

Dear Shouyi Zhong,

Thank you very much for your careful review of our manuscript and for your insightful comments. We appreciate the time and effort you have dedicated to providing valuable feedback. The replies are presented below.

#### 1. Comparison with Random Forest Model

In our earlier work, we combined LSTM and RF models and observed that the standalone RF model yielded lower prediction accuracy compared to LSTM and the LSTM-RF hybrid model (as shown in Figure 1). Based on these findings, we selected LSTM as a baseline model for accuracy validation in our current study. During the precision validation phase, we did not specifically compare the results of RF against proposed model in this study, as LSTM was already designated as a benchmark model within our study.

Given the relevant literature proposing the use of ConvGRU for SST prediction, ConvGRU has been added as a baseline model in our comparative analysis to provide a more comprehensive evaluation of predictive performance. This allows for a direct comparison between ConvGRU and our proposed ConvLSTM model, enriching the assessment of our methodology.

#### 2. Determination of Step Size (Timestep)

To illustrate our optimization process, we use the two South China Sea sub-regions (Region 1 & Region 2) as representative examples:

We used the feature variables constructed in paper of  $t$  continuous time steps as input data to predict the feature variable images constructed based on SST for the next 10 consecutive steps. To ensure that the ConvLSTM model we constructed achieved the best predictions for regions 1 and 2, we analyzed the variation in the magnitude of the loss function with the number of epochs in the ConvLSTM model for region 1 and region 2 predictions and selected the number of epochs in which the loss functions remained nearly constant (as shown in Fig. 2). The number of epochs predicted for regions 1 and 2 was set to 264 and 192, respectively. We also examined the variations in R2, MAE, and RMSE of the ConvLSTM model for region 1 and region 2 predictions with the timestep (as shown in Fig. 3), and the optimal timestep for region 1 and region 2 predictions was set to 18 and 21, respectively. This approach accounts for localized ocean dynamics, ensuring our ConvLSTM captures

spatially heterogeneous temporal dependencies.

Thank you again for your constructive feedback and welcome any further suggestions you may have.

