Dear Editor and Reviewer:

Thank you very much for your insightful comments concerning our manuscript "Simulation, Precursor Analysis and Targeted Observation Sensitive Area Identification for Two Types of ENSO using ENSO-MC v1.0" (ID: gmd-2021-396). Those comments are all very valuable and helpful for revising and improving our manuscript, and we have studied comments carefully and have made revisions. The manuscript has also been double-checked and modified, including the typos and grammar errors. The point-by-point responses are as following:

Major Comments

Comment 1

Comment 1.1: There are three parts of this manuscript, including ENSO prediction model named ENSO-MC, precursor analysis, and targeted observation sensitive area identification based on ENSO-MC. Accurately, the latter two parts are the application of ENSO-MC, and the conclusions are almost consistent with previous studies, such as Kumar et al. 2014, Duan and Hu 2016. Certainly, the method of precursor analysis and targeted observation sensitive area identification based on the deep neural network is an innovation, but I think the key point of this manuscript should be focusing on the ENSO-MC as the model description article. However, the description of ENSO-MC is inadequate, including the value of the weight in Loss function, discussions about the effects of the depth, structure, data for the neural network. I recommend authors add the discussion section.

Response: We gratefully thank you for the precious time the reviewer spent making constructive comments and suggestions on our manuscript. We also realize that the description of the model in the original manuscript is too brief. It only included a brief introduction of the neural network, data and loss function we used, and ignored the description of the idea of building the model for ENSO forecasting, the specific configuration and the effect of each component on the model performance.

As suggested by the reviewer, we have supplemented detailed descriptions of the ENSO-MC model structure, including its composition and hyperparameters. Specifically, the ENSO-MC model consists of an encoder, which is used to extract related features of ENSO, and a decoder, which is used to infer the development of SST anomalies in the equatorial Pacific in the future. And considering the important role of spatiotemporal interaction between atmosphere and ocean on ENSO, we use ConvLSTM as the main neural network architecture of encoder and decoder. In the encoder, each ConvLSTM layer followed by a 3D max-pooling layer is used to extract features at different spatial scales. Symmetric to the encoder structure, the decoder has upsampling layers followed by each ConvLSTM to restore the SST field. In addition, skip-layer connection and states connection, which make full use of the features extracted at each spatial scale of the encoder, are used to help the decoder recover the details of the forecast field. And the attention mechanism is designed to make the model automatically learn the temporal weights in the skip-layer connection. We also supplemented a figure (Fig 3(i)) in Section 3.3 to illustrate the skip-layer connection and its attention mechanism structure. For the hyperparameters in the model structure, the kernel size of ConvLSTM is 3×3, and the size of upsampling and downsampling is 2, which determines the field of feature extraction on the spatial scale. Besides, the depth of the model is 3, and the number of output filters of each layer is 8, 16, 32, which determines the degree of nonlinearity of the model.

Meanwhile, according to the variables, structure and loss function described in Section 2 of the manuscript, we also added a section to comprenhensively discuss the impact of these three factors on model performace, including the effects of the input sequence length, different predictors, each component of the model and different loss functions. Firstly, the input length represents the time dimension of ocean and atmosphere features that can be extracted by the model and plays an important role in ENSO forecast. In order to select the most suitable input length for ENSO-MC, we used the data of the past 3, 6, 9, 12 and 15 months as the input respectively, and compared the correlation skills and root mean square errors of their prediction results. Secondly, we designed an ablation experiment to verify the effectiveness of the multi-channel structure and examine the effect of each physical variable on the development of ENSO. Thirdly, we also supplemented experiments to prove the effectiveness of the skip-layer connection, states connection and attention mechanism structures for ENSO prediction, which are important components of model structure. Finally, for the combined loss functions, we detected the effect of each loss function by ablation experiment and determined the weight value of each function.

The detailed introduction about model structure and configurations have been supplemented in the Section 2.1 from the start of **line 105** as the blue text below:

