8-2022

Machine Learning for Earth Systems Modeling, Analysis and Predictability

Linsey Passarella
lpassare@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Linsey Passarella entitled "Machine Learning for Earth Systems Modeling, Analysis and Predictability." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Data Science and Engineering.

Salil Mahajan, Major Professor

We have read this dissertation and recommend its acceptance:

Russell Zaretzki, Kate Evans, Joshua Fu

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

. . .

To the Graduate Council:

I am submitting herewith a dissertation written by Linsey S. Passarella entitled "Machine Learning for Earth Systems Modeling, Analysis and Predictability." I have examined the final paper copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Data Science and Engineering.

Salil Mahajan, Major Professor

We have read this dissertation
and recommend its acceptance:

Katherine Evans

Joshua Fu

Russell Zaretzki

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

To the Graduate Council:
I am submitting herewith a dissertation written by Linsey S. Passarella entitled "Machine Learning for Earth Systems Modeling, Analysis and Predictability." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Data Science and Engineering.

Salil Mahajan, Major Professor

We have read this dissertation
and recommend its acceptance:

Katherine Evans

Joshua Fu

Russell Zaretzki

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Machine Learning for Earth Systems Modeling, Analysis and Predictability

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Linsey S. Passarella

August 2022

This work is dedicated to Latte, Pistachio and Shay. My very best friends.

Acknowledgements

I would like to thank the members of my committee, Katherine Evans, Joshua Fu, and Russell Zaretzki for their valuable comments and suggestions on my work. I would also like to extend my deepest gratitude to my advisor, Salil Mahajan, not only for his guidance and feed back on this dissertation, but for his endless support throughout this whole process.

I would not have been able to finish my degree without my fiance, Shay, who provided constant love, support and coffee. He continues to push me to be a better human and I would not be where I am today without him. I would also like to thank my parents and sister for doing more for me than I'll ever be able to comprehend. And for answering every phone call, even when they had no idea what I was talking about. Also to my grandparents for forming my official fan club and being a constant source of compliments, pep talks and love.

Abstract

Artificial intelligence (AI) and machine learning (ML) methods and applications have been continuously explored in many areas of scientific research. While these methods have lead to many advances in climate science, there remains room for growth especially in Earth System Modeling, analysis and predictability. Due to their high computational expense and large volumes of complex data they produce, earth system models (ESMs) provide an abundance of potential for enhancing both our understanding of the climate system as well as improving performance of ESMs themselves using ML techniques. Here I demonstrate 3 specific areas of development using ML: statistical downscaling, predictability using non-linear latent spaces and emulation of complex parametrization. These three areas of research illustrate the ability of innovative ML methods to advance our understanding of climate systems through ESMs.

In Aim 1, I present a first application of a fast super resolution convolutional neural network (FSRCNN) based approach for downscaling earth system model (ESM) simulations. We adapt the FSRCNN to improve reconstruction on ESM data, we term

the FSRCNN-ESM. We find that FSRCNN-ESM outperforms FSRCNN and other super-resolution methods in reconstructing high resolution images producing finer spatial scale features with better accuracy for surface temperature, surface radiative fluxes and precipitation.

In Aim 2, I construct a novel Multi-Input Multi-Output Autoencoder-decoder (MIMO-AE) in an application of multi-task learning to capture the non-linear relationship of Southern California precipitation (SC-PRECIP) and tropical Pacific Ocean sea surface temperature (TP-SST) on monthly time-scales. I 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. I also use a MTL method to expand on a convolutional long short term memory (conv-LSTM) to predict Nino 3.4 index by including multiple input variables known to be associated with ENSO, namely sea level pressure (SLP), outgoing longwave radiation (ORL) and surface level zonal winds (U).

In Aim 3, I demonstrate the capability of DNNs for learning computationally expensive parameterizations in ESMs. This study develops a DNN to replace the full radiation model in the E3SM.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Introduction

Applications in machine learning (ML) and artificial intelligence (AI) have seen rapid growth in recent years in many areas of scientific research. For example, the emergence of ML techniques applied to biological sciences has lead to many profound discoveries in genomics and proteomics. Despite having an abundance of potential applications and resources, climate science has not seen similar improvements. Emerging disciplines, such as climate informatics, seek to bridge the gap between climate scientists and machine learning researchers. These efforts are in the early stages of development with hopes that interdisciplinary research will vastly accelerate our understanding of climate science.

Earth Systems Models (ESMs) are the primary tool used by scientists to understand and predict the climate. The accuracy of ESMs has increased significantly

in recent years due to numerical modeling and computational ability developments. The advancement in high-performance computing has allowed for ESMs to be run at higher resolutions with more precision. Several limitations still exist, however. The computational expense of running these models on large scales and over long periods is enormous, even with access to super-computing capability. This cost has led to a shortage of reliable local and regional scale climate data crucial for determining climate change impacts on society. Even with the ability to run these models at high resolutions, a single 30 year run of a 0.25° model can produce up to 10TB of multivariate data making post-run analysis complicated and challenging to interpret. Thus, the growing abundance of ESM data offers many opportunities for implementing ML and AI techniques for both model improvement and interpretation.

The accelerating growth of ESM data availability and computational resources opens the door for vast improvements in data understanding. Early adoptions of ML and AI techniques have provided promising results in the areas of reducing run-time expense and in extracting useful information from the data once it is produced (49; 98). These applications are still in their infancy and need to be further developed and adapted to Earth sciences. I demonstrate here the usefulness of advance ML capabilities in three areas of prominent ESM research: statistical downscaling, predictability using non-linear latent spaces and emulation of complex parameterization.

1.1.1 Aim 1: Statistical Downscaling

ESMs are often limited in their spatial and temporal resolution due to computational cost, leading to a shortage of regional and local scale climate data. Downscaling techniques, statistical and dynamical, are often applied to provide higher resolution outputs of climate projections. Dynamical downscaling techniques are based on Regional Climate Models (RCMs), numerical climate predictions that are initialized with boundary conditions from a Global Climate Model (GCM) to simulate atmospheric and land surface processes at a higher resolution. For this proposal, however, I focus on statistical downscaling (SD). SD aims to learn a statistical mapping between low and high resolution climate data and are often more computationally efficient than dynamical downscaling methods. Recent studies have shown that ML techniques, like neural networks (117; 29) and support vector machines (34), for SD significantly outperform other traditional SD methods. Many of these traditional SD methods do not account for spatial dependencies within the climate model. ML techniques, like computer vision approaches, have been developed to capture and exploit these dependencies.

1.1.2 Aim 2: Predictability Using Non-Linear Latent Spaces

Teleconnections from natural modes of climate variability like El Nino Southern Oscillation (ENSO) cause significant impacts on infrastructure and human health. El Niño-Southern Oscillation (ENSO) is a large-scale ocean-atmosphere interaction

that occurs in the equatorial Pacific Ocean and global atmosphere and is one of the main drivers of inter-annual climate variability (4; 5; 93). El Niño (La Niña) is the ocean component of this phenomena that is characterized by a period increase (decrease) in sea surface temperature (SST). ENSO events are typically identified using sea surface temperature (SST) anomalies averaged in the tropical eastern or central Pacific ocean, such as the Niño 3 index (5° S - 5° N, 150° - 190° W), Niño 3.4 index (5° S - 5° N, 170° - 120° W), Niño 4 index (5° S - 5° N, 160° - 150° W) and Niño 1+2 index (0° - 10° S, 90° - 80° W). While these methods of defining ENSO events capture the canonical El Niño centered in the eastern equatorial Pacific, they fail to capture a secondary ENSO event centered in the central Pacific known as the Dateline El Niño (64), El Niño Modoki (2), central Pacific El Niño (53) or warm tongue El Niño (62). Although there is no consistent name for this phenomenon, here we will refer to it as the El Niño Modoki.

Research suggests that this failure to capture the full spectrum of ENSO events, in addition to climate change, could result in inaccurate predictions by seasonal forecast systems (112). This was evident in the 2015-2016 ENSO event, where statistical and dynamical models predicted above-average winter and spring precipitation over California similar to the 1982-1983 and 1997-1998 ENSO events (74). However, this event resulted in below-normal precipitation anomalies in the Western United States despite having a predicted Niño 3.4 index similar to events with heavy precipitation in that area.

Understanding the structure and developments of these atmospheric teleconnections in current and future climate scenarios is of great importance for extreme weather prediction and forecasting in general. However, such patterns can be difficult to detect on various spatial and temporal scales in large climate datasets. These patterns have been traditionally identified using linear models like principal components (PCs). However, the strict bi-orthogonality of the components can make them difficult to interpret and do not necessarily relate to a physical phenomenon. These analyses are also limited to capturing only linear dependencies and fail to offer any insight on non-linear relationships. Furthermore, extreme events are often identified using hand-picked thresholds based on human expertise and physics-based simulations with no universal definition. Recent advances in AI and ML techniques have shown great promise in the area of pattern recognition, spatial analysis, predictability and tracking problems, especially in large and complex datasets. Studies have demonstrated the usefulness of various ML methods, including CNNs, ANNs, self organizing maps (SOM), and autoencoders, to detect, track and predict extreme weather events or large-scale teleconnections in observational and model climate data. These approaches have also shown advantages in identifying non-linear modes of climate variability over other traditional methods (71; 56; 110; 43; 134).

1.1.3 Aim 3: Emulation of Complex Parameterization

Earth system processes are primarily described by non-linear differential equations that interact with other complex and non-linear components on differing spatial and

temporal, resolved and unresolved scales. The increase in the degrees of freedom with resolution implies that solving these multi-scale equations for high resolution models becomes computationally prohibitive, even on leadership-class high performance computing. Moreover, the unresolved processes also account for a large source of uncertainty in ESMs, as they are approximated from the resolved scales through parameterization schemes.

Recently, researchers have moved towards ML techniques to aid in parameterizations of a wide range of variables including but not limited to radiation (18; 19; 61; 89), ocean physics (59; 108), convection (60; 88) and ozone (87). ML can be used to create new parametrizations, or improve current ones, by training a statistical model to effectively emulate a complex physical phenomenon. ML techniques are capable of learning and accurately simulating nonlinear functions when trained on large datasets. Once trained, these models can provide fast calculation of sub-grid scale processes, particularly on hybrid architectures. Current ML tools used for parametrization tasks include neural networks (NN), deep learning, random forests (RF), and general adversarial networks (GAN).

1.1.4 Research Objectives

While there have been significant contributions to the application of ML techniques for ESM improvement, researchers are still in the beginning stages of what could be achieved. Described in this dissertation is the implementation of advanced ML techniques to improve the three areas of concern mentioned above: downscaling of

ESM model output data using a novel computer vision approach, identifying non-linear modes of climate variability in complex model data and reducing run-time computational expense through deep learning parametrization. My goal is to improve the performance and understanding of ESMs by using the latest ML algorithms. Through this work, I also hope to aid in closing the gap between data science and climate science.

Energy Exascale Earth Systems Model

In addition to using observational data, I also focus on applying ML techniques to data derived from the DOEs Energy Exascale Earth Systems Model version 1 (E3SMv1) (28) in aim 1 and 2. In aim3, I improve E3SM's computational performance using emulators to replace the full radiation scheme. fully coupled E3SMv1 was developed from the Community Earth System Model (CESM1) (50). Improvements from the CESM1 include:

- The E3SM Atmosphere Model (EAM) component which has a spectral-element dynamical core, replaced parametrizations of turbulence, cloud microphysics and shallow convection, enhanced aerosol parameterizations and a vertical resolution increased to 72 levels

- New ocean and sea ices components based on the Model for Prediction Across Scales (MPAS) framework that uses Spherical Centroidal Voronoi Tessellations (SCVTs) for multiresolution modeling

- The Model for Scale Adaptive River Transport (MOSART)
- The E3SM land model (ELM) that has options for representing soil hydrology and biogeochemistry

These changes are described in more detail in (36). The performance of the E3SMv1 has shown to be consistent with other ESMs, with some notable improvements over these different models, including increased representation of tropical variability and a more accurate ability to simulate the Madden Julian Oscillation (36). The E3SMv1 has also shown improvement in ocean and sea ice behavior and finescale atmospheric features in high-resolution simulations.

Chapter 2

Reconstructing High Resolution ESM Data Through a Novel Fast Super Resolution Convolutional Neural Network (FSRCNN)

Aim 1: Statistical Downscaling

A version of this chapter has been accepted for publication in Geophysical Research Letters:

Passarella, L. S., Mahajan, S., Pal, A., Norman, M. R. Reconstructing High Resolution ESM data through a novel Fast Super Resolution Convolutional Neural Network (FSRCNN). Geophysical Research Letters, e2021GL097571.

No revisions to this chapter have been made since the original publication.

2.1 Abstract

We present a first application of a fast super resolution convolutional neural network (FSRCNN) based approach for downscaling earth system model (ESM) simulations. Unlike other SR approaches, FSRCNN uses the same input feature dimensions as the low resolution input. This allows it to have smaller convolution layers, avoiding over-smoothing, and reduced computational costs. We adapt the FSRCNN to improve reconstruction on ESM data, we term the FSRCNN-ESM. We use high-resolution ($\sim$0.25°) monthly averaged model output of five surface variables over North America from the US Department of Energy's Energy Exascale Earth System Model's control simulation. These high-resolution and corresponding coarsened low-resolution ($\sim$1°) pairs of images are used to train the FSRCNN-ESM and evaluate its use as a downscaling approach. We find that FSRCNN-ESM outperforms FSRCNN and other super-resolution methods in reconstructing high resolution images producing finer spatial scale features with better accuracy for surface temperature, surface radiative fluxes and precipitation.

2.2 Plain Language Summary

High resolution global climate data is computationally expensive to run but crucial for assessing climate change effects at local and regional scales. Here, we adapt a new deep learning technique, called fast super-resolution convolutional neural network, to remap climate data from low resolution to high resolution grids. This approach is

faster and more accurate for statistical downscaling climate data compared to other prevalent methods.

2.3 Introduction

Accurate and reliable climate data is critical for assessing the risk of climate change to our society's well-being. Increases in temperature, sea-level, and extreme weather events can render many aspects of our society vulnerable including our health, natural resources and energy-systems (86; 111). Local and regional climate future projection data is the most crucial for planning and mitigating these risks, but is often the least reliable (101). Current Earth System Models (ESMs) used for simulating Earth's past climate and future projections are often computationally limited to coarse horizontal resolutions, generally between 1° to 3° (113). These low resolution models fail to accurately simulate important physical processes such as precipitation extremes (55). Recent advances in computing resources have allowed for global ESMs to be run at higher resolutions ($\sim$0.25°) for longer time periods and have shown to improve the simulations of regional mean climate as well as extremes (126; 76). However, these high resolution models remain prohibitively expensive.

A computationally less expensive approach to derive high resolution climate data over a region of interest is to map data from low resolution global model simulations to high resolution grids using dynamical or statistical downscaling techniques. Dynamical downscaling involves running high resolution regional dynamical models

to extrapolate large scale boundary conditions obtained from a coarser global ESM to finer resolutions on regional scales. Statistical downscaling (SD) aims to map coarse resolution data to high resolution projections using statistical methods like linear regression. Recent studies have shown that machine learning techniques, like neural networks (117; 29) and support vector machines (34), for SD significantly outperform other traditional SD methods. In this study, we use one such computer vision approach called super-resolution (SR), which generates a high resolution image from its low resolution equivalent. SR techniques attempt to generalize across images and have been shown to learn local scale patterns more efficiently than other downscaling methods (113).

One pioneering work in SR deep learning is a SR convolutional neural network (SRCNN). A convolutional neural network is a type of artificial neural network that convolves a kernel with the data to extract features for further use in the overall neural network architecture (66; 65; 37). The SRCNN was originally proposed by Dong et al. (26) and was shown to achieve significantly better performance over other traditional and state of the art SR methods. Vandal et al. (114) demonstrated the usefulness of using a stacked SRCNN, called DeepSD, to downscale ensemble ESMs and showed that it outperforms other methods including bias correction spatial disaggregation (BCSD), artificial neural networks (ANN), Lasso and support vector machines (SVM). Several recent studies have used similar super resolution approaches to downscale ESMs to higher resolutions and demonstrated significant skill. These include a super resolution general adversarial network (SRGAN) (104) to downscale wind and

solar radiation fields to 50x resolutions and a Laplacian pyramid super-resolution network, termed, ResLap, that uses a residual dense block to allow hierarchical feature extraction from the convolutional layers (16).

One common feature of these SR approaches is a pre-processing step where the low resolution images are pre-interpolated to the desired high-resolution output image size (say, using bilinear interpolation) before running the network. Dong et al. (27) developed a method, termed as fast SRCNN (FSRCNN), that alleviates this pre-processing and replaces it with a deconvolution layer at the end, facilitating mapping directly at the resolution of the low resolution image. Since the computational complexity of a CNN is proportional to the input image size, this lowers the computational cost of the network significantly - almost by a factor of n^2 compared to a SRCNN, where n is the downscaling factor. Further, the smaller input image size in FSRCNN implies that narrower filters can capture the same information, thus allowing for more filters for greater feature extractions while also lowering computational cost. This use of smaller kernel sizes to improve CNN models was proved in Simonyan and Zisserman (103). The FSRCNN has been shown to improve image reconstruction skill significantly compared to a general SRCNN for high downscaling factors, with the convolution-deconvolution structure reducing edge smoothing and loss of detail and improving feature reconstruction. Further, the same FSRCNN network can be used for different upscaling factors with only the deconvolution layer needing further tuning.

Here, we present a first attempt to apply FSRCNN for ESM downscaling and find that it is generally more skillful than DeepSD. Further, we improve upon the FSRCNN by adapting it to use additional SRCNN-like convolutional layers after the deconvolution step. By adding these additional convolutional layers, we are able to extract more information and finer spatial details in the high resolution images. We refer to this new adapted FSRCNN architecture as FSRCNN-ESM. Following previous validation studies (104) of the application of super resolution approaches to climate data, we also reconstruct high resolution data from a coarsened version of the same data. We evaluate the reconstruction results using an objective evaluation metrics like the mean absolute error, and find FSRCNN-ESM to be a promising downscaling method with superior skills as compared to both DeepSD and FSRCNN. Section 2 describes the ESM data used as well as the improved FSRCNN network architecture in more detail. We present the results of our objective and subjective evaluations in Section 3 and summarize our work in Section 4.

2.4 Methods

2.4.1 Data Collection

For this study we use monthly output of a 30-year segment of the 1950-control simulation with the global high resolution (0.25°) configuration of the Energy Exascale Earth Systems Model (E3SM) (28). Model data was obtained from the Earth System Grid Federation (ESGF) (20). It should be noted that E3SM data is bilinearly

interpolated from its native non-orthogonal cubed-sphere grid to an equivalent regular 0.25°x0.25° longitude-latitude grid. We call this model data, E3SM-HR. This high resolution data is interpolated to a 1°x1° grid using a bicubic method to obtain the corresponding low resolution input images, which lose the fine scale features present in the high resolution data. Our goal is to reconstruct the high resolution images back from these coarsened data using FSRCNN. When using a computer vision approach to gridded E3SM data, we can think of each grid point as a pixel in an image. For this study, we look at a subset of E3SM data corresponding to North America. The low-resolution images as a result are 60×60 pixels, while high-resolution images are four times larger across both dimensions and have a size of 240×240 pixels each.

We chose five variables to test the FSRCNN - surface temperature (TS), shortwave heat flux (FSNS), longwave heat flux (FLNS), precipitation convective rate (PRECC) and the large scale precipitation rate (PRECL). This results in 360 images for one variable, or 1,800 images in total. We use all the variables together, each normalized using the min-max scaler, in a single one-channel network when training so our algorithm can learn how to extract multiple spatial features. The addition of multiple variables in one network enhanced our reconstruction. For example, when training just PRECC on a single network, the average mean square error (MSE) for testing was $7.58e-7$. When using all variables together, the MSE on the testing dataset for PRECC is $2.79e-11$ (see table 1). It is common in computer vision to learn many different classifications of images in a single algorithm to improve feature

reconstruction. Before training the model, we split the last 3 years of data into a testing set, corresponding to 180 images or 36 images for a single variable.

We also explore the use of elevation as a second input channel to our methods. Several studies have shown the addition of elevation as an input is important for enhancing the reconstruction quality of precipitation data (113; 70).

2.4.2 Deep Learning

SRCNN and DeepSD

The SRCNN is trained to directly learn practical mappings between low resolution and high resolution images with little pre- and post-processing (26). The low resolution image must be interpolated to the desired output size before training using a bicubic method. The SRCNN consists of three operations: patch extraction and representation, non-linear mapping and reconstruction (26). The goal of an SRCNN is to take the low resolution image Y and generate a high resolution image G that is close to the ground truth image I.

Layer 1 of the SRCNN consists of 64 filters of 9x9 kernels, layer 2 has 32 filters of 1x1 kernels and the output layer has a single filter with a 5x5 kernel, the same as described in Dong et al. (26). We trained this SRCNN on 100 epochs with an adam loss optimizer, sigmoid activation function and an initial learning rate of 0.001 using tensorflow. The choice of a sigmoid activation function allows for back propagation

to return an output value between 0 and 1 which is useful in this context, since we normalize the images before training with a min-max scaler.

