



Granulation-based LSTM-RF combination model for hourly sea surface temperature prediction

Mengmeng Cao, Kebiao Mao, Sayed M. Bateni, Changyun Jun, Jiancheng Shi, Yongming Du & Guoming Du

To cite this article: Mengmeng Cao, Kebiao Mao, Sayed M. Bateni, Changyun Jun, Jiancheng Shi, Yongming Du & Guoming Du (2023) Granulation-based LSTM-RF combination model for hourly sea surface temperature prediction, International Journal of Digital Earth, 16:1, 3838-3859, DOI: [10.1080/17538947.2023.2260779](https://doi.org/10.1080/17538947.2023.2260779)

To link to this article: <https://doi.org/10.1080/17538947.2023.2260779>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 20 Sep 2023.



Submit your article to this journal [↗](#)



Article views: 161



View related articles [↗](#)



View Crossmark data [↗](#)



Granulation-based LSTM-RF combination model for hourly sea surface temperature prediction

Mengmeng Cao^{a,*}, Kebiao Mao^{a,b,c}, Sayed M. Bateni^d, Changhyun Jun^e, Jiancheng Shi^f, Yongming Du^b and Guoming Du^g

^aHulunbeir Grassland Ecosystem Research Station, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing, People's Republic of China; ^bState Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, People's Republic of China; ^cSchool of Physics and Electronic-Engineering, Ningxia University, Ningxia University, 750021, China; ^dDepartment of Civil and Environmental Engineering and Water Resources Research Center, University of Hawaii at Manoa, Honolulu, HI, USA; ^eDepartment of Civil and Environmental Engineering, College of Engineering, Chung-Ang University, Seoul, South Korea; ^fNational Space Science Center, Chinese Academy of Sciences, Beijing, People's Republic of China; ^gSchool of Public Administration and Law, Northeast Agricultural University, Harbin, People's Republic of China

ABSTRACT

Accurate predictions of sea surface temperature (SST) are crucial due to the significant impact of SST on the global ocean-atmospheric system and its potential to trigger extreme weather events. Many existing machine-learning-based SST predictions adapt the traditional iterative point-wise prediction mechanism, whose predicting horizons and accuracy are limited owing to the high sensitivity to cumulative errors during iterative predictions. Therefore, this paper proposes a novel granulation-based long short-term memory (LSTM)-random forest (RF) combination model that can fully capture the feature dependencies involved in the fluctuation of SST sequences, reduce the cumulative error in the iteration process, and extend the prediction horizons, which includes two sub-models (adaptive granulation model and hybrid prediction model). They can restack the one-dimensional SST time-series into multidimensional feature variables, and achieve a strong forecasting ability. The analysis shows that the proposed model can achieve more accurate prediction-hours in nearly all prediction ranges from 1 to 125 h. The average prediction error of the proposed model in 25–125 h is 0.07 K, similar to that (0.067 K) in the first 24 h, which exhibits a high generalization performance and robustness and isthus a promising platform for the medium- and long-term forecasting of hourly SSTs.

ARTICLE HISTORY



Received 22 March 2023
Accepted 13 September 2023

KEYWORDS

SST prediction; adaptive granulation method; LSTM; RF; error reciprocal method

1. Introduction

The sea surface temperature (SST), which is the water temperature at the interface between the ocean and atmosphere, is a key variable in the global air–sea system (Cao et al. 2021). This variable can be used to extract the ocean surface currents and location and strength of oceanic fronts, among

CONTACT Kebiao Mao  maokebiao@caas.cn  Hulunbeir Grassland Ecosystem Research station, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, People's Republic of China
*Authors Mengmeng Cao and Kebiao Mao contributed equally to this work and should be considered co-first authors.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

other information valuable for fisheries and allied aquacultures (Takahashi and Kawamura 2005). Furthermore, the SST plays an important role in the exchange of heat, moisture, momentum, and gases at the air–sea interface. Variations in the SST can considerably influence the regional and global climate and may even lead to extreme weather and climate events (e.g. droughts, heat waves, typhoons, and floods) (Bentamy et al. 2017; Minnett et al. 2019; Xiao et al. 2019). Therefore, future projections of the SST based on historical observations are of significance in the early warning of extreme events, study of climate dynamics, and planning of various offshore operations.

In general, the prediction of SST is challenging because of the large variations in the heat flux, radiation, and wind patterns near the sea surface (Patil, Deo, and Ravichandran 2016). The SST has been conventionally predicted using physics-based methods (i.e. numerical methods) or data-driven methods. Representative examples of the former are coupled ocean–atmosphere general circulation models and modular ocean models, which involve sophisticated mathematical models that couple kinetic and thermal equations to capture SST variations based on physical conditions and processes (Griffies and Greatbatch 2012; Noh et al. 2002). However, the results often involve uncertainties because these models typically involve multiple physical hypotheses and require a large amount of exogenous data input (Xu et al. 2020a). Because multiple parameters are simultaneously predicted, it is challenging for these models to precisely tune a single parameter (Patil and Deo 2018). In addition, physics-based models are engineered to provide average information in a large spatial area and are therefore better suited for obtaining predictions over large spatial regions with coarse resolutions than those in a specific location (Patil, Deo, and Ravichandran 2016; Xiao et al. 2019). In contrast, data-driven methods predict the SST by learning patterns and mining useful information from historical data. These methods are less sophisticated than physics-based numerical models and can be used to predict the SST in specific locations. Various types of data-driven methods, ranging from traditional statistical approaches to machine learning techniques, have been established. Traditional statistical approaches for SST forecasting include empirical canonical correlation analysis (Collins, Reason, and Tangang 2004; Tang et al. 2000), Markov models (Xue and Leetmaa 2000), linear regression (Kug et al. 2004), and empirical orthogonal functions (Neetu et al. 2011). Statistical approaches have excellent tractability in that they can address multiscale patterns of variations in the response variables in a probabilistic manner without requiring information regarding the physical mechanisms. Statistical approaches are especially advantages when information regarding the physical mechanisms of real-world processes is inadequate. With technological advancements, statistical approaches are being used to train artificial neural networks (ANNs) to achieve more accurate prediction results (Tangang, Hsieh, and Tang 1997). Compared with most linear statistical models, the nonlinear SST prediction model established using a multi-layer neural network can more effectively identify the trend and patterns of SST variations and exhibits a superior prediction performance (Aguilar-Martinez and Hsieh 2009; Wu, Hsieh, and Tang 2006). Variants of ANNs have been widely used to achieve excellent SST predictions, such as nonlinear autoregressive neural networks (Patil et al. 2013), wavelet neural networks (Patil and Deo 2017), ordinary feedforward neural networks (Aparna, D’Souza, and Arjun 2018), deep learning neural networks (Sarkar, Janardhan, and Roy 2020) and a type of recurrent neural network (RNN) named the long short-term memory (LSTM) network (Usharani 2022). Notably, although RNNs are effective at time-series prediction, they cannot capture the long-term dependencies involved in historical SSTs owing to the exploding and vanishing gradient problems. In contrast, LSTMs can store information of long time periods owing to the recurrent structure and gating mechanism and thus represent one of the most advanced prediction models for time-series-related problems (Chao et al. 2018).

Despite the potential and widespread application of LSTMs in different tasks, the existing LSTM-based methods, whose inputs pertain to single or multiple temporal dimension data, cannot gauge the internal dependencies of periodic features, for example, in situations in which the SSTs exhibit periodic characteristics (Shao et al. 2022; Yang et al. 2018; Yu et al. 2020). Bouktif et al. (2020) optimized the LSTM model based on metaheuristic search algorithms, enabling it to learn long-term

dependencies in the time series more accurately and improve the accuracy of predictions. However, these methods have been rarely applied to SST prediction. Furthermore, several LSTM-based SST predictions usually perform H-step ahead forecasting using the direct strategy and multiple-output strategy (Broni-Bedaiko et al. 2019; Chao et al. 2018; Xu et al. 2020b; Zhang et al. 2017). In the direct strategy, independent predictions are made for each horizon. Consequently, the predicted values may represent a broken curve, leading to a low forecast accuracy (Ben Taieb et al. 2012). Moreover, this strategy requires a considerable amount of calculation time because the number of prediction models to be learned is as large as the size of the horizon. In the multiple-output strategy, a multiple-output forecasting model is learned from historical SST observations, and the predictions are returned in one step by the forecasting model. Although this framework can eliminate the conditional dependency assumption involved in the direct strategy, the prediction model structure is consistent for all the horizons, which may result in a low forecast accuracy and reduced flexibility of the prediction method. These problems can be effectively avoided using the recursive strategy, in which predictions are produced iteratively by feeding back the output of the model as part of the inputs. However, the performance of these frameworks is limited owing to the high sensitivity to the cumulative errors during iterative predictions (Ben Taieb et al. 2012).

To address these gaps, this paper proposes a novel multiseries featural LSTM-RF combination model based on granular computing. The proposed model can fully exploit the periodic characteristics of SST sequences, reduce the iteration errors caused by the multi-step prediction process, and extend the prediction horizons. Validation against observations and cross-comparisons with state-of-the-art models were performed to demonstrate the reliability and suitability of the proposed prediction method for medium- and long-term hourly SST forecasting.

2. Study area and data

The SSTs retrieved from the Advanced Himawari Imager (AHI) were used as the data source. The AHI on the Himawari-8 geostationary meteorological satellite of Japan Meteorological Agency has 16 bands (3 visible (VIS) bands, 3 near-infrared (NIR) bands, and 10 infrared (IR) bands). Five of the infrared bands are used for near real-time SST retrieval. The AHI provides full disk images of the East Asia and Western Pacific regions as rapidly as every 10 min, with certain areas being observed more frequently at 2.5 min intervals, if required. Hourly SSTs from 2015/07/07 to 2021/12/31 (56,880 h) with a spatial resolution of 2 km were extracted from the website <https://www.eorc.jaxa.jp/ptree/index.html> (last access: 30 January 2022) and used for SST prediction.

Figure 1 shows the frequency of valid pixels in the hourly AHI SST data in 2021. Many missing pixels were observed, which could be attributed to atmospheric influences (Cao et al. 2021). We selected five pixels with few missing data during the study period (i.e. L1 to L5) as our study locations, as shown in Figure 1. Because the raw data of these five pixels also had missing points and outliers, we used the first-order difference method to eliminate outliers before establishing the dataset and then used the Savitzky–Golay filter and singular spectrum analysis-multitaper method (SSA-MTM) to reconstruct the missing values (Ghil et al. 2002; Vautard, Yiou, and Ghil 1992).

For each location, the first 80% of the prepared dataset (hourly SSTs from 2015/07/07 to 2020/08/31) was selected as the training data, and the remaining 20% was used as the test data, which covers the time period from 2020/09/01 to 2021/12/31. Furthermore, the 10% of the training dataset was split to serve as the validation set.