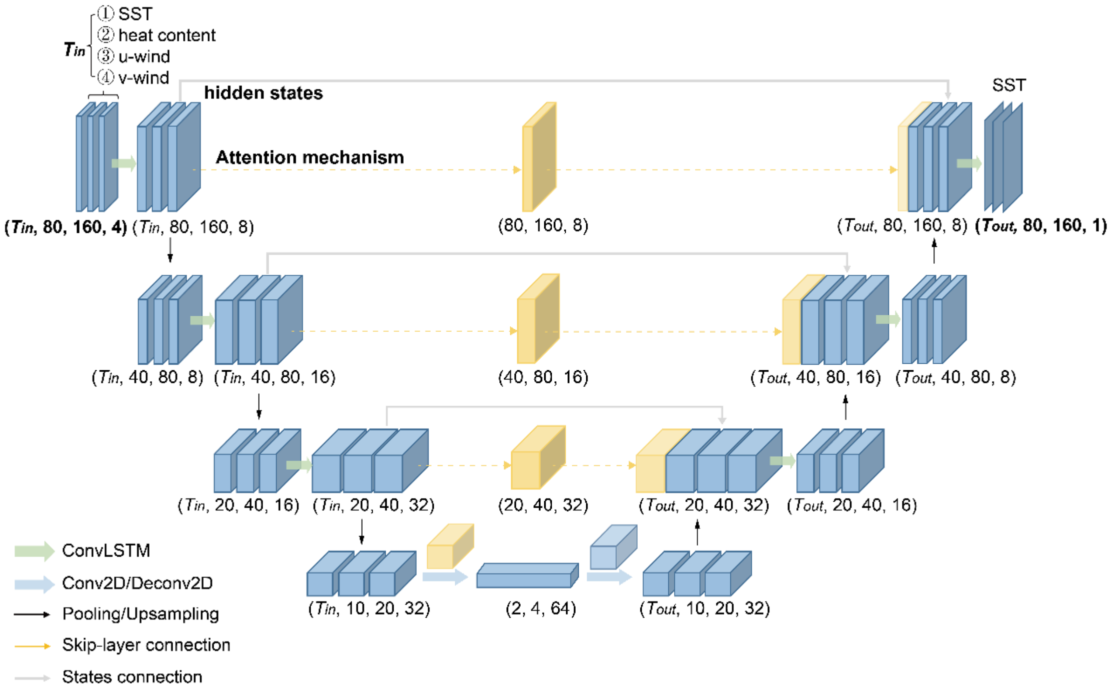

Figure 1 The encoder-decoder architecture of ENSO-MC for SST pattern sequence prediction. The encoder contains three ConvLSTM layers and each layer is followed by a pooling layer, and the last layer is a convolutional layer, which allows extracting spatial and temporal features. And the decoder comprises one deconvolutional layer, three ConvLSTM layers and three upsampling layer that restores the features to the same size as the initial spatial dimension (80×160). The model uses skip-connection with attention mechanism and state-connection between the encoder and the decoder to improve forecasting skills. The input variables are the SST, heat content, zonal wind and meridional wind for T_{in} *consecutive months over 40ºN-40ºS, 120ºE-80ºW (80×160, 1º×1º resolution) and each type of variable is input to the model as a channel. The output of the model is SST pattern sequence for the next* T_{out} *months.*

"*Here we develop a spatiotemporal model of multichannel structure named ENSO-MC to generate SST pattern sequence for ENSO forecasts. As shown in Fig. 1, the ENSO-MC is constructed based on the encoder-decoder architecture (Sutskever et al., 2014),* whose encoder extracts the feature representations associated with ENSO over the past period and decoder generates the sea surface temperature pattern in the future. Due to the diversity of ENSO in amplitude, spatial pattern and temporal evolution, several convolutional long short-term memory (ConvLSTM) layers (Xingjian et al., 2015) form the skeleton in the encoder-decoder architecture to learn its multiple spatial and temporal representations. The encoder is the first half of the architecture (Fig. 1). A ConvLSTM layer with kernel size 3×3 followed by a 3D max-pooling layer constitutes an encoding module. The max-pooling layer downsamples the input along the spatial dimensions to extract multi-scale spatial connections. We use three encoding modules to construct the encoder, which is the network depth that perform best in ENSO forecasting problem here. To balance model performance and computational cost, the output channels of ConvLSTM in the three modules are 8, 16, 32 in order. After three encoding modules, we use a convolution layer with kernel size 5×5 and stride 5×5, and the number of output channels is 64. The dimension of feature map output by each layer is shown in Fig. 1, and the final feature dimension of the encoder is 2×4×64. The structure of the decoder is symmetrical with the encoder. After a transposed convolution layer, there are three decoding modules. Each module consists of an upsampling layer with size 2 followed by a ConvLSTM layer to restore the original resolution of SST pattern, where the kernel size of the network and the number of output channels are the same as those in the encoder. And the final layer in the ENSO-MC model is an additional 3×3 ConvLSTM that generates a single feature map representing the SST pattern sequence predicted by the model."

And the model evaluation for the input sequence length, different predictors, each model component and different loss functions have also been discussed in the Section 3. This new section has been added from the start of **line 205** as the blue text below:

3 Model Performance evaluation

3.1 Influence of the input sequence length

Table 1: Correlation skill (Corr) and Root Mean Square Error (RMSE) of lead times with different input sequence lengths

Sequence length	3	6	9	12	15
	Corr/RMSE	Corr/RMSE	Corr/RMSE	Corr/RMSE	Corr/RMSE
3-month	0.78 / 0.67	0.59 / 1.02	0.58 / 1.24	**0.81 / 0.54**	0.68 / 1.10
6-month	0.52 / 0.96	0.46 / 1.15	0.47 / 1.16	**0.64 / 0.81**	0.48 / 1.72
12-month	0.29 / 1.00	0.36 / 1.39	0.28 / 1.14	**0.53 / 0.84**	0.31 / 2.06
18-month	0.17 / 1.06	0.25 / 1.72	0.15 / 1.17	**0.44 / 0.91**	0.30 / 2.15

Appropriate input sequence length is critical for ENSO prediction of the multichannel model. We use data from the past 3, 6, 9, 12 and 15 months as inputs to predict the development of ENSO in the next 18 months to examine the effects of different input sequence lengths on ENSO predictions. Table 1 shows the comparison of correlation skill and RMSE of lead times. For the correlation coefficient, the higher the value is, the higher the forecasting skill is. And the smaller RMSE represents the higher skill. The results show that the ENSO-MC model performs best with data from the past 12 months as input. This may be because the variables we select have long-term memory for the development of ENSO, such as oceanic heat content. A longer input sequence contains more information that is helpful to ENSO prediction, but also contains more noise that interferes with the prediction, so the improvement of forecasting skill is not positively correlated with the increase of input sequence length. But for different models and forecasting horizons, the most appropriate input sequence lengths may not be the same.

3.2 Physical variable sensitivity for the multichannel structure

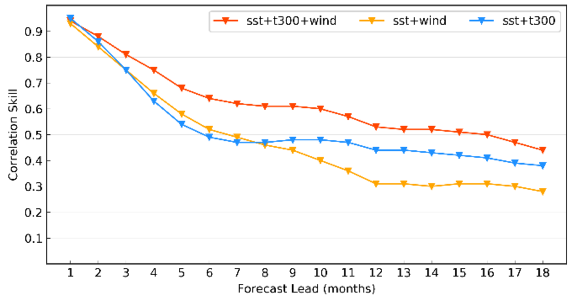

Figure 2. The correlation skill of Nino 3.4 index of ENSO-MC model with different predictors.

With the physical variables we selected in the Section 2, we construct multichannel input that takes into account the complicated spatiotemporal interactions in the ocean and atmosphere underpinning the onset and development of ENSO events. For the grid observations of monthly SST, heat content, zonal wind and meridional wind, we treat each type of the input variable as a channel in the first ConvLSTM layer and thus there are four channels. The number of channels is the depth of the matrices involved in the convolutions, so that the cross-correlation and transmission between these ocean-atmosphere data can be calculated in the convolution operation. We design an ablation experiment to examine the contribution of predictors and the effectiveness of the multichannel structure. In addition to SST, the most important predictor, we remove heat content (t300) and wind from the inputs respectively to detect their effects on the correlation skill of the Nino3.4 index. As shown in Fig. 2, the model using the three key ingredients of Bjerknes feedback (SST, heat content, wind) as input produces more favorable forecast skill than the ones that remove one of them, which indicates that the multichannel structure can help to learn the ocean-atmosphere coupling between several important predictors. For the models with two predictors, the model containing wind predictor shows slightly higher forecasting skill in the first few months, while the one containing heat content predictor performs more stable skill at lead times of more than eight months. It suggests that the memory of subsurface heat plays an important role in ENSO prediction on seasonal to interannual time scales, which is consistent with previous research.