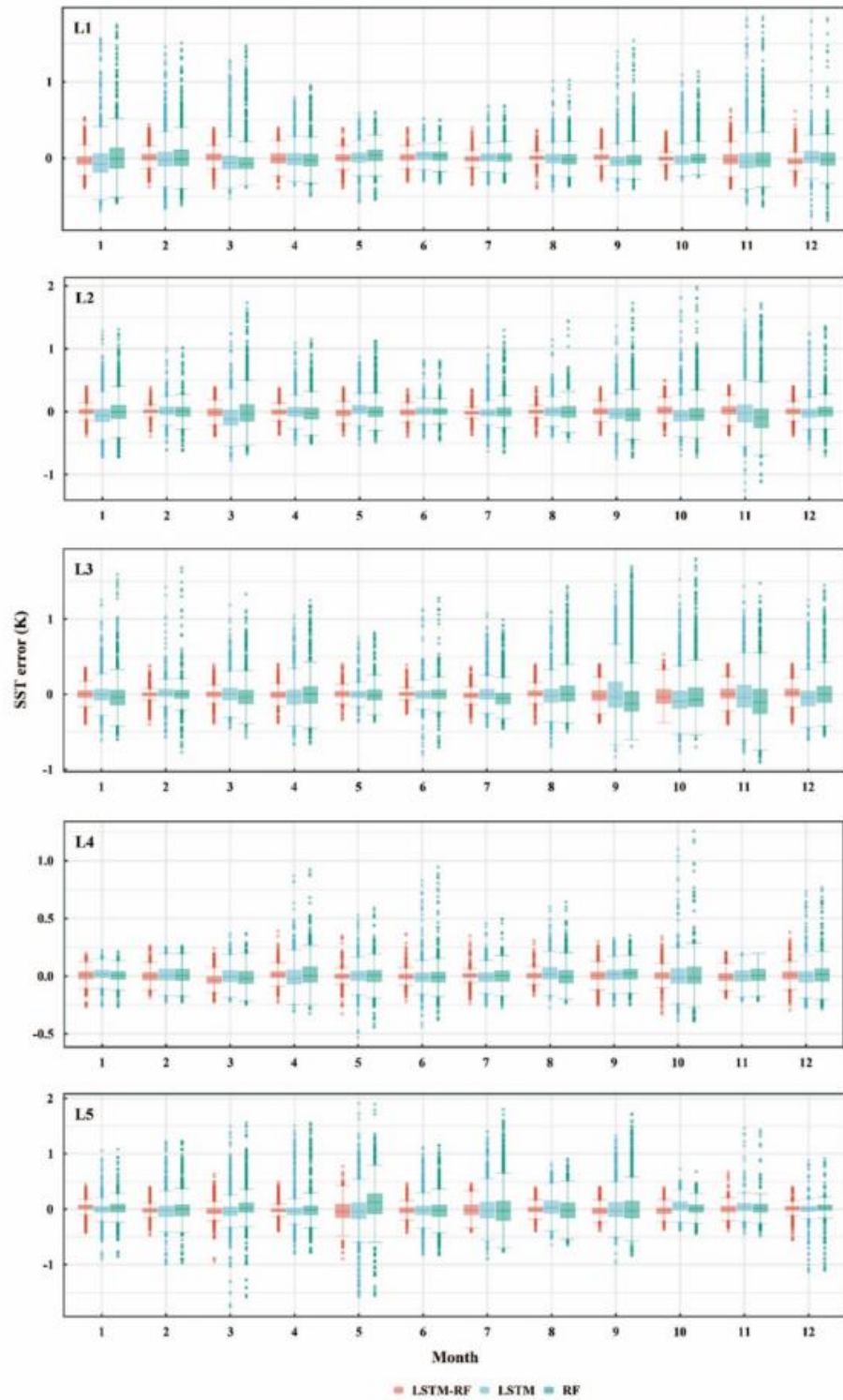


Fig. 1. Prediction errors for different prediction horizons (1-12 months) using the LSTM-RF, LSTM and RF methods at the five locations.

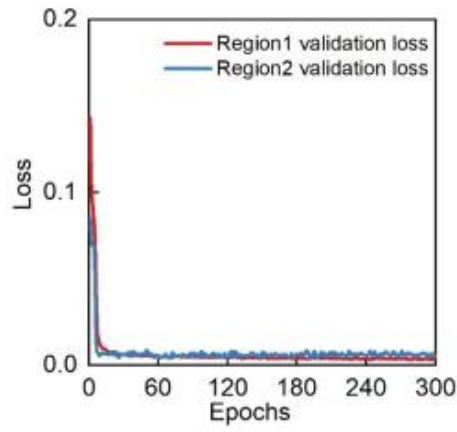


Fig.2 Variation of model loss with the number of epochs for different region prediction.

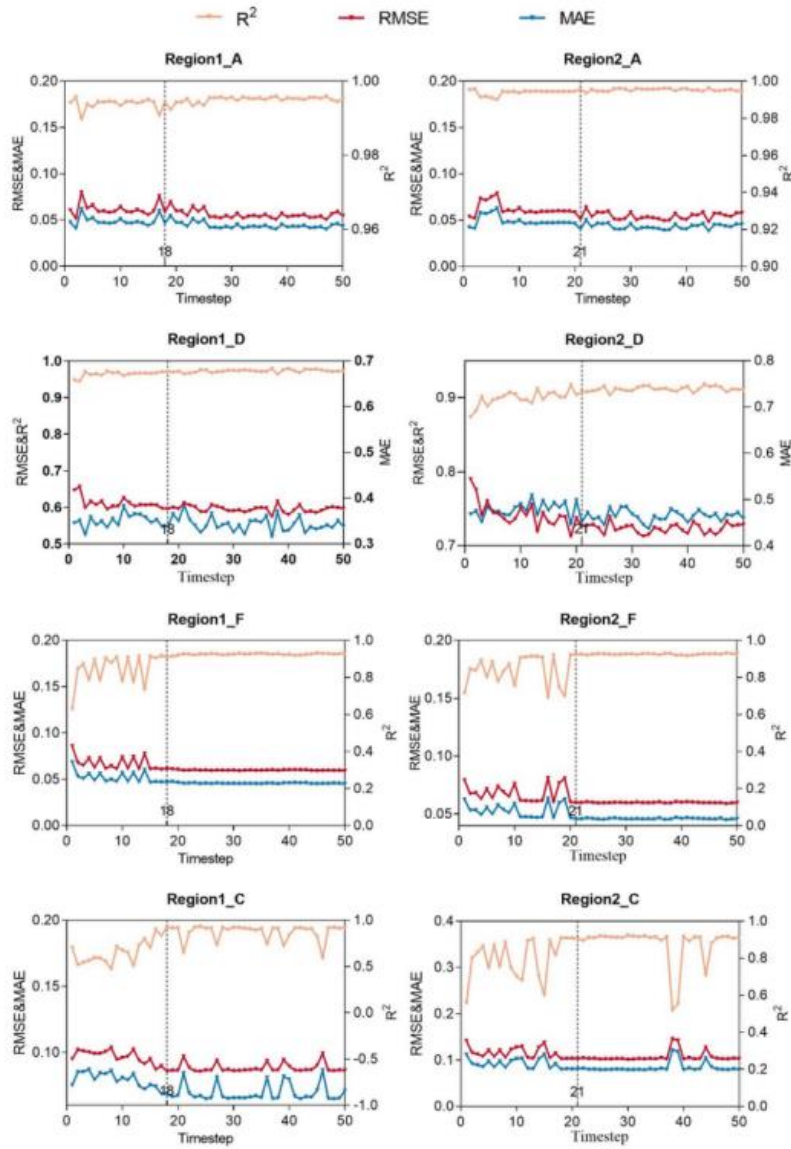


Fig.3 The relationship between prediction accuracy and timestep of feature variables in regions 1 and 2 based on the ConvLSTM model.