3. Model development

Figure 2 shows the working mechanism of the proposed granulation-based LSTM-RF combination model, which involves three stages. First, unequal-length temporal granules are generated by partitioning SSTs according to their semantic features. A template matching method is used to

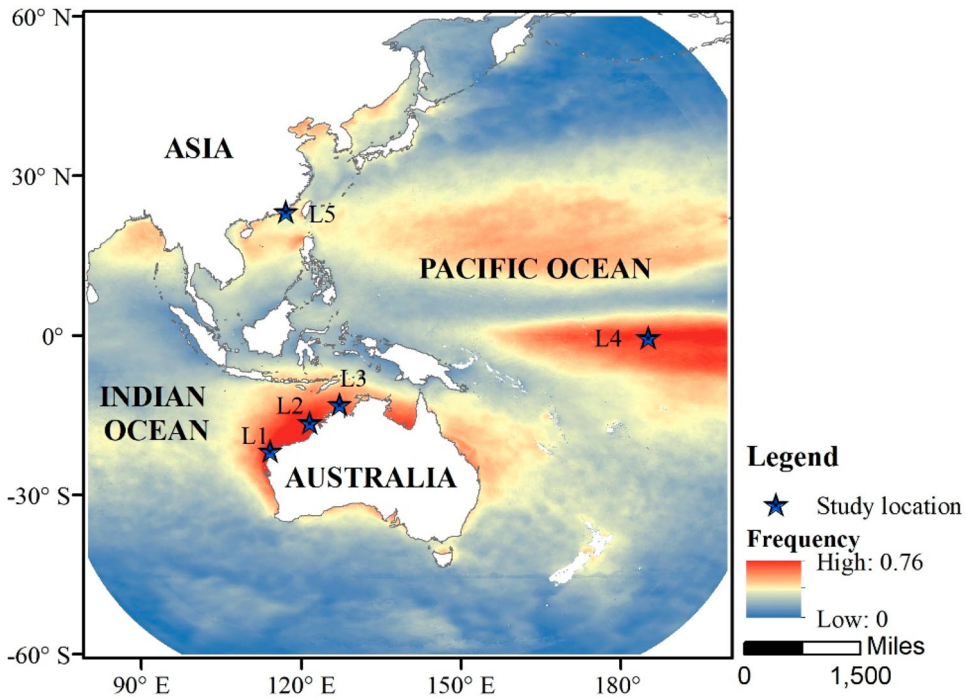


Figure 1. Frequency of valid pixels in the hourly AHI SST data from 01 January 2021 to 31 December 2021 and study locations. The coordinates (latitude, longitude) of the five selected locations are as follows: L1 (21.756°S, 114.124°E), L2 (16.349°S, 121.572°E), L3 (12.959°S, 127.104°E), L4 (0.458°S, 5.038°W), and L5 (23.316°N, 117.084°E).

adaptively extract specific periodic features of the information granules and restack the one-dimensional SSTs into multidimensional feature variables (Figure 2, Step 1). The extended feature variables are separately fed to the RF model and LSTM model, and their predictions are combined using the error reciprocal method. The combined prediction is fed back as part of the input sequence for the following prediction until the entire horizon has been predicted, and the resulting combined value is the final prediction for each horizon (Figure 2, Step 2). The SSTs are obtained by applying the degranulation process for these final predictions. The accuracy of the proposed prediction model is evaluated through a comparison with other models, i.e. the single LSTM and RF models, feedforward backpropagation neural network (BPNN) model, optimized support vector regression (SVR) model, and hybrid forecast model of LSTM and RF using the averaging strategy (A_LSTM-RF) (Figure 2, Step 3). Each step is described in the following subsections.

3.1. Formation of linguistic descriptors for data granules

3.1.1. Partitioning of the information granules

To granulate the original SST series X ($X = \{x_1, x_2, \dots, x_n\}$) according to the semantic features of the SSTs, the tendency feature of the data around x_i must be identified. The first-order dynamic X' ($X' = \{x'_1, x'_2, \dots, x'_{n-1}\}$) and second-order dynamic X'' ($X'' = \{x''_1, x''_2, \dots, x''_{n-2}\}$) of the SSTs, which are obtained by calculating the sequences of differences, show the monotonicity and concavity-convexity of the SST series, respectively.

$$x'_i = x_{i+1} - x_i \quad (i \in [1, 2, \dots, n-1]) \quad (1)$$

$$x''_i = x'_{i+1} - x'_i \quad (i \in [1, 2, \dots, n-2]) \quad (2)$$

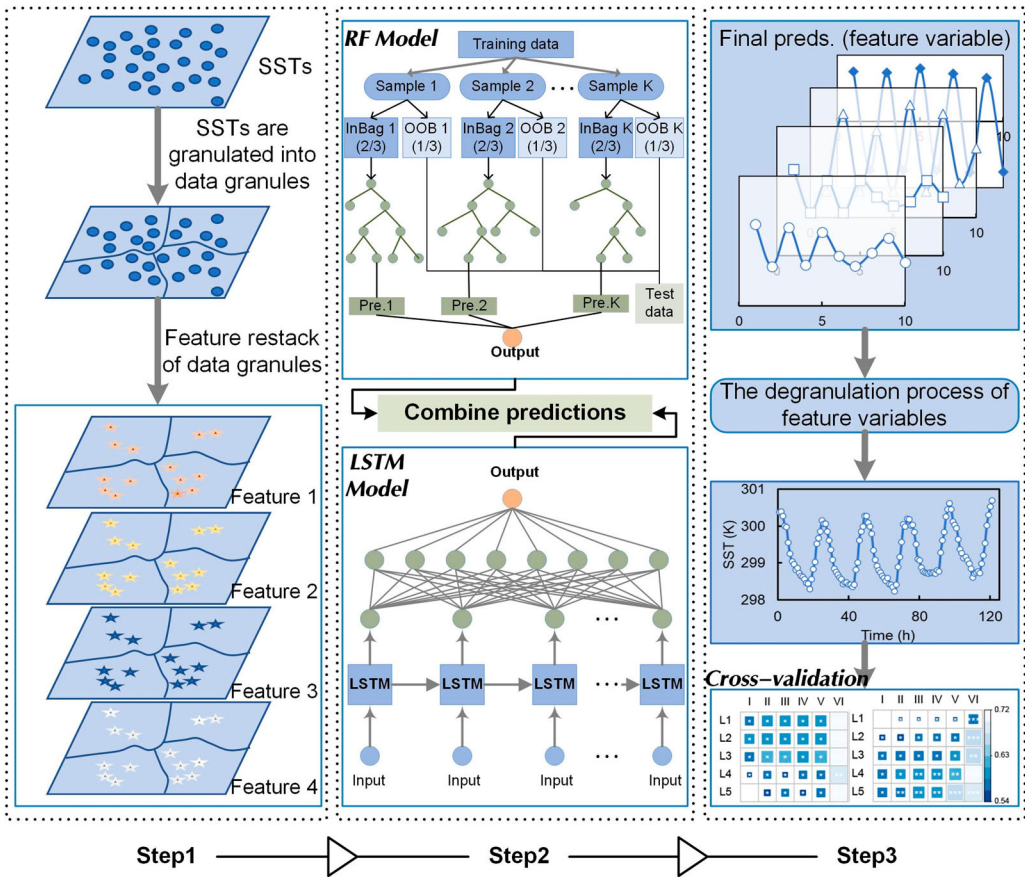


Figure 2. Visual representation of the proposed model. The cross-validation in Step 3 is based on the RMSE and (correlation coefficient) R^2 values of predictions obtained using different approaches (I, II, III, IV, V, and VI represent the proposed LSTM-RF, A_LSTM-RF, LSTM, RF, BPNN, and SVR methods, respectively). The color denotes the RMSE, and R^2 is represented by symbols: *** $R^2 < 0.985$, ** $0.985 < R^2 < 0.99$, * $0.99 < R^2 < 1$.

The monotonicity and concavity–convexity of the SST around x_i can be identified by the signs of $x'_i * x'_{i+1}$ and $x''_i * x''_{i+1}$. Subsequently, the SSTs are partitioned at the data point at which the sign changes. Specifically, given an SST series $X = \{x_1, x_2, \dots, x_t, x_{t+1}, \dots, x_n\}$, if $x'_i * x'_{i+1} < 0 \cup x''_i * x''_{i+1} < 0$, the monotonicity and/or concavity–convexity of the data point x_t change, and thus, the SST series X can be divided into $\{x_1, x_2, \dots, x_t\}$ and $\{x_{t+1}, \dots, x_n\}$. In addition, considering that the noisy characteristic of SSTs or small variations in the concavity–convexity may lead to dense partitioning, two constraints are added to each location in different months to address these problems through data analysis and multiple experiments: $|t - 1| > \theta_1 \cap |x_t - x_1| > \theta_2$, where θ_1 and θ_2 are the thresholds for each selected location in different months. In this manner, the SST series X is granulated into information granules with various trend characteristics.

3.1.2. Feature restacking of the information granules

To reflect the trend characteristics of the constructed unequal-length information granules and obtain an equalized representation, template matching is performed to describe the original granules and restack the one-dimensional SSTs of each selected location in the SST image into multi-series featural structure data. Specifically, quarter-cycle sine waves are used as templates, with each quarter-cycle sinusoid representing a typical combination of the concavity–convexity and

monotonicity. Subsequently, the template that has the same monotone and concave–convex property is stretched with appropriate horizontal and vertical scales to match these unequal-length granules to extract different tendency feature such as the amplitude (a), template type (I–IV) and time-domain (d) characteristics (Figure 3(A) and (B)). Finally, a 3-D feature space represented as (amplitude, template type, time domain) is used to describe the original granules. Considering that the duration of granules presenting a straight-line segment ($x'_i \cup x''_i = 0$) is typically short and this phenomenon rarely occurs, separate templates are not added and the corresponding granules are joined with the previous one. The original SSTs can be approximated through the extracted tendency feature of the information granules, as shown in Figure 3(B) and (C), where the value of each approximation point can be calculated by stretching the corresponding matching points in the templates. The stretching of the matching point can be obtained using the following equation:

$$x = a \sin\left(\frac{\pi i}{2d} + \vartheta\right) + b \tag{3}$$

where x is the approximate value of the i th SST of the data granule, and b is the value of its starting SST point. The parameter ϑ is determined by the type of matching template, e.g. the values of ϑ for templates I, II, III and IV are $-\pi/2$, 0 , $\pi/2$ and π , respectively. The parameters a and d denote the amplitude and duration of the matching templates, respectively.

Notably, directly describing the original granules with the extracted tendency feature leads to information redundancy, which serves to reduce the precision of subsequent prediction modeling and increase the computational complexity. Furthermore, the analysis of unequal-length granules illustrates that the data granule of the upward trend is typically first concave and then convex, whereas the downward trend is first convex and then concave. Thus, we combine the adjacent unequal-length granules with different concavities and convexities but the same monotonicity to identify a duration D and amplitude A , as shown in Figure 3(B), where the sign of A denotes

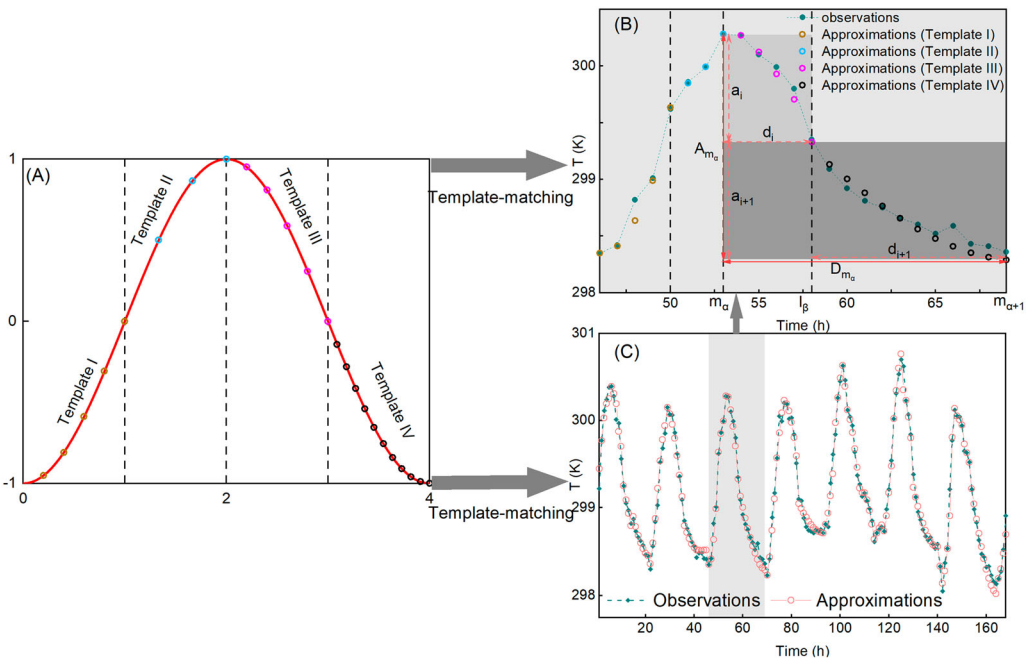


Figure 3. (A) Four quarter-period sinusoids used as templates. (B) Approximate SST values obtained through template matching, and feature description of the information granules. (C) Original SSTs and their approximation through the combination of stretched templates.

