Assessing Tropical Pacific–Induced Predictability of Southern California Precipitation Using a Novel Multi-Input Multioutput Autoencoder

LINSEY S. PASSARELLA and SALIL MAHAJAN

Computational Earth Sciences, Oak Ridge National Laboratory, Oak Ridge, Tennessee

(Manuscript received 18 January 2023, in final form 7 July 2023, accepted 11 July 2023)

ABSTRACT: We construct a novel multi-input multioutput autoencoder (MIMO-AE) to capture the nonlinear relationship of Southern California precipitation and tropical Pacific Ocean sea surface temperature. The MIMO-AE is trained on both monthly tropical Pacific sea surface temperature (TP-SST) and Southern California precipitation (SC-PRECIP) anomalies simultaneously. The covariability 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 yr of output from a historical simulation with the Energy Exascale Earth Systems Model, version 1, and a segment of observational data. We further use long short-term memory networks to assess subseasonal 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 4 months as compared with the Niño-3.4 index and the El Niño–Southern Oscillation longitudinal index.

SIGNIFICANCE STATEMENT: Traditional El Niño–Southern Oscillation indices, like the Niño-3.4 index, although well predicted themselves, fail to offer reliable subseasonal-to-seasonal predictions of western U.S. precipitation. Here, we use a machine learning approach called a multi-input, multioutput autoencoder to capture the relationship between tropical Pacific Ocean and Southern California precipitation and project it onto a new index, which we call the MIMO-AE index. Using machine learning–based time series predictions, we find that the MIMO-AE index offers enhanced predictability of Southern California precipitation up to a lead time of 4 months as compared with other ENSO indices.

KEYWORDS: Artificial intelligence; Data science; Deep learning; Dimensionality reduction; Neural networks

1. Introduction

Extracting subseasonal and seasonal predictability of precipitation over California from the leading mode of climate variability—the El Niño–Southern Oscillation (ENSO)—remains a challenge, even during strong ENSO episodes (L’Heureux et al. 2021; Pan et al. 2019; Wang et al. 2017). This was apparent during the 2015/16 El Niño event, which had below- and near-average precipitation anomalies over Southern and Northern California, respectively, despite having strong Niño-3.4 indices (Lee et al. 2018; Jong et al. 2018). This is in contrast to the predicted (Hoell et al. 2016) (also see CPC precipitation outlooks: https://www.cpc.ncep.noaa.gov/products/archives/longlead/) high likelihood of heavy precipitation, which occurred there during the 1982/83 and 1997/98 strong El Niño events (Cohen et al. 2017; L’Heureux et al. 2017). Studies suggest that atmospheric intrinsic variability (Niranjan Kumar et al. 2016; Cheng et al. 2021; L’Heureux et al. 2021), such as that of the jet stream, and remote influences like the Arctic Oscillation (Cohen et al. 2017) could have dampened California’s precipitation response in 2015/16. A recent study (Arcodia et al. 2020) also suggests that the ENSO impact on North American precipitation can be modulated by interference from Madden–Julian oscillation (MJO) signals, with the destructive interference in 2015/16 winter resulting in drying over California. However, some studies also suggest that the contrast in spatial pattern of tropical Pacific Ocean SST anomalies between the central Pacific (or Modoki) 2015/16 El Niño event and other canonical (eastern Pacific) events also contributed to the differing western U.S. precipitation response (Patricola et al. 2020).

ENSO indices, such as the static area-averaged Niño-3.4 index, do not capture this ENSO diversity in the spatial pattern of SST anomalies (Trenberth and Stepaniak 2001; Williams and Patricola 2018). Linear statistical modeling suggests that the Niño-3.4 index can only explain about 9%–25% of the variability of California precipitation on seasonal to interannual time scales, with the largest skill over Southern California (Jong et al. 2016; Huang and Ulrich 2017; Wang et al. 2021; Cheng et al. 2021). Also, ENSO teleconnections exhibit clear nonlinear behavior, with stronger absolute response during El Niño events than La Niña events and a nonlinear response to extreme ENSO events (Hoerling et al. 1997; Frauen et al. 2014). To capture ENSO’s spatial diversity, Williams and Patricola (2018) devised a nonlinear index that tracks the zonal shift of tropical Pacific SST anomalies above the deep convective threshold. This index, termed the ENSO longitude index (ELI), is also representative of the location of convective activity over the tropical Pacific that drives the extratropical wave train responsible for remote teleconnections during ENSO events. The ELI is thus found to be better associated with western U.S. precipitation than the Niño-3.4 index (Patricola et al. 2020).