3.3 Effectiveness of the model components

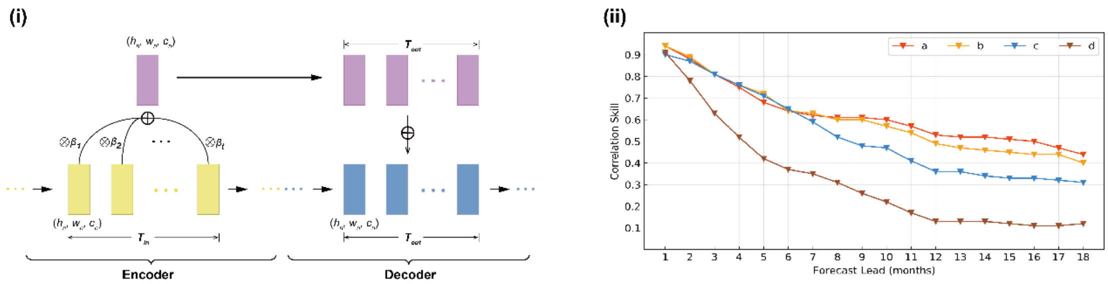

Figure 3. (i) The detailed structure of the skip-layer connection and attention mechanism between encoder and decoder at the n^{th} layer in ENSO-MC. (ii) The correlation skill of Nino 3.4 index of the forecast lead month in models with different structures: (a) ENSO-MC model of skip-layer connection with attention mechanism and states connection, (b) ENSO-MC model without attention mechanism, (c) ENSO-MC model without states connection, (d) ENSO-MC model without skip-layer connection.

The ENSO-MC model learns the feature of ENSO at different spatial scales with the convolution and max-pooling layers in the encoder, and gradually restores the spatial dimensionality of the original SST field in the decoder. With symmetrical structure design of the encoder and decoder as shown in Fig. 1, skip-layer connection is used to transfer features form the encoder to the decoder to recover spatial information lost during downsampling (yellow line in Fig. 1). Rather than transferring the original features of all time steps obtained from the encoder, we design an attention mechanism to enable the skip-layer to automatically learn the attention weights $\beta_1, \beta_2, \ldots, \beta_t$ on the temporal sequence because these air-sea features may have different effects on ENSO development at different time scales. As shown in Fig. 3(i), the encoder obtains the features $f_n \in \mathbb{R}^{T_{in} \times h_n \times w_n \times c_n}$ after max-pooling and convolution calculation at the n^{th} layer. Using a two-layer densely-connected neural network, we obtain the attention weight $\beta \in \mathbb{R}^{T_{in}}$ of each time step's features according to Eq. (1), where $f_n' \in \mathbb{R}^{T_{in} \times (h_n \times w_n \times c_n)}$ are reshaped from f_n:

$$\beta = \mathrm{softmax}\big(\mathbf{W}_{\beta\alpha} \tanh\big(\mathbf{W}_{\alpha f} f_n' + \mathbf{b}_{\alpha f}\big) + \mathbf{b}_{\beta\alpha}\big), \tag{1}$$

where $\mathbf{W}_{\alpha f}$, $\mathbf{W}_{\beta\alpha}$ are weight matrices created by the layer, and $\mathbf{b}_{\alpha f}$, $\mathbf{b}_{\beta\alpha}$ are the bias vectors. β represents the contribution of each time step to prediction. According to Eq. (2), the feature maps of each time step are multiplied by the corresponding weights, and the fused maps $\widetilde{f_n} \in \mathbb{R}^{h_n \times w_n \times c_n}$ are obtained by adding them along the time dimension.

$$\widetilde{f_n} = \sum\nolimits_{T_{in}} (\beta \circ f_n), \tag{2}$$

where $\widetilde{f_n}$ are the feature maps to be transmitted in the skip-layer connection, which are connected to the features of the corresponding layer in the decoder. Besides, we also add states connection between the encoder and the decoder (grey line in Fig. 1), where the hidden states output by the ConvLSTM layers in the encoder are reserved for the corresponding layer when the decoder is initialized. With the methods of skip-layer connection and states connection, the model can make full use of the information extracted from the encoder before ENSO events, which help stabilize training and convergence.

We remove the attention mechanism (model b), states connection structure (model c) and skip-layer connection structure (model d) respectively from the constructed ENSO-MC model (model a) to analyze their effects on model performance. As shown in Fig. 3(ii), the two connection structures, especially the skip-layer connection structure, have a great influence on the prediction results. In model b, we use average weights to replace the attention mechanism. Compared with model a, self-attention mechanism can play a greater role in long-term ENSO prediction.

3.4 Effects of different loss functions

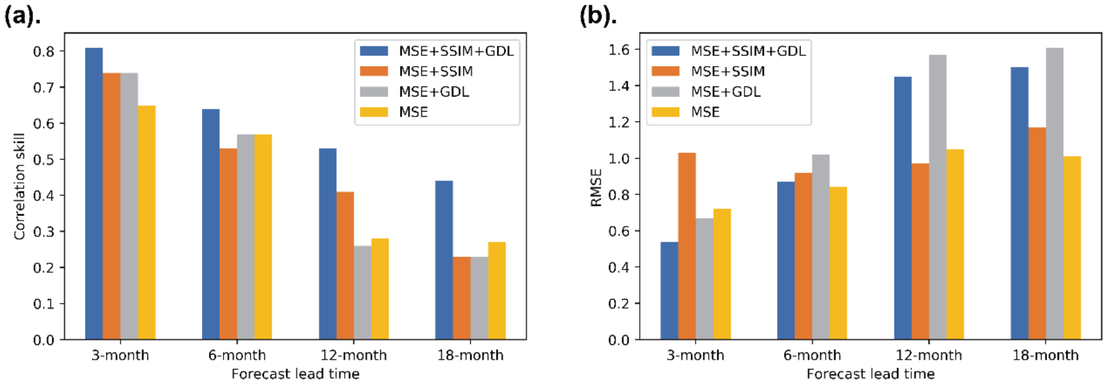

Figure 4. The performances of the ENSO-MC with different loss functions.