the monotonicity of the granules. Subsequently, the 3-D feature space consisting of the amplitude, template type, and time domain is replaced by a 4-D feature space $([A, D, R^1, R^2])$ consisting of the newly constructed amplitude (A), duration (D), curvature (R^1), and fluctuation (R^2) features. R^1 and R^2 can be calculated using the amplitude and duration of the original granules. An example of this calculation is shown in Figure 3(B). Suppose m_α and l_β are the time of monotonicity division and concavity–convexity division, respectively, that is, $m_\alpha \in \{i|x'_i * x'_{i+1} < 0\}$ and $l_\beta \in \{i|x''_i * x''_{i+1} < 0\}$. The 4-D feature space of the information granules between m_α and $m_{\alpha+1}$ can be represented by $[A_{m_\alpha}, D_{m_\alpha}, R^1_{m_\alpha}, R^2_{m_\alpha}]$, where $A_{m_\alpha} = x_{m_{\alpha+1}} - x_{m_\alpha}$ and $D_{m_\alpha} = m_{\alpha+1} - m_\alpha + 1$. $R^1_{m_\alpha}$ and $R^2_{m_\alpha}$ can be calculated using Equations 4 and 5, respectively.

$$R^1_{m_\alpha} = \frac{l_\beta - m_\alpha}{m_{\alpha+1} - m_\alpha} \tag{4}$$

$$R^2_{m_\alpha} = \frac{|x_{l_\beta} - x_{m_\alpha}|}{|x_{m_{\alpha+1}} - x_{m_\alpha}|} \tag{5}$$

3.2. Construction of LSTM-RF combination model

3.2.1. LSTM deep neural network model

LSTM is an RNN that learns the long-term dependencies between samples in a sequence by updating the state based on both the input for the current time step and network states of the output in the prior time step (Tsai, Kuo, and Tiwary 2020). LSTM involves repeating modules of neural networks, with each module composed of four interactive parts: a memory cell C , a forget gate f_t , an input gate i_t , and an output gate O_t , as shown in Figure 4.

As illustrated in the second repeating module in Figure 4, X_t is the input vector of the LSTM in which the gates f_t , i_t , and O_t and the candidate cell state C'_t are controlled by (X_t, h_{t-1}) . The cell state C_t is updated using f_t and i_t . O_t determines the amount of information propagated to the time step $t + 1$. These gates consist of a sigmoid fully connected neural network layer and point-wise multiplication operation. The working mechanism of these gates and information flow can be represented as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f) \tag{6}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i) \tag{7}$$

$$C'_t = \tanh(W_c \cdot [h_{t-1}, X_t] + b_c) \tag{8}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot C'_t \tag{9}$$

$$O_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o) \tag{10}$$

$$h_t = O_t \odot \tanh(C_t) \tag{11}$$

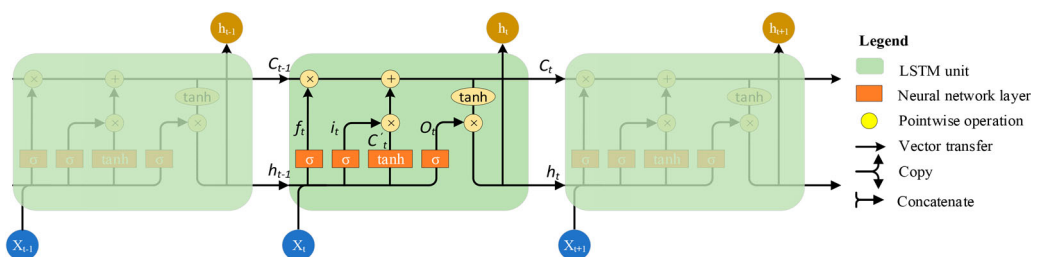


Figure 4. Structure of LSTM units.

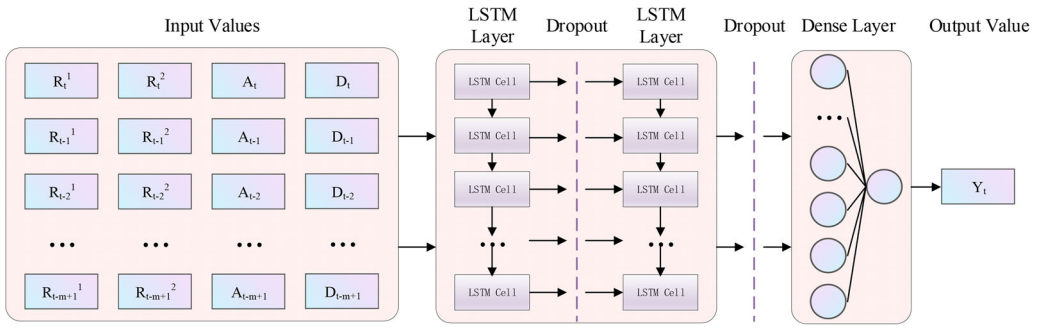


Figure 5. Proposed LSTM network for predicting the feature variables.

where the transformations σ from the inputs to i , f , and O are based on sigmoid functions. W and b contain the corresponding network weights and bias parameters, respectively. h represents the hidden states. The operation \odot is the element-wise multiplication (Hadamard product), and $\tanh(l)$ is the hyperbolic tangent function, which operates piecewise on each element of vector l .

The hyperparameters of an LSTM (e.g. the number of layers and number of neurons in each layer) must be tuned to improve its performance. In this study, the validation set was used to tune the LSTM hyperparameters via the Bayesian optimization method. We tested the performance associated with different numbers (1–5) of LSTM layers combined with 1, 2, and 3 dense layers (i.e. the fully connected layers). Considering the results, we selected a configuration with four layers (two LSTM layers and two fully connected layers) with 128, 128, 64, and 1 neurons in them, as shown in Figure 5. Batch normalization was introduced after each hidden layer in the network. The highest performance was obtained for a mini-batch size of 32. Moreover, to reduce overfitting during training, the dropout mechanism was applied to the inputs of the second LSTM layer and first fully connected layer, set as 0.3 and 0.2, respectively. The mean square error (MSE) was used as the loss function of the model. The Adam, RMSprop, AdaGrad, Nesterovs, stochastic gradient descent, and Adadelta optimization schemes were tested, and the Adam scheme was adopted as the optimizer. The proposed LSTM network was implemented using the Keras Python package on top of a TensorFlow backend.

In addition to the LSTM hyperparameters, the number of epochs and length of historical input data (timestep) affect the prediction of the multiseries featural structure data. We analyzed the variation in the magnitude of the loss function with the number of epochs in the LSTM model for A , D , R^1 , and R^2 predictions at different locations and selected the number of epochs in which the loss functions remained nearly constant (Table 1). The variations in the determination coefficient (R^2), mean absolute error, and root mean square error (RMSE) of the LSTM model for A , D , R^1 , and R^2 predictions at different locations with the timestep were examined, and the optimal timestep (N) for A , D , R^1 , and R^2 predictions at different locations was selected. Table 1 summarizes the optimal parameters.

3.2.2. RF ensemble learning model

Random forest regression (RFR) refers to an integrated learning algorithm that is useful for prediction problems. This algorithm applies bootstrap aggregation (bagging) and random feature selection

Table 1. Optimal parameters for the LSTM model for A , D , R^1 , and R^2 predictions at different locations.

Location	A		D		R^1		R^2	
	Epoch	N	Epoch	N	Epoch	N	Epoch	N
L1	289	62	291	56	272	47	264	69
L2	277	91	244	66	231	66	281	34
L3	257	58	300	48	286	64	232	73
L4	287	72	223	45	297	55	207	51
L5	274	79	212	48	294	59	283	63

to individual decision and/or regression trees to produce a final prediction with reduced variance and improved robustness (Breiman 2001; Hastie, Tibshirani, and Friedman 2009; Luo et al. 2021).

The RFR algorithm works by growing K different trees randomly, each built on a different subset that is generated using the bootstrapped method from the training dataset. The process of growing a decision tree starts at the root node of the tree, with the values of a single feature (predictor variable) used to split the bootstrapped subset into two smaller subsets that are as homogeneous as possible (i.e. sets of y (outcome variable) observations with the smallest variance among them). These smaller subsets are further divided into progressively smaller sets of homogeneous observations until either the number of data points in a node is smaller than S (i.e. the minimum number of data points to split) or the variance between y in a node falls below a pre-specified parameter. To ensure that the algorithm does not always choose the same feature when each split is performed, the feature is selected from a different random subset with J features in each node. The average of all the tree outcomes is used to assess the prediction.

The following hyperparameters must be tuned for the RF algorithm: number of regression trees (K), minimum number of data points to be split (S), minimum number of data entries at a leaf node (L), and number of randomly selected features to choose from at each node split (J). We exhaustively evaluated the possible combinations of settings for predicting the multiseres featural structure data, using the grid search method combined with five-fold cross-validation. The combination of hyperparameters that yield the highest predictive performance on the training data are retained for each predicted variable. Table 2 presents the optimal hyperparameter settings.

To further enhance the prediction performance, we evaluated the importance of each historical predictor in predicting the multiseres featural structure data. We sequentially fed predictors to the RF according to their importance rankings, until the RF explanatory power (r^2 between the predictions and the true sources) in the out-of-bag (OOB; similar to drop-one bootstrapping) test did not increase significantly, and the OOB error (MSE between the predictions and true sources) did not decrease significantly. Different numbers (le) of historical predictors were selected as the input of the final RF for predicting the four variables at five locations, as summarized in Table 2.

3.2.3. Combination of predictions from the LSTM and RF models

Previous studies have shown that hybrid prediction models tend to have smaller errors than single models (Deng et al. 2021; Xiao et al. 2019). In general, a complex LSTM framework can effectively model the long-term dependencies of SSTs but may be prone to overfitting. In contrast, the RF model has a strong predictive ability and does not easily overfit. Moreover, both models have different forecasting deviations. Therefore, to accurately predict the four variables (i.e. A , D , R^1 , and R^2) for all locations, we constructed a hybrid forecast model based on the LSTM and RF models. The error reciprocal method shown in Equation (12) was used to combine predictions from the LSTM deep neural network and RF models to construct the hybrid prediction model. This method assigns a higher weight to the model with a higher accuracy, thereby increasing the overall prediction accuracy.

$$y^i = \omega_l y_l^i + \omega_r y_r^i \tag{12}$$

$$\omega_l = \frac{\varepsilon_r}{\varepsilon_l + \varepsilon_r} \tag{13}$$

Table 2. Optimal hyperparameter setting for the RF model for A , D , R^1 , and R^2 predictions at different locations.

