# Assessing Tropical Pacific-induced Predictability of Southern California Precipitation Using a Novel Multi-input Multi-output Autoencoder

Linsey Sara Passarella[1] and Salil Mahajan[2]

[1]University of Tennessee
[2]Oak Ridge National Laboratory (DOE)

November 26, 2022

**Abstract**

We construct a novel Multi-Input Multi-Output Autoencoder-decoder (MIMO-AE) to capture the non-linear relationship of Southern California precipitation (SC-PRECIP) and tropical Pacific Ocean sea surface temperature (TP-SST). The MIMO-AE is trained on both monthly TP-SST and SC-PRECIP anomalies simultaneously. The co-variability of the two fields in the MIMO-AE shared nonlinear latent space can be condensed into an index, termed the MIMO-AE index. We use a transfer learning approach to train a MIMO-AE on the combined dataset of 100 years of output from a historical simulation with the Energy Exascale Earth Systems Model version 1 (E3SMv1) and a segment of observational data. We further use Long Short-Term Memory (LSTM) networks to assess sub-seasonal predictability of SC-PRECIP using the MIMO-AE index. We find that the MIMO-AE index provides enhanced predictability of SC-PRECIP for a lead-time of up-to four months as compared to Nino 3.4 index and the El Nino Southern Oscillation Longitudinal Index.

# Assessing Tropical Pacific-induced Predictability of Southern California Precipitation Using a Novel Multi-input Multi-output Autoencoder

**Linsey S. Passarella**[1], **Salil Mahajan**[1]

[1]Computational Earth Sciences, Oak Ridge National Laboratory, Oak Ridge, TN, USA

**Key Points:**

- We design a novel MIMO-AE to capture the non-linear relationships between tropical Pacific SSTs and Southern California precipitation.
- We use long-short term memory models of a MIMO-AE derived index to assess predictability of Southern California precipitation.
- MIMO-AE offers statistically significant improvement in predictive skill of Southern California precipitation on sub-seasonal scales.

Corresponding author: Linsey S Passarella, `passarellals@ornl.gov`

## Abstract

We construct a novel Multi-Input Multi-Output Autoencoder-decoder (MIMO-AE) to capture the non-linear relationship of Southern California precipitation (SC-PRECIP) and tropical Pacific Ocean sea surface temperature (TP-SST). The MIMO-AE is trained on both monthly TP-SST and SC-PRECIP anomalies simultaneously. The co-variability of the two fields in the MIMO-AE shared nonlinear latent space can be condensed into an index, termed the MIMO-AE index. We use a transfer learning approach to train a MIMO-AE on the combined dataset of 100 years of output from a historical simulation with the Energy Exascale Earth Systems Model version 1 (E3SMv1) and a segment of observational data. We further use Long Short-Term Memory (LSTM) networks to assess sub-seasonal predictability of SC-PRECIP using the MIMO-AE index. We find that the MIMO-AE index provides enhanced predictability of SC-PRECIP for a lead-time of up-to four months as compared to Niño 3.4 index and the El Niño Southern Oscillation Longitudinal Index.

## Plain Language Summary

Traditional El Niño Southern Oscillation indices, like the Niño 3.4 index, although well-predicted themselves, fail to offer reliable sub-seasonal to seasonal predictions of Western US precipitation. Here, we use a machine learning approach called a multi-input, multi-output autoencoder to capture the relationship between tropical Pacific and Southern California precipitation and project it onto a new index, which we call MIMO-AE index. Using machine learning based time-series predictions, we find that MIMO-AE index offers enhanced predictability of Southern California precipitation up-to a lead time of four months as compared to other ENSO indices.

## 1 Introduction

While El Niño-Southern Oscillation (ENSO) is a prominent predictor of precipitation over California, extracting sub-seasonal and seasonal predictability afforded from it remains a challenge (e.g. LHeureux et al., 2021; Pan et al., 2019; S. Wang et al., 2017). This was apparent during the 2015-16 Central Pacific (or Modoki) El Niño event, when California received just above average precipitation. This was in contrast to the forecast of heavy precipitation, which occurred there during the canonical (Eastern Pacific) 1982-83 and 1997-98 strong El Niño events (e.g. Cohen et al., 2017; Lee et al., 2018; LHeureux et al., 2017). Perfect model studies with dynamical models suggest that the potential predictability of Western US precipitation on sub-seasonal to seasonal timescales maybe larger than the observed forecast skills of dynamical and statistical models (Becker et al., 2014). Although, dynamical models capture the chaotic and non-linear nature of the climate system, their predictive skill is limited by systematic model biases (largely originating from the errors in the representation of sub-grid scale processes that grow rapidly) and from complications of model initialization from sparse observations of the coupled system.

Over California, statistical modeling suggests that tropical Pacific sea surface temperatures (TP-SSTs) offer predictability largely only for Southern California precipitation (SC-PRECIP) explaining about 20% of the variability there on seasonal to interannual timescales (e.g. Jong et al., 2016; X. Huang & Ullrich, 2017; G. Wang et al., 2021; Cheng et al., 2021). However, ENSO-induced predictability of regional climate using statistical models has largely been assessed from the linear relationship with ENSO, using linear regression or singular value decomposition, ignoring the inherent non-linearity of the climate system. Although, some studies have used non-linear machine learning approaches to study ENSO associated atmospheric teleconnections (e.g. Hsieh et al., 2006; Wu et al., 2005).

Further, traditional representations of ENSO in these linear statistical models, include spatial averages over specific regions of the tropical Pacific like the Niño 3.4 index, or use linear empirical orthogonal functions. These approaches prove to be inadequate in capturing the full spectrum of spatial variability of ENSO's SST pattern and the associated diversity of remote responses affecting regional climate predictability (e.g. Trenberth & Stepaniak, 2001; Williams & Patricola, 2018). Several recent studies have explored methods to better capture ENSO's variability, diversity and non-linearity (e.g. Williams & Patricola, 2018). However, these approaches largely devise indices that represent the oceanic or atmospheric variability over the tropical Pacific in isolation of its remote teleconnections.

Machine learning methods, like autoencoders, allow identification of dominant non-linear variability and co-variability patterns that might offer enhanced predictability. Autoencoders are artificial neural networks that regenerate the original data from efficient representations (encodings) of the data like principal component analysis (PCA). They, however, transform data to non-linear latent spaces via non-linear activation functions, thus imparting the additional capability of capturing the underlying non-linear relationships within the data (e.g. Y. Wang et al., 2016; Charte et al., 2018; Masti & Bemporad, 2018). Studies show that autoencoders can better detect dominant variability patterns over other techniques, like the PCA (e.g. Y. Wang et al., 2016; Zamparo & Zhang, 2015; Fournier & Aloise, 2019). Some studies (e.g. Tang & Hsieh, 2003; He & Eastman, 2020) have also demonstrated the use of autoencoders to effectively identify modes of climate variability, including those related to ENSO.