In order to balance the optimization speed of each loss in the training process, we set λ_{mse}=7, $\lambda_{ssim} = 9$ and λ_{gdl}=1. The effectiveness of combined loss function is validated. As shown in Fig. 4(a), although SSIM and GDL do not significantly improve the model performance when combined with MSE alone, the combination of MSE, SSIM and GDL loss functions achieve the best performance on the correlation skill. Besides, since GDL loss function tends to retain extreme values and MSE loss function tends to smooth all values, the presence of GDL inhibits the decrease of MSE, so the MSE errors of the models with GDL loss function are higher than the ones without GDL (Fig. 4(b)). And comparing the results of correlation skill and RMSE in Fig. 4(a) and (b), low RMSE values do not represent high correlation skills. Therefore, it is necessary to explore loss functions suitable for ENSO prediction other than MSE to balance the training of the model.

Comment 1.2: Moreover, it only showed the several cases and correlation results, but the more cases and RMSE also needed.

Response: Thank you for the above comments. In the original manuscript, we only selected three individual events of 2015/2016 EP El Niño, 2009/2010 CP El Niño and 1988/1989 La Niña to validate the forecast skills in spatial patterns and Niño 3.4 index time series. The simulation ability of ENSO-MC model for different types of events is not fully explained in the manuscript. As suggested by the reviewer, we have added the simulation results of three more cases in recent years for each type of event, namely, EP El Niño events of 1991/1992, 1997/1998 and 2006/2007, CP El Niño events of 1994/1995, 2002/2003 and 2018/2019, and La Niña events of 1984/1985, 1998/1999 and 2000/2001. We compare the predicted spatial patterns and observations of SST anomalies from the onset to the mature phase for these events. The simulation results show that ENSO-MC model can generally simulate the development and characteristics of SSTA for different types of events. We also summarized the classification results of the model for the two types of El Niño events occurring from 1984 to 2019, and calculated the RMSE of Niño3, Niño3.4 and Niño4 index for different lead months. The results show that ENSO-MC model has a higher classification accuracy for EP events, and smaller prediction errors for CP events in amplitude.

The related results and statements have been supplemented in the Section 4 from the start of **line 304** as the blue text below:

“In addition to the above three typical events in recent years, the prediction results of other events that occurred between 1984 and 2019 are also detected. For each event, we compare the spatial development of predicted and observed SST anomalies in the equatorial Pacific from the onset to the maturity stage.

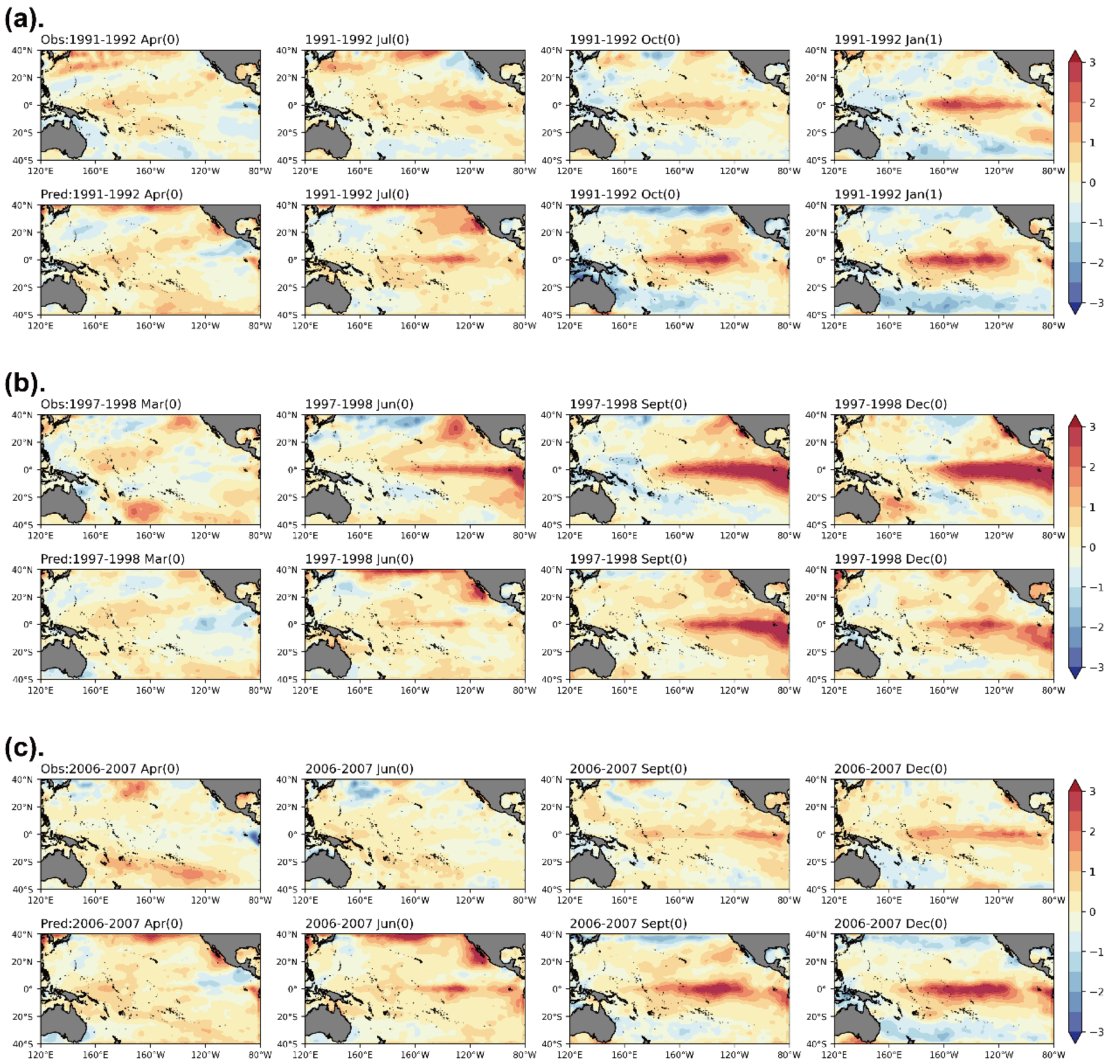

Figure 5. SSTAs of three EP El Niño events in (a) 1991/1992, (b) 1997/1998 and (c) 2006/2007 from the onset to the maturity stage, with observations in the first row and predictions in the second row for each event. The mature phases here are the months when the El Niño events peak. And “0” and “1” next to the calendar month denote the year when the El Niño event occurred and the following year, respectively.