To enhance regional seasonal predictability afforded by the low-frequency variability modes, it is thus important to account for the complexity of the modes as well as their teleconnections.
more fully. Here, in a novel approach, we adapt a multitask learning autoencoder technique to capture nonlinear covariability patterns of tropical Pacific SSTs and Southern California precipitation. Autoencoders are artificial neural networks, with the same input and output layers, that regenerate the original data from efficient representations (encodings) of the data, like principal component analysis (PCA). Autoencoders, however, transform data to nonlinear latent spaces via nonlinear activation functions, thus imparting the additional capability of capturing the underlying nonlinear relationships within the data (Wang et al. 2016; Charte et al. 2018; Masti and Bemporad 2018). Studies show that autoencoders can better detect dominant variability patterns over other techniques, like PCA (Wang et al. 2016; Zamparo and Zhang 2015; Fournier and Aloise 2019). Some studies (Tang and Hsieh 2003; He and Eastman 2020) have also demonstrated the use of autoencoders to effectively identify modes of climate variability, including those related to ENSO. Multitask learning, on the other hand, solves multiple learning tasks at the same time while exploiting commonalities and differences across tasks (Caruana 1997). It has been applied to many problems including natural language processing (Chen et al. 2021), speech recognition (Toshniwal et al. 2017), and computer vision (Cipolla et al. 2018) to improve the prediction accuracy and learning efficiency of task-specific models. Ghifary et al. (2015) implemented multitask learning with a multoutput autoencoder, for related cross-domain object recognition. It has a single input layer and multiple output layers corresponding to different domains, where a domain refers to the transformation of the object image, like rotation of the viewing angle.

Here, we expand on this approach by Ghifary et al. (2015) and design a new network, termed as the multi-input multioutput autoencoder (MIMO-AE), to effectively extract the most prominent shared features between monthly tropical Pacific sea surface temperatures (TP-SSTs) and Southern California precipitation (SC-PRECIP) anomalies and capture their underlying nonlinear relationships, using Earth system model (ESM) simulation data and observational data. A similar design, albeit for convolutional neural networks, was proposed by Raza et al. (2017) for cell features segmentation in biomedical microscopy images. Our architecture of the MIMO-AE yields a temporal index of the covariability 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 the MIMO-AE is a powerful tool to isolate important teleconnections and yield enhanced subseasonal regional predictability.

2. Methodology

a. Model simulations and data

We use a 165-yr-long historical simulation of the Energy Exascale Earth System Model, version 1 (E3SMv1; E3SM Project 2018), and utilize the first 100 yr of the simulation for training the MIMO-AE network. E3SMv1 is found to effectively capture the temporal variability of ENSO and reproduce ENSO-associated spatial SST patterns when compared with observational datasets, although with a larger westward extent of SST anomalies during El Niño events (Golaz et al. 2019). It also simulates the teleconnections of ENSO to U.S. winter-season precipitation well (Mahajan et al. 2021). We use observed precipitation data from NOAA’s Precipitation Reconstruction over Land (PRECL) at a 1° resolution (Chen et al. 2002). PRECL 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 Sea Ice and Sea Surface Temperature dataset (HadISST 1.1) available at a 1° resolution (Rayner et al. 2003).

b. 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 Fig. 1a, contains a hidden layer \( h \) that describes a representation of important attributes of the input (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 \( f(\cdot) \):

\[
h = f(xw_e),
\]

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

\[
x' = f(hw_d),
\]

where \( w_d \) are the decoder weights.

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 with a nonlinear activation function 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 spatiotemporal data, the temporal vectors at the bottleneck nodes are analogous to the 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 nonlinear 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 (Tang and Hsieh 2003).

c. MIMO-AE

Figure 1b illustrates our MIMO-AE architecture designed to extract the nonlinear relationship between TP-SSTs and SC-PRECIP on monthly time scales. 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 the tropical Pacific (6970 boxes in total)
(20°N–20°S, 120°E–70°W) and precipitation anomalies over each grid box in the boxed domain over Southern California (32°–35°N, 120°–114°W) (24 boxes in total) (Fig. 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 the data. These data are then passed to a single hidden node that is shared by the two input variables. This hidden node represents the shared nonlinear 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 nonlinear hyperbolic tangent or “tanh,” activation function for all the hidden layers.