Further, multitask learning (MTL) solves multiple learning tasks at the same time while exploiting commonalities and differences across tasks (e.g. Caruana, 1997). MTL has been applied to many problems including natural language processing (Collobert & Weston, 2008; Liu et al., 2017; S. Chen et al., 2021), speech recognition (Deng et al., 2013; Kim et al., 2017; Toshniwal et al., 2017; Shinohara, 2016) and computer vision (Girshick, 2015; Devries et al., 2014; Kendall et al., 2018) to improve prediction accuracy and learning efficiency of task-specific models. Recent studies have shown the usefulness of multi-input and/or multi-output networks for segmenting data and extracting useful information when there are multiple variables present (Raza et al., 2017; Yaguchi et al., 2020; Ghifary et al., 2015). For example, Ghifary et al. (2015) used a multi-output autoencoder, which they call a multi-task autoencoder (MTAE), for domain generalization. The MTAE has a single input variable with multiple outputs where the input-hidden weights represent variable shared parameters and the hidden-output weights represent domain-specific parameters. MTAE learns features shared across all domains.

Here, we expand on the MTAE approach and construct a novel multi-input multi-output autoencoder (MIMO-AE) to effectively extract the most prominent shared features between monthly TP-SSTs and SC-PRECIP anomalies and capture their underlying non-linear relationship using an Earth System Model (ESM) simulation and observational data. Our network architecture is designed to yield a temporal index of the co-variability of the two variables. We further use Long Short-term Memory (LSTM) models to predict this monthly index, which we decode to generate predicted SC-PRECIP, and evaluate its predictive skill. We show that MIMO-AE can be a powerful tool to isolate important teleconnections and wield enhanced sub-seasonal regional predictability.

### 1.1 Model Simulations and Data

We use a 165 years-long historical simulation of the Energy Exascale Earth System Model version 1 (E3SMv1) (E3SM Project, 2018), and utilize the first 100 years of the simulation for training the MIMO-AE network. E3SMv1 is found to effectively capture temporal variability of ENSO and reproduce ENSO associated spatial SST patterns when compared to observational datasets (Golaz et al., 2019), although with a larger westward extent of SST anomalies during El Nino events. It also simulates the teleconnections of ENSO to US winter season precipitation well (Mahajan et al., 2021). We use observed precipitation data from NOAA's PRECipitation REConstruction over Land (PREC/L) at 1° resolution (M. Chen et al., 2002). PREC/L is a global analysis of interpolated rain gauge observations from 1948 to 2020. We use observed SSTs for the same period from the Hadley Centre Global Sea Ice and Sea Surface Temperature (HadISST 1.1) dataset available at a 1° resolution (Rayner et al., 2003).

## 2 Methodology

### 2.1 Autoencoder

An autoencoder is an unsupervised neural network that is trained to learn an identity function, a function that returns the same value as its input. It aims to efficiently compress and encode data by minimizing the reconstruction error. A simple autoencoder, shown in figure 1a, contains a hidden layer $h$ that describes a representation of important attributes of the input (e.g. Goodfellow et al., 2016). The general autoencoder consists of two parts: an encoder and a decoder. The encoder maps input $x$ to $h$ by a chosen activation function $f()$,

$$h = f(x \cdot w_e) \tag{1}$$

where $w_e$ are the encoder weights. The decoder then maps $h$ to the reconstruction of $x$, represented by $x'$:

$$x' = f(h \cdot w_d) \tag{2}$$

where $w_d$ are the decoder weights.

By using a linear activation function, the single hidden layer autoencoder behaves similarly to a PCA (e.g. Bourlard & Kamp, 1988; Plaut, 2018). The number of hidden layers can also be increased to create a deep autoencoder, with the middle layer often referred to as the bottleneck layer. Tang and Hsieh (2003) used a simple autoencoder to extract the leading nonlinear mode of interannual variability of upper ocean heat content over the tropical Pacific, with a single node bottleneck hidden layer, to reveal an asymmetry in the spatial pattern between characteristic El Niño and La Niña episodes. For spatio-temporal data, the temporal vectors at the bottleneck nodes are analogous to principal components (PC) of PCA. The value of a temporal vector at a given time $t$ results from passing the spatial data at $t$ through the network. The non-linear activation functions imply that the spatial pattern derived from reconstructing the data using the decoder varies with the magnitude of the temporal vector at $t$, unlike PCs which yield a standing spatial pattern (e.g. Tang & Hsieh, 2003).

#### 2.1.1 MIMO-AE

Figure 1b illustrates our MIMO-AE architecture designed to extract the non-linear relationship between TP-SSTs and SC-PRECIP on monthly timescales. The encoder consists of two separate input temporal vectors (TP-SST and SC-PRECIP) that are passed through two hidden layers before concatenating and passing through a single hidden node. The input (and output) vectors represent SST anomalies at each grid box within the boxed domain over tropical Pacific (20°N to 20°S, 120°E to 70°W) and precipitation anomalies over each grid box in the boxed domain over Southern California (32°N to 35°N, 120°W to 114°W) (Figure 1b). The first hidden layer, consisting of 50 nodes each for the two variables, can be thought of as feature extraction of the original data. The next hidden layer then shrinks the data to 10 hidden nodes, again separately for the two variables, in order to reduce the computational complexity of data. This data is then passed to a single hidden node that is shared by the two input variables. This hidden node represents the shared non-linear latent structure of both the SST and precipitation vectors. The vectors are then split back into two from the shared hidden node and passed through the decoder, which is identical in structure (with different weights) to the encoder, to reconstruct back to the original shape in the output layer. We use the "tanh", or hyperbolic tangent, activation function for all the hidden layers.

We performed several iterations of the network design with different number of hidden layers, neurons and activation functions and chose the MIMO-AE architecture (described above) that exhibited a low value of the training loss function as well as explaining a large fraction of the variability of SC-PRECIP. The loss is calculated by using a mean squared error (MSE) using the following equation:

$$MSE = \frac{1}{N}\sum_{i}(P_i - T_i)^2 \tag{3}$$

where $P_i$ is the predicted value of the reconstructed data at point $i$ and $T_i$ is the true value of the data at point $i$, which here is the original input data. The input variables are scaled using a min-max scaler before training is performed. MIMO-AE was trained on first 100 years of the E3SM simulation data for 100 epochs with an AdaGrad loss optimizer using tensorflow on one CPU node on the National Energy Research Scientific Computing Center's (NERSC) Cori super computer. The training loss for the scaled TP-SST reconstruction (orange) and SC-PRECIP (blue) are shown in figure 1c. We refer to this MIMO-AE network as MIMO-AE-E3SM, hereafter.

