

OCNG/ATMO 651 Final Project: Linear Inverse Model of Tropical Sea Surface Temperatures

Jinjun Liu

December 15, 2022

Abstract

In this project, we employ the linear inverse model (LIM) to predict sea surface temperatures anomalies (SSTAs) in the global tropics. The LIM is based on the empirical orthogonal function (EOF) decomposition of the SSTAs. We use the monthly NCEP/NCAR reanalysis SST as the dataset to build the LIM and make predictions. The LIM is more skillful well when the leading time is not large. The tropical averaging correlation coefficient can reach to 0.54 when the leading time is one month. The LIM performs well in the Niño 3 region and Indian Ocean, but not very well in the tropical Atlantic and northwest Pacific. The LIM forecast based on the latest SST observations suggest that the La Niña event will likely continue in the next few months.

Contents

1	Introduction	1
2	Dataset and Method	2
3	Results	4
3.1	EOF decomposition of SSTs	4
3.2	Linear inverse model in EOF space	6
3.3	Prediction based on latest SSTs	8
4	Conclusion and discussion	10
5	Aknowledgements	10

1 Introduction

Sea surface temperatures (SSTs) are important for ocean and atmosphere because they play a significant role in shaping the Earth's climate and weather patterns. Many climate variability modes are directly related to the SSTs. For example, the El Niño-Southern Oscillation (ENSO) is a coupled ocean-atmosphere phenomenon that is characterized by the fluctuation of warming and cooling of the central and eastern equatorial Pacific Ocean [4]. The Atlantic Multidecadal Oscillation (AMO) is a long-term observed climate variability mode that is characterized by altering the warming and cooling phases of the North Atlantic SSTs in multi-decadal time scale [3]. These large-scale climate variability not only affect the global climate but also have a significant impact on the regional climate and weather patterns (e.g., IPCC AR5 [1]). Therefore, it is important to predict the SST future changes in order to make climate predictions.

Penland and Magorian [6] and Penland and Matrosova [7] proposed a linear inverse model (LIM) to predict the SSTs in the region of the equatorial Pacific Ocean and Tropical Atlantic. This method is based on the empirical orthogonal function (EOF) decomposition of the SSTs. They found that the LIM can predict the SSTs in the region of the equatorial Pacific Ocean with a root mean square error (RMSE) of half a degree Celcius when the leading time of 9 months is used. In the Caribbean Sea and north tropical Atlantic, Penland and Matrosova [7] found that the LIM can predict the SSTs better when using the global tropical SSTs as predictors compared only using tropical Atlantic SSTs as predictors. They also found that the SST anomaly in the Niño 3 region is a strong 6-month predictor of SST anomaly in the north tropical Atlantic. Penland [5] stated that the "optimal initial structure" predicted by LIM is approximately seven months before most of the major warm and cold (El Niño/La Niña) events in the data record.

Richter et al. [8] examine the impact of impacts of general circulation models (GCMs) errors on prediction skill as estimated by the LIM. It is found that the observation-derived LIM is more skillful than a simple persistence forecasts in most regions. Its skill also compares well with some GCM forecasts except in the equatorial Pacific, where the GCMs are superior. This indicates that the LIM performs well even compared with the complicated GCMs.

In this project, we employ the LIM to predict the SST anomalies (SSTAs) in the global tropics. The detailed method of the LIM is described in Section 2. The results of the LIM are shown in Section 3. The discussion and conclusion are presented in Section 4.

2 Dataset and Method

We use the NCEP/NCAR (National Centers for Environmental Prediction/National Center for Atmospheric Research) reanalysis SST as the dataset to build the LIM and make predictions. This dataset used in this project contains monthly-mean SSTs from January 1948 to September 2017. We divide the dataset into two parts: the training dataset and the test dataset. The training dataset contains the SSTs from January 1948 to December 1999, and the test dataset contains the SSTs from January 2000 to September 2017. We use the training dataset to build the LIM and use the test dataset to evaluate the performance of the LIM.

The latest SST observations (November 2022) is obtained from the monthly COBE SST (Daily Sea Surface Analysis for Climate Monitoring and Predictions) dataset [2]. Horizontal resolution of COBE SST is 1.0° . To use the LIM built from the NCEP/NCAR reanalysis SST, we interpolate the COBE SST to the same grid as the NCEP/NCAR reanalysis SST. The interpolation is done by using the bilinear interpolation method.