We performed several iterations of the network design with a 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-square error (MSE) using the following equation:

$$\text{MSE} = \frac{1}{N} \sum_i (P_i - T_i)^2,$$

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 are the original input data. The input variables are scaled using a minimum–maximum scaler before training is performed. The MIMO-AE was trained on the first 100 yr 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 (NERSC)'s Cori supercomputer. The training loss for the scaled TP-SST reconstruction (orange) and SC-PRECIP (blue) are shown in Fig. 2. We hereinafter refer to this MIMO-AE network as MIMO-AE-E3SM.

Figures 3a and 3b show the $R^2$ values (fraction of variance explained) between the reconstructions from the MIMO-AE-E3SM and the original data for the 100 yr 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 the MIMO-AE over the tropical Pacific is an artifact of our network design preference. Different network designs explain different fractions of variability for both TP-SST and SC-PRECIP. We chose ad hoc a network that explained a larger fraction of the variability of SC-PRECIP, since our goal was largely to assess the predictability of SC-PRECIP here. Networks that explain larger fractions of variability over the tropical Pacific did not explain SC-PRECIP variability as strongly. In the future, we would like to explore a loss function that also accounts for the fraction of explained variability of TP-SST and SC-PRECIP separately, say by using a penalty function, to design an optimal network that also maximizes predictive skill. Nonetheless, in a comparison of a linear regression of SC-PRECIP with the Niño-3.4 index reveals much weaker explained variability of SC-PRECIP by the Niño-3.4 index (Fig. 3c). A linear regression

![Fig. 2. Training losses for MIMO-AE-E3SM over 100 epochs using scaled data.](image)

![Fig. 3. The $R^2$ values between the MIMO-AE-E3SM reconstructed and original input data for (a) SC-PRECIP and (b) TP-SST. $R^2$ values of the linear regression of (c) SC-PRECIP and (d) TP-SST against the Niño-3.4 index, $R^2$ values of the linear regression of (e) SC-PRECIP and (f) TP-SST against the ELI, and $R^2$ values between the SVD1 reconstructed and original input data for (g) SC-PRECIP and (h) TP-SST.](image)
of TP-SST against the Niño-3.4 index reveals the large fraction of the variability of the deep tropical Pacific SST explained by the Niño-3.4 index (Fig. 3d), as expected. Here, the Niño-3.4 index is defined as the area-averaged SST anomalies over a region of the tropical Pacific (5°S–5°N, 120°–170°W) defined in Bamston et al. (1997).

We hereinafter refer to the temporal vector of the single-node bottleneck layer that represents the dominant nonlinear mode of covariability of TP-SSTs and SC-PRECIP as the MIMO-AE index. We apply the MIMO-AE-E3SM trained on 100 yr of E3SM historical simulation on the latter 65 yr of the run. As a form of transfer learning, we combine the first 100 yr of the E3SM simulation with 32 yr of observational data (1948–79) 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, to predict the Niño-3.4 index, they train a convolutional neural network (CNN) with observational SST and heat content data but with the network weights initialized from a network trained on historical simulations of 21 CMIP5 models. While not investigated in our exploratory study of the MIMO-AE here, we plan to apply this and other transfer learning methods to the MIMO-AE in the future.

d. ELI

We also evaluate the skill of the nonlinear ENSO index, ELI (Williams and Patricola 2018; Patricola et al. 2020), as briefly discussed in the introduction, at predicting California precipitation. The ELI is calculated for each month of the time period as follows. First, the tropical-averaged SST (5°S–5°N, 120°E–70°W) is computed for the month, which is assumed to be the SST threshold for initiating deep convection (Williams and Patricola 2018). This is based on the approximation that the entire tropical Pacific can be considered as a moist adiabat due to the rapid homogenization of heating from deep convective activity in the tropics to the entire tropical troposphere (Williams and Patricola 2018). A binary mask is then created by assigning 1 to grid points where SST is more than the threshold value and assigning 0 to grid points where SST is less than the threshold value for the month. Last, the ELI for the month is computed as the average of all longitudes over which this spatial mask is 1 (Williams and Patricola 2018).