Vandal et al. (113) uses a stacked SRCNN apporach called DeepSD and here we test that method against the FSRCNN-ESM. DeepSD uses elevation as a second input channel for the SRCNNs to train on. The first SRCNN is used to interpolate the images from a 1° resolution to 0.5°. The estimated 0.5° resolution images are then passed to the next SRCNN and interpolated to the final 0.25° images. Here, we only use 2 stacked SRCNN, as opposed to 3 in the original paper, since we are only downscaling by a factor of 4. It is important to note that Vandal et al. (113) used the DeepSD method to downscale one variable, precipitation, and here we are using it to downscale five.

General FSRCNN

The basic FSRCNN method was created to accelerate the SRCNN process and the redesign involved three features: (1) smaller convolutional kernels but more feature maps, (2) an added deconvolutional layer, and (3) the input feature dimension is the same as the original low resolution image (73). Because the FSRCNN takes the original 60×60 image as the input and does not have to interpolate it to a 240×240 image, it learns 16 times fewer weights than the SRCNN and as a result is much faster when training. The final reconstruction to the HR image then requires a deconvolutional layer at the end to remap the data from the low resolution reconstruction steps to a higher resolution grid. We note that the deconvolutional

layer is not the same as an unpooling plus convolutional layer, sometimes known as convolutional interpolation, which in its purest form resizes an image by copying pixels as many times as needed to achieve the desired image size before passing through a convolutional layer. A deconvolutional layer, also known as a transpose convolutional layer, instead pads the image with zeros to desired size before upsampling the image using learned kernels. dong2016accelerating found that replacing the deconvolutional layer with an unpooling layer resulted in a significant drop in reconstruction quality.

The FSRCNN can be broken down into five parts: feature extraction, shrinking, nonlinear mapping, extension, and deconvolution (27). The feature extraction step in Dong et al. (27) consists of a 5 by 5 filter size with d number of filters. The number of filters can be thought of as the number of desired learned features in the low resolution image. The shrinking step is a 1 by 1 filter with s number of filters, here $s < d$, that acts as a way to condense the number of features found in step 1. The nonlinear mapping step maps the features in step 2 nonlinearly to a new set of features. It uses multiple layers of nonlinear mapping with a filter size of 3 by 3. By selecting smaller convolutional kernels but more feature maps (large d), the FSRCNN learns more nonlinear features in the data and creates better SR reconstruction results compared to the SRCNN. The FSRCNN then moves on to the expansion layer, which acts like an inverse of the shrinking layer, to generate a larger number of feature maps to improve high resolution reconstruction. Finally, the FSRCNN uses the deconvolution step to achieve the final high resolution image.

FSRCNN-ESM Architecture

Here, we expand upon the basic FSRCNN method to maximize accuracy for image reconstruction for E3SM data. We include additional convolutional layers after the deconvolution step in the FSRCNN - an added patch extraction step consisting of 64 kernels, and a nonlinear mapping step with 32 kernals, similar to those in the SRCNN - both of which are applied to the high resolution data generated after the deconvolution step as shown in Fig. 1. Fig. 1 shows the SRCNN, original FSRCNN and the new configuration. We refer to this new network as FSRCNN-ESM. We found that these additions to the network further improve image reconstruction in our data, as determined by a loss function. The loss is calculated by using a pixel-wise MSE using the following equation:

$$MSE(I, G) = \frac{1}{N} \sum_{i,j} (I_{i,j} - G_{i,j})^2 (2.1)$$

where I is the original high resolution image and G is the generated high resolution image and i and j denote the location of the pixel. The training loss decreased by 50% with the added feature extraction step after the deconvolutional layer in the FSRCNN-ESM compared to the original FSRCNN architecture.

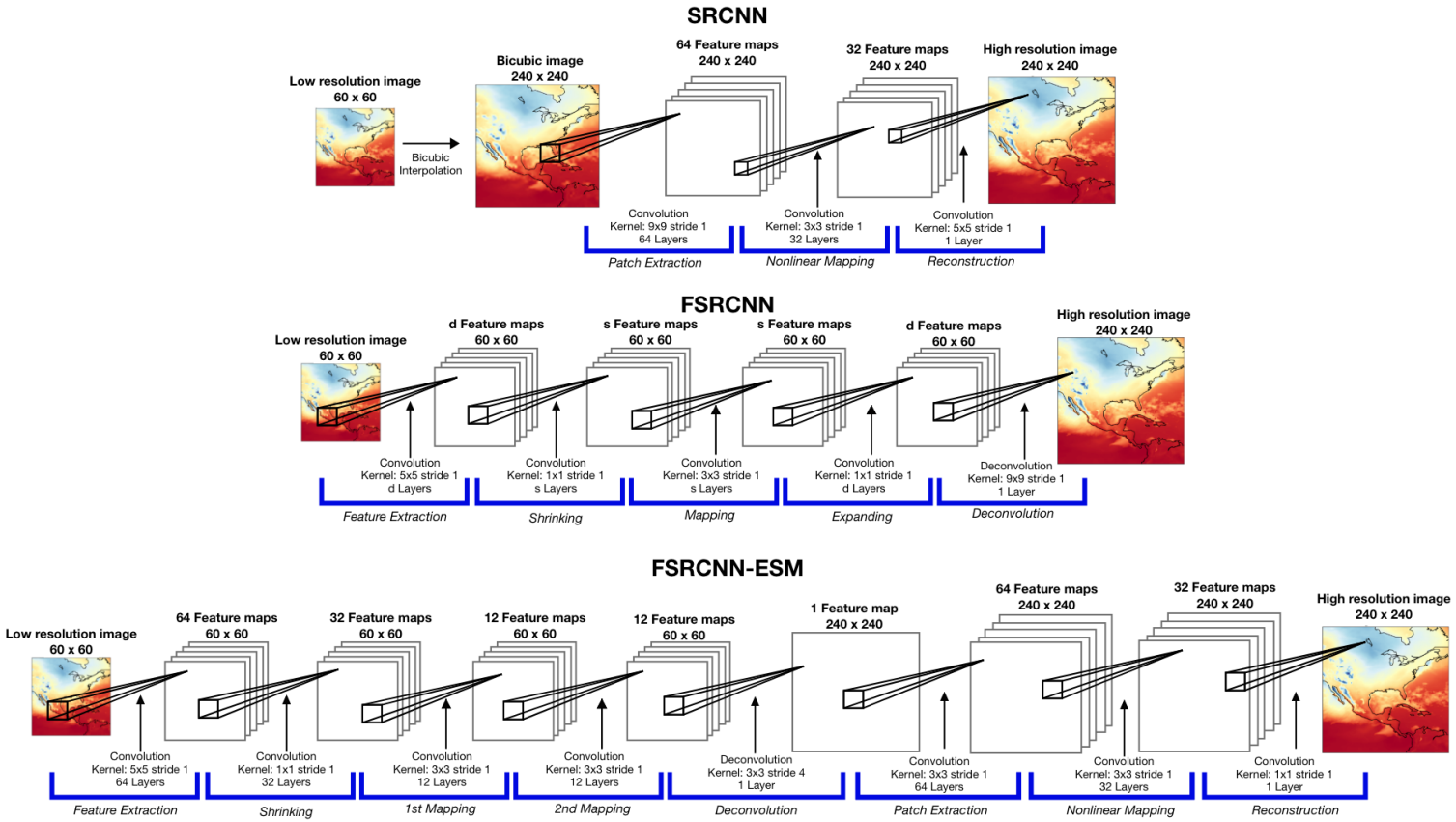


Figure 2.1: The network architecture of the SRCNN (24,513 trainable parameters), FSRCNN (15,681 trainable parameters) and our expanded FRSCNN-ESM (34,246 trainable parameters) used in this study. We add feature extraction and nonlinear mapping layers after the deconvolutional layer because it was shown to significantly enhance image reconstruction in our E3SM data.

Similarly, the average reconstruction loss in MSE for the FSRCNN was 0.000568 compared to the average MSE for the FSRCNN-ESM at 0.000261.

We also evaluate the impact of the addition of elevation as a second input on reconstruction quality. Table 1 shows the MSE for variables in the test set for the FSRCNN, FSRCNN plus elevation, FSRCNN-E3SM and FSRCNN-E3SM plus elevation. Elevation improved the simple FSRCNN method, but we did not see the same improvement for the FSRCNN-ESM. Therefore, we add elevation as an input channel to the FSRCNN but not our FSRCNN-ESM in our further evaluations with the testing data.

2.4.3 Evaluation Metrics

We evaluate the mean absolute error (MAE) for each reconstruction method on each variable the testing dataset. Before computing the MAE, we scale the variables back to their original values using min-max scaler. We define the MAE as follows:

$$MAE = \frac{1}{N} \sum_{1}^{N} |G_i - I_i| \tag{2.2}$$

where G_i is the predicted image i and I_i is the actual truth image. The MAE for each variable is expressed in the variable's units. Here N is 36, or the number of images per variable in the testing dataset.

We also evaluate the skill of each method's ability to reconstruct high resolution images by computing the L_1 and L_2 error for each sample as follows:

Table 2.1: The MSE for each variable in the test data set for the FSRCNN, FSRCNN plus elevation, FSRCNN-ESM and FSRCNN-ESM plus elevation.

	TS	FSNS	FSNT	PRECC	PRECL
FSRCNN	3.251e-5	4.377e-5	3.563e-5	3.299e-5	3.767 e-5
FSRCNN + elevation	**3.189e-5**	**4.285e-5**	**3.499e-5**	**3.246e-5**	**3.709e-5**
FSRCNN-ESM	**2.628e-5**	**3.582e-5**	**3.026e-5**	**2.796e-5**	**3.233e-5**
FSRCNN-ESM + elevation	3.336e-5	4.467e-5	3.841e-5	3.614e-4	4.167 e-5

$$L_1 = \frac{\sum_{i,j} |G_{i,j} - I_{1,j}|}{\sum_{i,j} |I_{1,j}|}, L_2 = \frac{\sum_{i,j} (G_{i,j} - I_{1,j})^2}{\sum_{i,j} (I_{1,j})^2} \tag{2.3}$$

We compute both the L_1 and L_2 errors and the MAE across each variable on the held out testing dataset.

2.5 Results

We compare the FSRCNN-ESM method as applied to E3SM data with the DeepSD method (26) and with a basic bicubic interpolation using the above stated metrics for determining image reconstruction quality: MSE, MAE, L1- and L2-error. We also compare the computational training time of the FSRCNN and FSRCNN-E3SM against DeepSD.

2.5.1 Reconstruction Evaluation

Figure 2 show samples of reconstruction of high resolution images over a part of Northern America from low-resolution images using bicubic interpolation, DeepSD, FSRCNN and FSRCNN-ESM approaches. We randomly pick one sample of a summer (June) month from 3 years of reconstruction test data for each of the five variables to visually illustrate the reconstruction quality. The inset plot shows a zoomed in portion to better visualize some of the finer spatial details of some region, generally over the Rocky Mountains and over Northern Andes - regions that show strong gradients in the high resolution images owing to the topography there. The plots (first column)

show the loss of fine-scale features as the high resolution images (last column) are interpolated to low-resolution images. To restate, the goal is then to reconstruct the high resolution image from these low resolution images. It is clear that a bicubic interpolation to downscale performs poorly. The DeepSD generally improves over the bicubic interpolation, but still lacks the finer-scale details noted in the high resolution images - this is apparent in most of the images for all variables. For example, the DeepSD is not able to discern the strong gradients over the Northern Andes clearly for convective precipitation (PRECC) (Figure 2, second and fourth rows). Similarly, it is not able to capture the fine scale features of surface temperature (TS), net surface shortwave radiation (FSNS), net surface longwave radiation flux (FLNS) and large-scale precipitation (PRECL) over North America in the summer (Figure 2, first, third and fifth rows). It is clear visually that FSRCNN-ESM better generates finer details in high resolution data when compared to the DeepSD and bicubic interpolation, which both tend to over-smooth the images. For the above stated examples, the FSRCNN-ESM is able to isolate the strong gradients in PRECC over the Northern Andes and in TS, FSNS, FLNS and PRECL over North America in the summer.

To quantify the reconstruction skill of these methods, we use the MAE metric over the testing dataset. Table 2 summarizes the results of our analysis for the three methods. We calculate the MAE for each variable image over each month in the testing dataset. Consistent with the visual illustrations (Figure 2), the FSRCNN-ESM generally outperforms the other methods in terms of this metric across most variables and months, with the exception of FSNS in which the FSRCNN had a

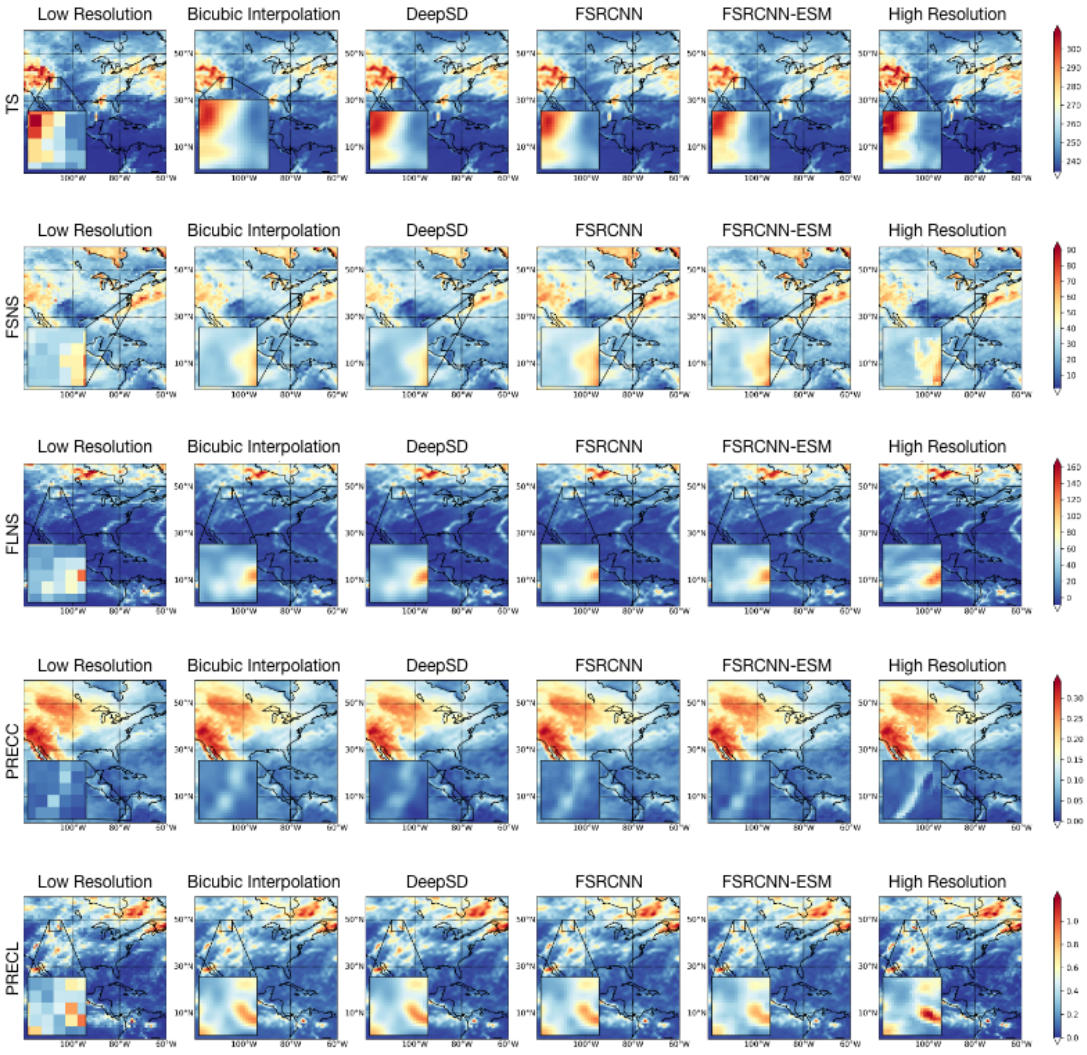


Figure 2.2: Super resolution reconstruction results for a sample in June.

higher reconsturction skill on average. Dong et al. (27) noted that a large fraction of the increase in reconstruction results came from replacing the bicubic interpolation step in the SRCNN with the deconvolution layer in FSRCNN when they partitioned their error metric into different steps. We see this here with the increase in skill from DeepSD to both the FSRCNN and FSRCNN-ESM. It is interesting to note that the bicubic interpolation skill is comparable to DeepSD for our data, and exceeds it for most variables. Most methods tend to underestimate precipitation variables. Precipitation occurs intermittently and is a highly non-linear process and results from multi-scale, multi-phase physical processes creating large spatial heterogeneity. The above suggest that image reconstruction becomes more difficult as spatial heterogeneity increases.

We also use the L_1 and L_2 error metrics to quantify reconstruction skill of the various methods on our held out testing dataset. Figure 3 shows the histogram of the L_1 and L_2 errors for FSRCNN-ESM (a,e), the original FSRCNN (b,f), DeepSD (c,g) and bicubic (d,h) respectively.The majority of our FSRCNN-ESM predicted samples (Figure 3a) have a L_1 error less than 10% and an L_2 error less than 1% suggesting the effectiveness of our FSRCNN-ESM for interpolating E3SM data to high resolution grids. Our FSRCNN-ESM also generally achieves better reconstruction skill compared to the other two methods based on both the metrics, consistent with the findings using the MAE metric.

Table 2.2: The mean absolute error (MAE) for surface temperature (TS K) , shortwave heat flux (FSNS w/m^2), longwave heat flux (FLNS w/m^2), precipitation convective rate (PRECC mm/day) and the large scale precipitation rate (PRECL mm/day). The PSNR for the variables is averaged across each month and presented here in terms of logarithmic decibel scale.

	Bicubic TS	DeepSD TS	FSRCNN TS	FSRCNN-ESM TS	Bicubic FSNS	DeepSD FSNS	FSRCNN FSNS	FSRCNN-ESM FSNS
Jan	0.377	0.499	**0.013**	0.061	3.869	3.208	1.179	**1.128**
Feb	0.409	0.511	0.106	**0.074**	3.116	2.527	0.949	**0.863**
Mar	0.314	0.437	0.117	**0.046**	1.736	1.116	**0.070**	0.520
Apr	0.294	0.417	0.069	**0.003**	0.318	**0.289**	0.620	0.788
May	0.380	0.504	0.048	**0.011**	**0.319**	0.978	0.617	0.998
Jun	0.432	0.558	0.063	**0.027**	**0.215**	0.847	0.479	0.499
Jul	0.469	0.586	0.141	**0.031**	**0.058**	0.567	0.313	0.312
Aug	0.430	0.547	0.125	**0.006**	0.606	**0.021**	0.238	0.056
Sep	0.394	0.514	0.048	**0.017**	1.304	0.676	0.523	**0.356**
Oct	0.260	0.392	0.070	**0.023**	2.502	1.886	1.162	**0.926**
Nov	0.188	0.335	0.047	**0.030**	3.392	2.715	1.029	**1.028**
Dec	0.327	0.453	0.072	**0.046**	4.088	3.421	1.124	**1.096**
	Bicubic FLNS	DeepSD FLNS	FSRCNN FLNS	FSRCNN-ESM FLNS	Bicubic PRECC	DeepSD PRECC	FSRCNN PRECC	FSRCNN-ESM PRECC
Jan	0.409	0.217	0.187	**0.164**	0.071	0.041	0.038	**0.033**
Feb	0.361	0.169	0.141	**0.115**	0.091	0.061	0.058	**0.053**
Mar	0.302	0.110	0.085	**0.057**	0.092	0.062	0.059	**0.054**
Apr	0.269	0.077	0.054	**0.024**	0.079	0.049	0.046	**0.041**
May	0.222	0.030	**0.009**	0.023	0.023	**0.006**	0.009	0.014
Jun	0.261	0.069	0.046	**0.015**	**0.016**	0.046	0.049	0.054
Jul	0.288	0.096	0.072	**0.042**	**0.033**	0.063	0.066	0.071
Aug	0.294	0.102	0.077	**0.048**	0.033	0.063	0.066	0.072
Sep	0.365	0.173	0.145	**0.120**	**0.041**	0.070	0.073	0.079
Oct	0.362	0.170	0.143	**0.117**	**0.031**	0.061	0.064	0.069
Nov	0.435	0.243	0.211	**0.190**	**0.013**	0.016	0.019	0.024
Dec	0.439	0.247	0.215	**0.194**	0.054	0.024	0.021	**0.016**
	Bicubic PRECL	DeepSD PRECL	FSRCNN PRECL	FSRCNN-ESM PRECL				
Jan	0.106	0.062	0.057	**0.049**				
Feb	0.137	0.093	0.088	**0.080**				
Mar	0.138	0.094	0.089	**0.081**				
Apr	0.119	0.074	0.069	**0.062**				
May	0.034	0.009	0.014	0.022				
Jun	**0.025**	0.069	0.074	0.082				
Jul	**0.050**	0.094	0.099	0.107				
Aug	**0.051**	0.095	0.100	0.107				
Sep	**0.061**	0.106	0.110	0.118				
Oct	**0.047**	0.091	0.096	0.104				
Nov	**0.020**	0.024	0.029	0.036				
Dec	0.081	0.036	0.032	**0.024**				