Figure 1d and e show the $R^2$ values (fraction of variance explained) between the reconstructions from the MIMO-AE and the original data for the 100 years of training data for TP-SSTs and SC-PRECIP respectively. The MIMO-AE explains more than 80% of the variability of Southern California for most grid points and about 20% of the variability of TP-SSTs over most of the domain. The relatively weaker explained variability of MIMO-AE over tropical Pacific is an artefact of our network design preference. We ad-hocly chose a network that explained a larger fraction of the variability of SC-PRECIP, while also capturing the tele-connections to TP-SST, since our goal was largely to assess predictability of SC-PRECIP here. Likewise, a network where TP-SST variability dominates can also be picked if needed. In the future, we plan to make the network design more systematic, for example, by adding a penalizing term for explained variability of each field in the loss function.

We refer to the temporal vector of the single node bottleneck layer that represents the dominant non-linear mode of co-variability of TP-SSTs and SC-PRECIP as the MIMO-AE index, hereafter. We apply the MIMO-AE-E3SM trained on 100 years of E3SM historical simulation on the latter 65 years of the run. As a form of transfer learning, we combine the first 100 years of the E3SM simulation with 32 years of observational data (1948-1979) to train another MIMO-AE network for application to remaining observational data (1980-2020), termed MIMO-AE-OBS. Although, we find that using MIMO-AE-E3SM on observational data imparted similar predictability skills (Results section) as MIMO-AE-OBS for observational data. Ham et al. (2019) also used a transfer learning approach, whereby, they train a convolutional neural network (CNN) with global SST and heat content data from historical simulations of 21 CMIP5 models. They retrained the network with observational data but with weights initialized from the CMIP5-trained network, which was then used to predict the observed Niño 3.4 index. While we have not investigated their approach to transfer learning in our exploratory study of MIMO-AE here, we plan to apply this and other transfer learning methods to MIMO-AE in the future.

## 2.2 LSTM

To study predictability, we also train long-short term memory (LSTM) recurrent neural networks as our time series prediction models. LSTMs are a special kind of recurrent neural network that learn long term dependencies whose cells are constructed with internal mechanisms called "gates" that control the flow of information through the cell (Hochreiter & Schmidhuber, 1997). There are three types of gates: forget, input and output. These allow for the model to learn what features in the data are important to keep or throw away before passing it down the line to the next cell. LSTM models have recently been shown to perform better at time series prediction over linear models for Niño 3.4 index (A. Huang et al., 2019; Mu et al., 2020; Broni-Bedaiko et al., 2019; Gupta et al., 2020), and we use them here to evaluate the predictability of MIMO-AE index as well as SC-PRECIP.

LSTM models are constructed individually for each of the time series of MIMO-AE index, Niño 3.4 index, ELI and regionally averaged SC-PRECIP anomalies using the first 100 years of the E3SM data. We train separate LSTMs for the above listed time series using the first 32 year segment (1948-1979) of observational dataset used. Given a predicted value of MIMO-AE index, predicted SC-PRECIP (and TP-SSTs) can be constructed by passing the index through the decoder of MIMO-AE. We optimize the LSTM architecture by choosing the number of hidden nodes that maintains a low training loss for all indices, found to be 100 nodes. We train separate LSTMs for each of the forecast lead times ranging from 1 to 12 months and evaluate their predictive skill on the remaining 65 years of E3SM data and the 41 years of observational data.

# 3 Results

## 3.1 MIMO-AE

Figure 2a shows the the three-month moving average of the standardized MIMO-AE index time-series for a segment (last 40 years, 1974-2013) of the 65 years of the E3SM testing data. MIMO-AE index was generated by passing the TP-SST and SC-PRECIP data through the MIMO-AE network trained on the prior 100 years of simulation. Also, shown are the three-month moving average time series of standardized Niño 3.4 index, ELI and domain averaged SC-PRECIP. The correlations of each time-series against domain averaged SC-PRECIP is also listed for the smoothed data. Fig. 2b shows the same but for a segment of the observational data (1980-2019) using MIMO-AE-OBS. To reiterate, MIMO-AE index for observations is computed by passing the observed data through the MIMO-AE-OBS network.

For both E3SM and observations, the correlation between SC-PRECIP and MIMO-AE index is higher than that between SC-PRECIP and Niño 3.4 index or ELI, given that precipitation data is fed in the generation of MIMO-AE index and MIMO-AE explains a large fraction of the SC-PRECIP variability. The correlation between MIMO-AE and both Niño 3.4 index and ELI is weak both for E3SM (0.35 and 0.27 respectively) and observational data (0.43 and 0.39 respectively). However, the correlation between MIMO-AE and Niño 3.4 is higher than the correlation between SC-PRECIP and Niño 3.4. The above correlations are indicative of the shared variability captured by MIMO-AE. Further, in observational data, all indices spike during the 1982-83 and 1996-97 El Niño events, but only the Niño 3.4 peaks during the 2015-16 El Niño event. Thus, MIMO-AE also categorizes the 2015/2016 event weaker than the Niño 3.4, similar to the ELI index (Williams & Patricola, 2018).

Fig. 2c and d show the probability density functions of the Niño 3.4, ELI, MIMO-AE index and domain averaged SC-PRECIP for E3SM testing data and observations. While the Niño 3.4 index tends to be symmetric, the ELI is skewed towards the left (westwards), both for E3SM data and observations as noted by Williams and Patricola (2018), with a thicker right tail (eastwards). ELI is a non-linear SST-based index and represents the average longitude of deep convective activity over the tropical Pacific governing the Rossby wave trains originating from there. MIMO-AE which represents the shared covariability between the TP-SSTs and SC-PRECIP also shows a leftwards skewed distribution with a larger number of strong positive events than strong negative events. The leftwards skewness may follow from the density function of precipitation that is naturally skewed leftwards, even for monthly average data (e.g. Mahajan et al., 2012). But, it could also be reflective of the skewed relationship between TP-SSTs and SC-PRECIP, with some events over the tropical Pacific triggering extreme positive anomalous events in SC-PRECIP. While, co-variability between the two remains weaker during strong negative SC-PRECIP anomalous events. The skewness of MIMO-AE index is noted to be stronger in E3SM than in observations.

Fig. 2e and f show the composite of reconstructions of TP-SST and SC-PRECIP during the strongest 10 positive and negative monthly MIMO-AE index values for both the E3SM testing data. Strong (negative) SC-PRECIP anomalies during those events are associated with strong positive (negative) anomalies in central tropical Pacific and northeast tropical Pacific and weak positive (negative) anomalies in the Eastern tropical Pacific. Similar patterns are noted for the strongest positive and negative MIMO-index values for reconstructions of TP-SST when observation data is passed through the MIMO-AE network (Fig. 2f).