A linear regression of SC-PRECIP against the ELI also exhibits weaker explained variability of precipitation there (Fig. 3e) relative to the MIMO-AE. The ELI only represents the zonal location of SST anomalies greater than the deep convective threshold; nonetheless, a regression of TP-SST on the ELI reveals explained variability of up to 50% over the eastern Pacific (Fig. 3f).

e. SVD

As a linear approach to capture the covariability of TP-SST and SC-PRECIP, we also conduct a singular value decomposition (SVD) of the covariance matrix of gridpoint data anomalies of SST over TP-SST and precipitation over the SC-PRECIP region using the training data. SVD is a commonly used method of extracting linear covariability patterns in multivariate climate data (Bretherton et al. 1992; Chang et al. 1997; Mahajan et al. 2010). The first SVD mode (SVD1) explains 28% of the squared covariance of TP-SST and SC-PRECIP. Figures 3g and 3h show the $R^2$ values of SVD1 reconstruction of SC-PRECIP and TP-SST against the original data. Although the SVD1 reconstruction can explain large fractions of variability of TP-SST (greater than 80% over the central Pacific), it exhibits weaker explained variability of SC-PRECIP relative to the MIMO-AE.

f. LSTM

To study predictability, we also train LSTM recurrent neural networks as our time series prediction models. LSTMs are a special kind of recurrent neural network that learn long-term dependencies (Hochreiter and Schmidhuber 1997). They have recently been shown to perform better at time series prediction of the Niño-3.4 index over linear models (Huang et al. 2019; Mu et al. 2020; Broni-Bedaiko et al. 2019; Gupta et al. 2020).

LSTMs are constructed with internal mechanisms called “gates” that control the flow of information through the cell (Hochreiter and 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 them down the line to the next cell.

LSTM models are constructed individually for each of the time series of the MIMO-AE index, Niño-3.4 index, ELI, SVD, and regionally averaged SC-PRECIP anomalies using the first 100 yr of the E3SM data. We train separate LSTMs for the above-listed time series using the first 32-yr segment (1948–79) of observational dataset used. Given a predicted value of the MIMO-AE index, predicted SC-PRECIP (and TP-SSTs) can be constructed by passing the index through the decoder of the 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 yr of E3SM data and the 41 yr of observational data.

3. Results

a. MIMO-AE

Figure 4a shows the 3-month moving average of the standardized MIMO-AE index time series for a segment (last 40 yr: 1974–2013) of the 65 yr of the E3SM testing data. The MIMO-AE index was generated by passing the TP-SST and SC-PRECIP data through the MIMO-AE network trained on the prior 100 yr of simulation. Figure 4a also shows the time series of the expansion coefficients of SST over the TP-SST domain (SVD1–SST) corresponding to SVD1 for the same segment of the testing period. SVD1 explains 28% of the squared covariance of TP-SST and SC-PRECIP. Figures 3a and 3b show the spatial patterns of the SVD1 weights for SC-PRECIP and TP-SST domains. Also shown in Fig. 4 are the Niño-3.4 index, ELI, and domain-averaged SC-PRECIP.
The correlations of each time series against domain-averaged SC-PRECIP are also listed for the smoothed data. Figure 4b shows the same but for a segment of the observational data (1980–2019) and using MIMO-AE-OBS. To reiterate, the MIMO-AE index for observations is computed by passing the observed data through the MIMO-AE-OBS network. SVD1-SST time series for the segment is also computed by projecting the TP-SST testing data on the SVD1 computed from a combination of E3SM and observational training data.

For both E3SM and observations, the correlation between SC-PRECIP and the MIMO-AE index is higher than that between SC-PRECIP and the Niño-3.4 index or ELI, given that precipitation data are fed in the generation of the MIMO-AE index, and the MIMO-AE explains a large fraction of the SC-PRECIP variability. The time series of the expansion coefficient of precipitation over the SC-PRECIP domain also shows a high correlation with SC-PRECIP for E3SM (0.98) and observations (0.98). The correlation between the SVD1-SST time series and SC-PRECIP is weaker for both data (about 0.17), pointing to the limitations of linear methods. The correlation between the MIMO-AE and Niño-3.4 is higher than the correlation between SC-PRECIP and Niño-3.4. The static Niño-3.4 index only captures the variability over a specific region over the tropical Pacific and is found to only weakly explain SC-PRECIP variability. The MIMO-AE, which accounts for the shared variability of SC-PRECIP and all of the tropical Pacific, however, is found to explain a larger fraction of SC-PRECIP variability. The correlation of SVD1-SST time series and the Niño-3.4 index (0.94) is higher than that of the MIMO-AE and Niño-3.4 indices, suggesting that linear methods like SVDS are dominated (and limited) by the variability over static regions when capturing relationships with the tropical Pacific. Further, in observational data, all indices spike during the 1982/83 and 1996/97 El Niño events, but only the Niño-3.4 index and SVD1-SST time series peak during the 2015/16 El Niño event. The MIMO-AE also categorizes the 2015/16 event as weaker than the Niño-3.4, similar to the ELI (Williams and Patricola 2018).