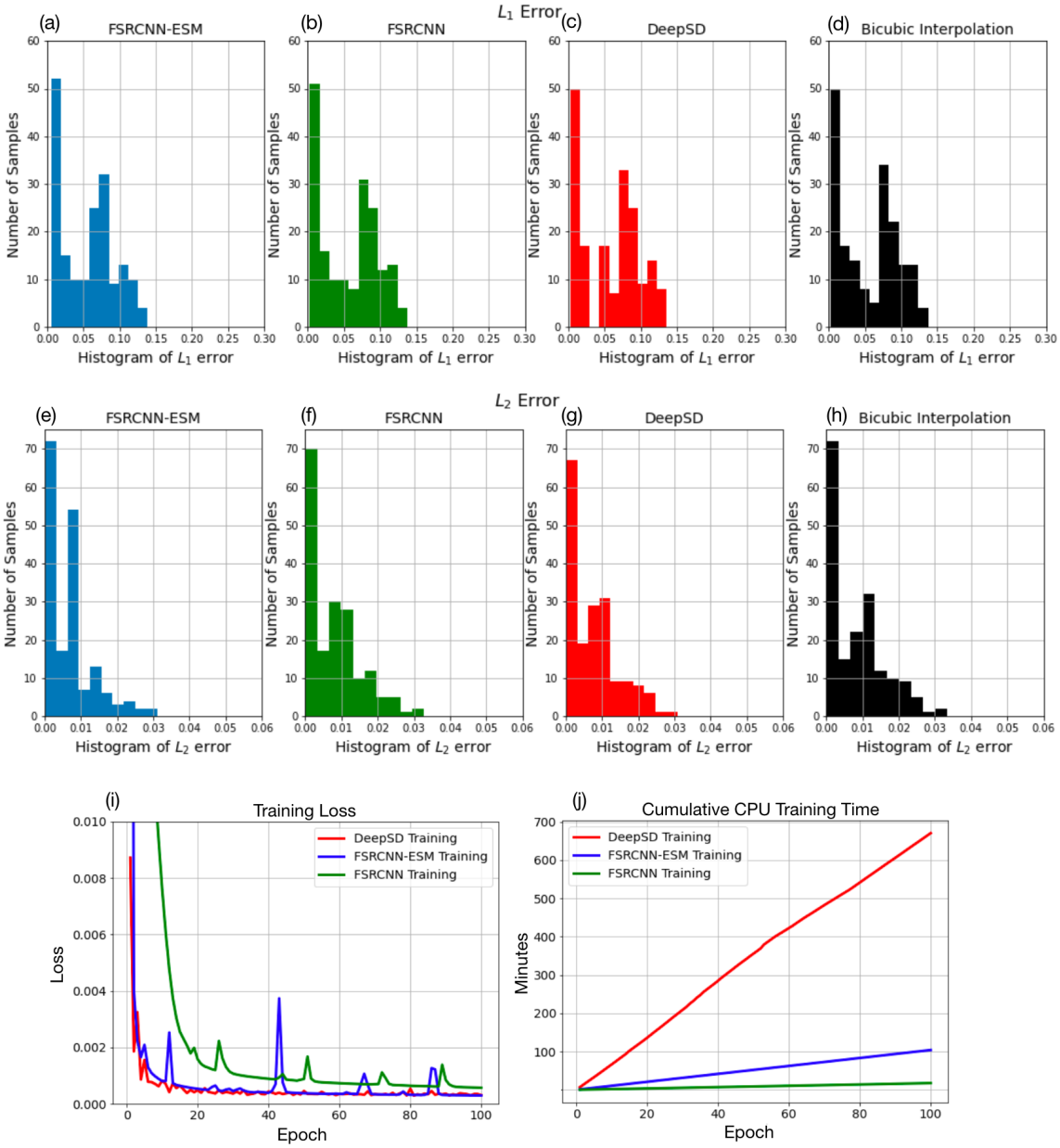


Figure 2.3: L_1 and L_2 error computed for all reconstruction methods on the held out testing dataset (a-h) and the cumulative training time in minutes (i) and the training loss (j) for the DeepSD, FSRCNN and FSRCNN-ESM over 100 epochs.

2.5.2 Computational Performance

Figure 3j shows the cumulative CPU computational time required to train FSRCNN-ESM (blue), DeepSD (red) and FSRCNN (green) on 100 epochs. FSRCNN-ESM completed training 600% faster than the DeepSD, comparable to the decrease in training time with the FSRCNN, while still maintaining similar loss values (Figure 3i). This is largely due to the fact that the DeepSD has a pre-processing step of interpolating low resolution images to the desired output size before training, while the FSRCNN-ESM uses the low resolution image size. As stated earlier this allows using narrower filters on the smaller images, reducing the total number of parameters and computations. Once trained, however, the DeepSD and FSRCNN-ESM take roughly the same amount of time, about 25 seconds, to downscale all 1,800 images.

Dong et al. (27) showed that an FSRCNN with similar number of steps as a SRCNN produced a speed-up of about 40x when upscaling by a factor of three. This was largely (about 30x) due to the narrower filters in the FSRCNN that led to the large difference between the number of trainable parameters in the nonlinear mapping step of SRCNN and that in the corresponding three steps (combination of shrinking, non-linear mapping and expanding steps) in FSRCNN. The use of the low-res image as input in the FSRCNN contributed to the remaining reduction in computational cost of network training. The computation of FSRCNN-ESM only provides a speed up of 6x over the DeepSD for a upscaling factor of four. The increase in cost of FSRCNN-ESM as compared to the base FSRCNN is due to the addition of new patch

extraction, nonlinear mapping and reconstruction steps after the deconvolution step in FSRCNN-ESM. The combination of these new steps is equivalent to a full SRCNN in itself that uses the full HR image size as an input, but with narrower filters. The use of narrower filters in these steps then still allows for a faster training of the overall FSRCNN-ESM network as compared to the DeepSD.

2.6 Summary and Discussion

We apply a novel super-resolution based approach for downscaling ESM data that uses a modified version of the FSRCNN method. We find that this FSRCNN-ESM is able to map low resolution climate images to a four times higher resolution with a better skill than DeepSD, FSRCNN and bicubic interpolation; for surface radiative fluxes and large scale and convective precipitation; while remaining computationally inexpensive to train. Our FSRCNN-ESM is also able to downscale images in a single-step process and without need for access to GPUs for training. The FSRCNN-ESM as a result is a more approchable method of downscaling using machine learning.

This study focused on reconstructing monthly averaged images. In the future, we will explore the application of FSRCNN-ESM for downscaling higher temporal resolution data like daily and sub-daily data, and target extreme events. Also, FSRCNN-ESM does not consider the concurrency of images of the different variables - the network is agnostic of the presence of data for other variables. In the future, we plan to use multi-channel approaches (114; 104; 16) using all the concurrent

variables simultaneously in the network for reconstruction allowing the network access to more correlated data, which may improve reconstruction skill. We also plan to evaluate the FSRCNN-ESM against other recent machine learning approaches that have been used for downscaling, for example, SRGAN (104), and ResLap (16), all of which have been shown to perform better than the SRCNN or other approaches for different variables, regions and scaling factors.This study is the first to use FSRCNN for ESM data and we applied the original version of FSRCNN here, and with some improvements (FSRCNN-ESM), to demonstrate its utility. We plan to explore the applicability of the latest advances in FSRCNN and SRCNN to ESM downscaling in the future. Some advanced applications include using GANs as a latent blank to improve image reconstruction on an SRCNN (12), using a student-teacher supervised learning approach to training (120), applying skip connections to alleviate the vanishing gradient problem (140), and the use of a multi-path residual to improve efficiency in the SRCNN (121).

The resolution of finer scales in the high resolution model and scale-agnostic nature of current sub-grid scale physical parameterizations used in climate models imply that a low-resolution model simulation is not statistically equivalent to coarsened data from a high-resolution model. For example, the simulation of precipitation extremes is found to be stronger in high resolution simulations than the low resolution simulations, even after conservative mapping of high-resolution simulation data to the low-resolution grid (127; 125; 76). In order to apply FSRCNN-ESM directly to low-resolution model output to generate high resolution images at the skill of a high

resolution simulation, we would thus require a bias-correction step. We plan to explore traditional bias-correcting methods as well as utilize machine learning approaches for it. We would also explore the use of nudged simulations or regionally refined models to generate equivalent pairs of low-resolution and high-resolution model simulations that could be used for training a FSRCNN-ESM network.

2.6.1 Acknowledgments

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/ (28).

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.

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript,

Chapter 3

Assessing ENSO-associated Predictability of Southern California Precipitation Using a Multitask Autoencoder

Aim II: Predictability Using Non-Linear Latent Spaces

A version of this chapter has been submission for publication in Geophysical Research Letters

3.1 Abstract

We construct a novel Multi-Input Multi-Output Autoencoder-decoder (MIMO-AE) in an application of multi-task learning to isolate the non-linear dependence of SC-PRECIP on tropical Pacific Ocean sea surface temperature (SST) on monthly time-scales. The MIMO-AE expands on the general autoencoder architecture and is trained on both monthly sea surface temperature (SST) anomalies over deep tropical Pacific Ocean (TP-SST) and Southern California precipitation anomalies (SC-PRECIP) simultaneously. This allows it to capture the co-variability of the two fields in a shared nonlinear latent space that can be condensed into an index, which we term as MIMO-AE index. We use a transfer learning approach by first training the MIMO-AE on 100 years of output from a historical simulation with the Energy Exascale Earth Systems Model version 1 (E3SMv1) and then passing observational data through the trained MIMO-AE model. 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 up-to a two-month lead-time as compared to Niño 3.4 index and the El Nino 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 isolate and maximize 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 two months as compared to other ENSO indices.

3.2 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 (75; 122). This was apparent during the 2015-16 Central Pacific (or Modoki) El Niño event, when California received just above average precipitation, in contrast to the forecast of heavy precipitation associated with the canonical (Eastern Pacific) 1982-83 and 1997-98 strong El Niño events (21; 67; 74). 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 (3). 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.

Statistical modeling, that also suffers from sparse data, suggests that over California 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 inter-annual timescales (52; 48; 119; 17). 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 (46; 130).

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 (112; 129). Several recent studies have explored methods to better capture ENSO's variability, diversity and non-linearity (129). However, these approaches largely devise indices that represent the oceanic or atmospheric variability over the tropical Pacific in isolation of its remote teleconnections.

Empirical orthogonal functions (EOFs), or principal component analysis (PCA), have generally been used to describe the two ENSO flavors, canonical and ENSO Modoki. The Niño 3.4 index and ENSO Modoki Index (EMI) are typically chosen to represent the 2 leading EOF patterns in SST data (2; 68). However, the strict

bi-orthogonality of the components can make them difficult to interpret and do not necessarily relate to a physical phenomenon (116). Several studies have also shown that these two types of ENSO events are not entirely orthogonal to begin with (62; 135; 105). As a result, the nonlinear nature of ENSO has been continuously explored (41; 1; 107; 129). williams2018diversity, for example, developed a new nonlinear ENSO index, the ENSO Longitude Index (ELI), that captures the response of deep convection associated with ENSO across the equatorial pacific. The ELI was also able to better capture remote climate impacts, in this case precipitation over the US, over tradition indices such as the Niño 3.4 index.

Explainable machine learning methods, like autoencoders, offer an avenue for identifying dominant nonlinear 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. They are typically used for dimensionality reduction, analogous to linear transformation methods like principal component analysis (PCA) (44; 123; 124). They, however, transform data to non-linear latent spaces via non-linear activation functions, thus offering the additional capability of capturing the underlying non-linear relationships within the data (124; 13; 81). Simplifying to a single hidden layer autoencoder with linear activation functions yields the PCA vector spaces. Studies show that autoencoders can better detect dominant variability patterns over other techniques, like the PCA (124; 139; 30). Some studies (106; 43) have also demonstrated the use

of autoencoders to effectively identify modes of climate variability, including those related to ENSO.

Multitask learning (MTL) is a sub-field of machine learning used to solve multiple learning tasks at the same time while exploiting commonalities and differences across tasks (11). MTL has been applied to many problems including natural language processing (22; 69; 15), speech recognition (23; 57; 109; 102) and computer vision (35; 24; 54) to improve prediction accuracy and learning efficiency of task-specific models. Recent studies have shown the usefulness of multi-input multi-output (MIMO) convolutional neural networks and autoencoders for segmenting data and extracting useful information when there are multiple variables present (97; 132; 33). For example, ghifary2015domain 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. 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 use data from

an Earth System model (ESM) as well as observational data to train and validate the MIMO-AE. We overcome the lack of available observational data by using a transfer learning approach, whereby, MIMO-AE is trained on the ESM data and applied to observational data. ham2019deep used a similar transfer learning approach to train a convolutional neural network (CNN) with global SST and heat content data from historical simulations of 21 CMIP5 models to predict the observed Niño 3.4 index.

3.2.1 Model Simulations and Data

We train and test the MIMO-AE on a historical 165 year run (100 years for training and 65 for testing) of the Energy Exascale Earth System Model version 1 (E3SMv1) (28). The atmosphere component model features a 110km nominal horizontal resolution and 72 vertical layers and the ocean model's nominal resolution ranges from 60 km in the mid-latitudes and 30 km near the equator and poles with 60 vertical layers. E3SMv1 is found to effectively capture temporal variability of ENSO and reproduce ENSO associated spatial SST patterns when compared to observational datasets (36), 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 (78).

We use observed SSTs from the Hadley Centre Global Sea Ice and Sea Surface Temperature (HadISST 1.1) dataset that spans 1870-2020 at a 1° resolution (96). Data can be downloaded from the web at https://www.metoffice.gov.uk/hadobs/hadisst/. We use observed precipitation data from NOAA's PRECipitation

REConstruction over Land (PREC/L) at 1° resolution (14). PREC/L is a global analysis of interpolated rain gauge observations from 1948 to 2020 and can be found open access at https://psl.noaa.gov/data/gridded/data.precl.html. We trim the observed SST data to range from 1948-2020 to match the available observed precipitation data.

3.3 Methodology

3.3.1 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 3.1a, contains a hidden layer h that describes a representation of important attributes of the input (37). The general autoencoder consists of two parts: an encoder and a decoder. The encoder maps input x to h by a chosen activation function $f()$,

$$h = f(x \cdot w_e) \tag{3.1}$$

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

$$x' = f(h \cdot w_d) \tag{3.2}$$

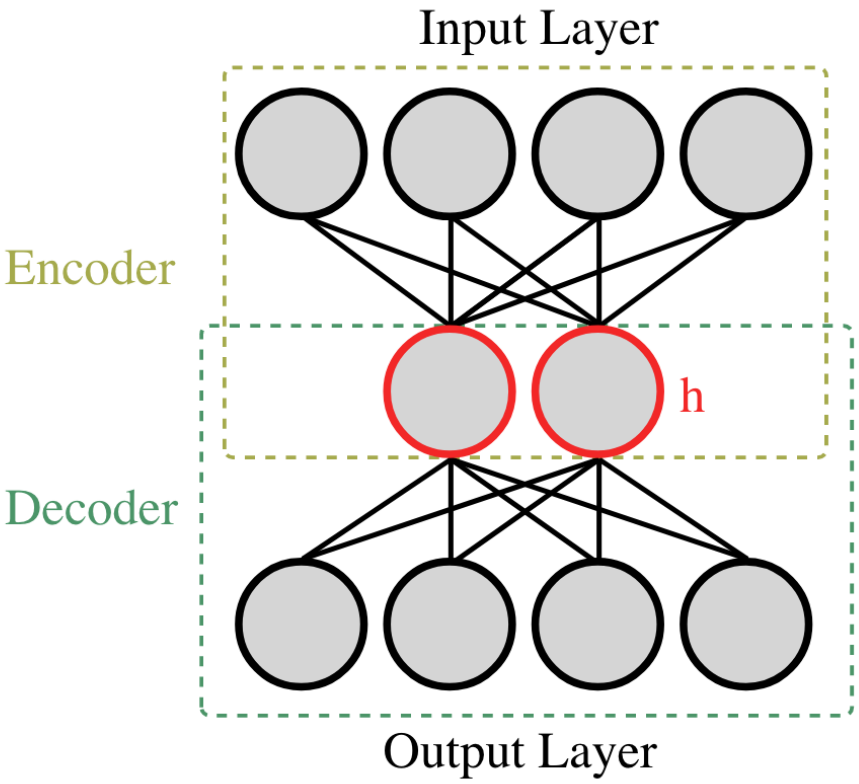


Figure 3.1: A simple autoencoder architecture.

where w_d are the decoder weights.

By using a linear activation function, the autoencoder behaves similarly to a PCA (7; 94). 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 (106) used such an 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. They found considerable asymmetry in the spatial pattern between characteristic El Niño and La Nina episodes from their network. For spatio-temporal data, the temporal vectors at the bottleneck nodes are analogous to PCs. 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 (106).

MIMO-AE

Figure 3.2 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 and precipitation anomalies over

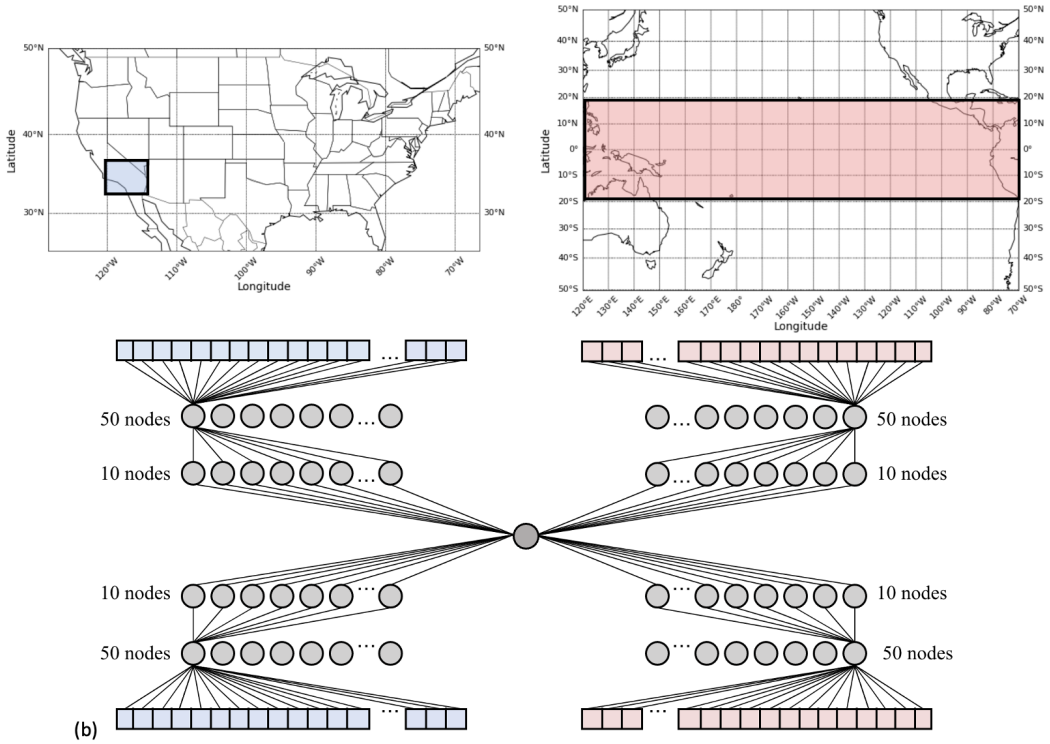


Figure 3.2: Our MIMO-AE Architecture.

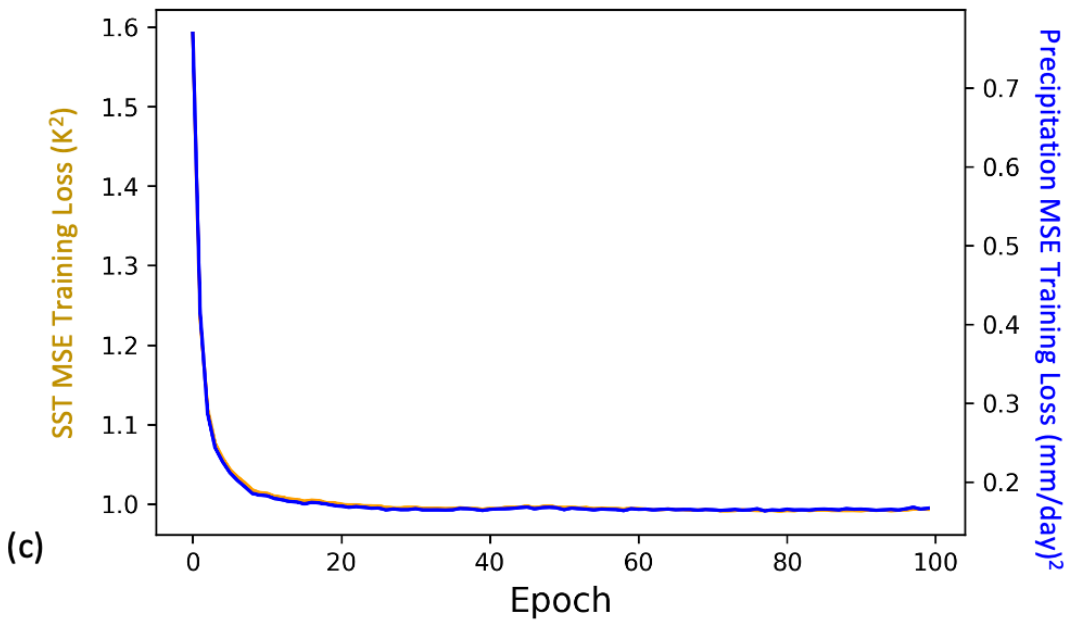


Figure 3.3: Training losses for MIMO-AE over 100 epochs using scaled data (c).