Fig. 2g shows the December to February average reconstructions for the three strongest El-Niño events (1981-82, 1997-98, 2015-16) in observations. It is apparent that the spatial patterns of these reconstructions are not standing - more clearly here than the composite plots (Fig. 2e,f) - with varying contour patterns of SST anomalies for each of the three events. The 2015-16 El Niño events is associated with weak positive anomalies in the MIMO-AE latent space for SC-PRECIP and TP-SST over much of tropical Pacific. In contrast, the 1981-82 and 1997-98 events are associated with strong positive anomalies both in SC-PRECIP and TP-SST, with the stronger 1997-98 SC-PRECIP anomalies associated with stronger TP-SST anomalies and varied contour delineations. When a separate MIMO-AE network is trained on all of the observation data (1948-2020) and with no E3SM data, the spatial pattern of the TP-SST during 1981-82 and 1997-98 exhibits a narrow band of strong anomalies over equatorial Pacific including coastal Eastern Pacific (not shown), illustrating the influence of E3SM model bias in MIMO-AE-OBS. We plan to explore ways to reduce the influence of model bias in MIMO-AE-OBS, for example by appropriately weighing the observational data used during training in the future.

### 3.2 Predictability of MIMO-AE Index

We use LSTM models to predict the MIMO-AE index for lead times of 1 to 12 months. Figure 3a shows the predictive skill of the LSTMs to predict the MIMO-AE index for the 65 years of E3SM testing data. The predictive skill is computed as the correlation between the LSTM predicted value and the true value of the MIMO-AE index when data is passed through the network. The predictive skill of Niño 3.4 index and ELI index using LSTMs are also shown. One standard deviation spread, computed using the Fisher transformation, are shown as color shadings. The MIMO-AE index exhibits a lower predictive skill than both the Niño 3.4 and the ELI index at all lead times longer than two months. This is likely due to the presence of noisy precipitation data in MIMO-AE, which demonstrates poor temporal auto-correlations on these time scales (e.g. Mahajan et al., 2012), offering little predictive skill.

This is evident in the Figure 3a, which also shows the predictive skill of domain averaged SC-PRECIP using LSTMs, and serves as a baseline for evaluation of predictive skill. Precipitation shows a high skill at a lead time of one month like the other indices, but offers poor predictive skill at longer lead times. MIMO-AE index provides more predictive skill than precipitation itself for two and three month lead times, likely due to the inclusion of TP-SSTs, which have higher predictive skill due to the thermal inertia of the oceanic mixed layer. But, MIMO-AE index provides poor skill for longer lead times. Figure 3b shows the LSTM skills as a function of the calendar month when the prediction is initialized for all indices and generally reflect Fig. 3a, while also showing the well-known spring predictability barrier associated with Nino3.4 and ELI.

The above results hold for the observational data too, with the the MIMO-AE index exhibiting poorer predictive skill when compared to Niño 3.4 index and ELI on these monthly time scales. Similar to E3SM data, MIMO-AE index demonstrates weaker skill at 2-months lead times and longer, while precipitation time series exhibits no skill at lead times longer than one month irrespective of the initial month of predictions (Figure 3d). Although, the skill of predicting MIMO-AE index is substantially higher than that of predicting SC-PRECIP.

### 3.3 SC-PRECIP Predictability from MIMO-AE Index

To evaluate the predictability of SC-PRECIP using the MIMO-AE index, we pass the predicted MIMO-AE index values through the decoder part of the MIMO-AE to construct spatio-temporal predictions of SC-PRECIP anomalies. Figure 3e shows the skill of predicted SC-PRECIP. The predicted spatial pattern of the SC-PRECIP constructed by the MIMO-AE decoder is domain averaged to compute predictive skill. For the Niño 3.4 and ELI, we predict domain average SC-PRECIP from LSTM predicted values of the indices by using linear regression models (also shown). The linear regression models were constructed using the training data for E3SM and observations separately. MIMO-AE generated predicted precipitation exhibits stronger skill than other indices for lead times of up to 3-months.

However, MIMO-AE index's skills at lead times of one to three months are statistically indistinguishable from that of SC-PRECIP index at the 95% confidence level based on a two-tailed Student's t-test of the Fisher transformations of the correlations. To account for the auto-correlation in the time-series', we use effective sample size for the null hypothesis tests. We calculate this effective sample size using the following equation $N_{effective} = \frac{N}{1+2\sum_i^N \gamma_i^2}$ where $\gamma_i$ is the auto-correlation of our SC-PRECIP time series at lag $i$ and N is our total number of samples (Livezey & Chen, 1983). Although, the improved skill is a significant improvement over that of Niño 3.4 index and ELI. MIMO-AE skills are weaker and also indistinguishable from that of SC-PRECIP for longer lags, and become statistically indifferent from zero at a lead time of 6-months and longer. The skill of Niño 3.4 and ELI is statistically insignificant at all lead times on these monthly scales. The enhanced predictive skill of precipitation from MIMO-AE up to a lead time of 3 months is noted for almost all initialization calendar months of the year as compared to other indices (Figure 3f).

Enhanced predictive skill of MIMO-AE of SC-PRECIP is also noted for the 41 years of observation testing data (Figure 3g). The improvement in MIMO-AE skill as compared to other indicies is statistically significant at two to four months lead times at the 95% confidence level. The high skill at 1-month lead time is statistically indifferent from that of SC-PRECIP. And, the skills are statistically zero for 6-months lead time and longer. Also, the enhanced skill of MIMO-AE is noted for almost all initialization calendar months of the year (Figure 3h). Similar to E3SM, the Niño 3.4 index and ELI demonstrate weak skill at all lead month lengths on monthly scales, although they are statistically different from zero for 1-month and 2-month lead times. This is consistent with other studies (e.g. LHeureux et al., 2021; Pan et al., 2019; S. Wang et al., 2017) that find poor skill from ENSO on noisier sub-seasonal time scales over Western US - largely due to atmospheric noise - in spite of significant correlations between SC-PRECIP and Niño 3.4 index on smoother seasonal and inter-annual time scales in observational data (e.g. Jong et al., 2016; X. Huang & Ullrich, 2017; G. Wang et al., 2021; Cheng et al., 2021, also Fig. 2b).

## 4 Summary and Discussion

In a novel approach, we apply MIMO-AE to extract the non-linear relationships between TP-SST and SC-PRECIP on monthly scales and find it to be a powerful tool to enhance sub-seasonal regional predictability. We design the network to yield a temporal index of the projection of these two data sets on the inherent non-linear space of the network. We use LSTMs of the MIMO-AE index to assess the predictability of SC-PRECIP afforded by MIMO-AE on monthly time scales. LSTM-predicted values of MIMO-AE index are decoded using the MIMO-AE decoder to yield predicted SC-PRECIP. We find that the MIMO-AE index offers statistically significant improvements in predictive skill of SC-PRECIP up to a lead time of up to four months for both E3SM and observations, as compared to that imparted by both Niño 3.4 and ELI.