	A					D					R^1					R^2				
	K	S	L	J	Le	K	S	L	J	Le	K	S	L	J	Le	K	S	L	J	Le
L1	500	2	1	6	36	500	3	5	7	46	369	4	7	6	44	429	6	3	6	38
L2	466	4	2	5	32	500	2	7	6	41	500	2	2	6	37	444	4	5	6	41
L3	500	2	7	6	38	500	4	6	6	36	424	6	6	6	35	500	2	7	6	35
L4	497	3	5	6	42	496	5	7	7	47	500	4	4	7	47	500	3	6	7	47
L5	486	2	7	5	32	384	2	8	6	38	500	5	8	8	73	488	2	7	6	39

$$\omega_r = \frac{\varepsilon_l}{\varepsilon_l + \varepsilon_r} \quad (14)$$

where y^i is the final prediction of the feature variable to be predicted at different locations, and i denotes the index of different variables to be predicted at different locations ($i \in [1, 20]$). y_l^i and y_r^i are the predictions obtained based on the LSTM and RF models, respectively. ω is the weight coefficient. ε_r and ε_l are errors on the training samples of the LSTM and RF models, respectively.

The hybrid forecast model uses the iterative prediction scheme shown in Figure 6. The RF and LSTM models are trained to learn the pattern and relationship between the y^i at time t and input sequence (X) using historical data. The future predictions from the two single models are obtained based the learned patterns. Subsequently, their predictions on each horizon are combined using the error reciprocal method to generate the final prediction. Next, the input sequence is updated using the value as the latest element of the inputs, which is then used to forecast y^i for the next step. This process can be repeated to obtain k-steps-ahead predictions. In this study, we obtained 10-steps-ahead forecasts for each of the four variables at five locations.

3.3. Degranelation of the obtained feature variables

The final predictions of the hybrid forecast model are multidimensional structure data of granules. The SSTs can be obtained by degranulating the final predictions. The method used for degranulation relies on the templates used to describe the granules. In this study, we used quarter-cycle sinusoids as templates, and therefore, the SSTs were determined using the sinusoidal formula:

$$x = a \sin\left(\frac{\pi i}{2d} + \phi\right) + b \quad (15)$$

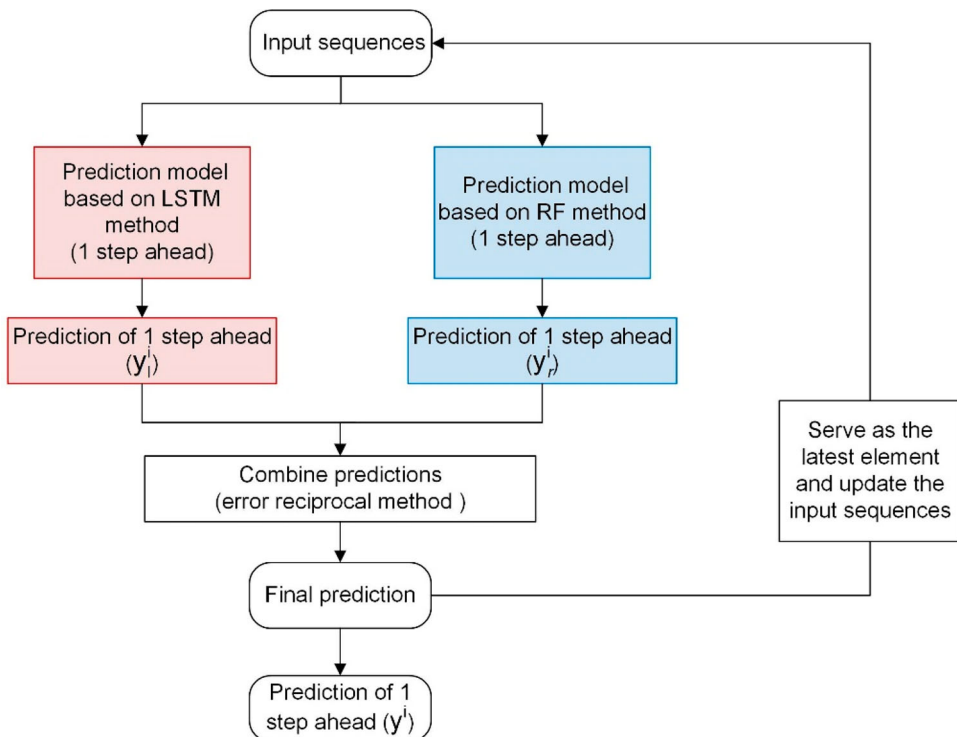


Figure 6. Architecture of the proposed hybrid forecast model based on the LSTM and RF approaches for predicting y^i .

where x is the i th SST of the data granule, and b is the value of its starting SST point. The parameter \emptyset is determined by the type of matching template, e.g. concave increasing and decreasing templates are used when \emptyset is 0 and $\pi/2$, respectively. a and d denote the amplitude and duration of the matching templates, respectively, which can be calculated using the final prediction of the feature variable. More specifically, a and d can be obtained according to

- ① if $(A < 0 \cap \emptyset = \frac{\pi}{2})$ or $(A > 0 \cap \emptyset = -\frac{\pi}{2})$, then $a = AR^2$, $d = DR^1$
- ② if $(A < 0 \cap \emptyset = \pi)$ or $(A > 0 \cap \emptyset = 0)$, then $a = A(1 - R^2)$, $d = D(1 - R^1)$.

4. Experimental results and discussion

4.1. Performance evaluation of the hybrid forecast model based on LSTM and RF

To investigate the performance of the proposed LSTM-RF combination model, we compared the prediction performance of the proposed hybrid forecast model and RF, LSTM, SVR, BPNN, and A_LSTM-RF models for the four variables using different statistical indicators and considering different perspectives. The grid search method was used to obtain the optimal hyperparameters of these algorithms. Predictions of the four variables at different prediction horizons were independently obtained for the five selected locations. All six approaches performed well in the 1-step ahead forecasting task of the four variables at five locations, as shown in Figure 7. The predictions obtained using the proposed LSTM-RF, A_LSTM-RF, RF, and LSTM models were more consistent with the actual values than those of the SVR and BPNN.

Figure 8 shows the statistical metrics for the 1-step-ahead forecasts of the test samples, obtained using different approaches. The proposed LSTM-RF combination method, based on the error reciprocal method, achieves the largest R^2 and lowest RMSE for the four variables at the five locations,

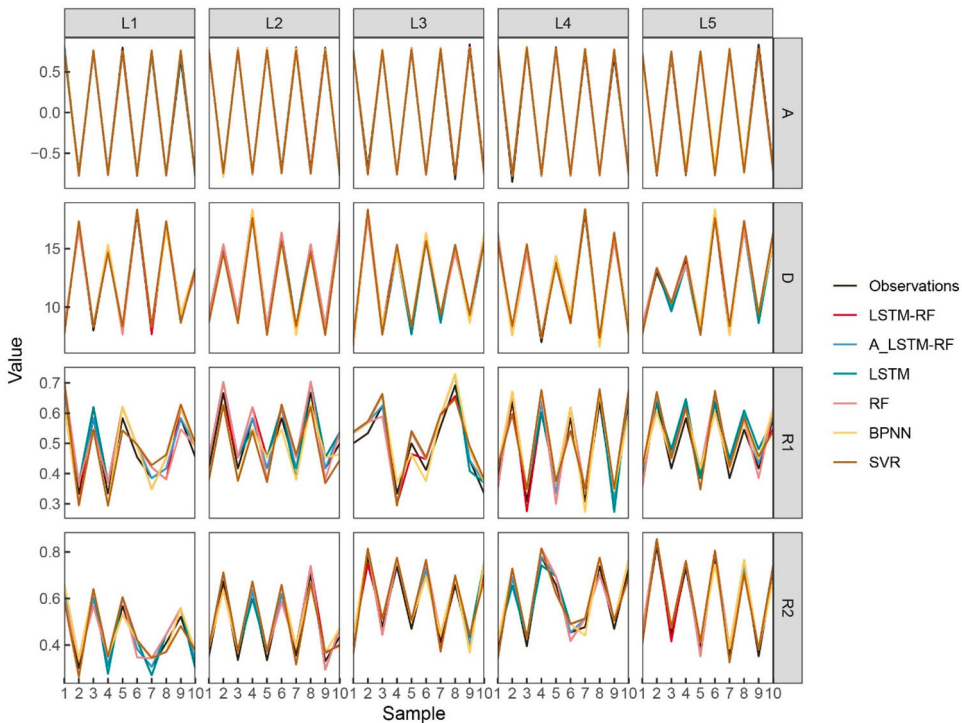


Figure 7. One-step-ahead prediction results of variables A, D, R^1 , and R^2 on the testing samples using the proposed LSTM-RF combination model and RF, LSTM, BPNN, SVR and A_LSTM-RF approaches at the five locations (L1–L5). LSTM-RF in all subsequent figures represents the proposed model, unless otherwise indicated.

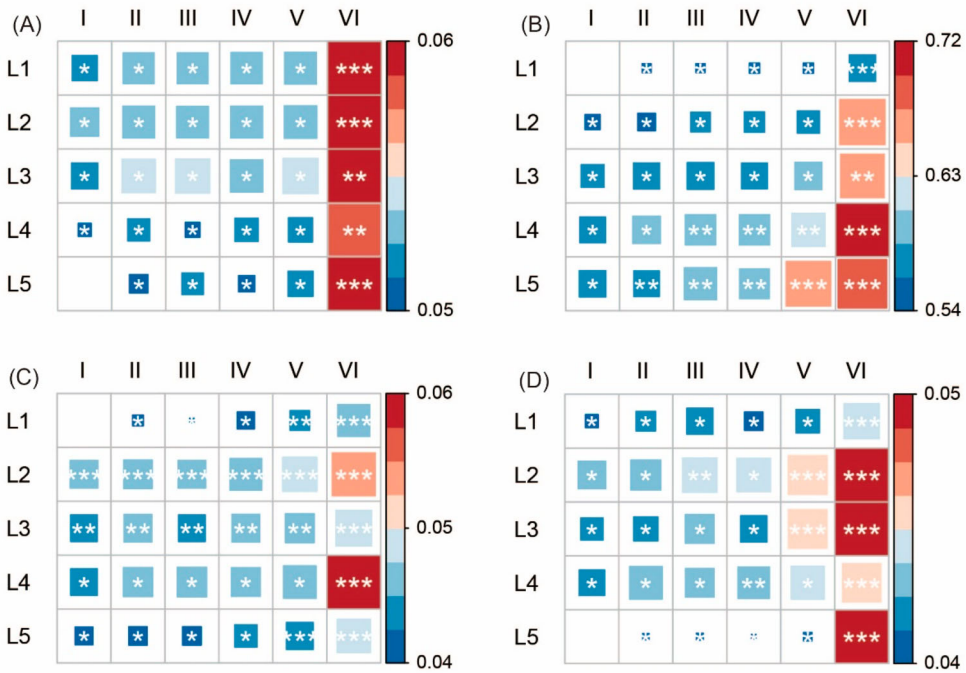


Figure 8. RMSE and correlation coefficients (R^2) values for the one-step-ahead prediction of variables (A) A, (B) D, (C) R^1 , and (D) R^2 on the testing samples using different approaches (I, II, III, IV, V, and VI denote the proposed LSTM-RF, A_LSTM-RF, LSTM, RF, BPNN, and SVR methods, respectively). The color and size of the square denote the RMSE (the blank white grids mean the smallest RMSE values), and R^2 values are represented by symbols: *** $R^2 < 0.985$, ** $0.985 < R^2 < 0.99$, * $0.99 < R^2 < 1$.

which performs much better than the RF, LSTM, SVR, BPNN, and A_LSTM-RF approaches in the 1-step-ahead prediction task. The proposed model performs well for most samples, although it does not guarantee the best accuracy for 1-step-ahead prediction for some samples. The SVR model exhibits the worst performance among all the models. In addition, the predictive performance of the LSTM was not always better than that of the RF model, i.e. the LSTM better predicted certain variables only at some locations, and the RF model achieved superior predictions at the other locations. This difference in the performance was our motivation for combining the LSTM and RF models. Although the A_LSTM-RF combined model also combines the advantages of the LSTM and RF models, the larger error in the averaging process inevitably reduces the overall prediction accuracy. Therefore, the prediction accuracy of this model was intermediate to those of the LSTM model and RF model in the one-step ahead prediction task. The reciprocal error method assigned a larger weight to the model with a smaller error, thereby achieving a prediction with a smaller error and enhancing the prediction accuracy.