The LIM assumes that the time evolution of the SSTs can be described as a linear differential equation:

$$\frac{d\mathbf{T}}{dt} = \mathbf{A}\mathbf{T} + \xi \quad (1)$$

where \mathbf{T} is the SST vector, \mathbf{A} is the linear operator, and ξ is the noise. The LIM assumes that the noise ξ is small and can be neglected. Then the evolution of \mathbf{T} from time t to time $t + \Delta t$ can be approximated as:

$$\mathbf{T}(t + \Delta t) = e^{\mathbf{A}\Delta t}\mathbf{T}(t) + \varepsilon \quad (2)$$

where $e^{\mathbf{A}\Delta t}$ is called the propagator matrix, and ε is the noise. By left-multiplying the tranpose of SST vector $T'(t)$ on both sides of Eq. (2), we have:

$$\langle T(t + \Delta t)T'(t) \rangle = e^{\mathbf{A}\Delta t} \langle T(t)T'(t) \rangle + \langle \varepsilon T'(t) \rangle \quad (3)$$

where $\langle \cdot \rangle$ denotes the time average. $T(t)T'(t) = \mathbf{C}$ is the covirance matrix of the SSTs. $T(t + \Delta t)T'(t) = C_{\Delta t}$ is the lag-covirance at Δt . If we neglect the noise term, we have:

$$C_{\Delta t} = e^{\mathbf{A}\Delta t}\mathbf{C} \quad (4)$$

Then the propagator can be determined as:

$$\mathbf{G} = e^{\mathbf{A}\Delta t} = C_{\Delta t}\mathbf{C}^{-1} \quad (5)$$

Once we have constructed the propagator matrix G from the observational data, we can use it to predict the SSTs in the future. The prediction of the SSTs at time $t + \Delta t$ is given by:

$$\mathbf{T}(t + \Delta t) = \mathbf{G}^{\frac{t_n - t_0}{\Delta t}} \mathbf{T}(t_0) \quad (6)$$

where t_0 is the initial time, t_n is the prediction time, and Δt is the lag time. According to Penland and Magorian [6], we take the lag time as 7 months. However, the physical space of SST could be too large, making it difficult to estimate the propagator matrix. Therefore, we construct the LIM in EOF space. The EOF decomposition of the SSTs can be written as:

$$\mathbf{T}(t) = \sum_{i=1}^M P_i(t) \mathbf{E}_i, \quad (7)$$

where $\mathbf{T}(t)$ is the SST vector, $P_i(t)$ is the principal components (PCs) of the i th EOF mode, and \mathbf{E}_i is the i th EOF mode. M is the total number of EOF modes. In EOF space, the LIM can be written as:

$$\mathbf{P}(t + \Delta t) = \mathbf{G}^{\frac{t_n - t_0}{\Delta t}} \mathbf{P}(t_0) \quad (8)$$

where $\mathbf{P}(t)$ is the PCs of the EOF modes. After predicting the PCs of the EOF modes, we can reconstruct the SST vector by Eq. (7).

The operational steps are listed as follows:

1. Smooth SST anomalies (SSTAs) using a 3-month running mean and then perform an EOF analysis on the smoothed SSTAs;
2. Reconstruct and truncate SSTAs using leading EOFs and the corresponding principal component time series as Eq. 7. Typically, M is less than 10 and can be determined by the explained variance, say 90%, by the first M EOFs;
3. Compute covariance $\check{\mathbf{C}}$ and lag-covariance $\check{\mathbf{C}}_{\Delta t}$ matrix in EOF space. $\check{\mathbf{C}} = \langle \mathbf{P}(t) \mathbf{P}'(t) \rangle$ and $\check{\mathbf{C}}_{\Delta t} = \langle \mathbf{P}(t + \Delta t) \mathbf{P}'(t) \rangle$;
4. Compute the propagator in EOF space, $\mathbf{G} = \check{\mathbf{C}}_{\Delta t} \check{\mathbf{C}}^{-1}$, where the lag Δt can be taken to be 7 months and \mathbf{G} has $M \times M$ dimension;
5. Take an initial SSTA state vector \mathbf{T} at any t_0 , regress $\mathbf{T}(t_0)$ onto \mathbf{E}_i to determine $\mathbf{P}_i(t_0)$ and then form $\mathbf{P}(t_0)$;
6. Make SST forecast in EOF space from t_0 to t_n : $\mathbf{P}(t_n) = \mathbf{G}^{\frac{t_n - t_0}{\Delta t}} \mathbf{P}(t_0)$;
7. Transform the forecast in EOF space, $\mathbf{P}(t_n)$, back to the physical space: $\mathbf{T}(t_n) = \sum_{i=1}^M \mathbf{P}(t_n) \mathbf{E}_i$.