Studies (e.g. LHeureux et al., 2021; S. Wang et al., 2017; G. Wang et al., 2021; Cheng et al., 2021) suggest enhanced sub-seasonal and seasonal predictability of Western US precipitation from atmospheric variables; like geopotential heights, upper level zonal winds, moisture transport, etc.; as well as Northern Pacific SSTs. While we have only utilized TP-SST here for demonstrating the use of multi-task learning for enhanced predictive skills, we plan to incorporate additional variables in the future within the MIMO-AE framework. Further, atmospheric noise is often associated with poor predictability of regional climate induced by modes of climate variability on seasonal timescales (e.g. S. Wang et al., 2017; Cheng et al., 2021). We plan to explore techniques, like de-noising autoencoders, that account for the presence of noise in data and the modeled system and may allow for the extraction of predictability through the atmospheric noise on seasonal and longer time scales. Also, we train our MIMO-AE here using a historical simulation from E3SMv1, which inherits E3SM's model bias. In the future, we plan to use multi-model simulations from the CMIP6 archive and use Bayesian and other transfer learning approaches to weight the available observations appropriately while training network architectures. We also plan to condense the MIMO-AE and LSTMs into one combined MIMO-AE-LSTM network to account for spatial and temporal variability simultaneously to assess predictability in a more coherent manner.

Our results demonstrate the promise of multi-task learning to enhance predictability afforded by remote teleconnections, supporting a focused exploration of other pertinent multi-task and multi-modal methods, like multi-task CNNs for such purposes. Dynamical models exhibit more skill at predicting tropical SSTs than precipitation. It would be interesting to explore hybrid models that utilize dynamical models predicted tropical SSTs with MIMO-AE-like networks that extract non-linear remote teleconnections to make regional climate predictions. Further, machine learning models have demonstrated significantly more skill at seasonal and multi-annual predictions of tropical SSTs than dynamical models (e.g. Ham et al., 2019). Combining such networks with multi-task learning methods provides the potential to further enhance predictability of regional climate.

## 5 Acknowledgements

This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231. This research was supported as part of the Energy Exascale Earth System Model (E3SM) project, funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research.

### 5.1 Data Availability Statement

The E3SMv1 data used in this study is freely available through the Earth System Grid Federation (ESGF) distributed archives via `https://doi.org/10.1029/2018MS001603` and is available through the ESGF interface `https://esgf-node.llnl.gov/projects/e3sm/` (E3SM Project, 2018).

Observational SST data from the HadISST 1.1 dataset (Rayner et al., 2003) can be downloaded from the web at `https://www.metoffice.gov.uk/hadobs/hadisst/`. Observed precipitation data from NOAA's PREC/L (M. Chen et al., 2002) can also be found open access at `https://psl.noaa.gov/data/gridded/data.precl.html`.

## References

Becker, E., van den Dool, H., & Zhang, Q. (2014). Predictability and forecast skill in nmme. *Journal of Climate*, *27*(15), 5891 - 5906. Retrieved from `https://journals.ametsoc.org/view/journals/clim/27/15/jcli-d-13-00597.1.xml` doi: 10.1175/JCLI-D-13-00597.1

Bourlard, H., & Kamp, Y. (1988). Auto-association by multilayer perceptrons and singular value decomposition. *Biological cybernetics*, *59*(4), 291–294.

Broni-Bedaiko, C., Katsriku, F. A., Unemi, T., Atsumi, M., Abdulai, J.-D., Shinomiya, N., & Owusu, E. (2019). El niño-southern oscillation forecasting using complex networks analysis of lstm neural networks. *Artificial Life and Robotics*, *24*(4), 445–451.

Caruana, R. (1997). Multitask learning. *Machine learning*, *28*(1), 41–75.

Charte, D., Charte, F., García, S., del Jesus, M. J., & Herrera, F. (2018). A practical tutorial on autoencoders for nonlinear feature fusion: Taxonomy, models, software and guidelines. *Information Fusion*, *44*, 78–96.

Chen, M., Xie, P., Janowiak, J. E., & Arkin, P. A. (2002). Global land precipitation: A 50-yr monthly analysis based on gauge observations. *Journal of Hydrometeorology*, *3*(3), 249–266.

Chen, S., Zhang, Y., & Yang, Q. (2021). Multi-task learning in natural language processing: An overview. *arXiv preprint arXiv:2109.09138*.

Cheng, R., Novak, L., & Schneider, T. (2021). Predicting the interannual variability of california's total annual precipitation. *Geophysical Research Letters*, *48*(7), e2020GL091465.

Cohen, J., Pfeiffer, K., & Francis, J. (2017). Winter 2015/16: A turning point in enso-based seasonal forecasts. *Oceanography*. Retrieved from `https://doi.org/10.5670/oceanog.2017.115`

Collobert, R., & Weston, J. (2008). A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on machine learning* (pp. 160–167).

Deng, L., Hinton, G., & Kingsbury, B. (2013). New types of deep neural network learning for speech recognition and related applications: An overview. In *2013 ieee international conference on acoustics, speech and signal processing* (pp. 8599–8603).

Devries, T., Biswaranjan, K., & Taylor, G. W. (2014). Multi-task learning of facial landmarks and expression. In *2014 canadian conference on computer and robot vision* (pp. 98–103).

E3SM Project. (2018, April). *Energy Exascale Earth System Model (E3SM).* [Computer Software] `https://dx.doi.org/10.11578/E3SM/dc.20180418.36`. Retrieved from `https://dx.doi.org/10.11578/E3SM/dc.20180418.36` doi: 10.11578/E3SM/dc.20180418.36

Fournier, Q., & Aloise, D. (2019). Empirical comparison between autoencoders and traditional dimensionality reduction methods. In *2019 ieee second international conference on artificial intelligence and knowledge engineering (aike)* (pp. 211–214).

Ghifary, M., Kleijn, W. B., Zhang, M., & Balduzzi, D. (2015). Domain generalization for object recognition with multi-task autoencoders. In *Proceedings of the ieee international conference on computer vision* (pp. 2551–2559).

Girshick, R. (2015). Fast r-cnn. In *Proceedings of the ieee international conference on computer vision* (pp. 1440–1448).

Golaz, J.-C., Caldwell, P. M., Van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., … others (2019). The doe e3sm coupled model version 1: Overview and evaluation at standard resolution. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2089–2129.

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1) (No. 2). MIT press Cambridge.

Gupta, M., Kodamana, H., & Sandeep, S. (2020). Prediction of enso beyond spring predictability barrier using deep convolutional lstm networks. *IEEE Geoscience and Remote Sensing Letters*.

Ham, Y.-G., Kim, J.-H., & Luo, J.-J. (2019). Deep learning for multi-year enso forecasts. *Nature*, *573*(7775), 568–572.

He, J., & Eastman, J. R. (2020). A sequential autoencoder for teleconnection analysis. *Remote Sensing*, *12*(5), 851.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Hsieh, W. W., Wu, A., & Shabbar, A. (2006). Nonlinear atmospheric teleconnections. *Geophysical Research Letters*, *33*(7). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005GL025471` doi: https://doi.org/10.1029/2005GL025471

Huang, A., Vega-Westhoff, B., & Sriver, R. L. (2019). Analyzing el niño–southern oscillation predictability using long-short-term-memory models. *Earth and Space Science*, *6*(2), 212–221.