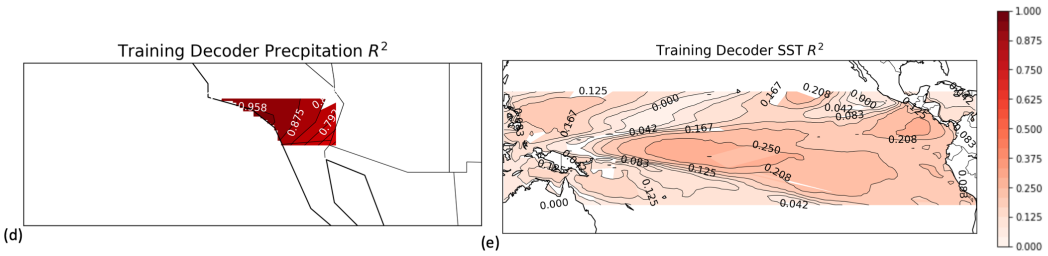


Figure 3.4: The average R^2 between the E3SM-only trained MIMO-AE reconstructed and original input data for Southern California precipitation (d) and Tropical Pacific SST (e).

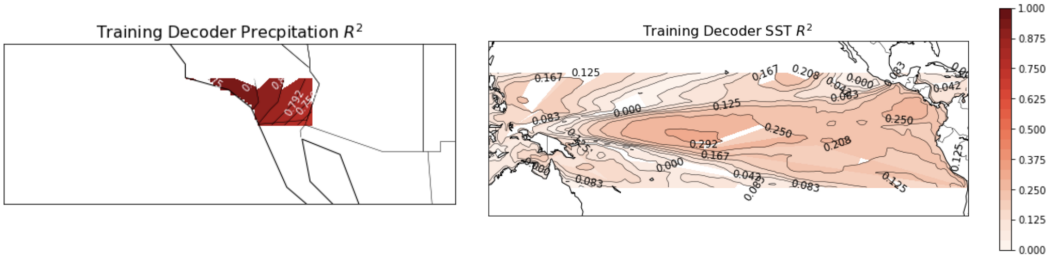


Figure 3.5: The average R^2 between the transfer learning trained MIMO-AE reconstructed and original input data for Southern California precipitation (d) and Tropical Pacific SST (e).

Table 3.1: R^2 values of MIMO-AE indices against traditional ENSO indices over different activation functions.

Activation Function	Niño 3.4	PC1	PC2	ELI	Precipitation
ReLU	0.856	0.843	0.027	0.461	0.589
SeLU	0.813	0.833	0.012	0.475	0.641
TanH	**0.892**	**0.837**	**0.035**	**0.511**	**0.651**
Leaky ReLU	0.928	0.906	0.002	0.491	0.483
eLU	0.766	0.841	0.029	0.491	0.614
sigmoid	0.868	0.964	0.029	0.546	0.306

each grid box in the boxed domain over Southern California (Figure 3.2). 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 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 found that using more than one hidden node resulted in the MIMO-AE learning either the precipitation pattern or the SST pattern and not the shared variability of both.

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:

$$MSE = \frac{1}{N}\sum_{i}(P_i - T_i)^2 \tag{3.3}$$

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. 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 encoder (orange) and decoder (green) are shown in figure 3.3.

Figure 3.4d 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 about 80% of the variability of Southern California for most grid points and about 30% 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 chose a network that explained a larger fraction of SC-PRECIP rather than (say) equally for both variables, since our goal was largely to assess predictability of SC-PRECIP. Likewise, a network can be designed to explain a larger fraction of TP-SST rather than SC-PRECIP.

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. ham2019deep 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.

3.3.2 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 (45). 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 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 of observational dataset used. LSTM models have recently been shown to perform better at time series prediction over linear models for Niño 3.4 index (47; 85; 10; 38), and we use them here to evaluate the predictability of these indices as well as SC-PRECIP. 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 30 years of observational data.

3.3.3 Pearson Correlation

The Pearson Correlation Coefficient (R) was used to evaluate the performance of prediction using the following equation

$$PCC = \frac{\sum_{i=1}^{N}(y_{true_i} - \overline{y}_{true_i}) \cdot (y_{pred_i} - \bar{y}_{pred_i})}{\sqrt{\sum_{i=1}^{N}(y_{true_i} - \overline{y}_{true_i})^2} \cdot \sqrt{\sum_{i=1}^{N}(y_{pred_i} - \bar{y}_{pred_i})^2}} \tag{3.4}$$

where y_{true_i} is the observed Niño 3.4 index and y_{pred_i} is the predicted index and N is the sample size.

3.3.4 Effective sample size

Auto-correlation and filtering causes a loss in temporal degrees of freedom when computing the prediction skill of our indices. To correct for this loss, we calculate the effective sample size using the following equation:

$$N_{effective} = \frac{N}{1 + 2\sum_{i}^{N} \gamma_i^2} \tag{3.5}$$

where γ_i is the auto-correlation of our SC-PRECIP time series at lag i Livezey and Chen (72).

3.3.5 Fisher Transformation

We also compute the Fisher transformation of Pearson's correlation for null hypothesis testing and to compute confidence intervals of the correlation skills. The Fisher transformation of Pearson's correlation is approximately normally distributed. The confidence intervals of R for, say, two standard normal deviates can be computed as follows:

$$Z = archtanh(R) \tag{3.6}$$

$$StandardError = \frac{1}{\sqrt{N_{effective} - 3}} \tag{3.7}$$

$$Delta = 1.96 * StandardError \tag{3.8}$$

$$LowerBound = tanh(Z - Delta) \tag{3.9}$$

$$UpperBound = tanh(Z + Delta) \tag{3.10}$$

where R is the Pearson correlation coefficient found in equation 3.4 and $N_{effective}$ is the effective sample size found in equation 3.5.

3.4 Results

3.4.1 MIMO-AE

Figure 3.6a shows the time evolution of the standardized MIMO-AE index for a segment (last 20 years, 1994-2013) of the 65 years of 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 time series of standardized Niño 3.4 index, ELI and regionally averaged SC-PRECIP. We apply a three month running average to each time series. Fig. 3b shows the same but for a segment of the observational data (2000-2019). To reiterate, MIMO-AE index is computed by passing the observed data through the E3SM trained MIMO-AE network. For both E3SM and observations, MIMO-AE index exhibits higher correlation with SC-PRECIP than Niño 3.4 index or ELI, given that precipitation data is fed in the generation of MIMO-AE index. The correlation between MIMO-AE is weak with both Niño 3.4 index and ELI both for E3SM and observational data. 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 capture 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, both ELI and MIMO-AE categorize the 2015/2016 event weaker than the Niño 3.4.

Fig. 3.7c 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 williams2018diversity, 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 co-variability 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 (77). 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.

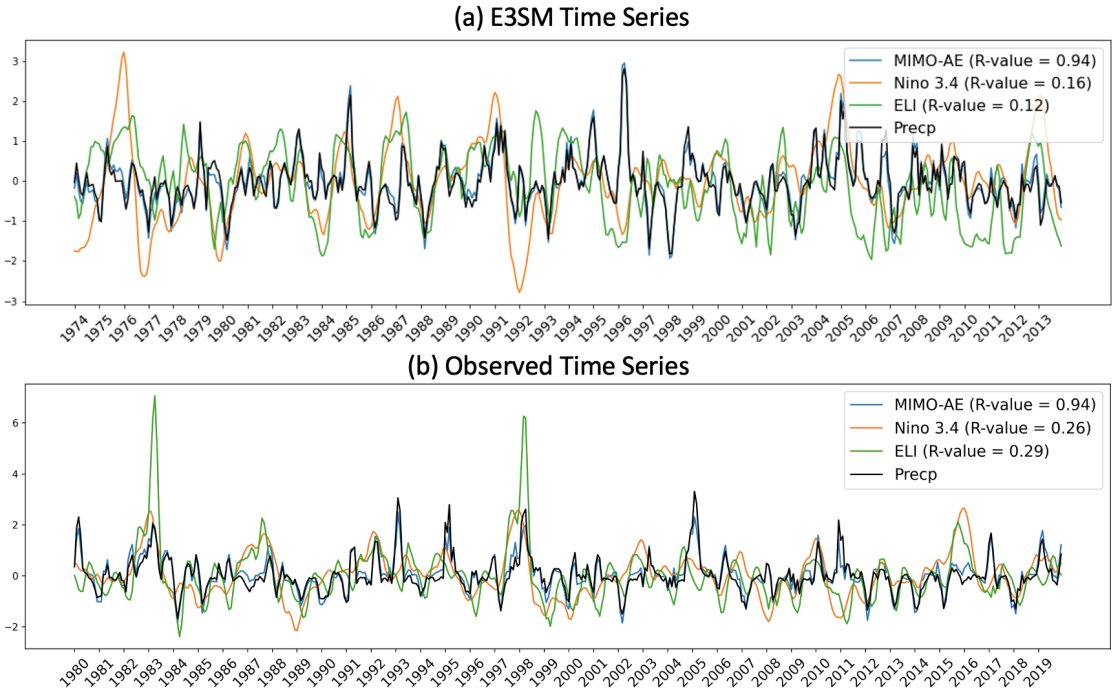


Figure 3.6: Standardized time series of MIMO-AE index (blue), Niño 3.4 (orange), ELI (green) and domain averaged SC-PRECIP (black) for a segment (last 40 years, 1974-2013) of the 65 years of testing data for the E3SM historical simulation (a); and a segment (last 40 years, 1980-2019) of the observed testing data (b).

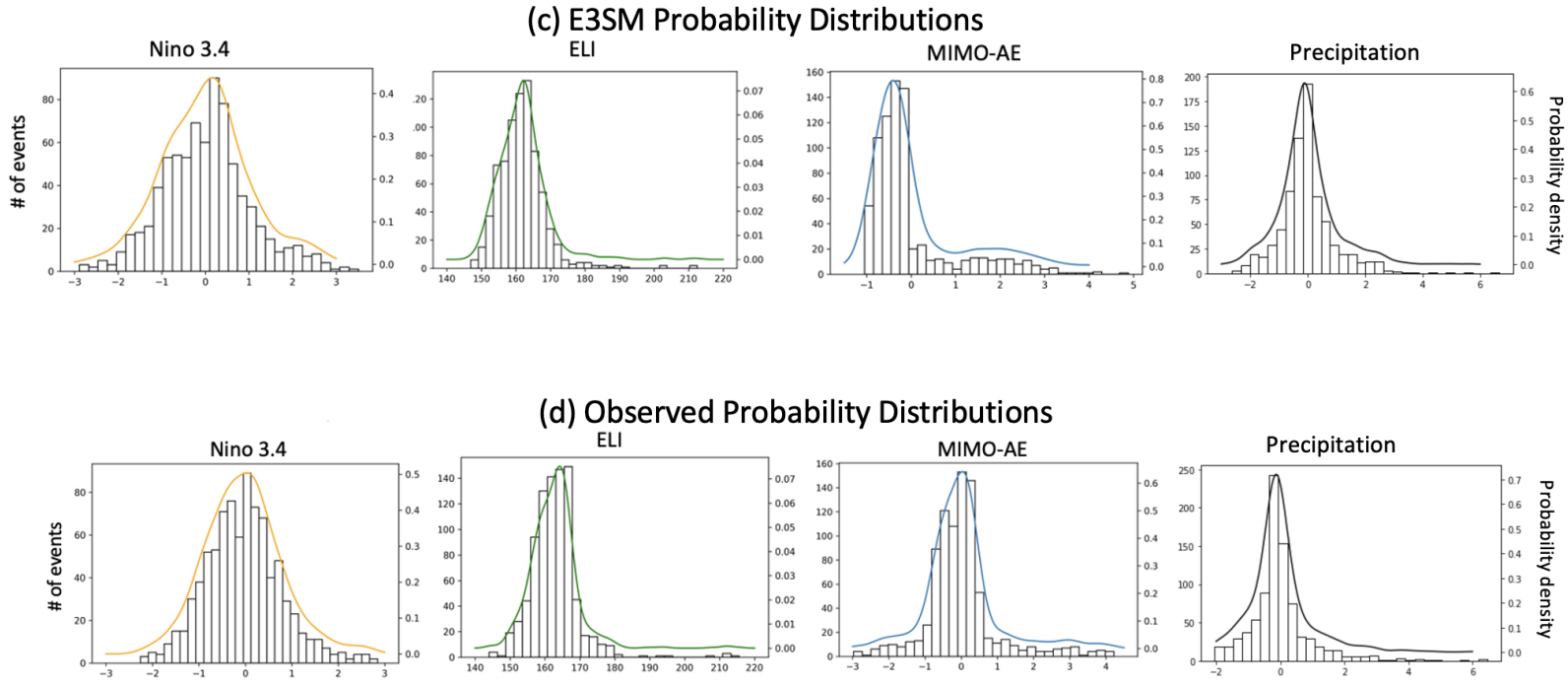


Figure 3.7: The probability density distributions for the Niño 3.4 index (orange), ELI (green), MIMO-AE index (blue) and SC-PRECIP (black) for E3SM data (c); and observational data (d).

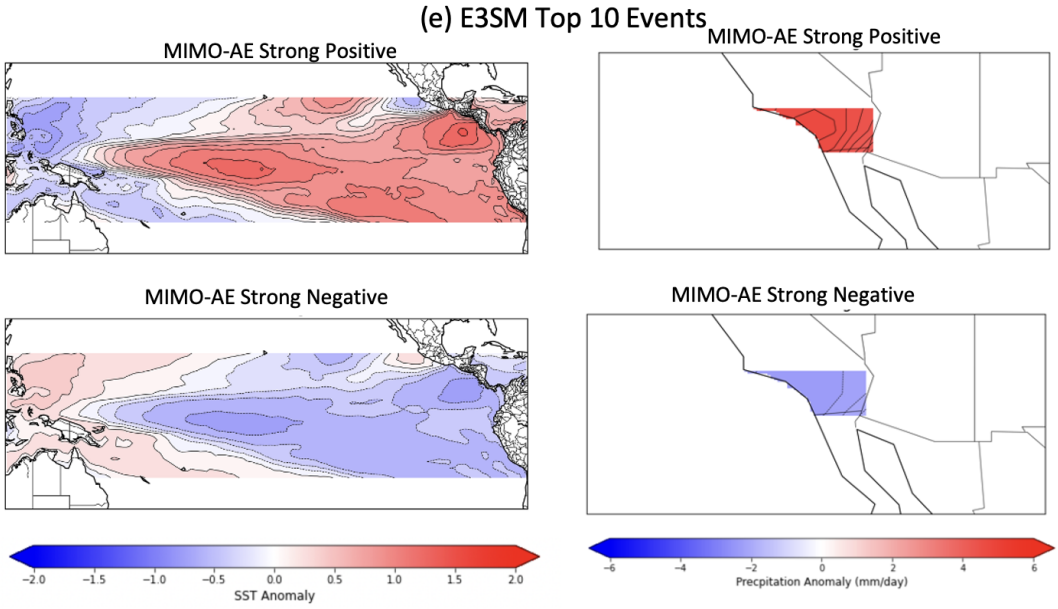


Figure 3.8: Composite of reconstructions of TP-SST and SC-PRECIP for the top 10 strongest positive and negative MIMO-AE events for E3SM testing data (e).

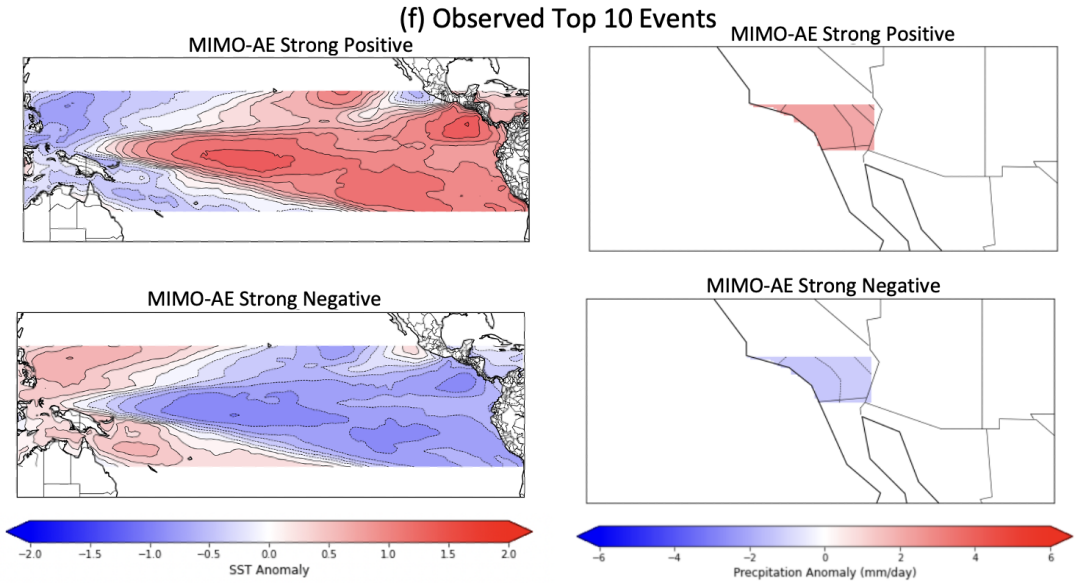


Figure 3.9: Composite of reconstructions of TP-SST and SC-PRECIP for the top 10 strongest positive and negative MIMO-AE events for observed testing data (f).

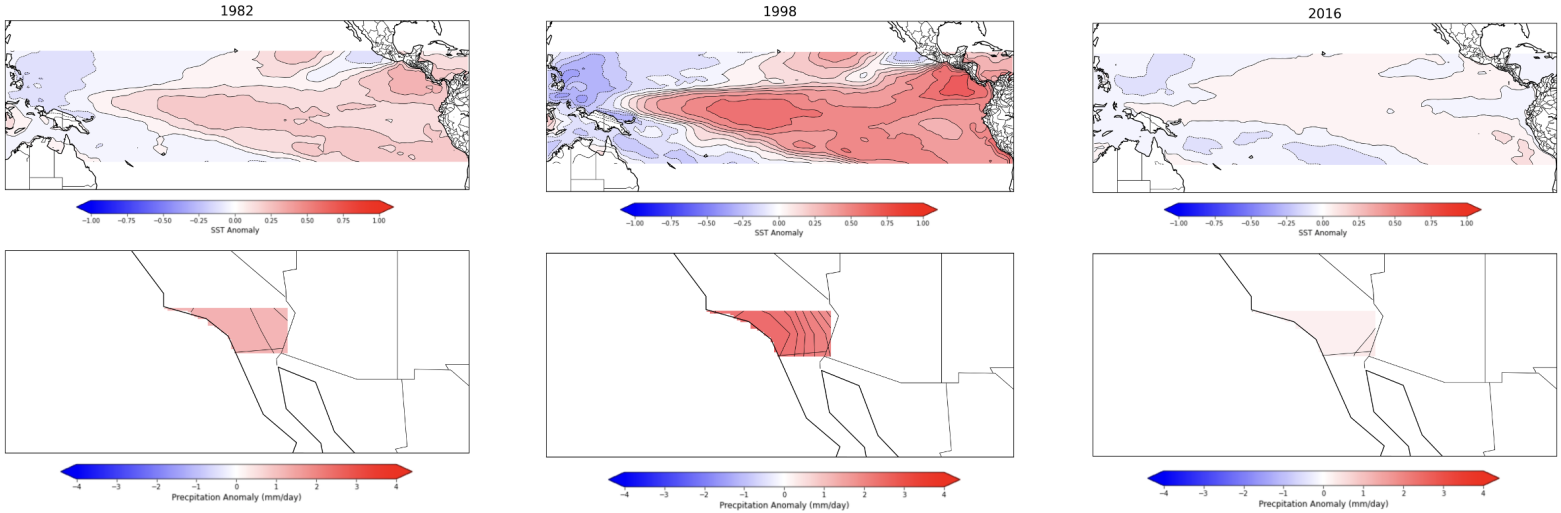


Figure 3.10: The December to February average reconstructions for the three strong El-Niño events of 1981-82, 1997-98, and 2015-16 in observations (g).