The Python script that processes the data and generates the figures is submitted

as a supplementary file.

3 Results

3.1 EOF decomposition of SSTs

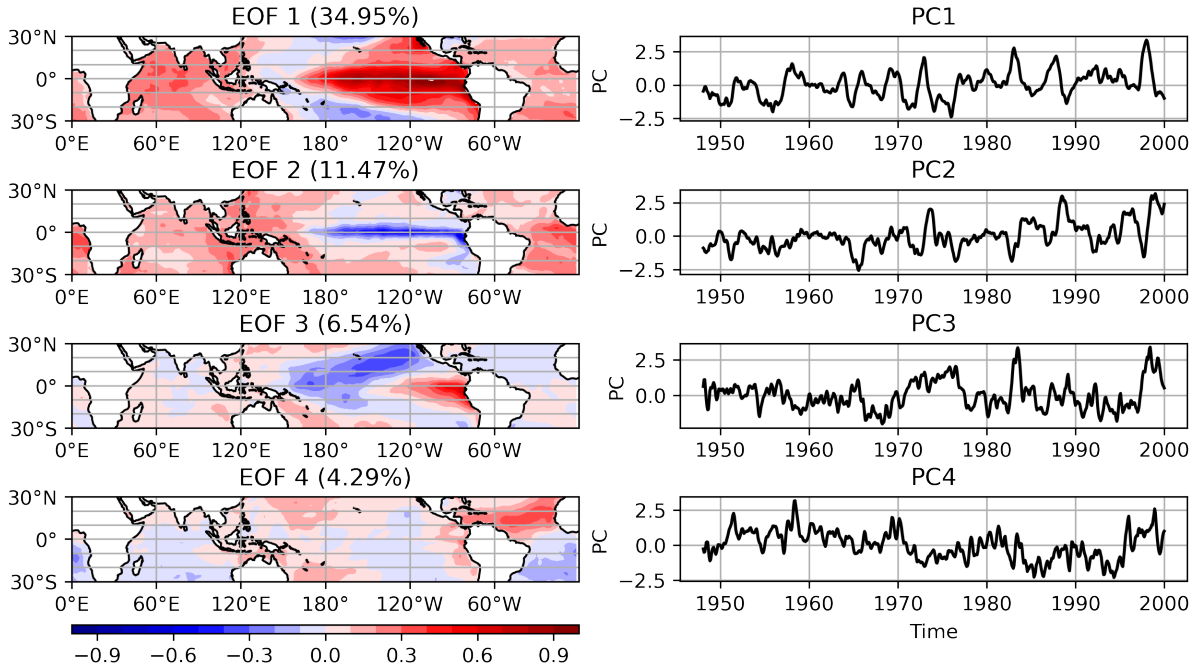


Figure 1: EOF decomposition of SSTs.

The leading four EOFs and PCs are plotted in Fig. 1. The first EOF mode explains 34.95% of the variance of the SSTAs, and the second EOF mode explains 11.47% of the variance of the SSTAs. The first 10 EOFs can explain 72.43% of the variance of the SSTAs, the first 15 EOFs can explain 79.45% of the variance of the SSTAs, and the first 20 EOFs can explain 83.89% of the variance of the SSTAs (Fig. 2). When the EOF number reaches 15, the cumulative explained variance is 79.45%, which is close to the 80% threshold. Therefore, we choose the first 15 EOFs to reconstruct the SSTAs.

To test if the first 15 EOFs can reconstruct the SSTAs, we reconstruct the SSTAs using the first 15 EOFs and corresponding PCs and plot the reconstructed SSTAs in Fig. 3. We selected four times spanning from 1949-01 to 1999-12 and varies in months. The reconstructed SSTAs are very similar to the original SSTAs, which indicates that the first 15 EOFs can reconstruct the SSTAs.