Huang, X., & Ullrich, P. A. (2017, 08). The Changing Character of Twenty-First-Century Precipitation over the Western United States in the Variable-Resolution CESM. *Journal of Climate*, *30*(18), 7555-7575. Retrieved from `https://doi.org/10.1175/JCLI-D-16-0673.1` doi: 10.1175/JCLI-D-16-0673.1

Jong, B.-T., Ting, M., & Seager, R. (2016). El niño's impact on california precipitation: Seasonality, regionality, and el niño intensity. *Environmental Research Letters*, *11*(5), 054021.

Kendall, A., Gal, Y., & Cipolla, R. (2018). Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 7482–7491).

Kim, S., Hori, T., & Watanabe, S. (2017). Joint ctc-attention based end-to-end speech recognition using multi-task learning. In *2017 ieee international conference on acoustics, speech and signal processing (icassp)* (pp. 4835–4839).

Lee, S.-K., Lopez, H., Chung, E.-S., DiNezio, P., Yeh, S.-W., & Wittenberg, A. T. (2018). On the fragile relationship between el nio and california rainfall. *Geophysical Research Letters*, *45*(2), 907-915. Retrieved from `https://`

agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL076197 doi: https://doi.org/10.1002/2017GL076197

Liu, P., Qiu, X., & Huang, X. (2017). Adversarial multi-task learning for text classification. *arXiv preprint arXiv:1704.05742*.

Livezey, R. E., & Chen, W. (1983). Statistical field significance and its determination by monte carlo techniques. *Mon. Wea. Rev*, *111*(1), 46–59.

LHeureux, M. L., Takahashi, K., Watkins, A. B., Barnston, A. G., Becker, E. J., Di Liberto, T. E., ... others (2017). Observing and predicting the 2015/16 el niño. *Bulletin of the American Meteorological Society*, *98*(7), 1363–1382.

LHeureux, M. L., Tippett, M. K., & Becker, E. J. (2021). Sources of subseasonal skill and predictability in wintertime california precipitation forecasts. *Weather and Forecasting*, *36*(5), 1815 - 1826. Retrieved from https://journals.ametsoc.org/view/journals/wefo/36/5/WAF-D-21-0061.1.xml doi: 10.1175/WAF-D-21-0061.1

Mahajan, S., North, G. R., Saravanan, R., & Genton, M. G. (2012). Statistical significance of trends in monthly heavy precipitation over the us. *Climate dynamics*, *38*(7), 1375–1387.

Mahajan, S., Tang, Q., Keen, N. D., Golaz, J.-C., & van Roekel, L. P. (2021). Simulation of enso teleconnections to precipitation extremes over the us in the high resolution version of e3sm. *Journal of Climate*, 1 - 62. Retrieved from https://journals.ametsoc.org/view/journals/clim/aop/JCLI-D-20-1011.1/JCLI-D-20-1011.1.xml doi: 10.1175/JCLI-D-20-1011.1

Masti, D., & Bemporad, A. (2018). Learning nonlinear state-space models using deep autoencoders. In *2018 ieee conference on decision and control (cdc)* (pp. 3862–3867).

Mu, B., Ma, S., Yuan, S., & Xu, H. (2020). Applying convolutional lstm network to predict el niño events: Transfer learning from the data of dynamical model and observation. In *2020 ieee 10th international conference on electronics information and emergency communication (iceiec)* (pp. 215–219).

Pan, B., Hsu, K., AghaKouchak, A., Sorooshian, S., & Higgins, W. (2019). Precipitation prediction skill for the west coast united states: From short to extended range. *Journal of Climate*, *32*(1), 161 - 182. Retrieved from https://journals.ametsoc.org/view/journals/clim/32/1/jcli-d-18-0355.1.xml doi: 10.1175/JCLI-D-18-0355.1

Plaut, E. (2018). From principal subspaces to principal components with linear autoencoders. *arXiv preprint arXiv:1804.10253*.

Rayner, N., Parker, D., Folland, C., Horton, E., Alexander, L., & Rowell, D. (2003). The global sea-ice and sea surface temperature (hadisst) data sets. *J. Geophys. Res.*

Raza, S. E. A., Cheung, L., Epstein, D., Pelengaris, S., Khan, M., & Rajpoot, N. M. (2017). Mimo-net: A multi-input multi-output convolutional neural network for cell segmentation in fluorescence microscopy images. In *2017 ieee 14th international symposium on biomedical imaging (isbi 2017)* (pp. 337–340).

Shinohara, Y. (2016). Adversarial multi-task learning of deep neural networks for robust speech recognition. In *Interspeech* (pp. 2369–2372).

Tang, Y., & Hsieh, W. W. (2003). Nonlinear modes of decadal and interannual variability of the subsurface thermal structure in the pacific ocean. *Journal of Geophysical Research: Oceans*, *108*(C3). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2001JC001236` doi: https://doi.org/10.1029/2001JC001236

Toshniwal, S., Tang, H., Lu, L., & Livescu, K. (2017). Multitask learning with low-level auxiliary tasks for encoder-decoder based speech recognition. *arXiv preprint arXiv:1704.01631*.

Trenberth, K. E., & Stepaniak, D. P. (2001). Indices of el niño evolution. *Journal of climate*, *14*(8), 1697–1701.

Wang, G., Zhuang, Y., Fu, R., Zhao, S., & Wang, H. (2021). Improving seasonal prediction of california winter precipitation using canonical correlation analysis. *Journal of Geophysical Research: Atmospheres*, *126*(17), e2021JD034848.

Wang, S., Anichowski, A., Tippett, M. K., & Sobel, A. H. (2017). Seasonal noise versus subseasonal signal: Forecasts of california precipitation during the unusual winters of 2015–2016 and 2016–2017. *Geophysical Research Letters*, *44*(18), 9513–9520.

Wang, Y., Yao, H., & Zhao, S. (2016). Auto-encoder based dimensionality reduction. *Neurocomputing*, *184*, 232–242.

Williams, I. N., & Patricola, C. M. (2018). Diversity of enso events unified by convective threshold sea surface temperature: a nonlinear enso index. *Geophysical*

*Research Letters*, *45*(17), 9236–9244.

Wu, A., Hsieh, W. W., & Shabbar, A. (2005). The nonlinear patterns of north american winter temperature and precipitation associated with enso. *Journal of Climate*, *18*(11), 1736 - 1752. Retrieved from `https://journals.ametsoc.org/view/journals/clim/18/11/jcli3372.1.xml` doi: 10.1175/JCLI3372.1

Yaguchi, K., Ikarigawa, K., Kawasaki, R., Miyazaki, W., Morikawa, Y., Ito, C., … Maeda, E. (2020). Human activity recognition using multi-input cnn model with fft spectrograms. In *Adjunct proceedings of the 2020 acm international joint conference on pervasive and ubiquitous computing and proceedings of the 2020 acm international symposium on wearable computers* (pp. 364–367).

