wang S, Gao M (2022) Prediction of GPS satellite clock offset based on an improved particle swarm algorithm optimized BP neural network. *Remote Sensing* 14(10):2407. <https://doi.org/10.3390/rs14102407>
- Malys S, Jensen PA (1990) Geodetic point positioning with GPS carrier beat phase data from the CASA UNO Experiment. *Geophys Res Lett* 17(5):651–654. <https://doi.org/10.1029/GL017i005p00651>

- Nie Z, Gao Y, Wang Z, Ji S, Yang H (2017) An approach to GPS clock prediction for real-time PPP during outages of RTS stream. *GPS Solutions* 22(1):14. <https://doi.org/10.1007/s10291-017-0681-y>
- Parikh AP, Täckström O, Das D, Uszkoreit J (2016) A decomposable attention model for natural language inference. arXiv preprint arXiv:1606.01933. <https://doi.org/10.48550/arXiv.1606.01933>
- Paulus R, Xiong C, Socher R (2017) A deep reinforced model for abstractive summarization. arXiv preprint arXiv:1705.04304. <https://doi.org/10.48550/arXiv.1705.04304>
- Qingsong AI, Tianhe XU, Dawei SUN, Lei REN (2017) The prediction of beidou satellite clock bias based on periodic term and starting point deviation correction. *Acta Geod Et Cartogr Sin* 45(2):132. <https://doi.org/10.11947/j.AGCS.2016.F034>
- Sayeed A, Lops Y, Choi Y, Jung J, Salman AK (2021) Bias correcting and extending the PM forecast by CMAQ up to 7 days using deep convolutional neural networks. *Atmos Environ* 253:118376. <https://doi.org/10.1016/j.atmosenv.2021.118376>
- Syam WP, Priyadarshi S, Roqué AAG, Conesa AP, Buscarlet G, Orso MD (2023) Fast and reliable forecasting for satellite clock bias correction with transformer deep learning. In: proceedings of the 54th annual precise time and time interval systems and applications meeting, Long Beach, California, pp. 76–96. <https://doi.org/10.33012/2023.18702>
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2023) Attention is all you need. arXiv preprint arXiv:1706.03762. <https://doi.org/10.48550/arXiv.1706.03762>
- Wang Y, Chen Y, Gao Y, Xu Q, Zhang A (2017a) Atomic clock prediction algorithm: random pursuit strategy. *Metrologia* 54(3):381. <https://doi.org/10.1088/1681-7575/aa6f62>
- Wang Y, Lu Z, Qu Y, Li L, Wang N (2017b) Improving prediction performance of GPS satellite clock bias based on wavelet neural network. *GPS Solutions* 21(2):523–534. <https://doi.org/10.1007/s10291-016-0543-z>
- Wang D, Guo R, Xiao S, Xin J, Tang T, Yuan Y (2019a) Atomic clock performance and combined clock error prediction for the new generation of BeiDou navigation satellites. *Adv Space Res* 63(9):2889–2898. <https://doi.org/10.1016/j.asr.2018.01.020>
- Wang L, Li Z, Ge M, Neitzel F, Wang X, Yuan H, (2019b) Investigation of the performance of real-time BDS-only precise point positioning using the IGS real-time service. *GPS Solutions* 23(3):66. <https://doi.org/10.1007/s10291-019-0856-9>
- Wang X, Chai H, Wang C, Xiao G, Chong Y, Guan X (2021) Improved wavelet neural network based on change rate to predict satellite clock bias. *Surv Rev* 53(379):325–334. <https://doi.org/10.1080/00396265.2020.1758999>
- Yu Y, Si X, Hu C, Zhang J (2019) A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput* 31(7):1235–1270. https://doi.org/10.1162/neco_a_01199
- Zhang L, Yang H, Yao Y, Gao Y, Xu C (2019) A new datum jump detection and mitigation method of real-time service (RTS) clock products. *GPS Solutions* 23(3):67. <https://doi.org/10.1007/s10291-019-0859-6>
- Zhao W, Liu G, Wang S, Gao M, Lv D (2021) Real-time estimation of gps-bds inter-system biases: an improved particle swarm optimization algorithm. *Remote Sensing* 13(16):3214. <https://doi.org/10.3390/rs13163214>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Chenglin Cai is a professor at Xiangtan University, Xiangtan, China. He earned his Ph.D. degree from the University of Chinese Academy of Sciences in 2010. His research interests encompass satellite navigation, precise point positioning, and wireless communications.



Mingyuan Liu graduated with a B.S. degree in Automation from Xiangtan University in 2022. He is currently pursuing an M.S. degree at the School of Automation and Electronic Information, Xiangtan University, Xiangtan, China. His current research mainly focuses on GNSS data processing and satellite navigation.



Pinchun Li graduated with a B.S. degree in Electronic Information Engineering from Xiangtan University in 2022. She is currently pursuing an M.S. degree at the School of Automation and Electronic Information, Xiangtan University, Xiangtan, China. Her current research mainly focuses on low-orbit satellite communication and navigation.



Zexian Li received the B.Eng. degree from the School of Information Engineering, Xiangtan University, Xiangtan, China, in 2019, where he is currently pursuing a Ph.D. degree. His research interests include computer vision and deep learning.



Kaihui Lv received his M.S. degree in Electronic Information from Xiangtan University, Xiangtan, China in 2023, where he is currently pursuing a Ph.D. degree. His research interests include precise single-point positioning and Real-time kinematic positioning.