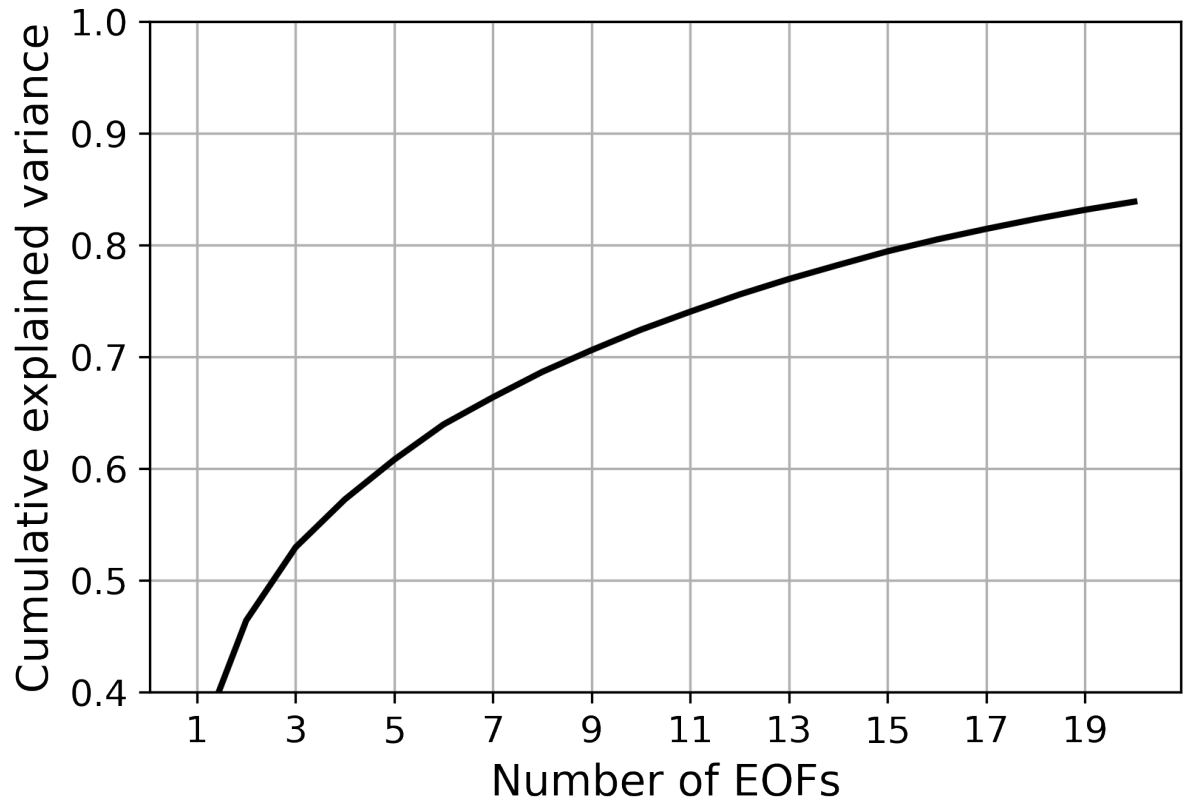


Figure 2: Cumulative explained variance of the first 20 EOFs.

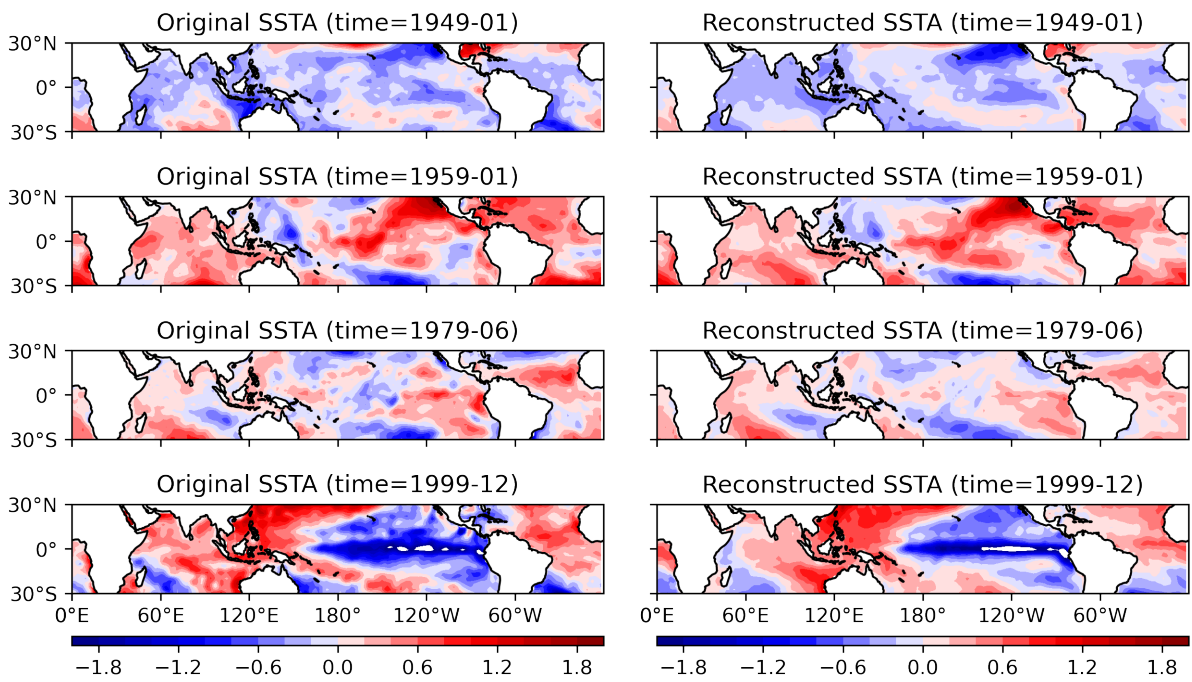


Figure 3: Reconstructed SSTAs using the first 15 EOFs.