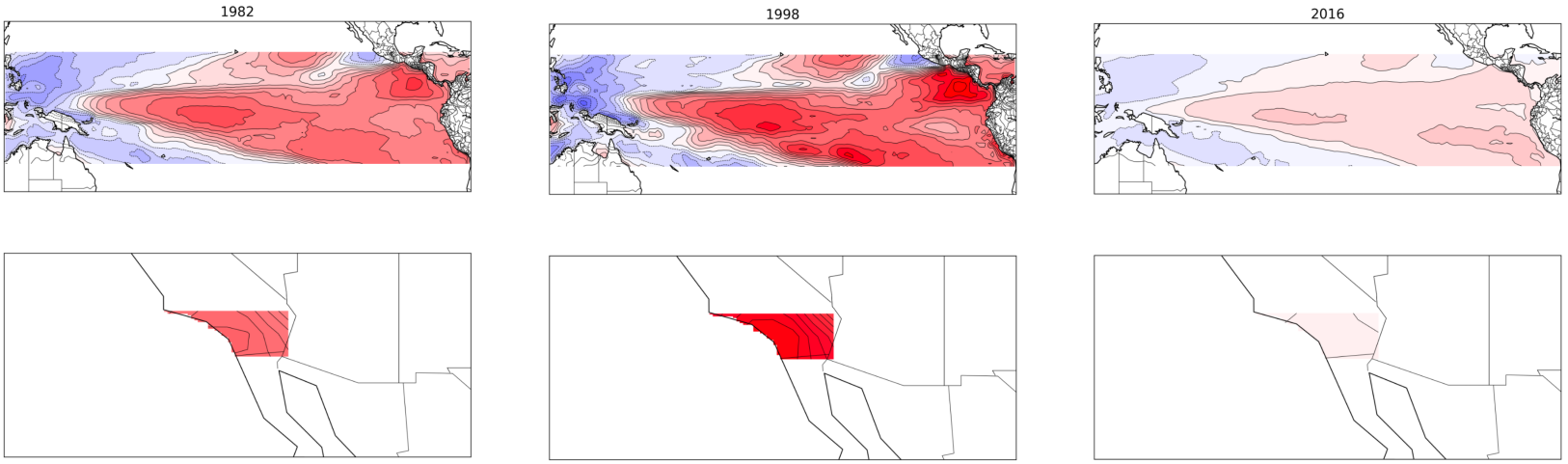


Figure 3.11: The December to February average reconstructions for the three strong El-Niño events of 1981-82, 1997-98, and 2015-16 in E3SM (g).

some events over the tropical Pacific triggering extreme positive anomalous events in SC-PRECIP.

Fig. 3.8e shows the composite of reconstructions of TP-SST and SC-PRECIP during the strongest 10 positive and negative 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 (figure 3.9g). Figure 3.10f shows the December to February average reconstructions for the three strongest El-Niño events (1981-82, 1997-98, 2015-16) in observations (and E3SM figure 3.11). It is apparent that the spatial patterns of the reconstructions are not standing, but rather vary for each of the years. 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 associated with stronger TP-SST anomalies. When a separate MIMO-AE network is trained on all of the observation data (1948-2020), 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 model

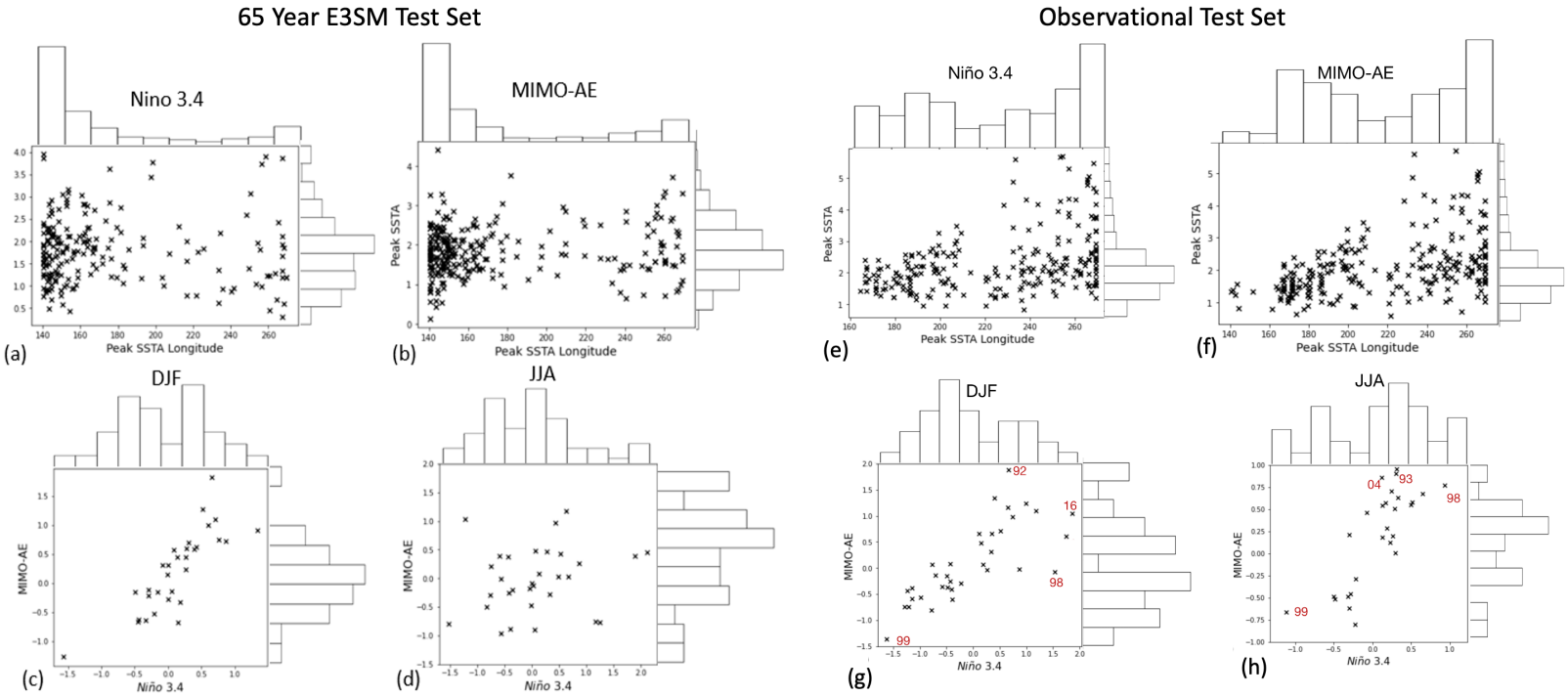


Figure 3.12: The peak SSTA along the equator and the corresponding longitude in which they occur for model and observational data respectively (ab,e,f) and compare the Nino 3.4 and MIMO-AE indicies averaged from December to February (DJF c,g) and June to August (JJA d,h) for each year in E3SM data and observed.

bias in E3SM-only trained network. Although, using some observational data along with E3SM data to train the network showed little impact on our predictability results here (next section), we plan to reduce the influence of model bias in our MIMO-AE by appropriately weighing the observational data used during training.

The scatter plots in figure 3.12(a,b,e,f) show the peak SSTA along the equator and the corresponding longitude in which they occur for model and observational data respectively. We compare the spatial distribution of ENSO events defined by both the Niño 3.4 index (figure 3.10a,i) and the MIMO-AE (figure 3.10b,j). For both the MIMO-AE and Niño 3.4, events in the western Pacific are generally weaker than those in the east. Similarly, the largest events always occur in the eastern Pacific for both indices. The MIMO-AE captures a greater spatial distribution of events over the Niño 3.4 as seen in the historgrams in figure 3.11(a,b,i,j). This is likely due to the spatial constraints used when calculating the Niño 3.4 index and suggests the MIMO-AE can effectively distinguish different flavors of ENSO events, such as the ENSO Modoki. This is seen in the high magnitude of events in the $160 - 200^{\circ}$ longitude region captured by the MIMO-AE compared to the Niño 3.4 in the observational data set (figure 3.10j).

The scatter plots in figure 3.12(c,d,g,h) compare the Nino 3.4 and MIMO-AE indicies averaged from December to February (DJF 3.12c,k) and June to August (JJA 3.12d,l) for each year. The MIMO-AE categorizes the 2015-2016 ENSO event in observational data as being a neutral ENSO event during boreal winter and a

stronger event in the summer, although not as strong as the 1997-1998 and 1982-1983 events where western US precipitation was far above average in observational data (figure 3.12k,i). The MIMO-AE also captures extreme precipitation events over Southern California, such as the years 1992-1993 and 2004-2005 which both saw roughly 200% above average rainfall in the area. These years are also associated with ENSO Modoki events which have been shown to have a positive influence on precipitation in the western US (137; 136). Additionally, there is a tendency for a higher MIMO-AE in boreal summer in both model and observations. A similar bias was found for the ELI in williams2018diversity. This tendency is associated with the seasonal cycle of cold-pool and warm-tongue phases of ENSO and is not captured in the Niño 3.4 index, but is effectively captured by the MIMO-AE and ELI (129).

Figure 3.13 shows the slope values of the regression of TP-SST anomalies and US precipitation anomalies on Niño 3.4 index as well as MIMO-AE for E3SM and observational data. We use the false discovery rate (FDR) approach (Renard et al.; 128; 115) to account for mutiple hypothesis being tested as once (in this case the number of pixel values M). We compute a constrained critical value, α_{FDR}, for a given global significance level, α, as follows:

$$\alpha_{FDR} = \max_{j=1,2,\ldots,M} \{p_j : p_j \leq \alpha(j/M)\} \tag{3.11}$$

where, p_j represents the p-value of the deviance statistic for grid-point j, and M is the total number of grid points. SC-PRECIP is more strongly related to MIMO-AE

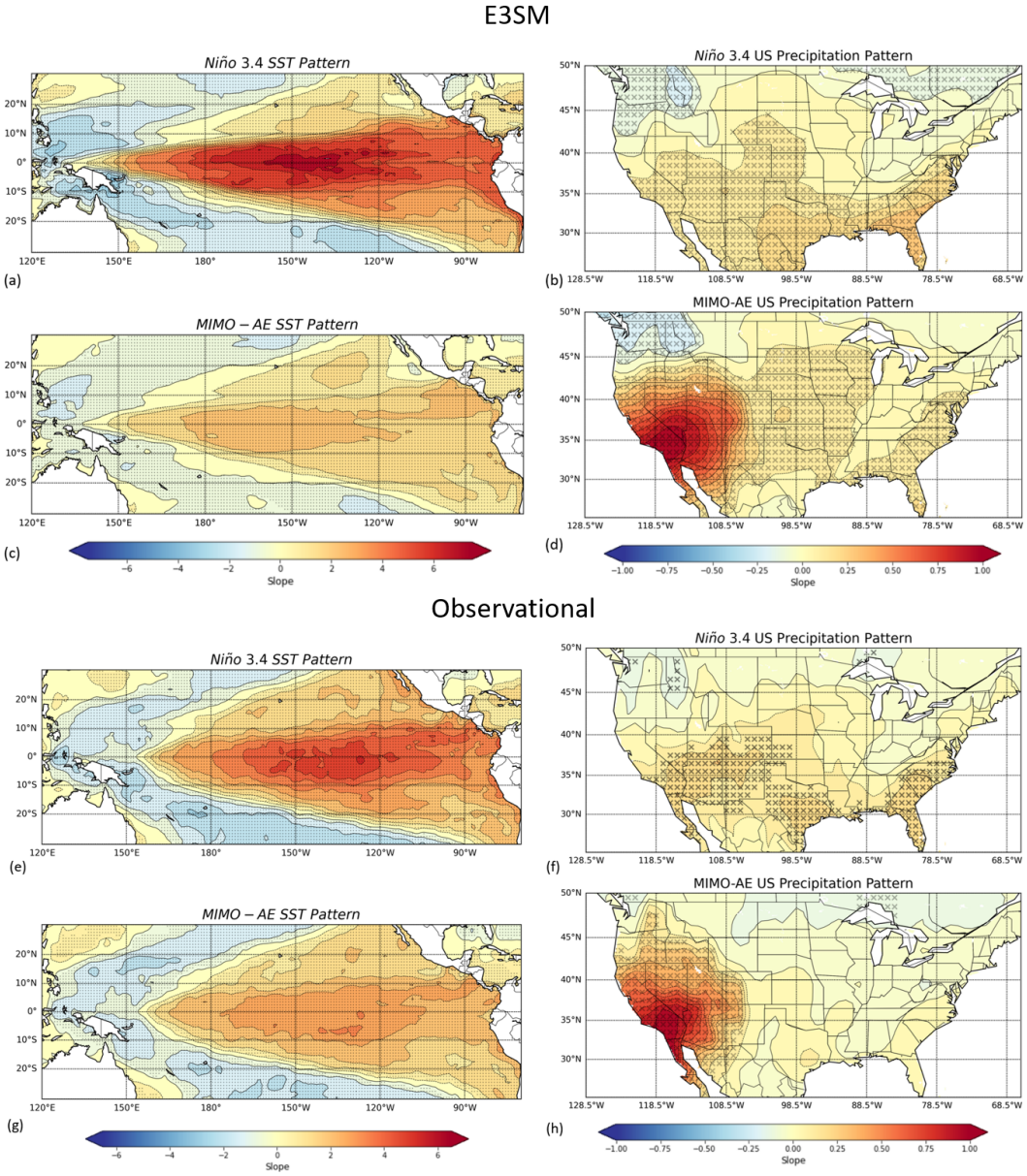


Figure 3.13: The slope values for Niño 3.4 and MIMO-AE regressed onto tropical pacific SST anomalies and US precipitation anomalies for the observational data from 1948-2020

than Niño 3.4, given that precipitation information is passed through the network. However, over the tropical Pacific, SSTs the MIMO-AE generates a weaker response than shown for Niño 3.4. The MIMO-AE not only picks up a strong precipitation response in the Western US, but also picks up the known weaker response in the Southeast in model data. The SST pattern is weaker for the MIMO-AE, however, this is expected given that the MIMO-AE is trying to maximize both the precipitation and SST patterns.

3.4.2 Predictability of Precipitation Using the MIMO-AE index

We use LSTM models to predict the MIMO-AE index for lead times of 1 to 12 months. Figure 3.14a 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 (77), offering little predictive skill.

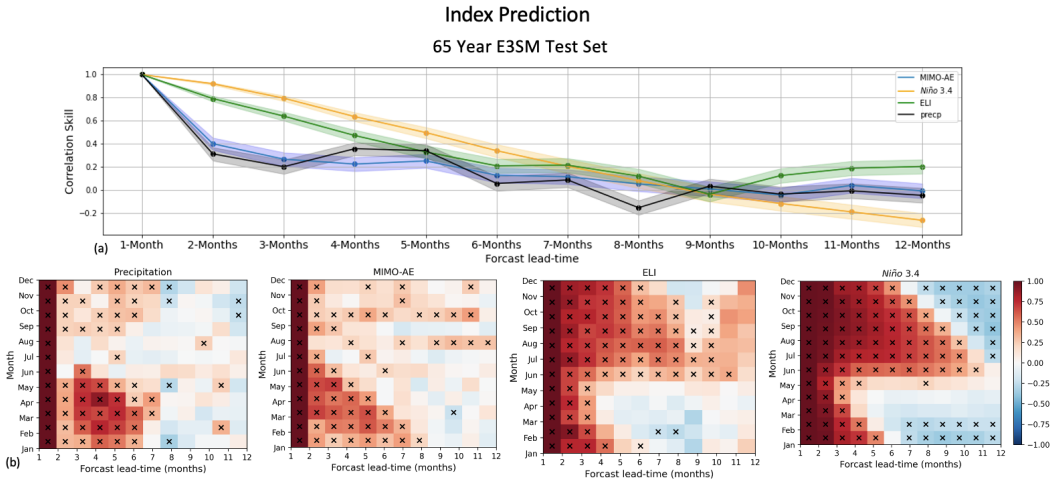


Figure 3.14: Predictive skill of MIMO-AE index (blue), Niño 3.4 index (orange) and ELI (green) and domain averaged SC-PRECIP (black) at forecast lead times of 1 to 12 months for E3SM testing data (a). Shading represents one standard deviation of the correlation coefficients. Predictive skill of domain 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 (b). Cross markings indicate values significant at the 95% confidence level.

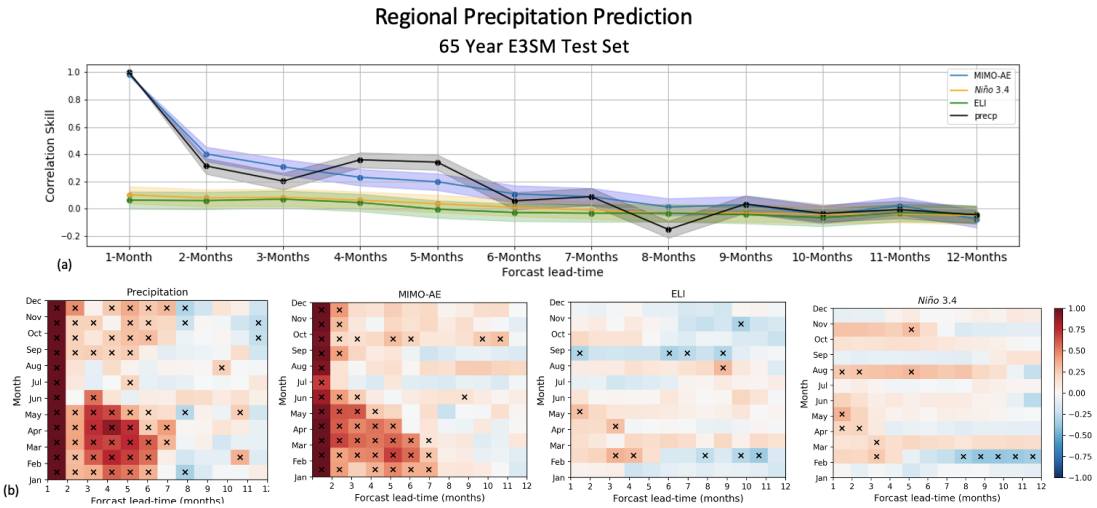


Figure 3.15: Predictability of SC-PRECIP using MIMO-AE. Predictive skill of MIMO-AE index (blue), Niño 3.4 index (orange) and ELI (green) at predicting domain averaged SC-PRECIP at forecast lead times of 1 to 12 months for E3SM testing data (a). Shading represents one standard deviation of the correlation 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 (b). Cross markings indicate values significant at the 95% confidence level.

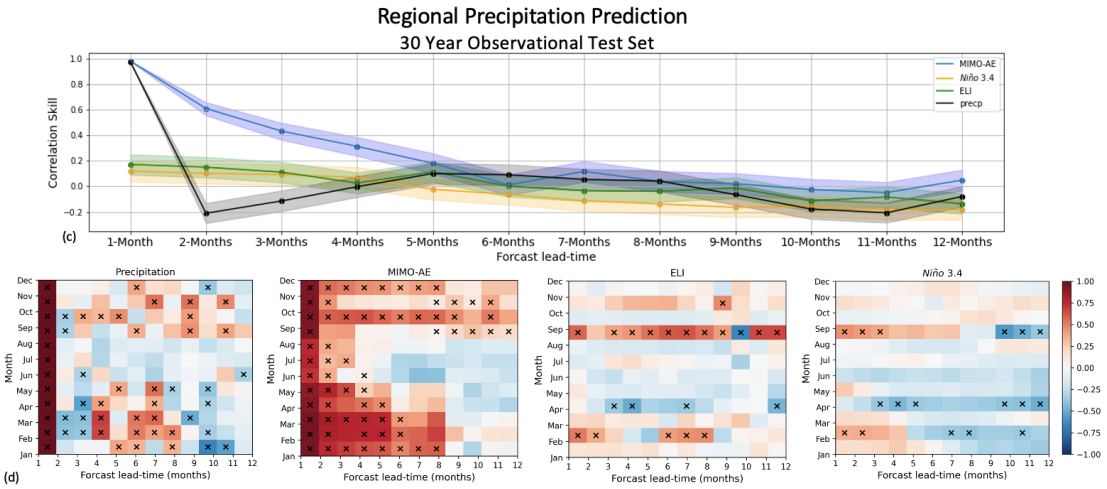


Figure 3.16: Predictability of SC-PRECIP using MIMO-AE. Predictive skill of MIMO-AE index (blue), Niño 3.4 index (orange) and ELI (green) at predicting domain averaged SC-PRECIP at forecast lead times of 1 to 12 months for observational testing data (c). Shading represents one standard deviation of the 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 observational testing data (d). Cross markings indicate values significant at the 95% confidence level.