Figure 9 shows the RMSE of the predictions for different prediction horizons on the held-out testing samples using the six methods. The RMSEs predicted by the proposed LSTM-RF and A_LSTM-RF were better than those predicted by the LSTM, RF, BPNN, and SVR methods over all prediction horizons for the different variables at five locations. The RMSEs associated with the proposed LSTM-RF were smaller than or equal to those pertaining to the A_LSTM-RF over all prediction horizons. Therefore, the hybrid forecast model based on LSTM and RF and using the reciprocal error method strategy outperformed the LSTM-RF hybrid model using the averaging strategy in short-term and medium-term feature variable prediction tasks. Furthermore, the prediction performance of the proposed model in the 3–10-step ahead prediction tasks was higher than that of the other four models (i.e. LSTM, RF, BPNN, and SVR). In other words, the proposed model effectively reduced the accumulated errors in the iterative prediction process. Tables 3–6 present the

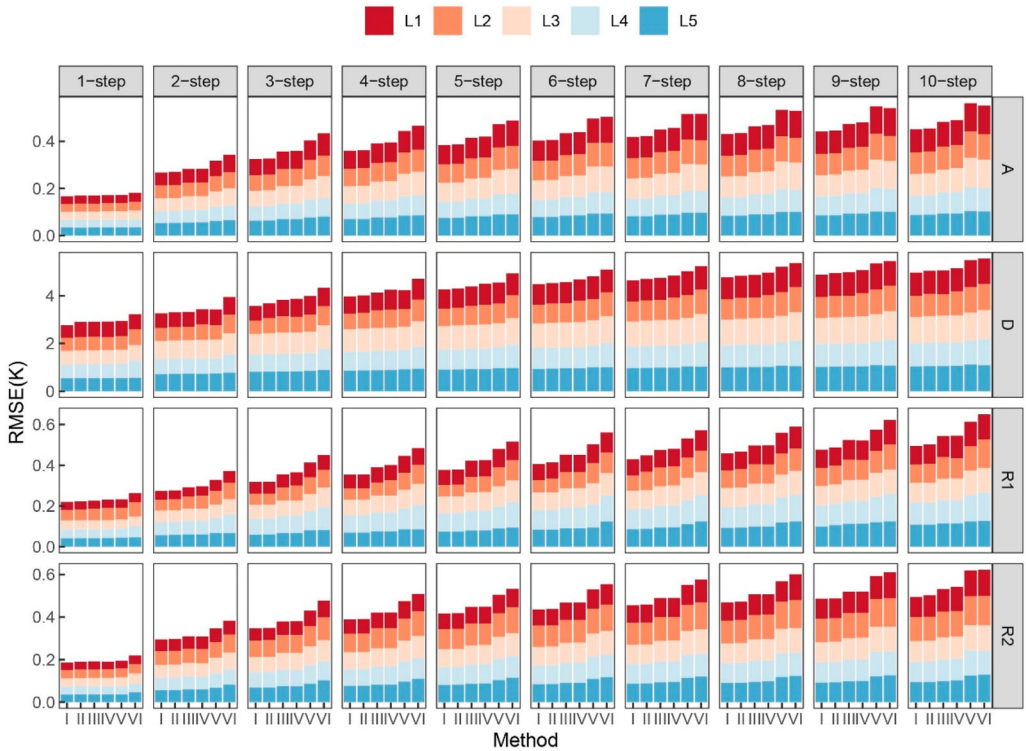


Figure 9. RMSEs of predictions for different variables using the proposed LSTM-RF (I), A_LSTM-RF (II), LSTM (III), RF (IV), BPNN (V), and SVR (VI) methods at five locations and different prediction horizons.

Table 3. Improvement percentages of the RMSE in variable A predicted by the proposed hybrid LSTM-RF model relative to that predicted by individual LSTM and RF models for each prediction horizon.

Relative to	Location	1-step	2-step	3-step	4-step	5-step	6-step	7-step	8-step	9-step	10-step
LSTM	L1	1.89	5.42	9.40	8.82	8.96	7.03	7.15	6.79	7.01	6.36
	L2	1.94	3.75	7.63	7.16	6.74	6.44	6.34	6.52	6.17	6.08
	L3	1.48	6.66	9.03	7.66	6.93	6.65	8.58	6.26	6.17	5.58
	L4	1.42	6.31	9.95	8.49	7.24	8.90	6.43	8.11	7.57	6.99
	L5	1.13	4.33	9.26	8.22	7.74	7.08	7.15	7.05	7.10	7.60
RF	L1	1.50	5.21	9.40	9.52	8.44	7.92	8.13	7.40	7.59	7.15
	L2	4.13	6.33	9.55	8.56	8.17	8.17	7.79	7.61	9.52	7.02
	L3	4.29	6.23	9.98	8.91	9.62	9.11	7.75	8.41	7.02	9.93
	L4	2.76	5.65	10.22	9.17	8.60	8.59	8.58	9.20	8.25	8.18
	L5	0.00	5.48	9.26	8.80	8.29	7.71	9.56	7.25	7.29	7.19

Table 4. Improvement percentages of the RMSE in variable D predicted by the proposed hybrid LSTM-RF model relative to that predicted by individual LSTM and RF models for each prediction horizon.

Relative to	Location	1-step	2-step	3-step	4-step	5-step	6-step	7-step	8-step	9-step	10-step
LSTM	L1	16.90	1.44	10.98	1.69	1.53	2.01	2.87	1.48	1.64	4.03
	L2	1.15	1.87	17.96	12.87	1.54	2.38	2.28	1.63	0.39	0.16
	L3	2.07	1.82	2.62	2.24	2.14	1.79	2.00	2.03	2.18	2.15
	L4	2.60	0.93	2.02	1.63	7.61	1.16	1.43	3.34	1.65	1.63
	L5	0.89	2.22	2.06	1.61	2.05	2.27	2.20	1.97	3.59	1.77
RF	L1	19.13	4.37	18.30	12.86	8.30	3.34	3.22	2.79	1.64	1.28
	L2	0.93	14.88	15.33	14.05	2.09	1.98	1.46	10.96	10.68	1.83
	L3	1.61	2.80	2.74	2.85	2.72	2.43	2.63	2.31	2.44	2.42
	L4	2.60	1.97	2.02	2.21	12.51	1.78	2.03	1.68	1.84	10.36
	L5	1.73	1.71	2.29	1.92	1.74	11.91	2.20	2.24	2.26	2.31

Table 5. Improvement percentages of the RMSE in variable R^1 predicted by the proposed hybrid LSTM-RF model relative to that predicted by individual LSTM and RF models for each prediction horizon.

Relative to	Location	1-step	2-step	3-step	4-step	5-step	6-step	7-step	8-step	9-step	10-step
LSTM	L1	0.29	9.67	10.57	8.31	7.12	6.95	7.19	7.19	7.06	11.30
	L2	7.98	4.37	15.32	7.57	27.88	6.94	20.41	6.44	11.31	6.06
	L3	0.03	4.54	8.71	8.03	7.08	6.62	6.22	6.03	6.75	6.62
	L4	2.11	5.05	8.64	7.89	6.80	6.63	6.75	11.98	6.64	16.26
	L5	0.39	6.37	10.24	15.05	7.21	25.96	8.57	6.91	13.21	5.21
RF	L1	4.07	7.12	10.63	8.81	7.35	23.92	7.24	7.92	8.12	17.59
	L2	12.67	15.69	8.53	26.64	27.05	7.01	21.86	6.91	7.01	8.84
	L3	0.83	4.32	8.71	8.45	8.29	7.51	7.21	6.64	7.33	7.41
	L4	1.59	4.73	26.36	8.17	7.80	7.35	7.56	6.91	7.10	6.71
	L5	2.58	6.55	10.56	7.77	8.06	6.96	10.56	10.23	13.54	5.21

Table 6. Improvement percentages of the RMSE in variable R^2 predicted by the proposed hybrid LSTM-RF model relative to that predicted by individual LSTM and RF models for each prediction horizon.

Relative to	Location	1-step	2-step	3-step	4-step	5-step	6-step	7-step	8-step	9-step	10-step
LSTM	L1	5.41	4.08	8.00	8.24	7.06	7.28	7.45	6.54	5.20	10.68
	L2	1.50	4.78	8.52	8.12	7.70	7.21	7.57	6.73	6.95	6.52
	L3	3.46	4.55	9.75	7.52	7.11	6.82	6.71	6.87	6.53	6.44
	L4	1.79	4.71	8.07	7.26	6.96	7.02	6.69	12.02	5.99	6.01
	L5	0.54	5.97	8.43	7.14	6.44	6.19	5.99	5.84	5.75	5.18
RF	L1	1.87	4.47	11.50	8.59	7.25	7.31	7.29	11.23	4.56	6.14
	L2	1.89	5.00	8.52	7.69	6.48	6.30	6.46	6.12	8.05	5.71
	L3	2.08	4.92	8.26	7.43	7.12	7.17	6.83	6.68	6.13	16.76
	L4	3.46	4.54	7.96	7.49	7.08	6.79	6.69	6.84	6.51	6.42
	L5	0.24	3.75	8.01	7.74	7.95	7.51	8.01	7.69	6.35	8.48

improvement percentages of the proposed hybrid LSTM-RF model over the individual LSTM and RF models in predicting the feature variables. The largest improvement relative to the LSTM and RF for each prediction horizon is boldfaced. The RMSEs of the forecasts of proposed hybrid LSTM-RF model were lower for nearly all locations and forecast horizons. In the 1–10 steps iterative prediction processes, the maximum improvements of the proposed model for predicting variables A, D, R^1 , and R^2 relative to the LSTM were 9.95%, 17.96%, 27.88%, and 12.02%, respectively, with means of 6.61%, 3.05%, 8.57%, and 6.51%. The corresponding values for the RF were 10.22%, 19.13%, 27.05%, and 16.76%, respectively, and the average improvements were 7.61%, 5.07%, 9.56%, and 6.71%.

4.2. Evaluation of the SST predictions

To demonstrate that the hybrid prediction model considering granular features can effectively extend the prediction horizon, a series of comparative experiments were conducted in which LSTM and RF were used to directly predict the SSTs. The parameters of these algorithms were determined through a grid search. The proposed hybrid prediction model considering granular features was established considering units of feature segments rather than individual SST values. Thus, the proposed prediction method could obtain 120–130 h ahead predictions of SST at locations L1–L5 after performing 10 steps of iterative prediction and degranulation. To fairly compare the SSTs predicted by the LSTM and RF models and those predicted by the proposed method, the number of iterations for both LSTM and RF was set to 130.