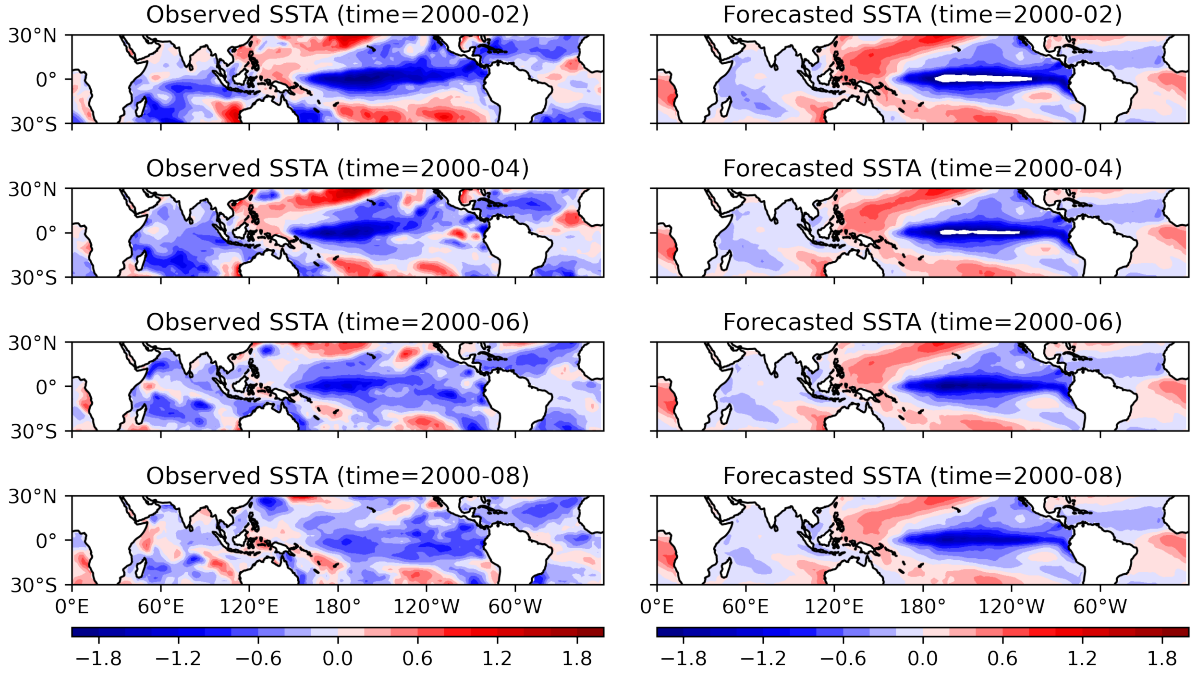


Figure 4: Forecasted SSTAs in 2000.

3.2 Linear inverse model in EOF space

We use the training data (1949-01 to 1999-12) to build the LIM. To test the performance of the model, we firstly picked 2000-01 as the initial time to apply LIM to predict the following 10 months SSTAs. The forecasted SSTAs are plotted in Fig. 4 (only 4 forecasted SSTAs are plotted). Compared to the observed SSTAs, the forecasted SSTAs are very similar to the observed SSTAs in the first a few (i 3) months. However, the forecasted SSTAs are not as good as the observed SSTAs in the later months. Therefore, the LIM is accurate when forecasting the near future SSTAs, but not accurate when forecasting the far future SSTAs.

To systematically evaluate how the leading time would influence the performance of the LIM, we calculated the correlation between the forecasted SSTAs and the observed SSTAs at different leading times. The correlation maps are plotted in Fig. 5. In general, the correlation is pretty high in the Niño 3 region even the leading time is large. The LIM also performs well in the Indian Ocean. However, the LIM performs poorly in the north Atlantic (especially in the Mexico Gulf) and tropical Atlantic. In the Northwest Pacific, the LIM also perform poorly.