Zamparo, L., & Zhang, Z. (2015). Deep autoencoders for dimensionality reduction of high-content screening data. *arXiv preprint arXiv:1501.01348*.

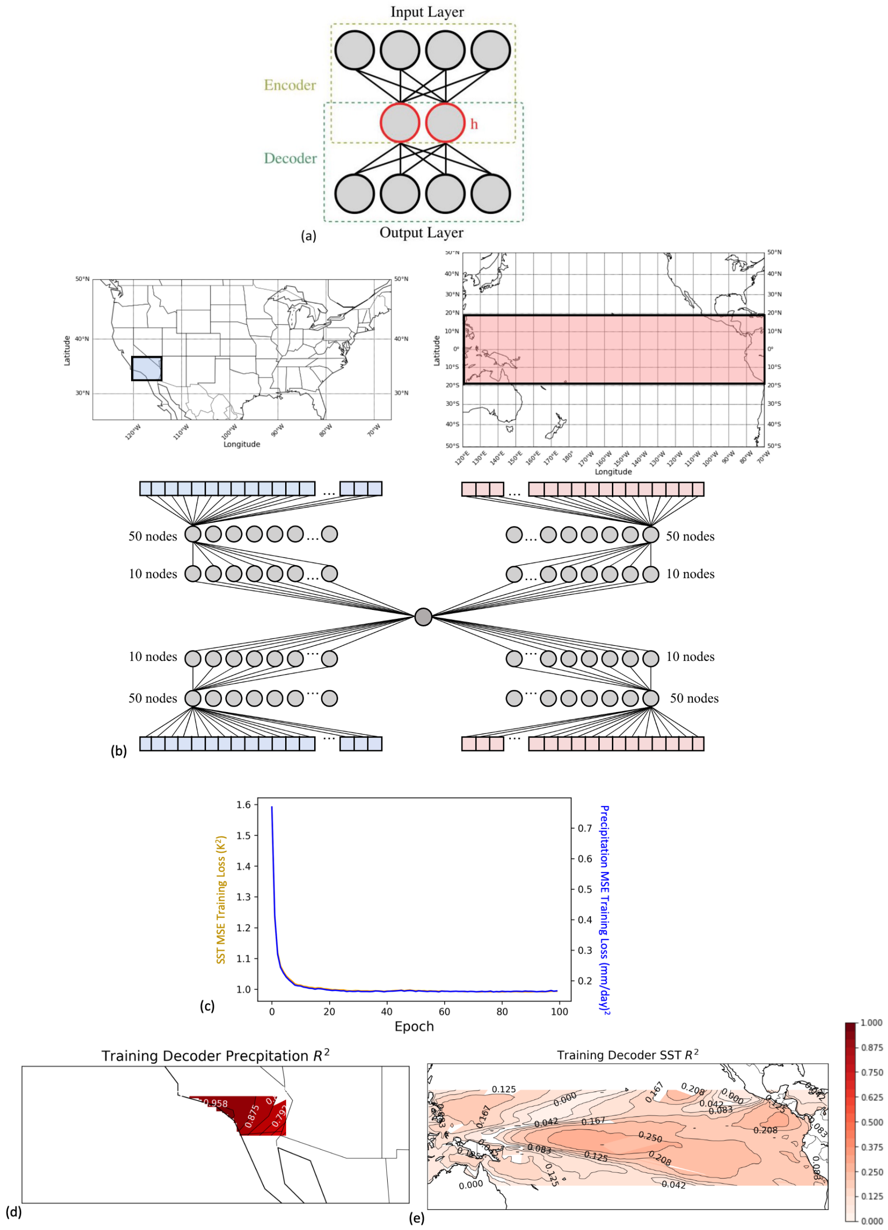


**Figure 1.** Network architecture. A simple autoencoder architecture (a), MIMO-AE architecture (b), training losses for MIMO-AE over 100 epochs using scaled data (c), and the average $R^2$ between the MIMO-AE reconstructed and original input data for Southern California precipitation (d) and Tropical Pacific SST (e). We note that the connections between neurons in (b) are shown selectively for clarity and in actuality all neurons are connected with every neuron in the adjacent layers.