Figures 5a and 5b show the probability density functions of the Niño-3.4 index, ELI, SVD1-SST time series, MIMO-AE index, and domain-averaged SC-PRECIP for E3SM testing data and observations. While the Niño-3.4 index and SVD1-SST time series tend to be symmetric, the ELI is skewed toward the left (westward), both for E3SM data and observations as noted by Williams and Patricola (2018), with a thicker right tail (eastward). The MIMO-AE, which represents the shared covariability between the TP-SSTs and SC-PRECIP, also shows a longer right tail with stronger positive events than negative events (Fig. 5). The leftwards skewness may follow from the density function of precipitation that is naturally skewed leftwards, even for monthly average data (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 the covariability between the two remains weaker during strong negative SC-PRECIP anomalous events. The skewness of the MIMO-AE index is noted to be stronger in E3SM than in observations seen in Fig. 5. This difference in the MIMO-AE density functions is plausibly suggestive of the differing relationship between TP-SST and SC-PRECIP in E3SM as compared with observations.

Figure 6a shows the composite of reconstructions of TP-SST and SC-PRECIP during the strongest 10 positive and negative monthly MIMO-AE index values for the E3SM testing data. Strong positive SC-PRECIP anomalies during those events are associated with strong positive anomalies in the central,

The correlations of each time series against domain-averaged SC-PRECIP are also listed for the smoothed data. Figure 4b shows the same but for a segment of the observational data (1980–2019) and using MIMO-AE-OBS. To reiterate, the MIMO-AE index for observations is computed by passing the observed data through the MIMO-AE-OBS network. SVD1-SST time series for the segment is also computed by projecting the TP-SST testing data on the SVD1 computed from a combination of E3SM and observational training data.

For both E3SM and observations, the correlation between SC-PRECIP and the MIMO-AE index is higher than that between SC-PRECIP and the Niño-3.4 index or ELI, given that precipitation data are fed in the generation of the MIMO-AE index, and the MIMO-AE explains a large fraction of the SC-PRECIP variability. The time series of the expansion coefficient of precipitation over the SC-PRECIP domain also shows a high correlation with SC-PRECIP for E3SM (0.98) and observations (0.98). The correlation between the SVD1-SST time series and SC-PRECIP is weaker for both data (about 0.17), pointing to the limitations of linear methods. The correlation between the MIMO-AE and Niño-3.4 is higher than the correlation between SC-PRECIP and Niño-3.4. The static Niño-3.4 index only captures the variability over a specific region over the tropical Pacific and is found to only weakly explain SC-PRECIP variability. The MIMO-AE, which accounts for the shared variability of SC-PRECIP and all of the tropical Pacific, however, is found to explain a larger fraction of SC-PRECIP variability. The correlation of SVD1-SST time series and the Niño-3.4 index (0.94) is higher than that of the MIMO-AE and Niño-3.4 indices, suggesting that linear methods like SVDS are dominated (and limited) by the variability over static regions when capturing relationships with the tropical Pacific. Further, in observational data, all indices spike during the 1982/83 and 1996/97 El Niño events, but only the Niño-3.4 index and SVD1-SST time series peak during the 2015/16 El Niño event. The MIMO-AE also categorizes the 2015/16 event as weaker than the Niño-3.4, similar to the ELI (Williams and Patricola 2018).

Figures 5a and 5b show the probability density functions of the Niño-3.4 index, ELI, SVD1-SST time series, MIMO-AE index, and domain-averaged SC-PRECIP for E3SM testing data and observations. While the Niño-3.4 index and SVD1-SST time series tend to be symmetric, the ELI is skewed toward the left (westward), both for E3SM data and observations as noted by Williams and Patricola (2018), with a thicker right tail (eastward). The MIMO-AE, which represents the shared covariability between the TP-SSTs and SC-PRECIP, also shows a longer right tail with stronger positive events than negative events (Fig. 5). The leftwards skewness may follow from the density function of precipitation that is naturally skewed leftwards, even for monthly average data (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 the covariability between the two remains weaker during strong negative SC-PRECIP anomalous events. The skewness of the MIMO-AE index is noted to be stronger in E3SM than in observations seen in Fig. 5. This difference in the MIMO-AE density functions is plausibly suggestive of the differing relationship between TP-SST and SC-PRECIP in E3SM as compared with observations.