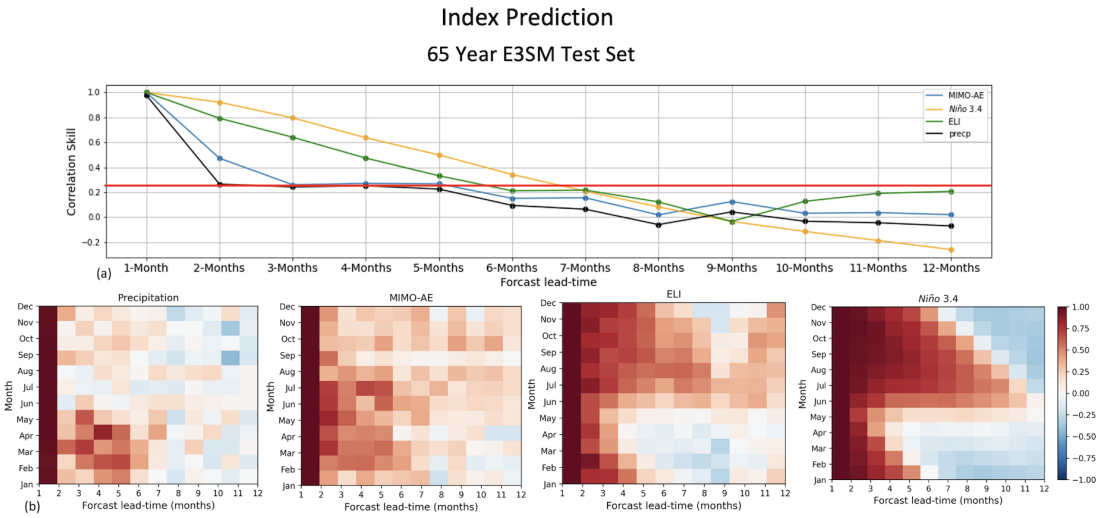


Figure 3.17: Predictive skill of E3SM-only trained MIMO-AE index (blue), Niño 3.4 index (orange) and ELI (green) and domain averaged SC-PRECIP (black) at forecast lead times of 1 to 12 months for E3SM testing data (a). Predictive skill of domain average SC-PRECIP as a function of initialization calendar month and forecast lead time from domain average SC-PRECIP, MIMO-AE index, ELI and Niño 3.4 index for E3SM testing data (b).

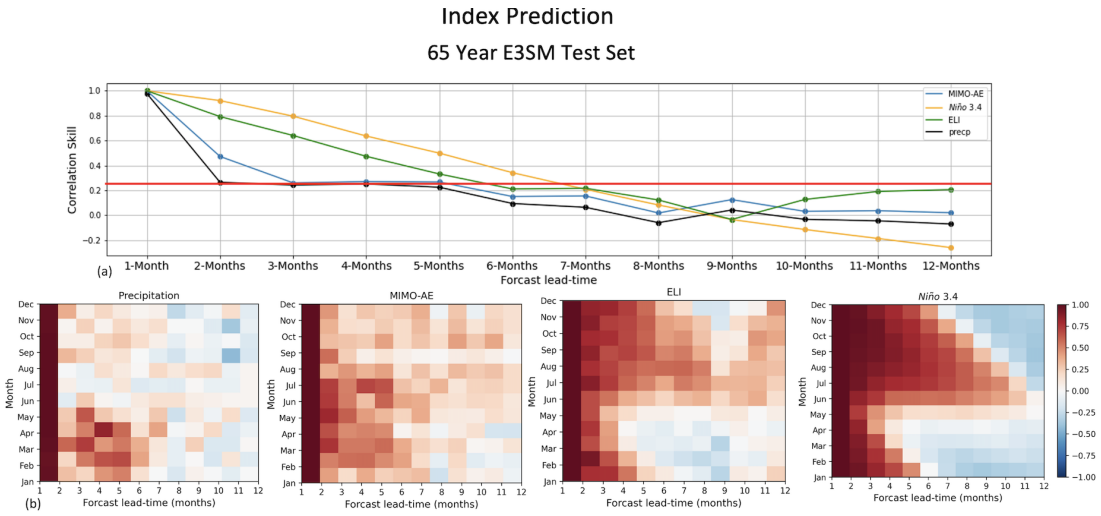


Figure 3.18: Predictive skill of E3SM-only trained MIMO-AE index (blue), Niño 3.4 index (orange) and ELI (green) and domain averaged SC-PRECIP (black) at forecast lead times of 1 to 12 months for observational testing data (c). Predictive skill of domain average SC-PRECIP as a function of initialization calendar month and forecast lead time from domain average SC-PRECIP, MIMO-AE index, ELI and Niño 3.4 index observational testing data (d).

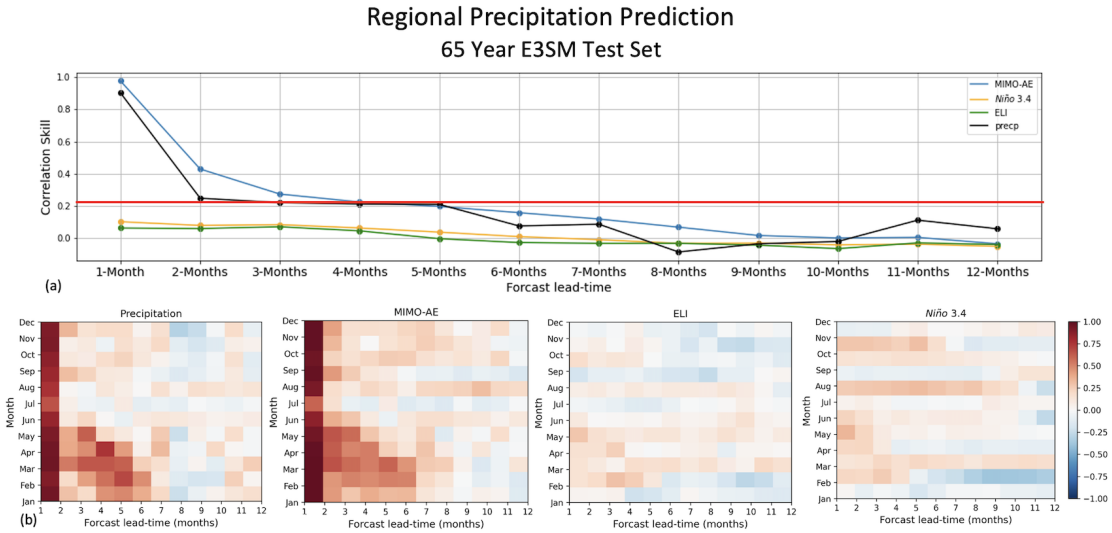


Figure 3.19: Predictability of SC-PRECIP using E3SM-only trained MIMO-AE. Predictive skill of MIMO-AE index (blue), Niño 3.4 index (orange) and ELI (green) at predicting domain averaged SC-PRECIP at forecast lead times of 1 to 12 months for E3SM testing data (a). Shading represents one standard deviation 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 (b). Cross markings indicate values significant at the 95% confidence level.

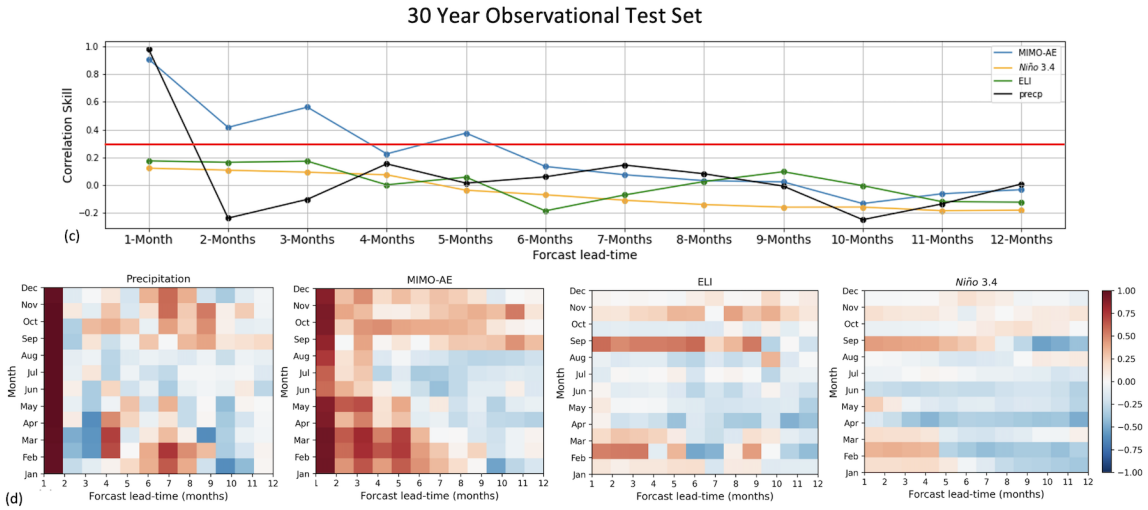


Figure 3.20: Predictability of SC-PRECIP using E3SM-only trained MIMO-AE. Predictive skill of MIMO-AE index (blue), Niño 3.4 index (orange) and ELI (green) at predicting domain averaged SC-PRECIP at forecast lead times of 1 to 12 months for observational testing data (c). Predictive skill of domain average 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 observational testing data (d).

This is evident in the Figure 3.14a and 3.15c, 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 3.14b shows the LSTM skills as a function of the calendar month when the prediction is initialized for all indices and generally reflect Fig. 3.14a, while also showing the well-known spring predictability barrier associated with Nino3.4 and ELI. MIMO-AE does not appear to suffer from the spring barrier that affects Niño 3.4 index and ELI - both of which demonstrate substantially weaker predictive skill in the Spring months as compared to other months.

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 statistically significant skill (critical value = 0.32 for 28 degrees of freedom at the 0.05 significance level for a one-tailed test) at a lead time of 2 months, while precipitation time series exhibits no skill at lead time longer than one month irrespective of the initial month of predictions (Figure 3.15d).

3.4.3 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 4a 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. 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 3.15b).

Enhanced predictive skill of MIMO-AE of SC-PRECIP is also noted for the 41 years of observation testing data (Figure 3.16c). 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 4d). 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 (75; 91; 122) 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 (52; 48; 119; 17).

3.5 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. MIMO-AE is trained on first 100 years of an E3SMv1 historical simulation (1850-2014) and applied on the latter 65 years. Another MIMO-AE network is trained on a combined dataset of 100-years of E3SM data and 32 years of observational data and applied to the latter 41 years of of the record (1980-2020). To assess the predictability of SC-PRECIP afforded by MIMO-AE on monthly time scales, we use LSTMs of the MIMO-AE index. LSTMs are trained separately for the E3SM simulation and observational data. 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 (75; 122; 119; 17) 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 (122; 17). We plan to explore techniques, like denoising 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.

Our MIMO-AE approach can be applied to assess the predictability of regional climate across the globe, both where linear correlations are known to exist - like the Southeast US, where ENSO influence is strong - and where the signal to noise ratio is low. Additionally, 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. Also, here we utilize two separate networks, MIMO-AE to extract the non-linear relationship between TP-SST and SC-PRECIP, and LSTMs to make predictions. We plan to condense these two into one combined MIMO-AE-LSTM network to account for spatial and temporal variability simultaneously to assess predictability in a more coherent manner.

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 (39). Combining such

networks with multi-task learning methods provides the potential to further enhance predictability of regional climate.

3.6 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.

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan(http://energy.gov/downloads/doe-public-access-plan).

3.6.1 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/ (28).

Observational SST data from the HadISST 1.1 dataset (96) can be downloaded from the web at https://www.metoffice.gov.uk/hadobs/hadisst/. Observed precipitation data from NOAA's PREC/L (14) can also be found open access at https://psl.noaa.gov/data/gridded/data.precl.html.

Chapter 4

Exploring the Use of a Multi-Input Conv-LSTM to predict Niño 3.4 index.

Aim II: Predictability Using Non-Linear Latent Spaces

4.1 Abstract

The advance prediction of El Niño Southern Oscillation (ENSO) is a challenging and important scientific problem due to its complex nature and wide ranging impacts on global short-term climate. Many prediction models fail to capture the spatiotemporal aspects of ENSO variability. Recent advances in ENSO prediction have show the usefulness of a deep learning approach called a convolutional long short term memory (conv-LSTM) for capturing both the spatial and temporal dependencies of ENSO. Here we expand on previous works using a conv-LSTM for ENSO prediction by including multiple input variables known to be associated with ENSO variability and predictability. We use this multi-input conv-LSTM to predict tropical Pacific sea surface temperature (SST) patterns and subsequent Niño 3.4 index in Earth systems model (ESM) and observational data.

Plain Language Summary

Predicting ENSO is a complex scientific problem made even more difficult by climate model uncertainty. Recent advances in machine learning have improved prediction including a technique called ConvLSTM. Here, we use the ConvLSTM approach to predict Niño 3.4 index with a combination of climate variables in ESM and observational data.

4.2 Introduction

El Niño-Southern Oscillation (ENSO) is a large-scale ocean-atmosphere interaction that occurs in the equatorial Pacific Ocean and global atmosphere (93). ENSO is one of the main drivers of inter-annual climate variability and as a result has a vast impact on global short term climate. El Niño (La Nina) is the ocean component of this phenomena that is characterized by a period increase (decrease) in sea surface temperature (SST) while the Southern Oscillation (SO) describes the bimodal variation in atmospheric pressure between the western and eastern Pacific. The Climate Prediction Center of NOAAs National Centers for Environmental Prediction defines an El Niño event as 5-month running SST means in the Niño 3.4 region (5° N - 5° S, 120° - 170° W) exceed 0.4°C for 6 months or more.

ENSO has been shown to influence a number of meteorological events ranging from Atlantic tropical cyclones and Pacific typhoons (100) to severe droughts in Southeast Asia and Southern Africa (25). Because of these wide reaching impacts, the problem of ENSO forecasting and prediction is one of great importance. However, the prediction and understanding of ENSO remains a challenge for scientists as there is no unified theory of how the inner oscillations mechanism of ENSO works (118). Furthermore, the relationship between global warming and ENSO is uncertain and is difficult to assess due to climate model uncertainty. zheng2016intermodel found that the change in ENSO SST amplitudes showed large uncertainty in ESMs while cai2014increasing found that ENSO response could double due to future climate

warming. To advance our understanding of the relationship between a warming climate and ENSO, improved GCM simulation and prediction of ENSO is required.

Predicting ENSO is complicated by two complex dependencies: spatial and temporal. This poses a challenge when building a model to efficiently predict ENSO variability. Three general approaches exist for prediction of ENSO: statistical-based methods, machine learning (ML) methods and a hybrid approach (133). Statistical methods generally have a hard time extracting nonlinear features and therefor fall short when predicting El Niño and La Nina events. Recently, ML techniques have developed to handle this issue. The most common ML methods for predicting ENSO including Support Vector Regression (90), Random Forests (RF) (90) and Long Short Term Memory (LSTM) (131). However, these methods fail to tackle the spatiotemporal dependencies mentioned above. Instead, they learn a time series on a certain latitude and longitude point to preform prediction which ignores the spatial component of ENSO prediction (42).

Several deep learning methods have been introduced to address this issue. A Temporal Convolutional Network (TCN) was proposed by yan2020temporal and was proven to work better than the LSTM approach for predicting ENSO at 1, 3, 6, and 12 month lead times. Similar to the TCN, both he2019dlenso and mu2019enso use a convolutional LSTM (ConvLSTM) to predict El Niño and La Nina events. The ConvLSTM was first proposed by xingjian2015convolutional as a method to improve the LSTM by adding convolutional structures to both the input-to-state and state-to-state transitions. This allows for more accurate learning of

spatiotemporal data. he2019dlenso proved the effectiveness of using this method to predict ENSO by creating a sequence to sequence model whose encoder and decoder are both ConvLSTM. Their model was shown to be more efficient than traditional LSTM and deterministic models when using reanalysis climate data. mu2019enso uses a ConvLSTM with added rolling mechanism (ConvLSTM-RM) which improved prediction performance in a real-world dataset over multiple other prediction methods including SVR, feed forward neural networks (FNN), convolutional neural networks (CNN), LSTM and auto-regressive moving average (ARIMA) models. Both of these papers prove the ability of the Conv-LSTM to capture the spatial and temporal components of ENSO as well as learn the non linear features over traditional statistical and common ML methods.

In this study, we expand on sa simple Conv-LSTM method to include multiple input variables to predict Niño 3.4 in the latest version of the Energy Exascale Earth Systems Model (E3SM v1) and observational data.

4.3 Methods

4.3.1 Model and observational data

For this study we use monthly E3SM v1 (28) data taken from the Earth System Grid Federation (ESGF) (20) at 0.25 degrees. The E3SM v1 was shown to effectively capture temporal variability of ENSO and outperformed other ESMs when reproducing ENSO spacial patterns when compared to observational datasets (36).

We chose four variables over the Niño 3.4 region to train our ConLSTM-FC in addition to sea surface temperature anomalies (SSTs) including sea level pressure (PSL), surface zonal wind component (U) and outgoing longwave radiation (OLR). These variables have all been related to ENSO variability. We compare different combinations of these variables at different forecasting lead times to determine the best model for ENSO prediction in the E3SM. We take the area average of these variables before training and running the models.

4.3.2 Convolutional Long-Short Term Memory with Fully Connected Layer (ConvLSTM-FC)

ConvLSTM builds on the original LSTM model, introduced by hochreiter1997long, in order to solve 3D prediction problems. LSTM 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 (45). 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 has been shown to perform well for many temporal learning tasks, however it does not solve spatial aspects of a problem. To resolve this issue, xingjian2015convolutional introduced a convolution structure to the LSTM. The cell inputs, outputs and hidden states are all composed of 3D tensors where the last two dimensions are spatial rows and columns. The ConvLSTM

predicts the future state of a point in the grid by using the inputs and past states of its neighboring points. For more information on the structure of the ConvLSTM see xingjian2015convolutional.

In this study, we make use of the ConvLSTM with an added FC layer to predict Niño 3.4 index at 6 9 and 12 month lead times. By adding the FC layer after the ConvLSTM, we can better predict the index directly. Figure 1 shows the architecture of our ConvLSTM-FC model. The input consists of a time series of variables from $t = 1$ to k, where k is the number of lead months. The inputs are then passed to the ConvLSTM layer. The output from the ConvLSTM layer is then flattened and passed through a fully connected layer where the output is the next k number of Niño 3.4 indices predicted.

We test a combination of different input variables and lead times to determine which model works best for predicting ENSO in the E3SM. The combinations are as follows:

- ConvLSTM-FC-1var: SST at 6, 9 and 12 lead times
- ConvLSTM-FC-2var: SST and PSL at 6, 9 and 12 lead times
- ConvLSTM-FC-3var: SST, PSL and U at 6, 9 and 12 lead times
- ConvLSTM-FC-4var: SST, PSL, U and ORL at 6, 9 and 12 lead times

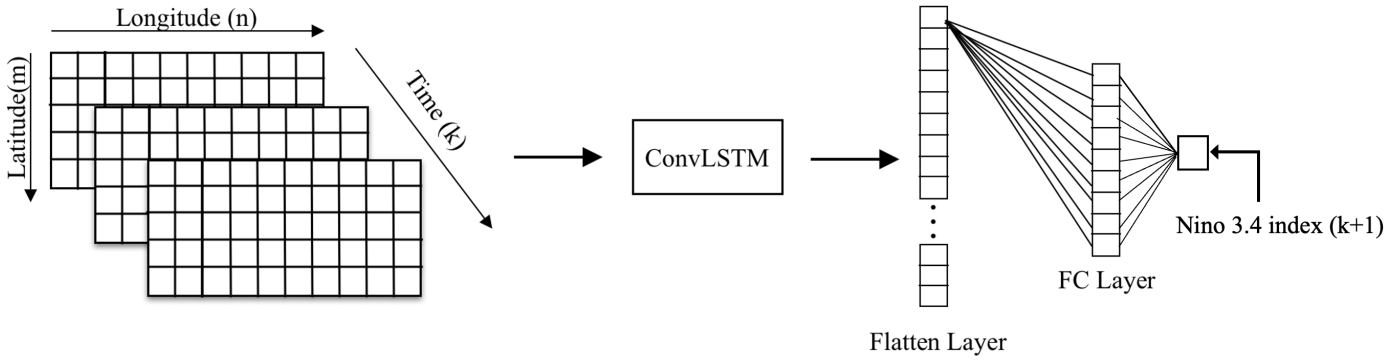


Figure 4.1: The general architecture for the multi-input conv-LSTM to predict future SST patterns.

We use 27 years of monthly E3SM data for training each model configuration and 3 years for testing the fully trained models. We also initialize lead times over all months to improve the skill of forecasting and to reduce the effect of the Spring prediction barrier. Each model configuration was trained 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. AdaGrad, Adaptive Gradient, proposed by duchi2011adaptive is a modified stochastic gradient decent method that increases or decreases its learning rate based on parameters of the data being trained.

4.3.3 Evaluation Metrics

Mean Squared Error (MSE) and Pearson Correlation Coefficient (PCC) were used to evaluate the performance of each trained model on the held out test set. A smaller MSE and larger PCC indicate better performance of the model. These metrics are defined as follows:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_{obs_i} - y_{pred_i})^2 \tag{4.1}$$

$$PCC = \frac{\sum_{i=1}^{N}(y_{obs_i} - \overline{y}_{obs_i}) \cdot (y_{pred_i} - \bar{y}_{pred_i})}{\sqrt{\sum_{i=1}^{N}(y_{obs_i} - \overline{y}_{obs_i})^2} \cdot \sqrt{\sum_{i=1}^{N}(y_{pred_i} - \bar{y}_{pred_i})^2}} \tag{4.2}$$

where y_{obs_i} is the observed Niño 3.4 index and y_{pred_i} is the predicted index.