When the leading time increases, the correlation decreases. Fig. 6 shows the space mean correlation between the forecasted SSTAs and the observed SSTAs at different leading times. The space mean correlation monotonically decreases with the leading time. The correlation is the highest when the leading time is 1 month (about 0.54), and

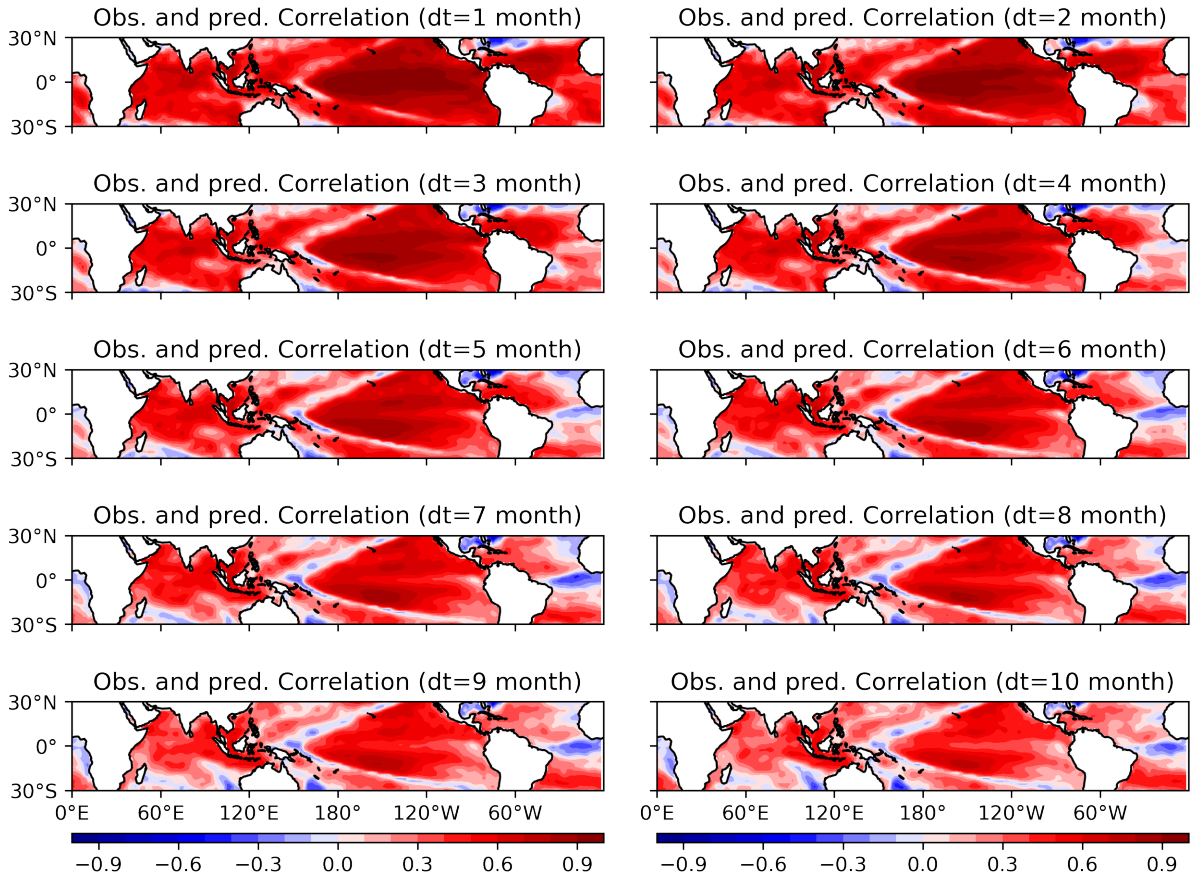


Figure 5: Correlation between the forecasted SSTAs and the observed SSTAs at different leading times.

