1. **DL is widely used in various fields as well as in geosciences studies.**

Q: Here, any other DL model in ENSO prediction?

Ans: I have added the application of machine learning in ENSO prediction in the manuscript. Actually, the works mainly depend on the traditional networks such as ANN, ELM, ARIMA and k-nearest neighbor method. I haven't find related work using the deep learning algorithms. The detailed information has also be added in the paper.

2. **The pentad global operational datasets since 1980, produced by the Global Ocean Data Assimilation System (GODAS) are provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their website at http://www.esrl.noaa.gov/psd/**

Q: Are you sure this is the correct website for the GODAS data? I cannot find GODAS data here. Are you using the following? https://www.esrl.noaa.gov/psd/data/gridded/data.godas.html

Ans: I download the data from http://apdrc.soest.hawaii.edu/las/v6/dataset?catitem=17020

3. **The indices of upper ocean heat content in the equatorial Pacific are derived from the archive of https://www.pmel.noaa.gov/tao/**

Q1: Can I confirm this dataset? TAO array data starts from 1993 instead of 1982. How do you use this data? I believe you are not using TAO to calculate WWV and TP300, right? I guess you use everything based on GODAS, right?

Ans: The data is acquired from https://www.pmel.noaa.gov/tao/wwv/data/. I do not calculate the WWV or TP300 index by myself and I get the index directly from the website. The methods for calculating is also list on the website. As the readme.txt shows that the data are based on temperature profiles from TAO moorings, Argo floats and XBTs.

The COEF ranking also reveals the better prediction performance of TP300 than other index, thus indicating the key role of tropical Pacific upper heat content in ENSO prediction. The general ranking of WWV reveals that the

批注 [u1]: Are you sure this is the correct website for the GODAS data? I cannot find GODAS data here. Are you using the following? https://www.esrl.noaa.gov/psd/data/gridded/data.godas.html

difference between machine learning predictions and observations may reduce if we take the warm water volume over whole tropical Pacific into consideration.

Q2: Yes, it is better to say why TP300 is better than WWV. WWV and TP300 normally behavior very similar. For WWV, may be try the region, I suggested. The reason may be due to the warm pool fronts I mentioned here. We can choose the specific regions to examine this further.

Ans: Ok, we can discuss about this later.

4. **Five-day moving average is applied to the daily NECP datasets to remove the high-frequency signal. The pentad datasets from GODAS are processed into the daily format by linear interpolation. Each field is interpolated into the same-dimensional matrix with a spatial resolution of 1.0º.**

Q1: Why do you interpolate everything into the daily output? Why not use pentad time-series? Why I ask this is because a lot of data you have is based on the monthly data (like all indices) or pentad data (GODAS/TAO). Then, you also do the five-day moving average to smooth out the daily NCEP data (note not NECP). Essentially, you don't use any daily information. There is no meaning to interpolate everything into daily, right? You mentioned "to obtain enough data, we forcast the daily Nino3.4 SST". I don't think you can get any daily information even if you use DL, right? Just interpolation……I feel pentad data may be the more reasonable choice than daily.

Q2: Also, here, you interpolate everything into 1degree. So why do you use ERSST 1/4 degree data? Do you check the sensitivity of different dataset in the DL? Many reviewers like to ask this kind of questions to ensure the quality of the forecast is stable.

Ans1: We interpolate the data into daily format to get more samples for training. As the previous study shows if a sufficient amount of data is available, DL model can be better trained to describe the evolution of non-linear process. If we use pentad time series, the amount of the data will be reduced to one fifth, thus the model will be harder to be stable and the results for the model will be worse. We have used the monthly data directly before and the correlation skill decrease to about 0.6 at 6-mon lead time. We may confirm the performance for the pentad data again later.

Ans2: We interpolate everything into 1degree because most of data from GODAS is with such resolution and we want to get the precursors with same format. We haven't checked the sensitivity of the model to the resolution of data. I think it may be a topic worth discussing.

5. **Our focus here is on the whole ocean bounded by the region 30ºS to 30ºN.**

Q: Why do you choose this region? Why I ask this is mainly because the ocean key domain is different from the atmosphere key domain. To enhance the hindcast skill at the longer lead time, the extratropical SLP should play a dominant role. See Dr. Ding's work on this http://140.112.69.65/coda/research/papers/2017_Ding_NP_SP_ENSO_impact.pdf

For the ocean part, it is the tropical region (the Recharged diagram you mentioned here).

Ans: Considering ENSO can be effected by the signals outside tropical Pacific, we chose larger area in the initial study and hope machine can extract the useful information by itself. However, ENSO may be sensitive to different variable at different location as you mentioned. And machine may not as clever as we think. Too large information fed into the model tends to become the interruption.

6. **Table 1 shows the detailed information of CNN-ResNet module, designed for processing the spatio-temporal data.**

Q: Could you include the information about the output size? Why input is 360x60x3? Also, what about the other size 180, 89, 30, 64, 66 etc. How do they correspond to the model?

As I mentioned earlier, could we use different domain for ocean and the atmospheric variable since they have different impacts? I believe we can, right?

7. **As the convolution operators in the model are selected randomly by machine, we train model with same input parameters for 10 times to avoid uncertainty.**

Q: This is interesting, since the convolution operators can be randomly selected, is "10 times" enough?

Actually, the number can be set as large as possible. However, if we choose the number too large, the model will be trained with too long time with the similar performance. But we may discuss this in the paper if others also have such puzzlement.