Figure 6a shows the composite of reconstructions of TP-SST and SC-PRECIP during the strongest 10 positive and negative monthly MIMO-AE index values for the E3SM testing data. Strong positive SC-PRECIP anomalies during those events are associated with strong positive anomalies in the central,
lies over the northwestern Pacific anomalies in the eastern tropical Pacific; and negative anomalies over the northwestern Pacific. Conversely, strong positive SC-PRECIP anomalies during those events are associated with strong negative anomalies in the central, north-central, and northeast tropical Pacific; weak negative anomalies in the eastern tropical Pacific; and positive anomalies over the northwestern Pacific. Similar patterns are noted for the strongest positive and negative MIMO index values for reconstructions of TP-SST when observation data are passed through the MIMO-AE network (Fig. 6b). SC-PRECIP anomalies during the strongest MIMO-AE events are stronger in E3SM than in observations, possibly suggesting a stronger-than-observed association between TP-SST and SC-PRECIP in E3SM. We note that this could also be due to the smaller sample size of the observations training dataset (32 yr) as compared with that of E3SM (100 yr).

Figure 7a shows the December–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 SST anomalies substantially vary between each of the three El Niño events (Fig. 6). The 2015/16 El Niño event is associated with weak positive anomalies in the MIMO-AE latent space for SC-PRECIP and TP-SST over much of the tropical Pacific. In contrast, the 1981/82 and 1997/98 events are associated with strong positive anomalies in both 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 the equatorial Pacific including coastal eastern Pacific (not shown), illustrating the influence of E3SM model bias in MIMO-AE-OBS. In the future, we plan to explore ways to reduce the influence of model bias in MIMO-AE-OBS, for example, by appropriately weighing the observational training data.

b. Predictability of MIMO-AE index

We use LSTM models to predict the MIMO-AE index for lead times of 1–12 months. Figure 8a shows the predictive skill of the LSTMs in predicting the MIMO-AE index for the 65 yr 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 are passed through the network. The predictive skill of the Niño-3.4 index, ELI, and SVD1-SST using LSTMs is also shown. A spread of 1 standard deviation, computed using the Fisher transformation, is shown as color shadings. The MIMO-AE index exhibits a lower predictive skill than the Niño-3.4 index, ELI, and SVD1-SST at all lead times longer than 1 month. This is likely due to the dominance of noisy precipitation data in the MIMO-AE, which demonstrates poor temporal autocorrelations on these time scales (Mahajan et al. 2012), offering little predictive skill.

This is evident in Fig. 8a, 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 1 month like the other indices but offers poor predictive skill at longer lead times. The MIMO-AE index provides more predictive skill than precipitation itself for 2- and 3-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 the MIMO-AE index provides poor skill for longer lead times. Figure 8b shows the LSTM skills as a function of the calendar month when the prediction is initialized for all indices and generally reflect Fig. 8a, while also showing the well-known spring predictability barrier associated with the Niño-3.4 and the ELI.

The above results hold for the observational data too, with the MIMO-AE index exhibiting poorer predictive skill when compared with the Niño-3.4 index, ELI, and SVD1-SST on these monthly time scales. Similar to E3SM data, the MIMO-AE index demonstrates weaker skill at 2-month lead times and longer, while precipitation time series exhibits no skill at lead times longer than 1 month irrespective of the initial month of predictions (Fig. 3d), although the skill of predicting the MIMO-AE index is substantially higher than that of predicting SC-PRECIP.

c. 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 spatiotemporal predictions of SC-PRECIP anomalies. Figure 9a 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 index, ELI, and SVD1-SST, we predict domain-averaged SC-PRECIP from LSTM predicted values of the indices by using linear regression models (Fig. 9a). The linear regression models were constructed using the training data for E3SM and observations separately. Domain-averaged SC-PRECIP computed from reconstructions of SC-PRECIP from LSTM of SVD1-PRECIP predictions yields similar skill (not shown) as the SC-PRECIP index, given their high correlation. MIMO-AE-generated predicted precipitation exhibits stronger skill than other indices for lead times of up to three months.