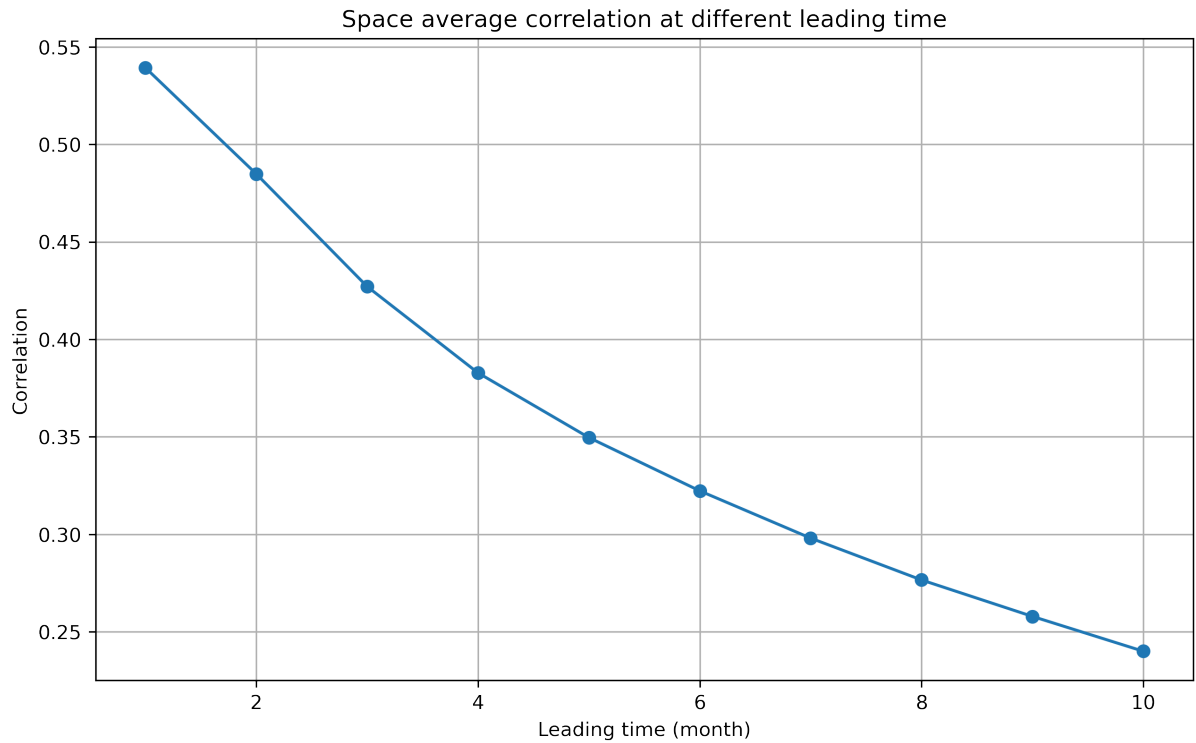


Figure 6: Space mean correlation between the forecasted SSTAs and the observed SSTAs at different leading times.

the correlation is the lowest when the leading time is 10 months (about 0.23).

3.3 Prediction based on latest SSTs

Fig. 7 shows the forecasted SSTAs based on the latest SSTs. The initial SSTA are taken from the COBE SST dataset. To get the anomalies, we use the latest 20 years data to calculate the climatology. From the forecasted SSTA map, we can see we are in a La Niña event. And the in next 10 months, the La Niña event is likely to continue. But due to the LIM is not accurate when forecasting the far future SSTAs, we do not have much confidence on the SSTAs in the later spring in next year. Furthermore, because of the existence of the "ENSO spring barrier", it is hard to predict the SSTAs in the later spring in next year.