4.4 Experimental Results

4.4.1 Comparison of Different Models to Predict Niño 3.4

We first compare the ability of the ConvLSTM-FC to predict the Nino 3.4 index in the E3SM against other commonly used machine learning time series regression models including FNN, CNN, and LSTM. The models were trained only using SST in the Niño 3.4 region as an input variable for 27 years of data. SST variability in the tropical Pacific Ocean is generally caused by ENSO events, so we use this simple input for testing these different methods (51). The models are all trained over 100 epochs, with an AdaGrad loss optimizer using tensorflow.

Table 1 shows the MSE and PCC for each model on the unseen 3 year test set for 6, 9 and 12 month lead times. From table 1 we see that (1) the ConvLSTM-FC model outperforms all other models on both evaluation metrics across the three different lead times; (2) the LSTM significantly under performs compared to the other models which is in line with previous findings that LSTM fails to capture spatiotemporal details and also confirms the crucial need for a model to effectively capture both of those aspects in ENSO prediction; (3) all of the models show a consistent decrease in skill with an increase in lead times with the highest skill being the 6 month lead time and the lowest with the 12 month.

Table 4.1: Performance Comparison of Different Machine Learning Models

Lead Time	Metric	FNN	CNN	LSTM	ConvLSTM-FC
6-month	MSE	0.5415	0.5954	1.978	**0.4343**
	PCC	0.8577	0.8389	0.2794	**0.8858**
9-month	MSE	0.7589	0.6693	1.991	**0.5853**
	PCC	0.8107	0.8090	0.0938	**0.8672**
12-month	MSE	0.9115	0.8547	2.002	**0.7378**
	PCC	0.7245	0.7339	0.0192	**0.7760**

4.4.2 Exploration of Different ConvLSTM-FC Architecture

We further investigate the use of the ConvLSTM-FC to capture and predict the Niño 3.4 index in our E3SM data by testing 4 different combinations of input variables. We build upon the simple SST model by adding variables also known to be influence by ENSO variability namely PSL, U and ORL. The combinations are again defined as follows:

- ConvLSTM-FC-1var: SST at 6, 9 and 12 lead times
- ConvLSTM-FC-2var: SST and PSL at 6, 9 and 12 lead times
- ConvLSTM-FC-3var: SST, PSL and U at 6, 9 and 12 lead times
- ConvLSTM-FC-4var: SST, PSL, U and ORL at 6, 9 and 12 lead times

We compare prediction skill of the Niño 3.4 index for these combinations on the unseen test set after training. From figure 2 and 3, (1) the skill of index prediction generally decreased with advance prediction time; (2) for a 6 month lead time, the 2 variable structure outperformed the other configurations with an MSE of 0.4194 and a PCC of 0.9044; (3) the 3 variable structure for the 9 month lead time had the highest prediction skill with an MSE of 0.4247 and PCC of 0.9051; (4) both the 1 variable and 3 varible structure performed well for the 12 month lead predictions with the 1 variable structure being slightly more skillful with an MSE of 0.7378 and PCC of 0.7760; (5) the lowest PCC value for all models was still above 0.5 which strongly

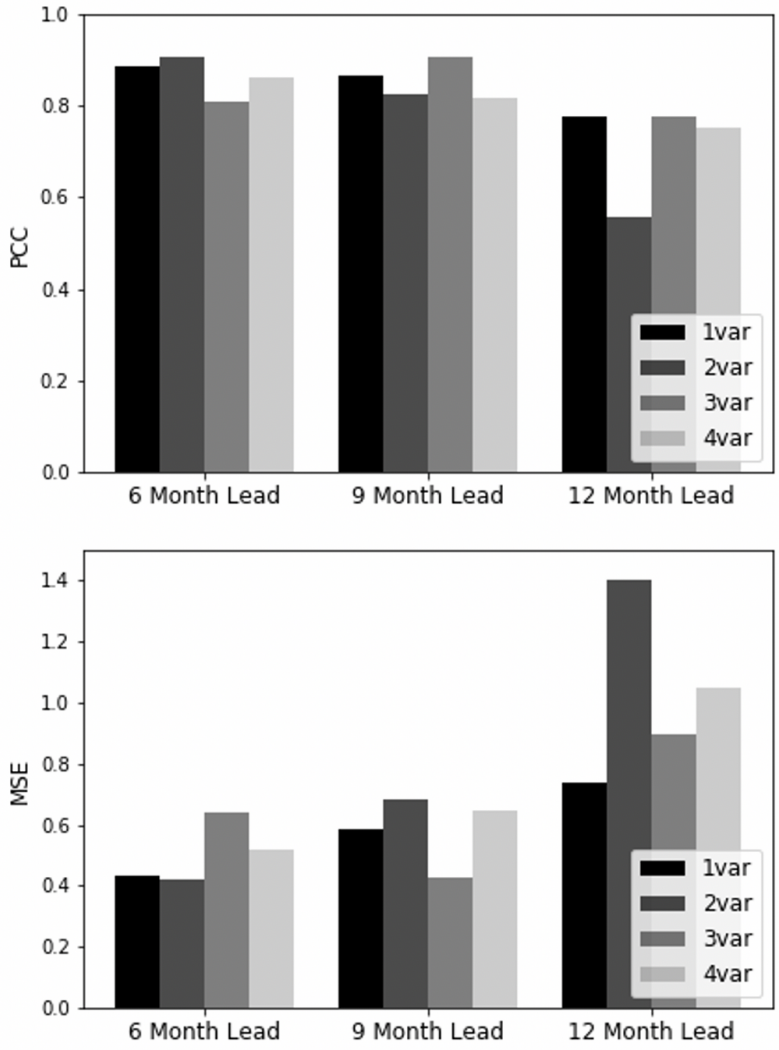


Figure 4.2: Performance comparison of different variable combinations in the ConvLSTM-FC.

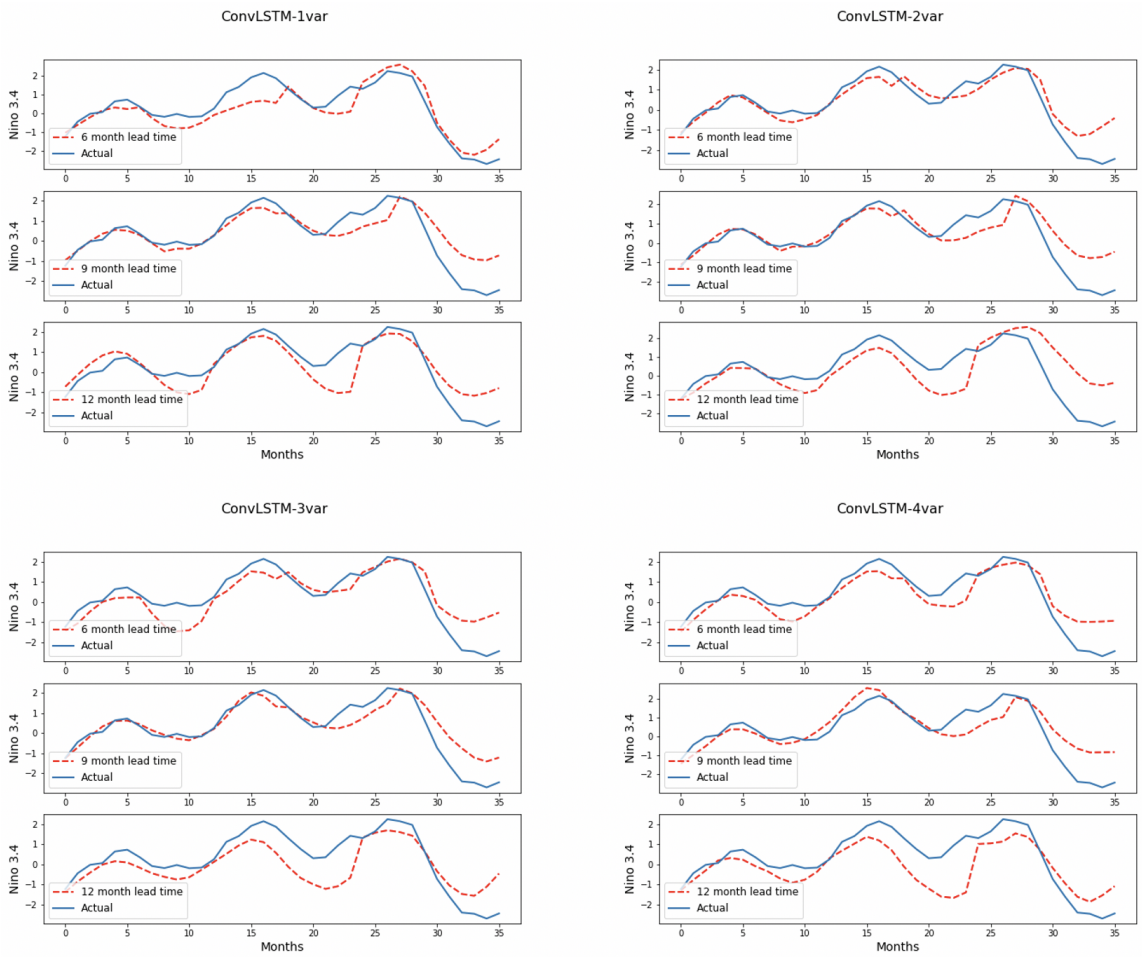


Figure 4.3: ConvLSTM-FC architecture used in this study.

suggests an overall effectiveness of using a ConvLSTM approach to predicting Niño 3.4 index in the E3SM.

4.5 Conclusions

In this paper, we present a ConvLSTM-FC approach to predicting Niño 3.4 indices in a GCM, namely the E3SM. We found this method outperforms other commonly used ML time series regression models. This is the first time a ConvLSTM method has been used to forecast ENSO variability in a GCM to our knowledge. We also explore the optimal input variables for Niño 3.4 index prediction using a ConvLSTM-FC at three different lead times.

The ConvLSTM model has been proved as an effective way of capturing the spatial and temporal details of ENSO in a GCM. Future workwill investigate the impact of model resolution on the ConvLSTM-FC's ability to accurately capture ENSO events.

4.6 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.

4.6.1 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/ (28).

Chapter 5

Simulation of Full Radiation Transfer Calculations in the E3SM Using a Deep Neural Network

Aim III: Emulation of Complex Parameterization

5.1 Abstract

This study develops a Deep Neural Network (DNN) to replace the full radiation model in the Department of Energys (DOE) Energy Exascale Earth System Model (E3SM). This study will build upon research that demonstrated the ability of a DNN to imitate just the shortwave and longwave radiative transfer calculations in a super-parameterized version of E3SM with an accuracy of 90-95% while also proving to be qualitatively similar to the original parametrizations in year-long simulations (89). We have run the E3SM at an ultra-low resolution (800 km) for a year while saving the input and output variables of the radiation module for each grid box at each time step. We use this data to train a dense, fully-connected, feed-forward DNN. We validate the impact of using the DNN on the model climate in a year long run of a single column model.

5.2 Plain Language Summary

The calculation of radiative transfer equations in a global climate model are often the most computationally expensive due to their complex nature. Here, we learn these equations through the use of a deep neural network (DNN) to reduce this computational time while maintaining high accuracy.

5.3 Introduction

The representation of radiative forcing in atmospheric general circulation models are among the most computationally expensive physical properties due to their complex nature (80; 83). In a typical general circulation model (GCM), such as the National Center for Atmospheric Research (NCAR) Community Atmospheric Model (CAM), the computation of model physics for a moderate resolution accounts for 70% of computational expense. The radiative transfer calculations for CAM take up 60% of that time required to solve those parameterizations (61). Even for weather prediction models such as the GME, the German Weather Services operational global numerical weather prediction model, the radiation calculations took over 25% of total computational time (79). These radiative computations are critical for providing accurate simulations of Earth-atmosphere systems and analyzing climate change forcing. As such, this raises an important need for reducing computational time while retaining high accuracy.

While recent developments in computing capabilities have generally improved the process of solving these radiative transfer calculations, they still remain incredibly time consuming. Direct radiative models that compute the radiative fluxes at each grid point for every time step are generally used for reference only as they are too computationally expensive for full GCMs (82). Several techniques have been used to replace this line-by-line model in order to reduce computational expense including the correlated- k method (63) and reducing the spacial and temporal resolution

(84). More recently, machine learning (ML) based methods have been explored for applications in parameterization including shallow neural networks (18; 61), random forests (RF) (138) generative adversarial networks (GAN) (31), and deep learning (95; 89; 40; 32).

Many papers exploring the use of ML techniques for parameterization only evaluate the method's offline ability, meaning the ability of the model to predict variables on a pre-computed dataset. Offline training generally fails to account for feedback loops between input and output variables and therefor accuracy measured in offline testing does not directly correlate with online accuracy, in this case using the trained model to replace parameterizations in the GCM itself. This often leads to online runs crashing within a few days of simulation due to numerical instability of the trained model (8). This instability is commonly caused by flawed choices of ML architecture and hyperparameters as well as inputs generated by the coupled GCM that are much different than any the ML model saw during training (9). Despite these obstacles, however, recent papers have proven that these ML methods can be used to accurately emulate parameterization schemes in online GCM simulations. Pal et al. (89) trained two deep neural networks (DNN) to emulate the long-wave (LW) and short wave (SW) radiation equations in a super parametrized version of the Department of Energys (DOE) Energy Exascale Earth System Model (E3SM) and successfully ran an online simulation with the trained DNNs for 1 year.

This study builds upon the work of Pal et al. (89), but here we use a novel approach of training a single DNN on the full radiation model in an ultra low-resolution run

of the E3SM version 1 (v1). Previous studies that have emulated GCM radiation schemes using NNs have used two separate networks to learn both the SW and LW equations (58; 32; 89). We investigate the ability of one DNN to approximate a portion of both the SW and LW radiation equations in both offline and online testing. We first evaluate the DNN's skill on a held-out testing dataset and then replace the partial radiation equations in the ultra-low resolution E3SMv1 with our DNN. We run the DNN surrogate model for 1 year on a single column model to explore the accuracy of our full radiation DNN.

5.4 Methods

5.4.1 E3SM

We train a DNN to emulate the full radiation transfer equations in the ultra-low resolution Enegry Exascale Earth Systems Model version 1 (E3SMv1) (28). The E3SM's Atmosphere Model (EAM), is based on the Community Atmosphere Model Version 5 (CAM5) with some additional modifications and improvements. Enhancements from the CAM5 include the EAM using a default spectral-element dynamical core, replaced parametrizations of turbulence, cloud microphysics and shallow convection, enhanced aerosol parameterizations and a vertical resolution increased to 72 levels.

5.4.2 Single Column Model

We run a prognostic simulation of our trained DNN in the E3SM on the single column model (SCM). The E3SM SCM is based off of the Community Atmosphere Model (CAM) and is useful for testing and evaluating the new parameterization and is widely used to identify weaknesses and biases in model simulation and can be found online at https://github.com/E3SM-Project/scmlib/wiki/E3SM-Single-Column-Model-Home (6).

5.4.3 Data Collection and Preparation

We gather our data for training the DNN by running the E3SM V1 at an ultra-low resolution for 1 year while saving the input and output variables of a subset of the original radiative calculations at each grid space for every time step. From this configuration, around 3 million samples were generated for the DNN to train on. Table 3.1 provides a list of the input and output variables used in this study. We select a small sample of variables to investigate the ability to learn both LW and SW equations in one DNN, 1,234 inputs and 154 outputs. The full radiation model includes the computation of cloud optics, conversion of aerosol masses to bulk optical

Table 5.1: The input and output variables used for training the DNN.

Input Variable	Vertical Levels	Output Parameter	Vertical Levels
0.2-0.7 micro-meter surface albedo: direct radiation	1	Surface Net LW FLux	1
0.7-5.0 micro-meter surface albedo: direct radiation	1	Outgoing Net LW FLux	1
0.2-0.7 micro-meter surface albedo: diffuse radiation	1	Downwelling LW FLux at Surface	1
0.7-5.0 micro-meter surface albedo: diffuse radiation	1	Downwelling SW FLux at Surface	1
Fractional Cloud Cover	72	Surface Absorbed Solar Flux	1
Cloud Fraction of Snow Clouds	72	Surface Net SW FLux	1
Convective Cloud Fraction	72	Outgoing Net SW FLux	1
Ice Fraction	1	LW Heating Rate	72
Land Fraction	1	SW Heating Rate	72
Land Fraction Ramp	1	Direct Solar Radiation on Surface (< 0.7)	1
Snow Depth	1	Direct Solar Radiation on Surface ($>= 0.7$)	1
Upward Longwave Flux	1	Diffuse Solar Radiation on Surface (< 0.7)	1
Pressure	72	Diffuse Solar Radiation on Surface ($>= 0.7$)	1
Temperature	72		
Ice Radiative Effective Diameter in Clouds	72		
In-cloud Cloud Ice Water Path	72		
In-cloud Cloud Liquid Water Path	72		
Upward LW Radiation	1		
Prognosed Value of Mu for Clouds	72		
Wet Diameter for All Modes	72 (4 modes)		
Aerosol Water (g/g) for All Modes	72 (4 modes)		

properties for SW and LW radiation, and SW and LW radiation calculation (89). By selecting a subset of these, we are able to test the ability of a DNN to emulated some LW and SW calculations before moving on to the more computationally heavy calculations of the full radiation model.

After obtaining the original E3SMv1 data for 1 year, we spilt it into three sets:

- Training set: A random 80% subset of the data used to train the DNN
- Testing set: The remaining random 20% subset of the data held-out for testing the accuracy of the trained DNN
- Validation set: A random 10% split of the training set used for validation during each epoch

The DNN uses a sigmoid method for mapping variables into a desired range. This functions allows for back propagation to be useful and takes any range real number and returns an output value between 0 and 1. We also use a min-max scaler to standardize the radiation variables before training using the following equations:

$$\hat{X_j} = \frac{X_j - \min_i(X_j)}{\max_i(X_j) - \min_i(X_j)} \tag{5.1}$$

$$\hat{Y_j} = \frac{Y_j - \min_i(Y_j)}{\max_i(Y_j) - \min_i(Y_j)} \tag{5.2}$$

where X_js are the input variables, Y_js are output variables, i is the training sample number and j is the specific variable. For the ultra-low resolution E3SM and after

reducing the number of input variables further, we end up with 1594 input radiation variables and 154 output radiation variables.

5.4.4 DNN Parameter Exploration, Training and Implementation

We construct a fully connected, feed-forward DNN to learn the radiation parameters. Our DNN is written using the Python library keras (https://keras.io/) and trained on one GPU node on the National Energy Research Scientific Computing Center's (NERSC) Cori super computer. We carried out parameter exploration on a subset of the training data to determine the optimal number of hidden layers, neurons and batch size. The loss is computed using mean squared error (MSE) between the original parameterization output variable Y_j and its corresponding DNN prediction output variable Y_{jDNN}. The equation used for MSE is as follows:

$$MSE = \frac{\sum_{i=1}^{N}(Y_j - Y_{jDNN})^2}{N} \tag{5.3}$$

where N is the total number of output samples. We use MSE over root mean squared error (RMSE) because variables were scaled prior to training and therefor RMSE would be redundant.

Figure 5.1a shows the results of several different combinations of hidden layers and neurons on training and validation loss across 100 epochs. We chose a 2 layer, 64 neuron DNN for this study as it minimized the loss in both training and validation

sets while not overfitting during validation. Figure 5.1b shows the results of batch size on the chosen DNN configuration. We chose 1024 samples for the batch size as it still performed well and cut down on computational time significantly.

5.5 Results

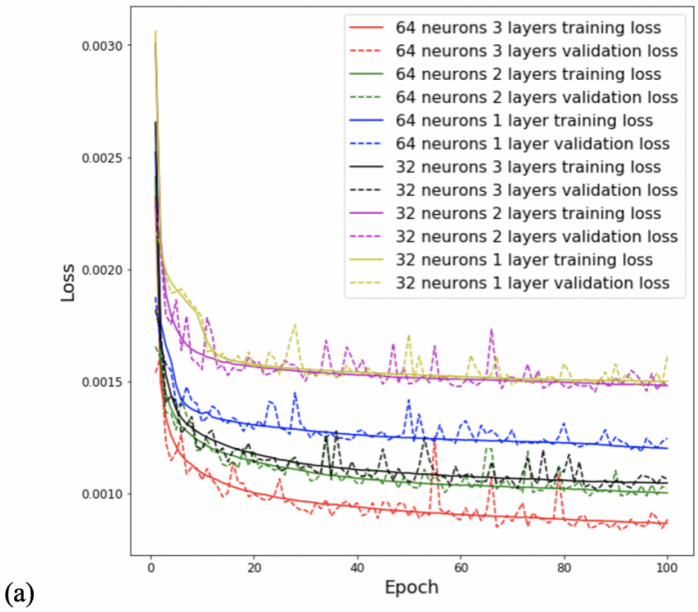
5.5.1 Offline Verification

We investigate the offline accuracy by evaluating the DNN's performance on the held-out testing dataset. We compute the L_1 and L_2 errors (figure 2) between the predicted and actual scaled output parameters. The L_1 and L_2 errors are computed as follows:

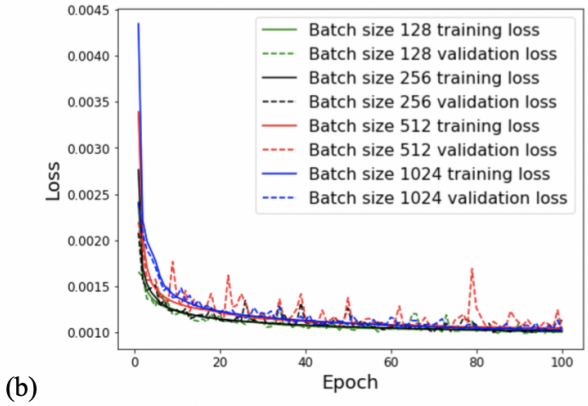
$$L_1 = \frac{\sum_{i,j} |G_{i,j} - I_{1,j}|}{\sum_{i,j} |I_{1,j}|} \tag{5.4}$$

$$L_2 = \frac{\sum_{i,j} (G_{i,j} - I_{1,j})^2}{\sum_{i,j} (I_{1,j})^2} \tag{5.5}$$

A majority of the L_1 and L_2 testing sample errors are bellow 0.1, or 10%, verifying the effectiveness of the trained DNN to successfully emulate the radiation equations in an offline, static dataset. We also visualize this effectiveness by plotting a random sample of scaled predicted and actual output parameters and their corresponding difference (figure 5.3). We can see from figure 5.3 that a majority of the difference between predicted and observed parameters occur in the short and long wave heating



(a)



(b)

Figure 5.1: Parameter exploration for (a) number of layers and neurons and (b) batch size.

rates (parameters 72 to 144). This suggests the DNN is not learning parameters at different levels in the atmosphere as effective as static variables.