Fig. 5 shows the simulation results of the ENSO-MC model for three EP El Niño events of 1991/1992, 1997/1998 and 2006/2007, with observations in the first row and predictions in the second row for each group. The results show that the model can simulate the occurrence and development of SSTA for each event. However, for some events with less significant EP type characteristics (for example, that of 1991/1992), the SSTA center of predictions is closer to the central Pacific than observed. In addition, for some super strong events (for example, that of 1997/1998) and weak events (for example, that of 2006-2007), the amplitude of the predicted results at mature phase may be lower or higher than the observed.

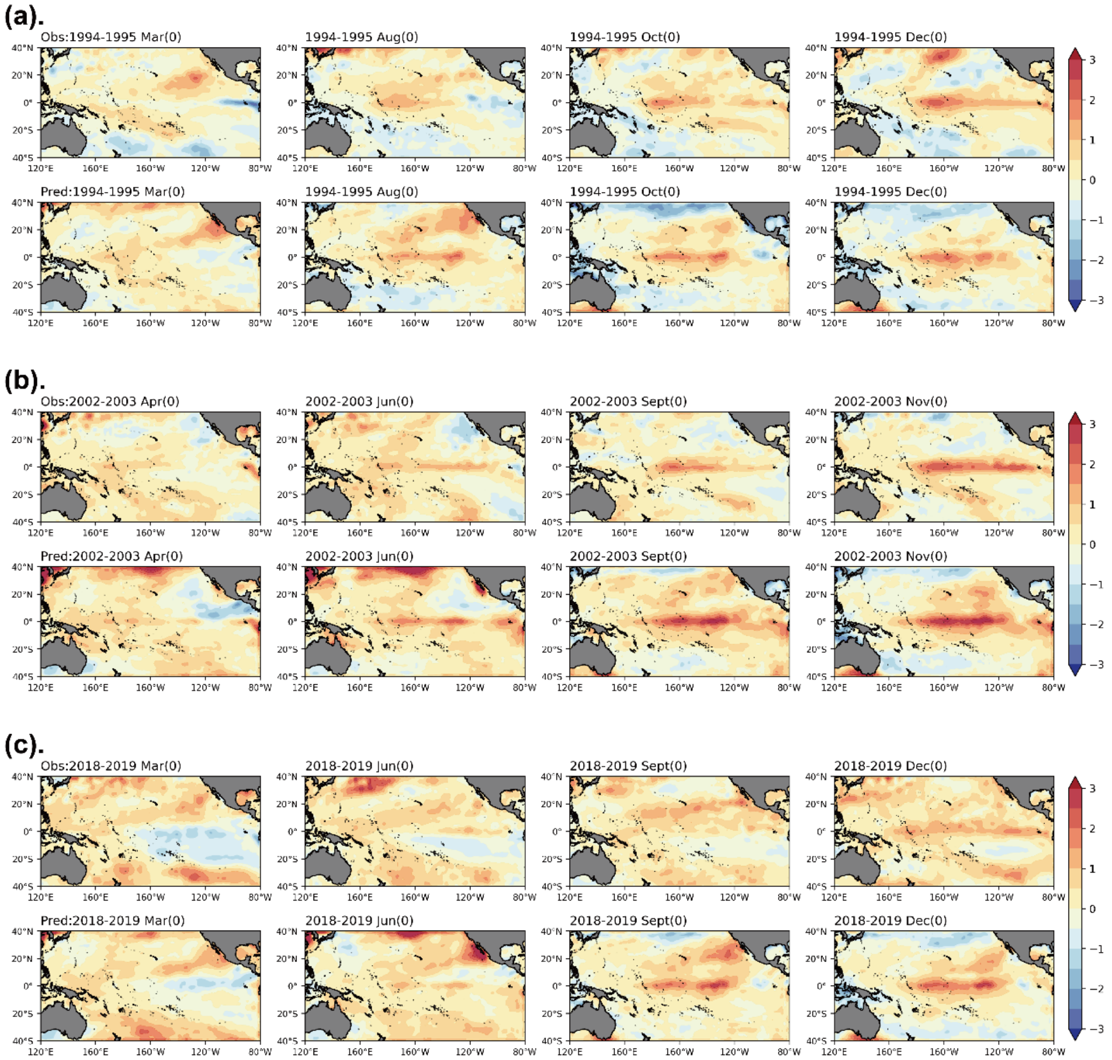

Figure 6. As in Fig. 5, but for the three CP El Niño events in (a) 1994/1995, (b) 2002/2003 and (c) 2018/2019.

The prediction results for three CP El Niño events in 1994/1995, 2002/2003 and 2018/2019 are displayed in Fig. 6. For the events of 1994/1995 and 2018/2019, the model can simulate the process that the SST anomalies in the northeast Pacific propagate to the southwest and finally contribute to the occurrence of CP events. The amplitude and center location of the predicted anomalies are also in agreement with the observations. However, the meridional distribution of predicted SSTA is not as broad as observed in the mature stage. The observed SSTA extends eastward to 80ºW, while the predicted value extends roughly between 100ºW and 120ºW.