However, the MIMO-AE index's skills at lead times of 1–3 months are statistically indistinguishable from that of the 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 autocorrelation in the time series, we use an effective sample size for the null hypothesis tests. We calculate this effective sample size using the following equation:

Fig. 6. Composite of reconstructions of TP-SST and SC-PRECIP for the top 10 strongest positive and negative monthly MIMO-AE events for (a) E3SM testing data and (b) observational testing data.
where $g_i$ is the autocorrelation of our SC-PRECIP time series at lag $i$ and $N$ is our total number of samples (Livezey and Chen 1983). Although, the improved skill is a significant improvement over that of the Niño-3.4 index, ELI, and SVD1-SST, the 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 six months and longer. The skill of Niño-3.4 and the ELI is statistically insignificant at all lead times on these monthly scales. A two-channel LSTM method that is commonly used in areas such as text classification (Liang et al. 2021) and that uses both the Niño-3.4 index and SC-PRECIP as inputs to predict both of the indices and to capture the nonlinear shared variability of the two indices also performs poorly in comparison with the MIMO-AE index (Fig. 10). This result further suggests that tropical Pacific regions, rather than regions such as the Niño-3.4, play a role in enhancing the predictability of the MIMO-AE. The enhanced predictive skill of precipitation from the MIMO-AE up to a lead time of three months is noted for almost all initialization calendar months of the year as compared with other indices (Fig. 9b).

The enhanced predictive skill of the MIMO-AE of SC-PRECIP is also noted for the 41 yr of observation testing data (Fig. 9c). The improvement in the MIMO-AE skill as compared with other indices is statistically significant at 2–4-month lead times at the 95% confidence level. The high skill at a 1-month lead time is statistically indifferent from that of SC-PRECIP. In addition, the skills are statistically zero for a 6-month lead time and longer. Also, the enhanced skill of the MIMO-AE is noted for almost all initialization calendar months of the year (Fig. 9d). Similar to E3SM, the Niño-3.4 index, ELI, and SVD1-SST demonstrate weak skill at all lead month lengths on monthly scales, although they are statistically different from zero for 1- and 2-month lead times. Precipitation time series are known to be noisy, and this is reflected in the poor skill of LSTMs at predicting precipitation for lead times longer than a month. The skill is statistically zero for observational data (Fig. 9c); the 2 standard deviations (the shadings in the plots show 1 standard deviation) of the skill estimate includes zero for almost all lead times greater than 1 month. The likely unphysical oscillation in skill (increase in skill at longer lead times) is likely an artifact of the low sample size used for training and validating the LSTM. The slightly improved LSTM skill of SC-PRECIP for E3SM data at lags longer than a month is likely due to the larger sample size of the E3SM testing (100 yr) and validation (62 yr) data used to train the LSTM model than that used for observational data. These LSTM skill plots suggest that SC-PRECIP is a poor predictor of itself. However, tropical Pacific SSTs can modulate SC-PRECIP precipitation, and SSTs have a longer memory because of the large thermal inertia of the oceanic mixed layer. The MIMO-AE, which includes a nonlinear weighted combination of TP-SST, is found to explain larger variability of SC-PRECIP and is thus able to provide more predictive skill.

**d. Underlying mechanistics**

The underlying nonlinear relationships, captured by the MIMO-AE, are missed when variability over the tropical Pacific is represented by broad indices based on area averages like Niño-3.4 and linear transformations like SVDs and precipitation predictions are made using linear regressions. The MIMO-AE identifies a distinctive combination of SST anomalies over the central, north-central, northeastern, eastern, and northwestern tropical Pacific that are associated with SC-PRECIP anomalies, as illustrated by the SST reconstructions during strong MIMO-AE-defined events (Figs. 6 and 7). SC-PRECIP anomalies are driven by extratropical storm systems and jet streams, as well as atmospheric rivers that are influenced by wave trains emanating from the tropical Pacific forced by TP-SSTs (Lee et al. 2018; Ellis and Barton 2012). Figure 11 shows the linear regression of vertically integrated
moisture transport anomalies (integrated vapor transport [IVT])—a proxy for processes impacting precipitation—against the standardized Niño-3.4 index, ELI, and SVD1-SST, as well as against the SC-PRECIP index for the 100 yr of E3SM testing data. The weak IVT response associated with these indices indicates that they fail to adequately represent the processes associated with SC-PRECIP (Fig. 11) and thus are poor predictors. Since the spatial pattern of SST anomalies associated with the MIMO-AE is not standing, we use the principal component (PC1) of the leading EOF of the MIMO-AE monthly SST reconstructions over the tropical Pacific as an index to represent the MIMO-AE-derived SST anomaly pattern. The leading EOF explains almost all of the variability of the MIMO-AE SST reconstructions. A regression of IVT against the standardized PC1 reveals a strong IVT response over Southern California, about 70% of the moisture transport response associated with SC-PRECIP anomalies (Fig. 11). This indicates that MIMO-AE-learned SST anomaly patterns, unlike those associated with linear approaches like SVD that are dominated by variability over the Niño-3.4 region (seen in Figs. 3a,b), strongly influence processes that result in SC-PRECIP anomalies, allowing the MIMO-AE to provide enhanced predictive skill there.