Figures 10–12 show the error statistics of the SST derived from three models according to the testing samples at the five locations. The proposed method outperformed the LSTM and RF models in the first 25 h of prediction (circles in Figures 10–12). The RMSEs and MAEs of SSTs predicted by the proposed method in the first 25 h were less than or equal to those of the LSTM and RF models at

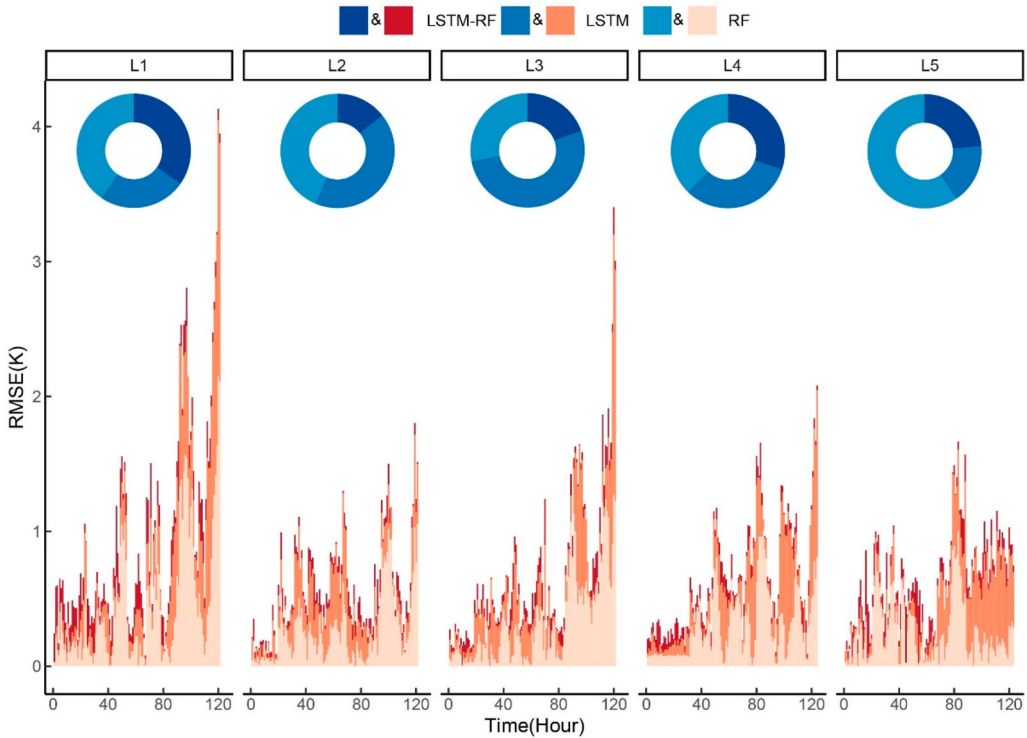


Figure 10. RMSE of SST predictions for different prediction horizons, obtained using the proposed LSTM-RF, LSTM and RF approaches at the five locations. The inset donut plots display the relative magnitude of the total RMSE for 1–25 h ahead predictions using the different methods.

all five locations. The correlation coefficients (R^2) of SSTs predicted by the proposed method in the first 25 h were greater than or equal to those of the LSTM and RF models at all five locations. With the increase in the prediction horizon, the RMSE and MAE of the proposed method became considerably lower than those of the LSTM and RF models and did not increase significantly. The correlation coefficients of the proposed method were also higher than those of the LSTM and RF models. Therefore, the proposed method outperformed the LSTM and RF models for all prediction horizons at all five locations. In addition, the proposed method consumed less time, and the number of iterations required for prediction was lower for the same prediction horizon.

Figure 13 investigates the prediction performance of different methods based on the differences between the predicted and observed values (error distribution). The predictions obtained using the proposed LSTM-RF model are closer to the actual values than those of the LSTM and RF models, indicating that the proposed method exhibits the highest prediction accuracy in each month at all five locations. To further assess the quality of SST predictions in different months, we calculated statistical indicators of the proposed method's SST predictions in different months at all locations, as well as statistical values of historical SST observations for different months (Figure 14). The statistical values of historical SST observations include the standard deviation of the averaged diurnal variation of SST for different months and the standard deviation of the diurnal variation of the amplitude of the diurnal cycle of SST. From Figure 14, it can be observed that the proposed model exhibits strong prediction performance for various months at all five locations, with a maximum RMSE and MAE of 0.3 and 0.19 K, respectively, and a minimum R^2 value of 0.98. The proposed method exhibits the highest prediction accuracy for location L4 and the lowest prediction accuracy for location L5 in 125-hour ahead predictions of SST.

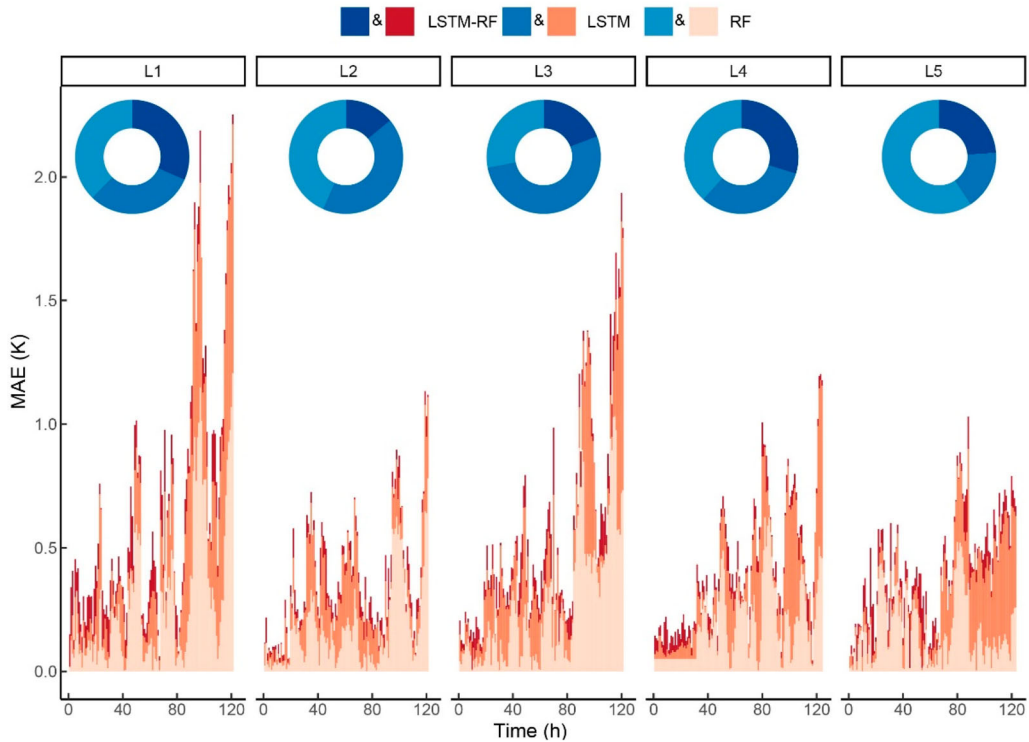


Figure 11. MAE of SST predictions for different prediction horizons, obtained using the proposed LSTM-RF, LSTM and RF approaches at the five locations. The inset donut plots display the relative magnitude of the total MAE for 1–25 h ahead predictions using the different methods.

Additionally, in terms of SST forecasts for different months, the proposed method exhibits the lowest prediction accuracy for locations L1, L2, L3, L4, and L5 in the months of November, November, October, September, and May, respectively. This discrepancy could be attributed to the varying standard deviations of SST observations and the standard deviations of the amplitude of the SST diurnal cycle at different locations. A smaller standard deviation suggests that the dataset contains fewer outliers or extreme values. By reducing the influence of noisy or anomalous data points, the model can focus on the main trends and patterns in the SST data, thereby enhancing prediction accuracy.

To further evaluate the capability of the proposed model in predicting SST variations at all five locations in different months, we conducted a comprehensive comparison of predictions and corresponding observations for all testing samples across prediction horizons ranging from 1 to 125 h. Figure 15 displays the results of this comparison. Due to space constraints, we randomly selected a sample with lower prediction accuracy from all the test samples for presentation. The LSTM and RF obtained accurate predictions in only the first 24–35 h owing to the accumulation and propagation of iteration errors with the increasing prediction horizon. As the number of iterations increased, the RF could not effectively follow the fluctuating trend of the SST observations at L2, L4, and L5, resulting in inaccurate predictions. Although the RF could follow the trend of SST observations better at L1 and L3, it could not track the amplitude and phase of the periodic fluctuation, resulting in large deviations in the predicted values near the peak. Compared with the RF, the LSTM was better at following fluctuation trends similar to real data; however, it could not capture the amplitude and phase of the data fluctuations and thus did not exhibit a stable prediction performance. Because the modeling process in this study focused on the main periodic

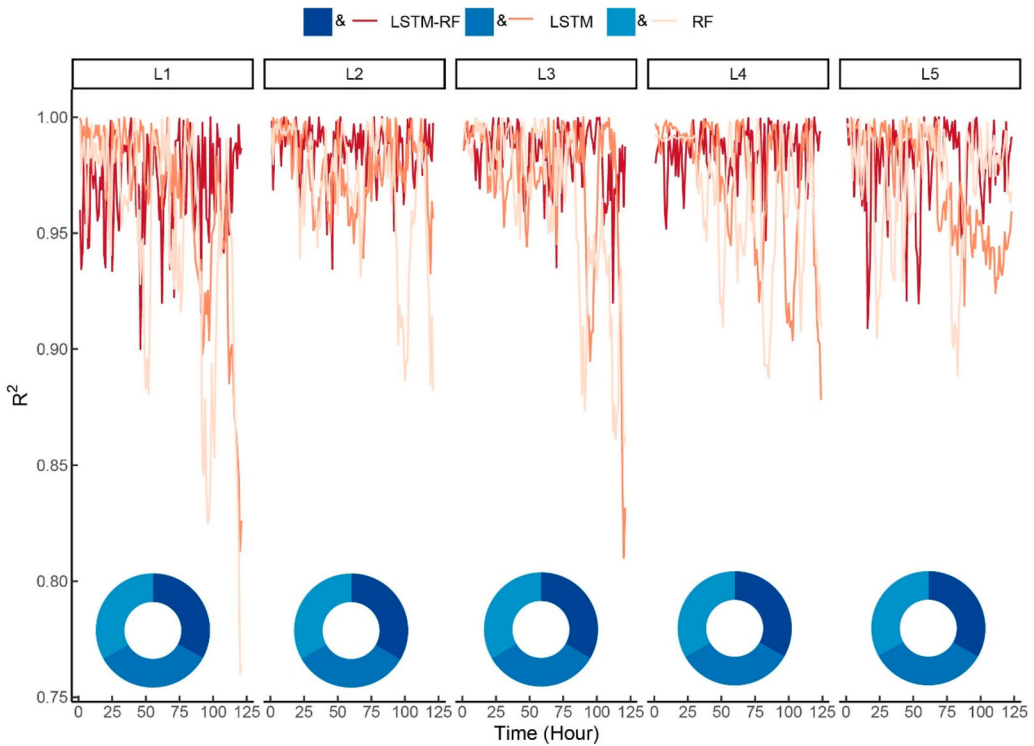


Figure 12. Correlation coefficients (R^2) of SST predictions for different prediction horizons, obtained using the proposed LSTM-RF, LSTM and RF approaches at the five locations. The inset donut plots display the relative magnitude of the total R^2 for 1–25 h ahead predictions using the different methods.

characteristics such as the fluctuation amplitude and duration, the forecasting results were closer to the fluctuation trend and amplitude of real data at different time scales. Therefore, the proposed approach achieved better forecasting results and was preferable for medium- and long-term forecasting.

5. Conclusions

The SST is an important geophysical parameter that considerably influences the ocean systems, climate change, and precipitation distribution, and may trigger extreme weather events such as droughts, typhoons, and floods. To achieve accurate long-term predictions of SST, this paper proposes an adaptive granulation method that comprehensively considers the concavity–convexity and monotonicity of the SST series to reconstruct the one-dimensional SSTs into multidimensional feature variables. Multiseries feature variable data are fed to the hybrid prediction model, and the feature-temporal pattern is used to obtain predictions.