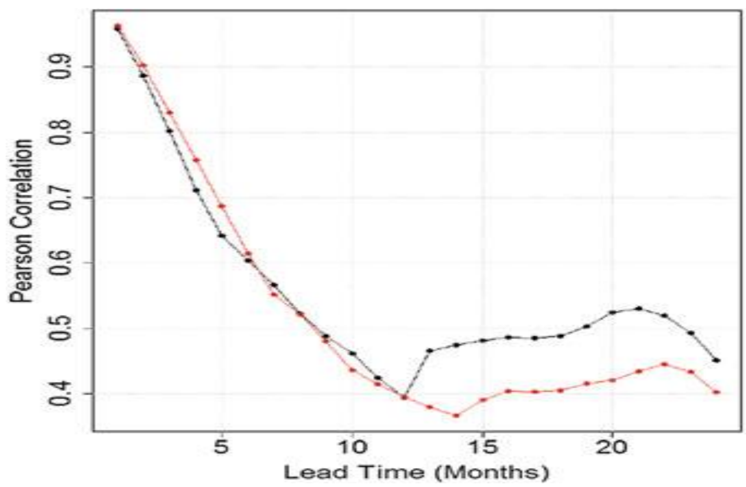
8. Figure 2 shows the COEF and the RMSE between the monthly observation of Niño3.4 index and the hindcasts from models as the individual function of different lead times.

Q: Figure 2: I think we should discuss the correlation is rising for 10 and 11 month lead time. The decrease of COEF (increase of RMSE) from lead time 1-9 is mainly related to the ocean. This could be possibly due to the extratropical signal we propose (may come from SLP). This can be quantified by a similar plot as Fig. 4 if we include all variables used here.

At the longer lead time from 9 to 12 months, the performance of DL models tends to improve, which is different from traditional prediction models.

Q: Yes, we may emphasize this further because this is the key to overcome the problem of Spring barriers.

Ans: Interestingly, we also find the similar results from other study as below shows. But we don't know the reason for this. We may discuss this in detail on Sunday.



Lima, C. H. R., Lall, U., Jebara, T., & Barnston, A. G. (2015). Machine Learning Methods for ENSO Analysis and Prediction. Machine Learning and Data Mining Approaches to Climate Science.

9. Figure 3 shows time series of the Niño3.4 index and the corresponding predictions by multiple models at different lead times.

Q: This plot, do you remove the monthly climatology? ie. Does the seasonal cycle exist? If so, the climatology period should be mentioned earlier.

Ans: We remove the monthly climatology. The seasonal cycle does not exist. I will mention this clearly in the paper.

10. However, the intensity of the La Niña events tends to be amplified based on NMME models, but well captured by DL models. Relatively, DL models are prone to predict La Niña events accurately, rather than El Niño.

Q: Here, we may discuss more about the behavior of the DL models. However, it is not easy to see the change here.

Ans: We find NMME models tend to amplify the intensity of La Niña events. DL models shows the better performance when predict La Niña at the lead time of 3 or 6 months. However, this may due to the deficiency of the NMME models themselves. I think we should think more about this.

11. **Twelve Indices are used here including Niño3.4, MEI, PDO, PNA, QBO, NAO, WWV, WWVW, WWVE, TP300, TPW300, and TPE300. Eleven physical fields include SST, 10-m near surface winds (UW, VW), SLP, surface ocean currents (UVEL, VVEL), sea surface height (SSH), isothermal depth (ISOD), mixed layer depth (MLD), ocean potential temperature averaged over upper 300-m (PTD), and net heat flux through sea surface (QFLUX).**

Q1: Why do you choose these 12 indices? Any reasoning? We need to say a bit about the indices we choose.

Q2: Similarly, why do we choose these? It seems we don't present them all in Fig. 4. Why do you choose these in Fig. 4?

Ans： We intend to find factors as much as possible and let the machine to learn and choose the best precursors from these parameters without human intervention. The number for the physical fields and index used here is related to atmosphere and ocean and is all that I can find. Maybe you have some advice for the selection.

12. **Seeing that ENSO is the signal of SST, we treat the DL model only with the fundamental input of early SST and Niño3.4 index as the control experiment here.**

Q: Why do you combine SST and Nino3.4? They are dependent predictors, right? I suggest to have one for Nino3.4 first and another for SST. It is interesting to discuss the difference.

Ans: ok, maybe we can try next.

13. **This means proper addition of multiple variables can improve the prediction skill and deep learning technique may be prone to grasp the interaction between several precursors.**

Q: Yes, I believe we can use this to say more about why.

Ans: Maybe we can compare the results from different experiments. Experiment can be set as only involve SST, SST+Nino3.4, and SST plus other variables in detail.

14. **The pre-influence of equatorial zonal flow associated with long-wave Rossby wave physics on ENSO events has been discussed by Zhang and Clarke (2017). Wang et al. (2017) have also mentioned the "goal shot" for ENSO development and the effective precursor of zonal current for the ENSO prediction.**

Q: We may elaborate this more. Rossby wave physics can only precondition the off-equator status. It is not the key in the equatorial region (5S-5N). So zonal current has different impacts at different latitudes and longitudes. We can clarify this more.

Ans: Yes. We feel puzzled about why zonal flow shows better performance than other horizontal variables. I think maybe we can also see the evolution of zonal current UVEL as the figure 9 and 10 shows in your study (An ENSO prediction approach based on ocean conditions and ocean–atmosphere coupling). Thus we can explore the relationship between UVEL and other variables such as warm water in central Pacific and wind during different phase of ENSO and give better explanations.

15. **The correlation coefficient is 0.96 at 1-month lead time and can even reach 0.73 at 12-month lead time. The RMSE is only 0.3°C at 1-month lead time, and no more than 0.7°C at 12-month lead time. The performance of UVEL is next to TP300, beats that based on WWV. Such result is also consistent with Wang et al. (2017).**

Q: The excellent performance of the 12-month lead time is a good point to strength further in this manuscript. We should say more about why.

Ans: Yes. This is one of the light spot of this study and we will mention on this more.