4. Summary and discussion

In a novel approach, we apply a multi-input multioutput autoencoder (MIMO-AE) to extract the nonlinear relationships between the TP-SST and SC-PRECIP on monthly scales. Using LSTMs, we find the MIMO-AE to be a powerful tool to enhance subseasonal-to-seasonal regional predictability that offers statistically significant improvements in predictive skill of SC-PRECIP up to a lead time of 4 months, as compared with that imparted by the Niño-3.4 index, ELI, and SVD1-SST. The poor skill imparted by these indices is consistent with other studies (L’Heureux et al. 2021; Pan et al. 2019; Wang et al. 2017) that find poor skill from ENSO on noisier subseasonal
time scales over the eastern United States—largely due to atmospheric noise—despite significant correlations between SC-PRECIP and the Niño-3.4 index on smoother seasonal and interannual time scales in observational data (Jong et al. 2016; Huang and Ullrich 2017; Wang et al. 2021; Cheng et al. 2021). Using the number of rainy days, with a threshold-based definition of rainy days, as a predictor of total precipitation, Cheng et al. (2021) found that ENSO can explain 20% (correlation

**FIG. 9.** Predictive skill of MIMO-AE index (blue), Niño-3.4 index (orange), SVD1 (purple), and ELI (green) at predicting domain-averaged SC-PRECIP at forecast lead times of 1–12 months for (a) E3SM testing data and (c) observational testing data. Shading represents 1 standard deviation of the correlation coefficients. Also shown is predictive skill of domain-averaged SC-PRECIP as a function of initialization calendar month and forecast lead time from domain-averaged SC-PRECIP, MIMO-AE index, ELI, SVD1, and Niño-3.4 index for (b) E3SM testing data and (d) observational testing data. Crosses in (b) and (d) indicate values that are statistically different from zero at the 95% confidence level.

**FIG. 10.** Predictive skill of LSTMs of nonlinear MIMO-AE index (blue), two-channel LSTM (green), and SVD1 (purple) at predicting domain-averaged SC-PRECIP at forecast lead times of 1–12 months for E3SM testing data. Shading represents 1 standard deviation of the correlation coefficients.
skill of about 0.45) of the variability of winter-season precipitation over Southern California. This is lower than the average skill of the MIMO-AE up to a lead time of three months (about 0.6). We find that MIMO-AE-learned SST anomaly patterns associated with SC-PRECIP better capture the moisture transport into the region, which imparts to the MIMO-AE its improved predictive skill.

Studies (L’Heureux et al. 2021; Wang et al. 2017, 2021; Cheng et al. 2021) suggest enhanced subseasonal and seasonal predictability of the western U.S. precipitation from atmospheric variables, such as geopotential heights, upper-level zonal winds, moisture transport, and northern Pacific SSTs. We plan to explore the benefits of including more variables in the MIMO-AE framework. We also plan to explore techniques, like denoising autoencoders, that systematically account for system noise, often associated with poor predictability of the climate system. Our MIMO-AE approach can also be applied to assess the predictability of regional climate across the globe, both where linear correlations are known to exist and where the signal-to-noise ratio is low. Future work will expand the precipitation domain from Southern California to all of the conterminous United States to explore if the MIMO-AE can explain a significant fraction of variability over other regions like the southeastern United States. Further, dynamical models exhibit more skill at predicting tropical SSTs than dynamical models (e.g., Ham et al. 2019). Combining such networks with multitask learning methods provides the potential to further enhance the predictability of regional climate. Our results demonstrate the promise of multitask learning to enhance predictability afforded by remote teleconnections, supporting a focused exploration of other pertinent multitask and multimodal methods (like multitask CNNs) for such purposes using the wealth of multimodel simulation ensembles.

Acknowledgments. 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 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.

Data availability statement. The E3SMv1 data used in this study are freely available through the Earth System Grid Federation (ESGF) distributed archives (https://doi.org/10.1029/2018MS001603) and are 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 internet (https://www.metoffice.gov.uk/hadobs/hadisst/). Observed precipitation data
from NOAA’s PREC/L (Chen et al. 2002) can also be found online as open access (https://psl.noaa.gov/data/gridded/data.preci.html).

REFERENCES


McK. Tippett, and E. J. Becker, 2021: Sources of subseasonal skill and predictability in wintertime California precipitation