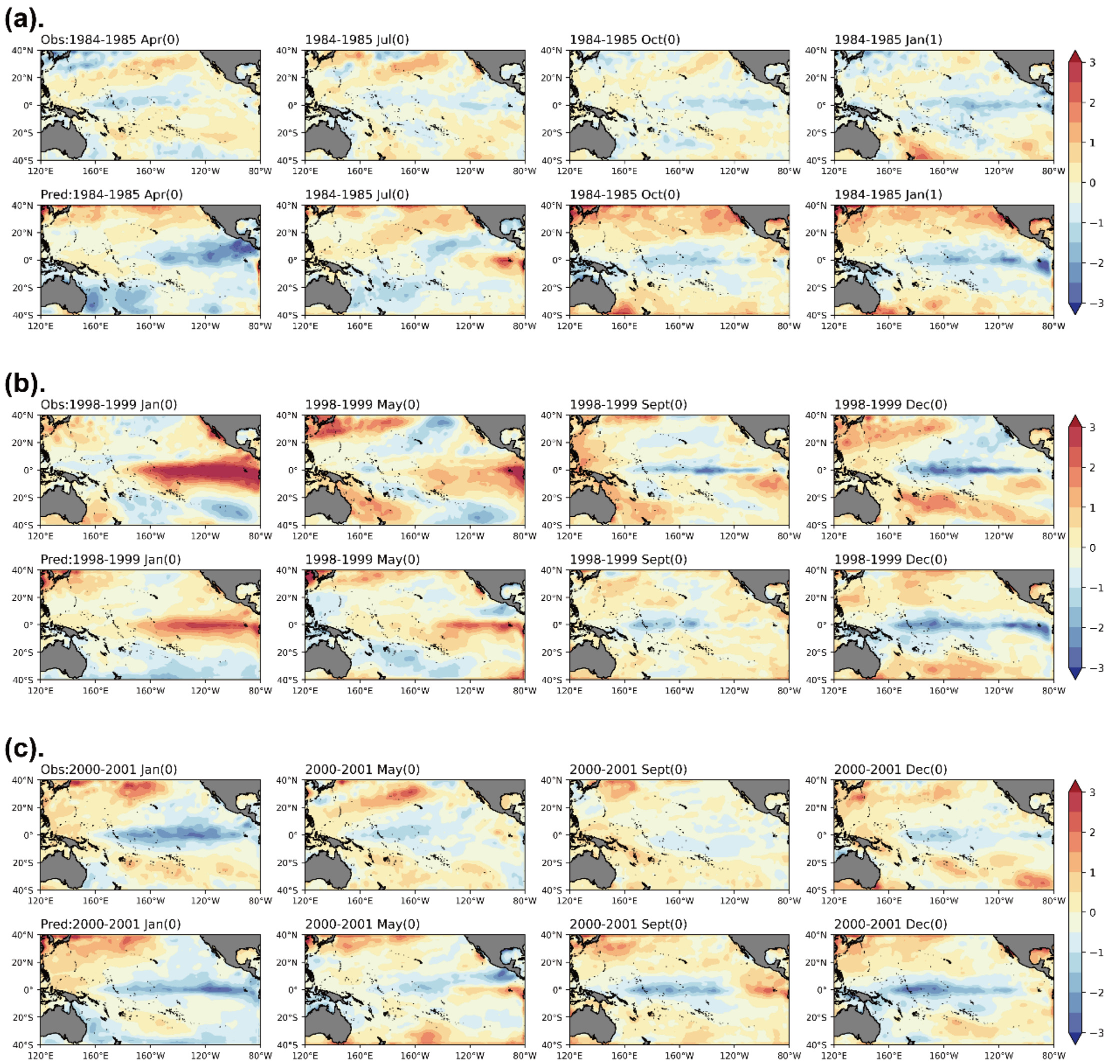

Figure 7. As in Fig. 5, but for the three La Niña events in (a) 1984/1985, (b) 1998/1999 and (c) 2000/2001.

And Fig. 7 shows the predictions of three La Niña events in 1984/1985, 1998/1999 and 2000/2001. These three events occurred under different conditions. The events of 1984/1985 and 1998/1999 were preceded by strong El Niño events, and the 1998/1999 event occurring more rapidly. The 2000/2001 La Niña was another weaker event after the previous La Niña event ended. Compared with observations, the model can simulate the occurrence, development and phase transition or persistence of La Niña events.

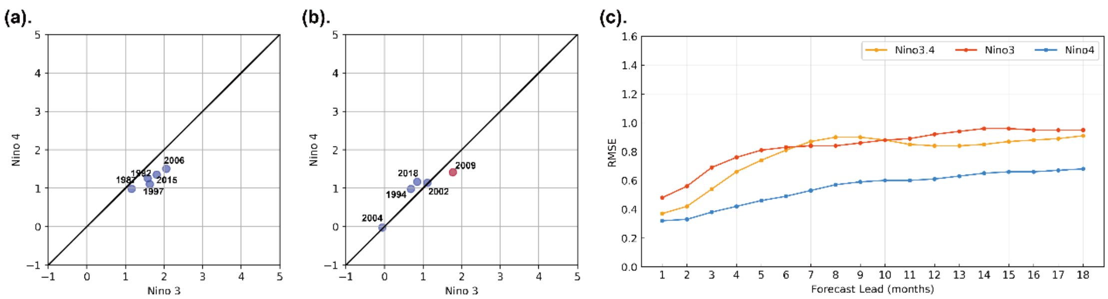

Figure 8. Scatterplots in Nino3-Nino4 index plane of 12-month-lead predictions for all (a) EP El Niño events and

(b) CP El Niño events during peak phase from 1984 to 2019. (c) Root Mean Square Error (RMSE) of the Nino3.4, Nino3 and Nino4 indexes between the forecast results of the ENSO-MC model and observations during validation period.

Table 2: Root Mean Square Error (RMSE) of Nino3/4 index for all EP/CP El Niño events during peak phase from 1984 to 2019.