We applied the proposed model to predict SST in various months at five locations with different latitudes. The results demonstrated that the proposed model successfully captured the seasonal information and exhibited high prediction accuracy for SST across different months at all five locations. By comparing with the other five models including RF, LSTM, SVR, BPNN, and A_LSTM-RF models, we found that the proposed method achieved more accurate predictions across nearly all prediction horizons from 1 to 125 h, with minimal increase in prediction error as the prediction horizon increases. The average prediction error of the proposed model within the range of 25–125 h is 0.07 K, which is comparable to the error of 0.067 K observed within the first 24 h. Overall, the proposed adaptive feature granulation-based prediction model that

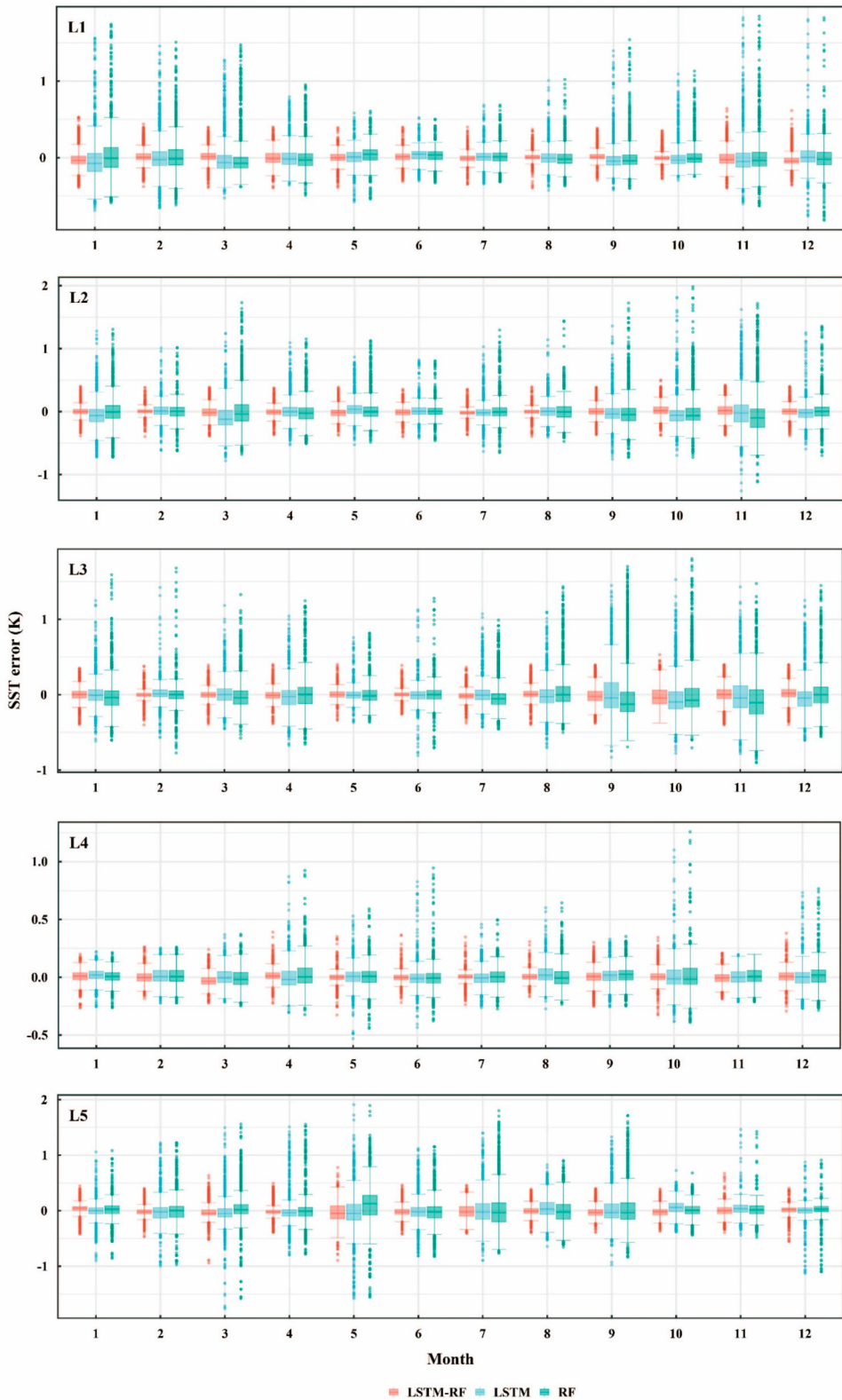


Figure 13. Prediction errors for all prediction horizons (1–125 h) using the proposed LSTM-RF, LSTM and RF methods at the five locations.

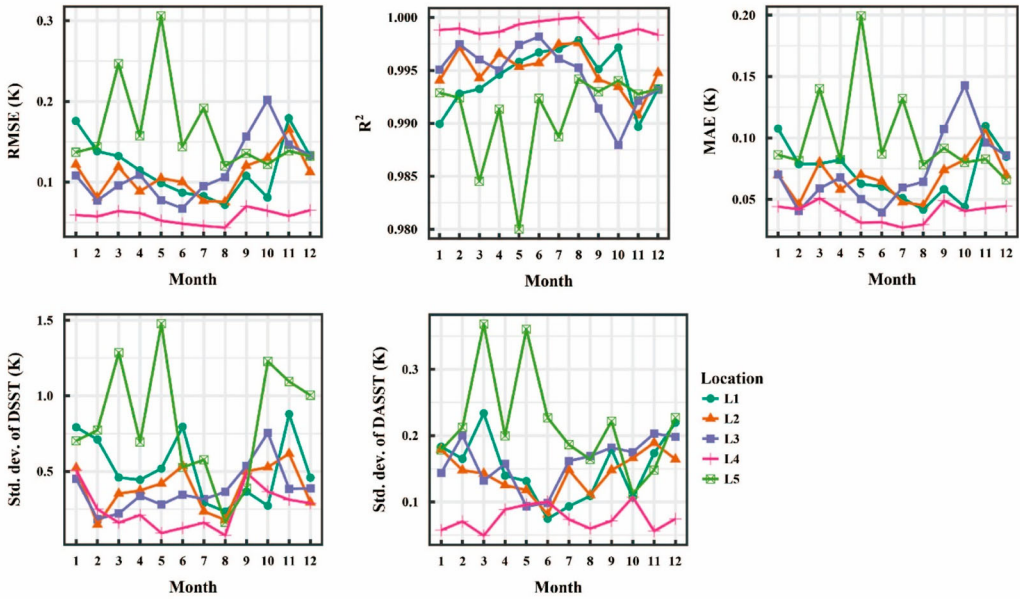


Figure 14. Values of RMSE, MAE and R^2 of SST predictions for different months, obtained using the proposed LSTM-RF approach at the five locations, as well as statistics of the SSTs for different months at the five locations. Std. dev. of DSST and td. dev. of DASST denote the standard deviation of the averaged diurnal variation of SST for different months and the standard deviation of the diurnal variation of the amplitude of the diurnal cycle of SST.

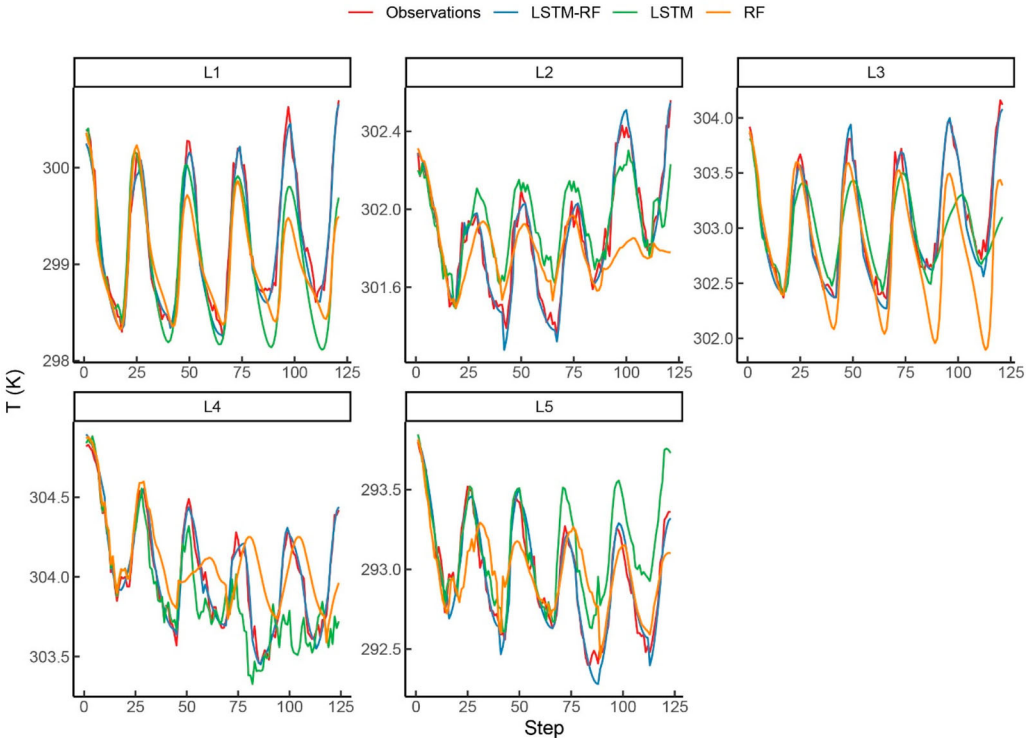


Figure 15. Prediction results using the proposed LSTM-RF, RF, and LSTM methods for the SST from 01 December to 07 December 2021 at five locations (L1–L5).

hybridizes the LSTM and RF models exhibits strong robustness and generalization ability, making it suitable for mid and long-term hourly SST forecasting. Notably, in this study, the multivariable analysis approach based on individual SST sequences was used. This technique can mine the multi-dimensional characteristic dependence involved in the time series and overcome the problem that the LSTM and RF models based on the iterative strategy are highly sensitive to the errors accumulated during iterative predictions. However, the variation in the SST resulting from the interaction of multiple variables cannot be considered. In this context, the prediction performance of the proposed model can be further enhanced by feeding it with meteorological factors with high spatial and temporal resolutions. This site-specific SST prediction can be used to fill the missing values of SST and aid in studying ocean dynamics, guiding fisheries decisions based on temperature preferences, and supporting effective coastal resource management by monitoring climate impacts and managing coastal ecosystems.

Acknowledgments

We are grateful to the three anonymous reviewers and the editor for their constructive comments and suggestions on this paper. We thank JAXA for providing the Himawari-8 AHI skin SST (<https://www.eorc.jaxa.jp/ptree/>) and the SSA-MTM Group for providing the SSA-MTM Toolkit (<http://research.atmos.ucla.edu/tcd/ssa/>).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Second Tibetan Plateau Scientific Expedition and Research Program (STEP) – ‘Dynamic monitoring and simulation of water cycle in Asian water tower area’ [grant number 2019QZKK0206]; Open Fund of the State Key Laboratory of Remote Sensing Science [grant number OFSLRSS202201]; Ningxia Science and Technology Department Flexible Introduction talent project [grant number 2021RXTDLX14] and Fengyun Application Pioneering Project [grant number FY-APP-2022.0205].

Author contributions

MC and KM designed the research and developed the methodology. MC, KM and Bateni contributed to the analysis and discussion of the results. MC drafted the manuscript, and all the authors revised the manuscript.

Data availability statement

The data that support the findings of this study are openly available in the Open Science Framework data repository.

References

- Aguilar-Martinez, Silvestre, and William W. Hsieh. 2009. “Forecasts of Tropical Pacific Sea Surface Temperatures by Neural Networks and Support Vector Regression.” *International Journal of Oceanography* 2009: 167239. <https://doi.org/10.1155/2009/167239>.
- Aparna, S. G., Selrina D’Souza, and N. B. Arjun. 2018. “Prediction of Daily Sea Surface Temperature Using Artificial Neural Networks.” *International Journal of Remote Sensing* 39 (12): 4214–4231. <https://doi.org/10.1080/01431161.2018.1454623>.
- Ben Taieb, Souhaib, Gianluca Bontempi, Amir F. Atiya, and Antti Sorjamaa. 2012. “A Review and Comparison of Strategies for Multi-Step Ahead Time Series Forecasting Based on the NN5 Forecasting Competition.” *Expert Systems with Applications* 39 (8): 7067–7083. <https://doi.org/10.1016/j.eswa.2012.01.039>.