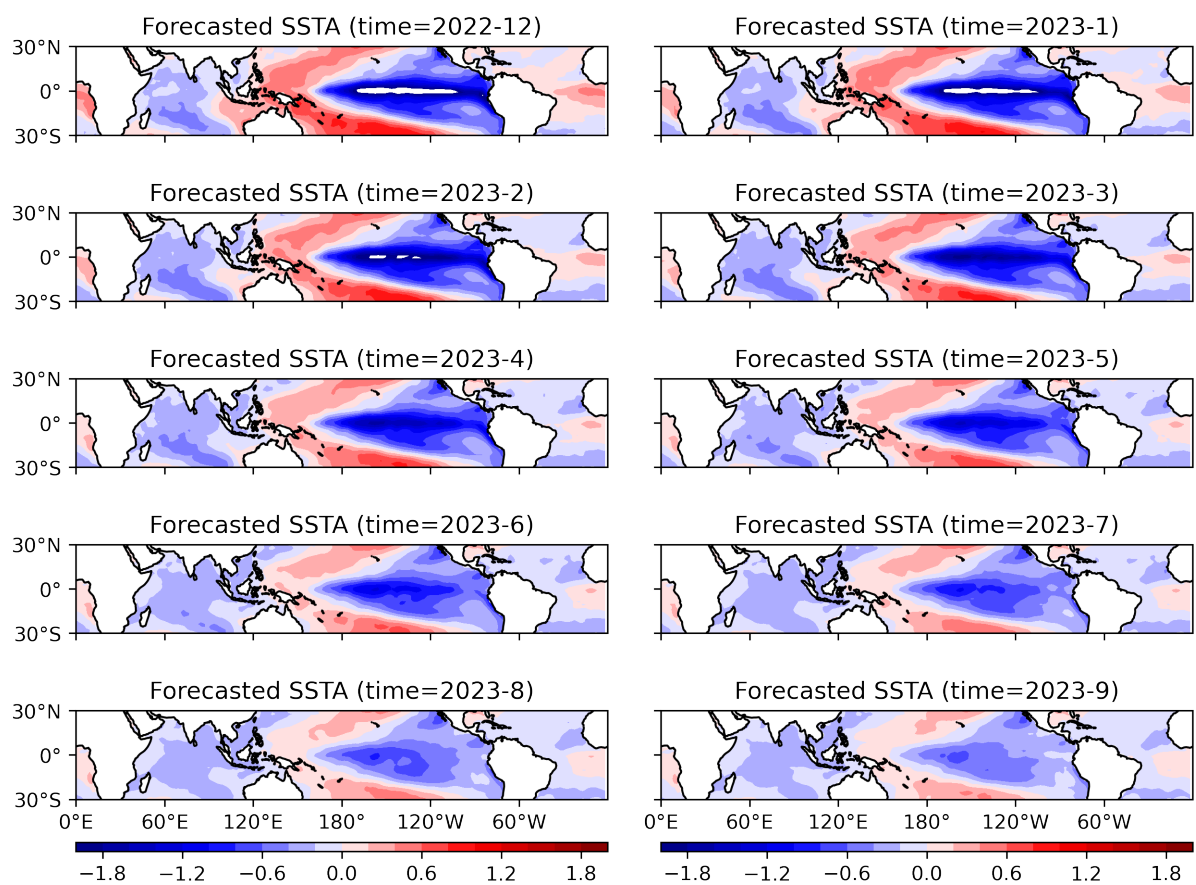


Figure 7: Forecasted SSTAs based on the latest SSTs.

4 Conclusion and discussion

In this project, we build a linear inverse model (LIM) to predict the SSTAs in the global tropics. The LIM is built in the EOF space. The LIM is accurate when forecasting the near future SSTAs, but not accurate when forecasting the far future SSTAs. The LIM is accurate in the Niño 3 region and the Indian Ocean, but not accurate in the north Atlantic (especially in the Mexico Gulf) and tropical Atlantic. In the Northwest Pacific, the LIM also perform poorly. This is because the LIM is built based on EOF decomposition, and the strongest EOF mode is the ENSO. Therefore, the LIM can capture the ENSO signal well, but cannot capture the other signals well.

5 Acknowledgements

Thanks to Dr. Ping Chang for providing the datasets and the guidance. Thanks to Ms. Qiuying Zhang for helping me with the debugging of the Python script.

References

- [1] J. H. Christensen, K. K. Kanikicharla, E. Aldrian, S. I. An, I. F. A. Cavalcanti, M. de Castro, W. Dong, P. Goswami, A. Hall, J. K. Kanyanga, et al. Climate phenomena and their relevance for future regional climate change. In *Climate change 2013 the physical science basis: Working group I contribution to the fifth assessment report of the intergovernmental panel on climate change*, pages 1217–1308. Cambridge University Press, 2013.
- [2] M. Ishii, A. Shouji, S. Sugimoto, and T. Matsumoto. Objective analyses of sea-surface temperature and marine meteorological variables for the 20th century using icoads and the kobe collection. *International Journal of Climatology*, 25:865–879, 6 2005.
- [3] J. R. Knight, C. K. Folland, and A. A. Scaife. Climate impacts of the atlantic multi-decadal oscillation. *Geophysical Research Letters*, 33:17706, 9 2006.
- [4] M. J. McPhaden, S. E. Zebiak, and M. H. Glantz. Enso as an integrating concept in earth science. *Science*, 314:1740–1745, 12 2006.
- [5] C. Penland. A stochastic model of indopacific sea surface temperature anomalies. *Physica D: Nonlinear Phenomena*, 98:534–558, 11 1996.

- [6] C. Penland and T. Magorian. Prediction of niño 3 sea surface temperatures using linear inverse modeling. *Journal of Climate*, 6:1067–1076, 1993.
- [7] C. Penland and L. Matrosova. Prediction of tropical atlantic sea surface temperatures using linear inverse modeling. *Journal of Climate*, 11:483–496, 1998.
- [8] I. Richter, P. Chang, and X. Liu. Impact of systematic gcm errors on prediction skill as estimated by linear inverse modeling. *Journal of Climate*, 33:10073–10095, 2020.