Lead time	3-month	6-month	9-month	12-month
EP (Nino3)	1.18	1.23	1.04	1.07
CP(Nino4)	0.45	0.99	0.91	0.93

In addition to comparing the detailed spatial distribution of SSTA, the related indices and metrics are calculated to further evaluate the simulation performance of the ENSO-MC model. The Niño 3 index (average SST anomalies over 5ºN-5ºS, 150ºW-90ºW) and the Niño 4 index (average SST anomalies over 5ºN-5ºS, 160ºE-150ºW) are commonly used to define two types of El Niño events. Events with Niño 4 index greater than Niño 3 are regarded as CP El Niños, and events with Niño 3 index greater than Niño 4 are classified as EP El Niños. Figure 8a, b shows the distribution for Niño 3 and Niño 4 indexes calculated from the model's one-year-lead predictions of the peak periods for all EP events (Fig. 8a) and CP events (Fig. 8b) from 1984 to 2019. The results show that the model can correctly classify five EP events (1987/1988, 1991/1992, 1997/1998, 2006/2007, 2015/2016) and three CP events (1994/1995, 2002/2003, 2018/2019) in the past 30 years, but misjudge the event of 2009/2010 as EP type and no El Niño event occurred in 2004 (Niño 3=0, Niño 4=0). The CP event of 2004/2005 is much weaker than other CP ones, making it more difficult for the model to capture its development. We also make statistics on the RMSE between the predictions and official records of Nino3/4 index for all EP/CP El Niño events at mature phase (Table 2) and for the whole validation period (Fig. 8c), and find that although the model has a higher classification accuracy for EP events, the index error of predictions for EP events is larger than that for CP. It may be because most of the strong El Niños are EP-type events, and the prediction skills of the model for such extreme events need to be improved. The SSTA distribution in Fig. 5 also shows that for some EP events, there is a difference in amplitude between predictions and observations for the maturity stage of the event, while that of the CP events is consistent with the observations (Fig. 6)."

Minor Comments

Comment 1: Usually we use the years covering the ENSO process, such as 1985-1986 defining the ENSO year, rather than the ENSO peak occurring year.

Response: Thank you so much for your professional comments. We have read through the full manuscript and corrected all the related statements in the text as well as in the legend of Figures in the revised version.

Comment 2: Line 20: "periodically", in fact, usually we called the ENSO is an irregular signal with 2-7 years period.

Response: Thank you so much for your professional attitude and helping us find a mistake. We have updated the corresponding statements at **line 20** as the **blue** text below:

"**El Niño-Southern Oscillation (ENSO) is an irregular climate signal with a period of 2-7 years in the tropical Pacific Ocean** *and often grows up to be exceptionally strong under unstable air-sea interactions, causing large global climatic anomalies and hence affecting many regions even far from the tropical area.*"

Comment 3: Why did author select the combination of SODA and GODAS/ERA-Interim data? I'm very interested that whether the results keep consist if use other data, e.g., SODA and GODAS/ERA-5 (ERA-5 is better than ERA-Interim), or just using one datasets CERA-20C from 1901 to 2010 with almost same resolutions to SODA.

Response: We thank the reviewer for pointing out this issue. This is indeed a valuable question for investigating the effects of data from different sources on the ENSO-MC forecast model. In the manuscript, we chose the data based on Ham's article (Ham et al., 2019), in which SODA and GODAS/ERA-Interim data were used. As suggested by the reviewer, we have investigated the forecasting skills of ENSO-MC model for other data. Figure 9 shows the correlation skill of Nino 3.4 index of ENSO-MC model with data from different sources. The red line represents the skills using data from the original manuscript, namely SODA and GODAS/ERA-Interim data, the blue line is the result of ERA-Interim data being replaced by ERA-5, and the yellow line is the result with CERA-20C data.

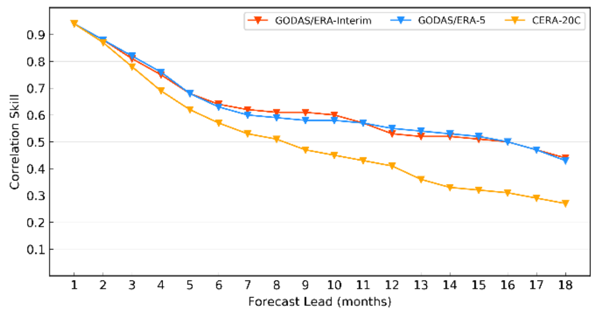

Figure 9. The correlation skill of Nino 3.4 index of ENSO-MC model with data from different sources: (a) SODA and GODAS/ERA-Interim, (b) SODA and GODAS/ERA-5 and (c) CERA-20C.

The results show that replacing ERA-Interim data with ERA-5 data has little effect on forecast skills. However, only using CERA data will reduce correlation skills. We think the possible reason is that the data volume of CERA data is less than that of SODA, and the resolution is not as high as SODA. The CERA data is from 1901 to 2010, which is nearly a quarter less than the soda data. The resolution of the SODA data is 0.5, which is twice that of the CERA data. Therefore, the amount and resolution of data can affect the training process of deep learning models and thus the forecasting skills of ENSO-MC.

References

Ham, Y.-G., Kim, J.-H., and Luo, J.-J.: Deep learning for multi-year ENSO forecasts, Nature, 573, 568–572, 2019.

Comment 4: Line 176: "although there are stronger anomalies in the eastern tropical Pacific during the growth phase", please double check it. I did not see stronger anomalies.

Response: We thank the reviewer for pointing out this issue. This is a mistake in our expression. What we originally intended to express was that, in the growth phase, the predicted SST anomalies distributed in the central Pacific were smaller in amplitude and scope than the observations, which made the anomalies distributed in the eastern tropical Pacific appear more pronounced in the prediction results, and the CP characteristics of ENSO were not obvious. We feel sorry for our poor expression. The modified statements have been updated at **line 291** as the **blue** text below:

"*The model captures most of these features,* **although the anomalies distributed over the central equatorial Pacific are smaller in amplitude and scope than observed during the growth and mature phase, and the CP characteristics of ENSO are not as pronounced as the actual ones.**"

Comment 5: Lines 195-196: please clarify what is the one- and multi-step strategy.