- Bentamy, A., J. F. Piollé, A. Grouazel, R. Danielson, S. Gulev, F. Paul, H. Azelmat, et al. 2017. "Review and Assessment of Latent and Sensible Heat Flux Accuracy Over the Global Oceans." *Remote Sensing of Environment* 201: 196–218. <https://doi.org/10.1016/j.rse.2017.08.016>.
- Bouktif, Salah, Ali Fiaz, Ali Ouni, and Mohamed A. Serhani. 2020. "Multi-Sequence LSTM-RNN Deep Learning and Metaheuristics for Electric Load Forecasting." *Energies* 13 (2). <https://doi.org/10.3390/en13020391>.
- Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Broni-Bedaiko, Clifford, Ferdinand Apietu Katsriku, Tatsuo Unemi, Masayasu Atsumi, Jamal-Deen Abdulai, Norihiko Shinomiya, and Ebenezer Owusu. 2019. "El Niño-Southern Oscillation Forecasting Using Complex Networks Analysis of LSTM Neural Networks." *Artificial Life and Robotics* 24 (4): 445–451. <https://doi.org/10.1007/s10015-019-00540-2>.
- Cao, M., K. Mao, Y. Yan, J. Shi, H. Wang, T. Xu, S. Fang, and Z. Yuan. 2021. "A New Global Gridded Sea Surface Temperature Data Product Based on Multisource Data." *Earth System Science Data* 13 (5): 2111–2134. <https://doi.org/10.5194/essd-13-2111-2021>.
- Chao, Zeyi, Fangling Pu, Yuke Yin, Bin Han, and Xiaoling Chen. 2018. "Research on Real-Time Local Rainfall Prediction Based on MEMS Sensors." *Journal of Sensors* 2018: 6184713. <https://doi.org/10.1155/2018/6184713>.
- Collins, D. C., C. J. C. Reason, and F. Tangang. 2004. "Predictability of Indian Ocean sea Surface Temperature Using Canonical Correlation Analysis." *Climate Dynamics* 22 (5): 481–497. doi:10.1007/s00382-004-0390-4.
- Deng, H., F. Yan, H. Wang, L. Fang, Z. Zhou, F. Zhang, C. Xu, and H. Jiang. 2021. "Electricity Price Prediction Based on LSTM and LightGBM." Paper Presented at the 2021 IEEE 4th International Conference on Electronics and Communication Engineering (ICECE), 17–19 December 2021.
- Ghil, M., M. R. Allen, M. D. Dettinger, K. Ide, D. Kondrashov, M. E. Mann, A. W. Robertson, et al. 2002. "Advanced Spectral Methods for Climatic Time Series." *Reviews of Geophysics* 40 (1): 3-3-1-3-41. <https://doi.org/10.1029/2000RG000092>.
- Griffies, Stephen M., and Richard J. Greatbatch. 2012. "Physical Processes that Impact the Evolution of Global Mean sea Level in Ocean Climate Models." *Ocean Modelling* 51: 37–72. <https://doi.org/10.1016/j.ocemod.2012.04.003>.
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. "Random Forests." In *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, edited by Trevor Hastie, Robert Tibshirani, and Jerome Friedman, 587–604. New York, NY: Springer New York.
- Kug, Jong-Seong, In-Sik Kang, June-Yi Lee, and Jong-Ghap Jhun. 2004. "A Statistical Approach to Indian Ocean Sea Surface Temperature Prediction Using a Dynamical ENSO Prediction." *Geophysical Research Letters* 31 (9). <https://doi.org/10.1029/2003GL019209>.
- Luo, Xiangzhong, Trevor F. Keenan, Jing M. Chen, Holly Croft, I. Colin Prentice, Nicholas G. Smith, Anthony P. Walker, et al. 2021. "Global Variation in the Fraction of Leaf Nitrogen Allocated to Photosynthesis." *Nature Communications* 12 (1): 4866. <https://doi.org/10.1038/s41467-021-25163-9>.
- Minnett, P. J., A. Alvera-Azcárate, T. M. Chin, G. K. Corlett, C. L. Gentemann, I. Karagali, X. Li, et al. 2019. "Half a Century of Satellite Remote Sensing of sea-Surface Temperature." *Remote Sensing of Environment* 233: 111366. <https://doi.org/10.1016/j.rse.2019.111366>.
- Neetu, R. Sharma, S. Basu, A. Sarkar, and P. K. Pal. 2011. "Data-Adaptive Prediction of Sea-Surface Temperature in the Arabian Sea." *IEEE Geoscience and Remote Sensing Letters* 8 (1): 9–13. <https://doi.org/10.1109/LGRS.2010.2050674>.
- Noh, Yign, Chan Joo Jang, Toshio Yamagata, Peter C. Chu, and Cheol-Ho Kim. 2002. "Simulation of More Realistic Upper-Ocean Processes from an OGCM with a New Ocean Mixed Layer Model." *Journal of Physical Oceanography* 32 (5): 1284–1307. [https://doi.org/10.1175/1520-0485\(2002\)032<1284:SOMRUO>2.0.CO;2](https://doi.org/10.1175/1520-0485(2002)032<1284:SOMRUO>2.0.CO;2).
- Patil, Kalpesh, and Makaranand Chintamani Deo. 2017. "Prediction of Daily Sea Surface Temperature Using Efficient Neural Networks." *Ocean Dynamics* 67 (3): 357–368. <https://doi.org/10.1007/s10236-017-1032-9>.
- Patil, Kalpesh, and M. C. Deo. 2018. "Basin-Scale Prediction of Sea Surface Temperature with Artificial Neural Networks." *Journal of Atmospheric and Oceanic Technology* 35 (7): 1441–1455. <https://doi.org/10.1175/JTECH-D-17-0217.1>.
- Patil, Kalpesh, M. C. Deo, Subimal Ghosh, and M. Ravichandran. 2013. "Predicting Sea Surface Temperatures in the North Indian Ocean with Nonlinear Autoregressive Neural Networks." *International Journal of Oceanography* 2013: 302479. <https://doi.org/10.1155/2013/302479>.
- Patil, Kalpesh, M. C. Deo, and M. Ravichandran. 2016. "Prediction of Sea Surface Temperature by Combining Numerical and Neural Techniques." *Journal of Atmospheric and Oceanic Technology* 33 (8): 1715–1726. <https://doi.org/10.1175/JTECH-D-15-0213.1>.
- Sarkar, Partha Pratim, Prashanth Janardhan, and Parthajit Roy. 2020. "Prediction of Sea Surface Temperatures Using Deep Learning Neural Networks." *SN Applied Sciences* 2 (8): 1458. <https://doi.org/10.1007/s42452-020-03239-3>.
- Shao, Q., W. Li, G. Hou, G. Han, and X. Wu. 2022. "Mid-Term Simultaneous Spatiotemporal Prediction of Sea Surface Height Anomaly and Sea Surface Temperature Using Satellite Data in the South China Sea." *IEEE Geoscience and Remote Sensing Letters* 19: 1–5. <https://doi.org/10.1109/LGRS.2020.3042179>.
- Takahashi, Wataru, and Hiroshi Kawamura. 2005. "Detection Method of the Kuroshio Front Using the Satellite-Derived Chlorophyll-a Images." *Remote Sensing of Environment* 97 (1): 83–91. <https://doi.org/10.1016/j.rse.2005.04.019>.

- Tang, Benyang, William W. Hsieh, Adam H. Monahan, and Fredolin T. Tangang. 2000. "Skill Comparisons Between Neural Networks and Canonical Correlation Analysis in Predicting the Equatorial Pacific Sea Surface Temperatures." *Journal of Climate* 13 (1): 287–293. [https://doi.org/10.1175/1520-0442\(2000\)013](https://doi.org/10.1175/1520-0442(2000)013).
- Tangang, F. T., W. W. Hsieh, and B. Tang. 1997. "Forecasting the Equatorial Pacific Sea Surface Temperatures by Neural Network Models." *Climate Dynamics* 13 (2): 135–147. <https://doi.org/10.1007/s003820050156>.
- Tsai, Sun-Ting, En-Jui Kuo, and Pratyush Tiwary. 2020. "Learning Molecular Dynamics with Simple Language Model Built upon Long Short-Term Memory Neural Network." *Nature Communications* 11 (1): 5115. <https://doi.org/10.1038/s41467-020-18959-8>.
- Usharani, Bhimavarapu. 2022. "ILF-LSTM: Enhanced Loss Function in LSTM to Predict the Sea Surface Temperature." *Soft Computing*. <https://doi.org/10.1007/s00500-022-06899-y>.
- Vautard, Robert, Pascal Yiou, and Michael Ghil. 1992. "Singular-Spectrum Analysis: A Toolkit for Short, Noisy Chaotic Signals." *Physica D: Nonlinear Phenomena* 58 (1): 95–126. [https://doi.org/10.1016/0167-2789\(92\)90103-T](https://doi.org/10.1016/0167-2789(92)90103-T).
- Wu, Aiming, William W. Hsieh, and Benyang Tang. 2006. "Neural Network Forecasts of the Tropical Pacific Sea Surface Temperatures." *Neural Networks* 19 (2): 145–154. <https://doi.org/10.1016/j.neunet.2006.01.004>.
- Xiao, Changjiang, Nengcheng Chen, Chuli Hu, Ke Wang, Jianya Gong, and Zeqiang Chen. 2019. "Short and Mid-Term Sea Surface Temperature Prediction Using Time-Series Satellite Data and LSTM-AdaBoost Combination Approach." *Remote Sensing of Environment* 233: 111358. <https://doi.org/10.1016/j.rse.2019.111358>.
- Xu, Lingyu, Yifan Li, Jie Yu, Qin Li, and Suixiang Shi. 2020a. "Prediction of Sea Surface Temperature Using a Multiscale Deep Combination Neural Network." *Remote Sensing Letters* 11 (7): 611–619. doi:10.1080/2150704X.2020.1746853.
- Xu, Lingyu, Qin Li, Jie Yu, Lei Wang, Jiang Xie, and Suixiang Shi. 2020b. "Spatio-Temporal Predictions of SST Time Series in China's Offshore Waters Using a Regional Convolution Long Short-Term Memory (RC-LSTM) Network." *International Journal of Remote Sensing* 41 (9): 3368–3389. <https://doi.org/10.1080/01431161.2019.1701724>.
- Xue, Yan, and Ants Leetmaa. 2000. "Forecasts of Tropical Pacific SST and Sea Level Using a Markov Model." *Geophysical Research Letters* 27 (17): 2701–2704. <https://doi.org/10.1029/1999GL011107>.
- Yang, Y., J. Dong, X. Sun, E. Lima, Q. Mu, and X. Wang. 2018. "A CFCC-LSTM Model for Sea Surface Temperature Prediction." *IEEE Geoscience and Remote Sensing Letters* 15 (2): 207–211. <https://doi.org/10.1109/LGRS.2017.2780843>.
- Yu, Xuan, Suixiang Shi, Lingyu Xu, Yaya Liu, Qingsheng Miao, and Miao Sun. 2020. "A Novel Method for Sea Surface Temperature Prediction Based on Deep Learning." *Mathematical Problems in Engineering* 2020: 6387173. <https://doi.org/10.1155/2020/6387173>.
- Zhang, Q., H. Wang, J. Dong, G. Zhong, and X. Sun. 2017. "Prediction of Sea Surface Temperature Using Long Short-Term Memory." *IEEE Geoscience and Remote Sensing Letters* 14 (10): 1745–1749. <https://doi.org/10.1109/LGRS.2017.2733548>.