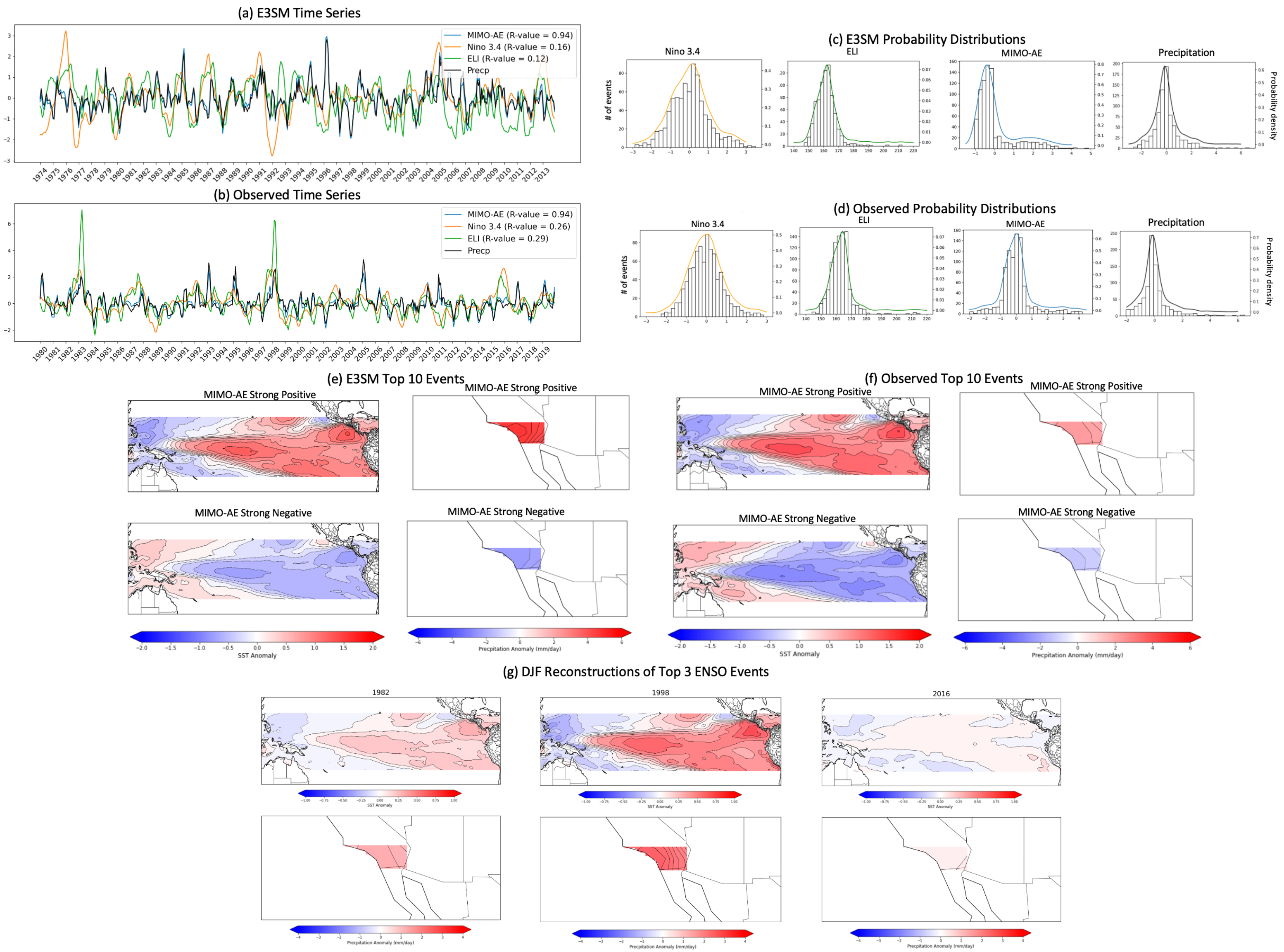


**Figure 2.** MIMO-AE. Three-month moving average of the standardized time series of MIMO-AE index (blue), Niño 3.4 (orange), ELI (green) and domain averaged SC-PRECIP (black) for a segment (last 40 years, 1974-2013) of the 65 years of testing data for the E3SM historical simulation (a); and a segment (last 40 years, 1980-2019) of the observed testing data (b). Correlations of each time-series against domain averaged SC-PRECIP is also listed. The probability density distributions for the Niño 3.4 index (orange), ELI (green), MIMO-AE index (blue) and SC-PRECIP (black) for E3SM data (c); and observational data (d). Composite of reconstructions of TP-SST and SC-PRECIP for the top 10 strongest positive and negative monthly MIMO-AE events for E3SM testing data (e); and observational testing data (f). The December to February average reconstructions for the three strong El-Niño events of 1981-82, 1997-98, and 2015-16 in observations (g).

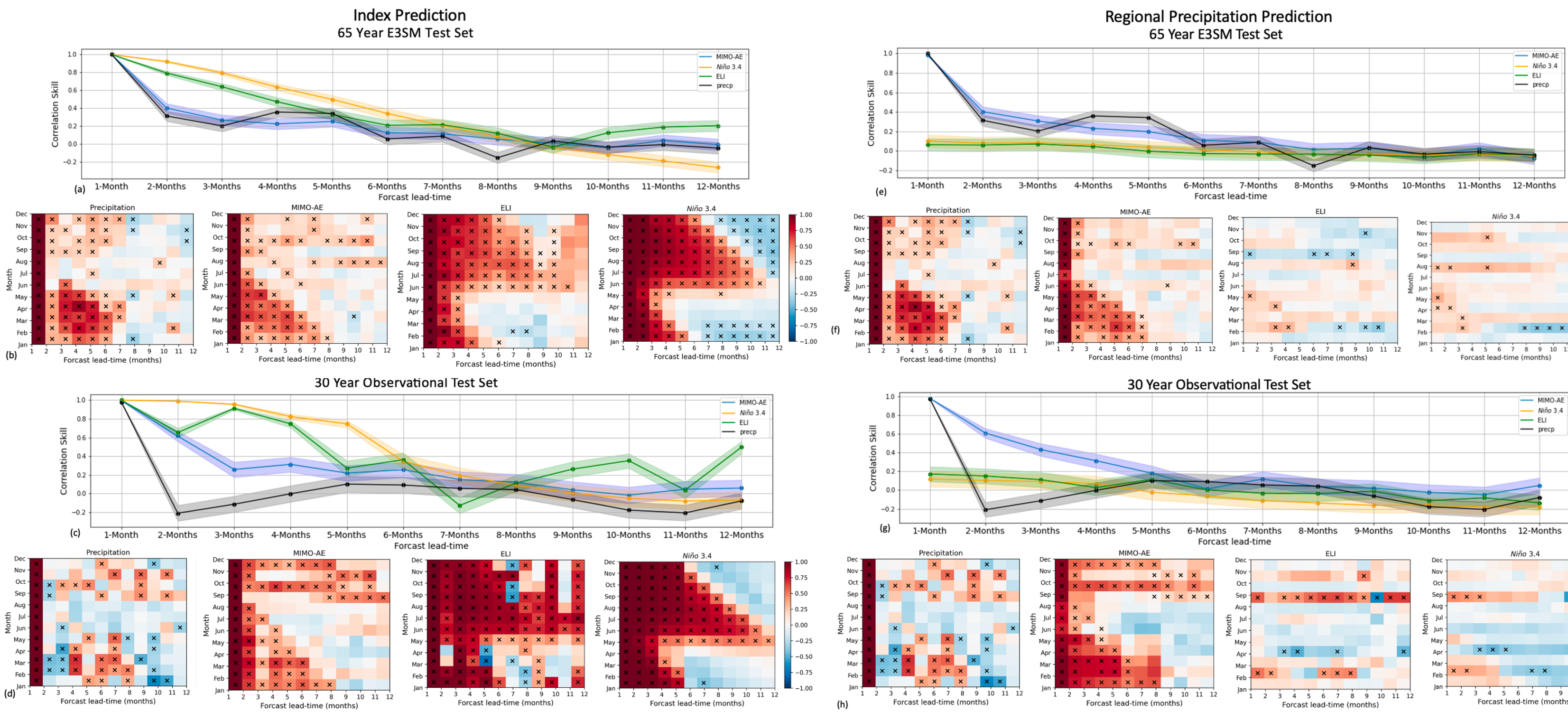


**Figure 3.** Predictability of MIMO-AE index. Predictive skill of LSTMs of MIMO-AE index (blue), Niño 3.4 index (orange), ELI (green) and domain averaged SC-PRECIP (black) at forecast lead times of 1 to 12 months for E3SM testing data (a); and observational testing data (c). Shading represents one standard deviation of the correlation coefficients. Predictive skill of LSTMs as a function of initialization calendar month and forecast lead time from domain average SC-PRECIP, MIMO-AE index, ELI and Niño 3.4 index for E3SM testing data (b); and observational testing data (d). Cross markings indicate values significant at the 95% confidence level. Predictability of SC-PRECIP using MIMO-AE. Predictive skill of MIMO-AE index (blue), Niño 3.4 index (orange) and ELI (green) at predicting domain averaged SC-PRECIP at forecast lead times of 1 to 12 months for E3SM testing data (e); and observational testing data (g) Shading represents one standard deviation of the correlation coefficients. Predictive skill of domain average SC-PRECIP as a function of initialization calendar month and forecast lead time from domain average SC-PRECIP, MIMO-AE index, ELI and Niño 3.4 index for E3SM testing data (f); and observational testing data (h). Cross markings indicate values significant at the 95% confidence level.