Response: Thank you for spotting our crucial neglects in description of forecasting strategies. Since managing policy responses requires robust long-term forecasting results of ENSO, multiple time steps must be predicted. In general, there are two main strategies for the multi-step time series forecasting problem, direct multi-step forecast strategy and recursive multi-step forecast strategy. In our work, one-step strategy refers to the multiple use of a one-step model, in which the prediction of the previous time step is used as the input for the prediction of the next time step, i.e. recursive multi-step forecast. While multi-step strategy is to build a model that can forecast the entire prediction sequence in a one-shot manner, which belongs to a direct multi-step forecasting method.

Considering the content and structure of each section of the manuscript, the figure and detailed descriptions of the one- and multi-step strategies have been additionally supplemented in the Section 2 at **line 134** as the blue text below:

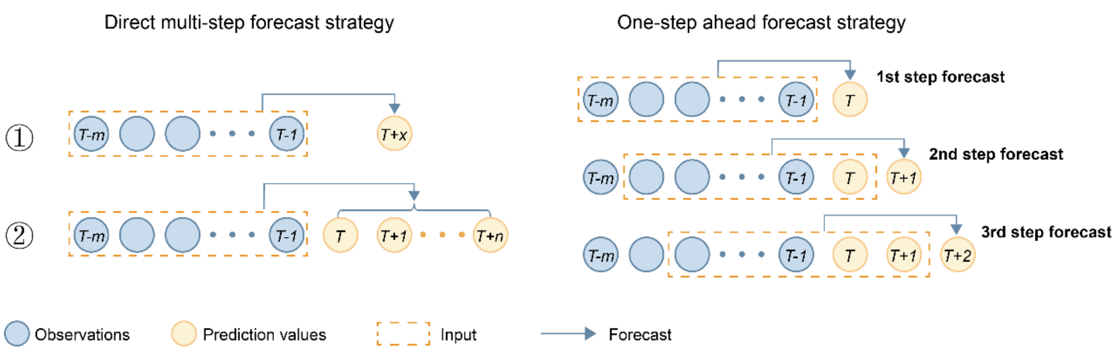

Figure 10. Two main strategies of multi-step forecasting for ENSO prediction, direct multi-step forecast strategy and one-step ahead forecast strategy. And direct multi-step forecast strategy has two methods, the second one is used in this paper.

"In addition, ENSO prediction requires a multiple-step forecasting strategy to achieve long-term prediction. There are two main strategies, direct multi-step forecast and one-step ahead forecast.

As shown in Fig. 10, the inputs for direct multi-step forecast are fixed observations $X_{T-m}, \dots, X_{T-1}$. To achieve multi-step prediction, one of the methods is to build a separate model G_{T+x} for each prediction time step $T + x$. In the case of predicting SST for the next n months, n models $G_{T+1}, \dots, G_{T+n}$ need to be constructed. This is also the strategy used in Ham's paper (Ham et al., 2019), which produced skillful prediction results one and a half years in advance. Another approach is to build a model that can forecast the entire prediction sequence $Y_{T+1}, \dots, Y_{T+n}$ in a one-shot manner to achieve direct multi-step forecasting, which has the advantage of significantly reducing computational and maintenance costs. While one-step ahead forecast strategy refers to the multiple use of a one-step model, in which the prediction of the previous time step is used as the input for the prediction of the next time step, i.e., recursive multi-step forecast. In general, one-step ahead forecasting models are more stable and easier to train (Shi and Yeung, 2018). However, because predictions are used instead of observations, the one-step ahead strategy allows prediction errors to accumulate, resulting in a rapid decline in performance as the prediction time increases, while direct multi-step forecasting has more accurate results in long-term prediction (Taieb et al., 2012). In this paper, direct multi-step forecast strategy and one-step forecast strategy are both used for prediction and comparison. Considering the computational cost, we use the second direct forecast method instead of the first method of constructing several individual models."

References

Ham, Y.-G., Kim, J.-H., and Luo, J.-J.: Deep learning for multi-year ENSO forecasts, Nature, 573, 568–572, 2019.

Shi, X. and Yeung, D.-Y.: Machine learning for spatiotemporal sequence forecasting: A survey, arXiv [preprint], arXiv:1808.06865 2018.

Taieb, S. B., Hyndman, R. J., et al.: Recursive and direct multi-step forecasting: the best of both worlds, vol. 19, Citeseer, 2012.

Comment 6: Section 4 and 5, author should give the details of the initial perturbation distribution, at least the magnitude of the perturbation.

Response: Special thanks to you for your professional comments. We have added the detailed description of the perturbation distribution in Section 5 at **line 399** and Section 6 at **line 453** as the blue text below:

"*The precursor maps from 12-month lead to 1-month lead of EP-type El Niño, CP-type El Niño and La Niña are shown in the Fig. 6, and we present results every few months for each type to see more clearly how precursors change over time.* Since the composite maps are the sum of saliency maps of each event, here we focus on the distribution of perturbation without considering the intensity, and the saliency values in Fig. 6 are the standardized results of the scale between 0 and 1. For EP El Niño (Fig. 6(a)), the subsurface temperature component presents large anomalies in the equatorial Pacific, especially in the central and western Pacific, the subtropical northeast Pacific and the subtropical South Pacific. With the occurrence of El Niño, the anomalies weaken in the equatorial region and slightly intensify in the subtropical area. And the large anomalies of the precursory perturbation for CP El Niño (Fig. 6(b)) are concentrated in the subtropical northeast Pacific. As El Niño approaches, the disturbance tends to spread to the southeast. Regarding the subsurface disturbance for La Niña (Fig. 6(c)), their anomalies are concentrated in equatorial regions and propagates from the western Pacific to the eastern."

"*Then all the saliency maps of SST are added up to obtain the composite saliency map of the surface (Fig. 7(a)), and that of the subsurface (Fig. 7(b)) is obtained in the same way.* The saliency values in the figure are the standardized results of the scale between 0 and 1. The anomalies of surface disturbance are mainly distributed in the north central Pacific, the central Pacific and the western Pacific. The anomalies in the north Pacific are more intense than those in the south Pacific. For the subsurface temperature precursory, the perturbation possesses a wide range of anomalies in the equatorial Pacific and the North Pacific, and the anomaly values are more intense than that of surface disturbance."

Thank you again for your positive comments and valuable suggestions to improve the quality of our manuscript.

On behalf of all the co-authors, best regards,
Yuehan Cui