We can also see this in figure 5.4 which plots the MSE for each parameter. Although there is a spike in MSE across variables 90 through 110 which correspond to longwave heating rates, the overall MSE for the entire dataset remains at 0.025. The model's difficulty in predicting longwave heating rates at mid-levels could be due to the ultra-low resolution and subsequent absence of clouds. However, the skill of the DNN on this held out testing dataset suggests the ability of using one DNN to emulate both SW and LW parameters. From figure 5.3 and 5.4 , we can also see that the model learns efficiently across parameters for both SW and LW variables without a bias towards one or the other. The DNN learned parameters skillfully across all global grid points across a calendar year.

5.6 Prognostic Simulation Full Global Model

Next, we replace the radiative transfer equations in the E3SM with our trained DNN and run the model for 30 years. We compare the DNN model (RADDNN) with a reference 30 year run of the ultra low resolution model (e3sm_low_res) for online analysis. The following equations show the calculation of output parameters (P_{out}) given the input parameters (I_{in}) and the trained DNN weights (W) and biases (b):

$$h_1 = f(I_{in}W_1 + b_1) \tag{5.6}$$

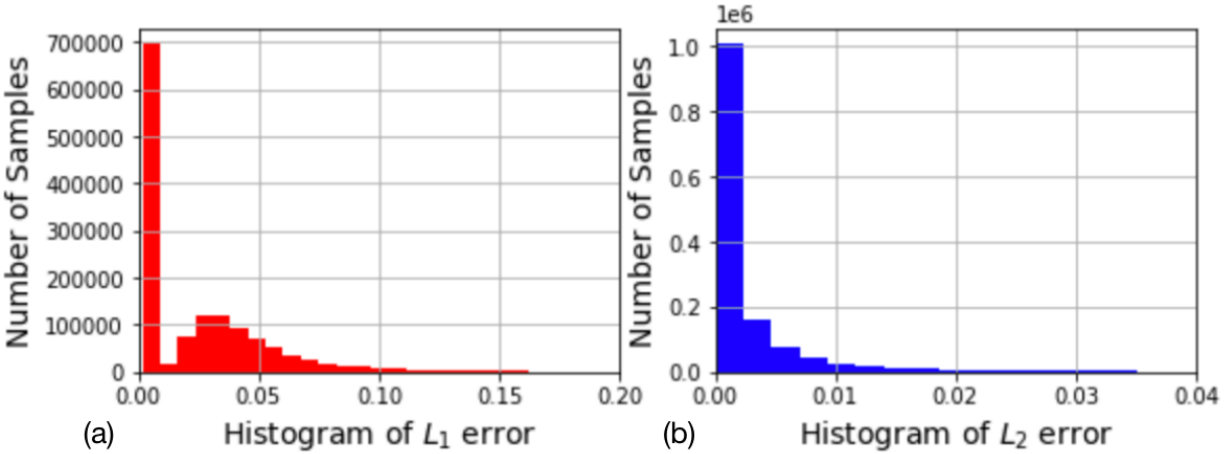


Figure 5.2: The mean squared error for each output variable from the trained DNN model.

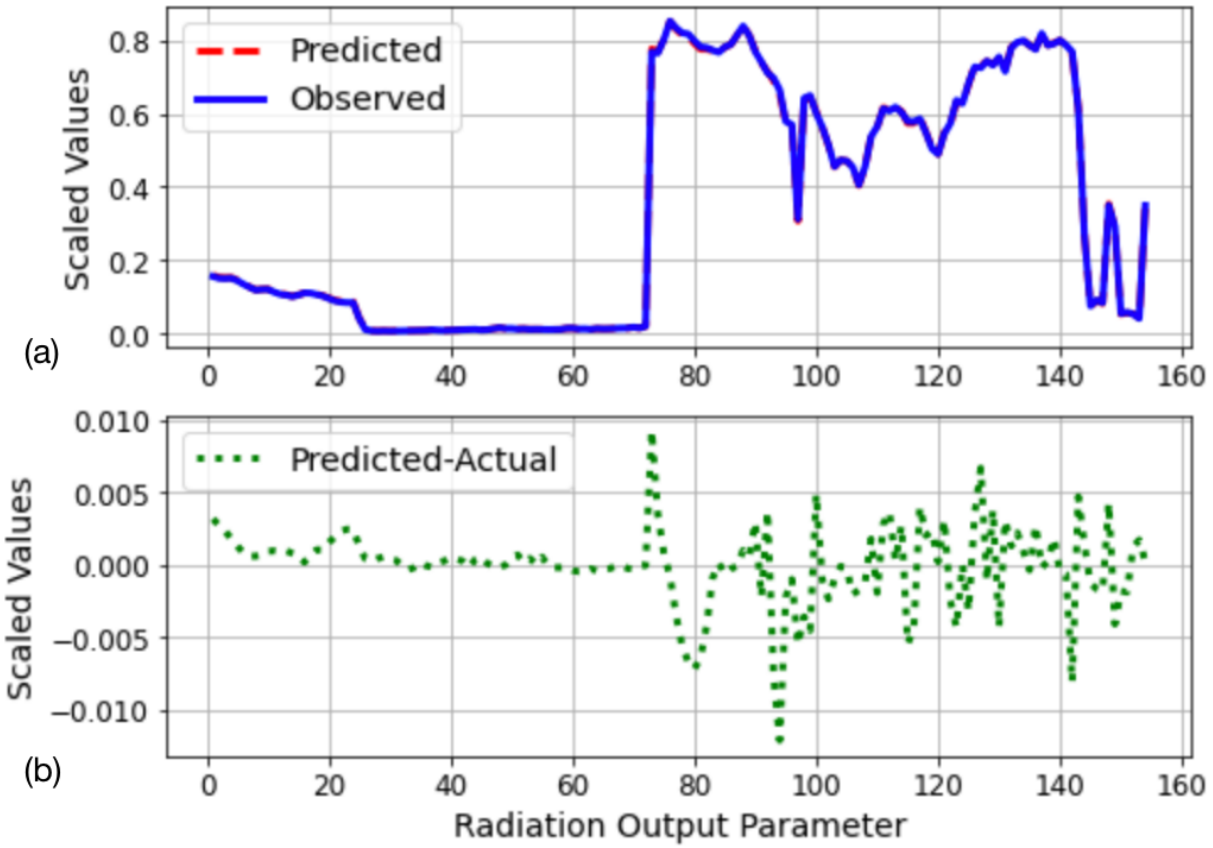


Figure 5.3: An example of predicted and actual output parameters (a) and the resulting difference (b).

$$h_2 = f(h_1 W_2 + b_2) \tag{5.7}$$

$$P_{out} = f(h_2 W_3 + b_3) \tag{5.8}$$

where $f()$ is the activation function, in this instance a sigmoid

From figures 5.4 and 5.5, we see that by the end of the first year the RADDNN model is already diverging from the reference dataset quickly. We investigate this variation by looking at the surface net LW flux (FLNS) and SW flux (FSNS) predicted in the first 9 months at each column number (ncol) compared to the original E3SM reference model. We can see that the DNN model starts to diverge around month 5 and 5 for both FLNS and FSNS. This highlights the issues addressed by brenowitz2020machine, that DNN models tend to become numerically unstable due to the nature of feedback loops in couple models providing inputs unlike those in the training dataset. Although our model did not become fully unstable and collapse, these issues did lead to unrealistic predictions of variables. This lead to our predicted output parameters of the DNN to converge into a single predicted number, such as FLNS in figure 5.4, or continuously get larger as seen in the prediction of FSNS in figure 5.5.

We also replace the full radiative transfer parameterization with the trained DNN and ran a simulation of the SCM for a year and found similar results of instability across the years run.

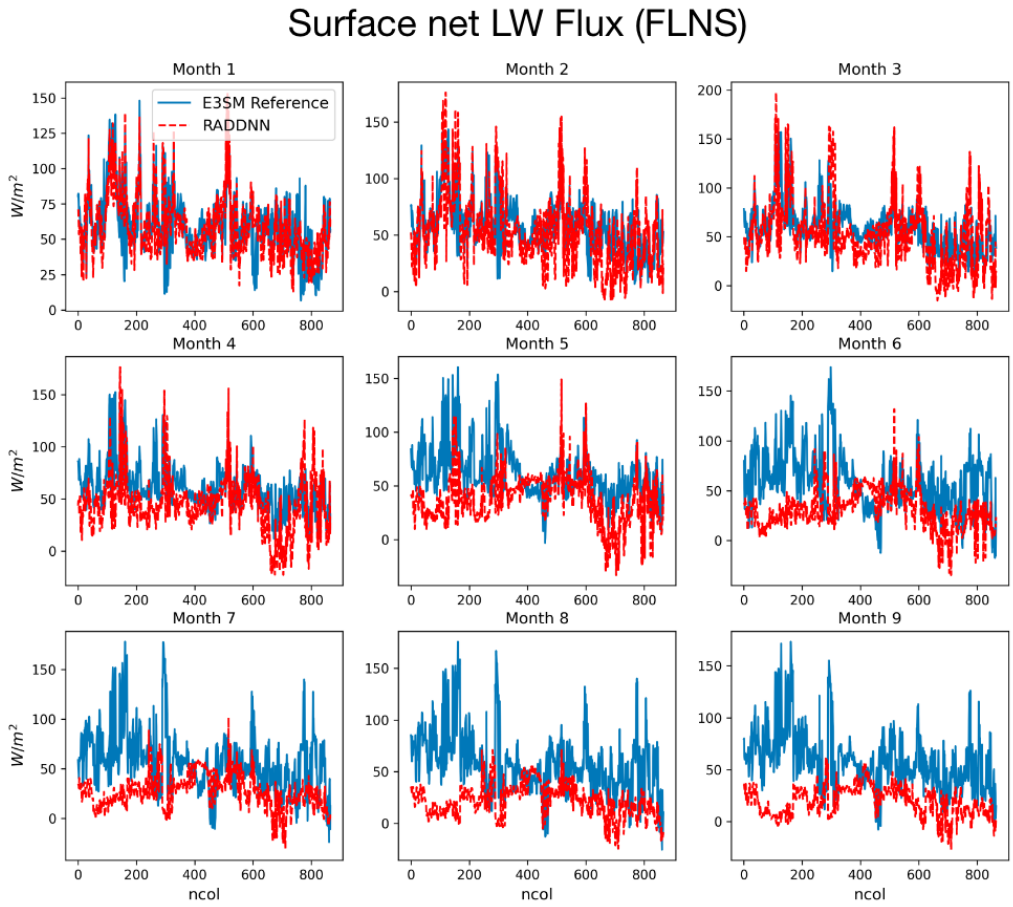


Figure 5.4: Predicted FLNS (red) and reference FLNS (blue) across every column (ncol) for the first 9 months of simulation.

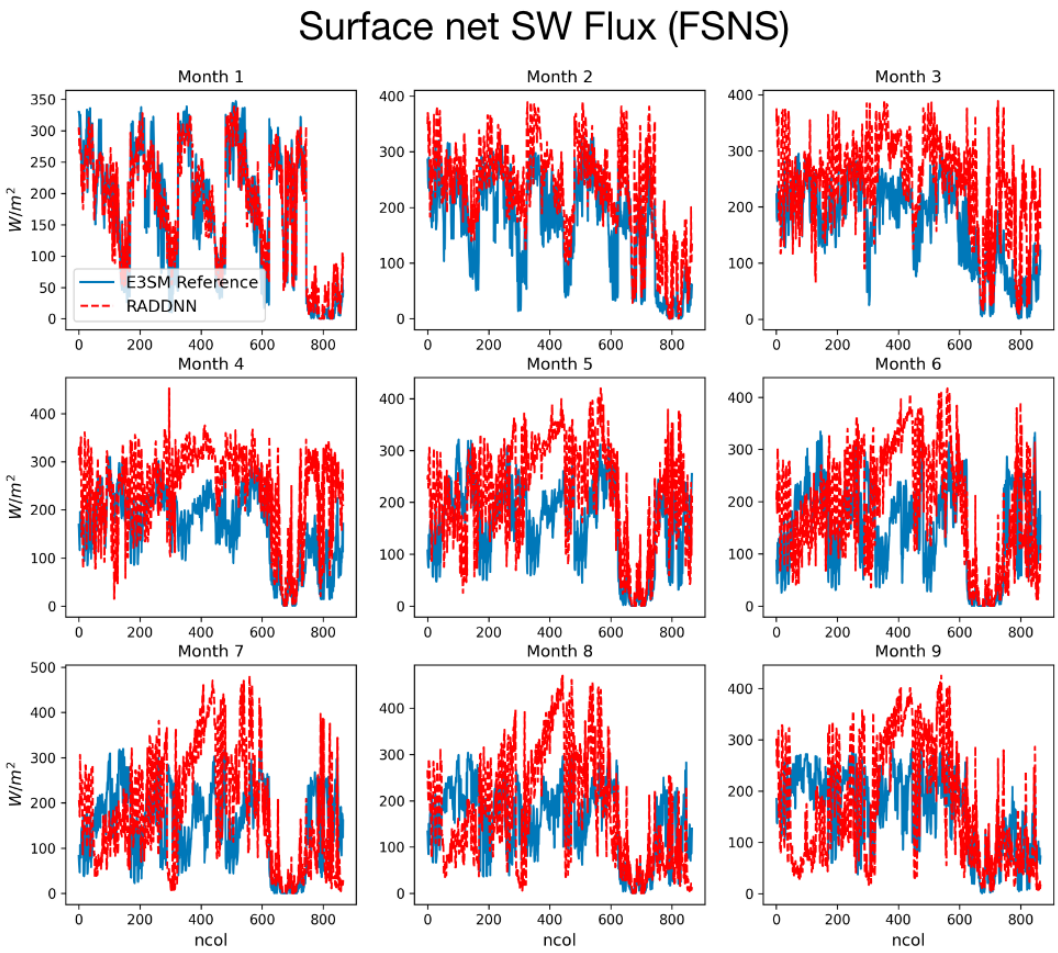


Figure 5.5: Predicted FSNS (red) and reference FSNS (blue) across every column (ncol) for the first 9 months of simulation.

5.7 Future Work

The ability of ML techniques to learn GCM parameterizations is continuously explored. Previous studies have proved the usefulness of NNs to learn LW and SW radiation parameterizations separately. Here, we show that DNNs are capable of learning a subset of the full radiation model in an offline ultra low resolution E3SM simulation. Offline simulations of the subset of the full radiation equations show high accuracy, $90 - 95\%$. By replacing the radiation equations with the trained DNN and running the model for 30 years, the E3SM ran 133% faster than the original model. While the RADDNN model still ran for 30 years without crashing, the model did suffer from unreasonable predictions of output parameters. This caused several issues to persist in the overall model, such as higher than expected temperatures at all pressure levels and extreme fluctuations in short wave and longwave heat fluxes.

Future work will look into methods to bias correct or account for the errors mention above. A general adversarial network (GAN) is a ML method that is commonly used to generate more data from a given sample (92). GANs may be useful for overcoming the lack of data available for training DNNs to emulate parameterizations and could create a more robust training dataset. This may improve the DNNs ability to predict out of sample inputs and lead to a more stable online simulation.

5.8 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.

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan(http://energy.gov/downloads/doe-public-access-plan).

5.8.1 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/ (28).

Chapter 6

Conclusion

This dissertation explored three areas of machine learning applicability to Earth Systems Models and demonstrated the usefulness of these methods in advancing our understanding of ESM modeling, analysis and predictability, namely statistical downscaling, predictability using non-linear latent spaces and emulation of complex parameterization.

In chapter 2 aim 1, I demonstrate the ability of a novel super-resolution based approach for downscaling ESM data that uses a modified version of the FSRCNN method. I find that this FSRCNN-ESM is able to map low resolution climate images to a four times higher resolution with a better skill than DeepSD, FSRCNN and bicubic interpolation; for surface radiative fluxes and large scale and convective precipitation; while remaining computationally inexpensive to train. The FSRCNN-ESM is also able to downscale images in a single-step process and without need for

access to GPUs for training. The FSRCNN-ESM as a result is a more approachable method of downscaling using machine learning.

Chapter 3 aim 2 part 1 establishes the ability of ML techniques to improve analysis and predictability of complex, non-linear latent spaces. ML can be used as an effective tool for isolating a well tropical Pacfic SST variability. In a novel approach, I 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. I 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. MIMO-AE is trained on first 100 years of an E3SMv1 historical simulation (1850-2014) and applied on the latter 65 years. Another MIMO-AE network is trained on a combined dataset of 100-years of E3SM data and 32 years of observational data and applied to the latter 41 years of of the record (1980-2020). To assess the predictability of SC-PRECIP afforded by MIMO-AE on monthly time scales, we use LSTMs of the MIMO-AE index. LSTMs are trained separately for the E3SM simulation and observational data. 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. The use of advanced ML techniques and multi-talk learning to enhance predictability of tropical Pacific variability and teleconnections is further

shown in chapter 4 aim 2 part 2. By adding additional variables to a conv-LSTM model, I can more efficiently predict the Niño 3.4 index in the E3SM.

Finally, chapter 5 aim 3 demonstrates the capability of DNNs for learning computationally expensive parameterizations in ESMs. I show the skill of prediction in offline verification for a single DNN emulation of the full radiation transfer equations in the E3SM. Future work will continue to improve the ability of online simulation and reduce the instability the DNN introduces to the model.

While there continues to be emerging challenges in the application of ML to Earth systems data, I demonstrate through this dissertation the vast ability of ML to advance the performance and understanding of ESMs using the latest ML algorithms.

Bibliography

[1] An, S.-I. and Jin, F.-F. (2004). Nonlinearity and asymmetry of enso. *Journal of Climate*, 17(12):2399–2412. 40

[2] Ashok, K., Behera, S. K., Rao, S. A., Weng, H., and Yamagata, T. (2007). El niño modoki and its possible teleconnection. *Journal of Geophysical Research: Oceans*, 112(C11). 4, 39

[3] Becker, E., van den Dool, H., and Zhang, Q. (2014). Predictability and forecast skill in nmme. *Journal of Climate*, 27(15):5891 – 5906. 38

[4] Bjerknes, J. (1966). A possible response of the atmospheric hadley circulation to equatorial anomalies of ocean temperature. *Tellus*, 18(4):820–829. 4

[5] Bjerknes, J. (1969). Atmospheric teleconnections from the equatorial pacific. *Mon. Wea. Rev*, 97(3):163–172. 4

[6] Bogenschutz, P. A., Tang, S., Caldwell, P. M., Xie, S., Lin, W., and Chen, Y.-S. (2020). The e3sm version 1 single-column model. *Geoscientific Model Development*, 13(9):4443–4458. 105

[7] Bourlard, H. and Kamp, Y. (1988). Auto-association by multilayer perceptrons and singular value decomposition. *Biological cybernetics*, 59(4):291–294. 45

[8] Brenowitz, N. D. and Bretherton, C. S. (2018). Prognostic validation of a neural network unified physics parameterization. *Geophysical Research Letters*, 45(12):6289–6298. 103

[9] Brenowitz, N. D., Henn, B., McGibbon, J., Clark, S. K., Kwa, A., Perkins, W. A., Watt-Meyer, O., and Bretherton, C. S. (2020). Machine learning climate model dynamics: Offline versus online performance. *arXiv preprint arXiv:2011.03081*. 103

[10] Broni-Bedaiko, C., Katsriku, F. A., Unemi, T., Atsumi, M., Abdulai, J.-D., Shinomiya, N., and Owusu, E. (2019). El niño-southern oscillation forecasting using complex networks analysis of lstm neural networks. *Artificial Life and Robotics*, 24(4):445–451. 54

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

[12] Chan, K. C., Wang, X., Xu, X., Gu, J., and Loy, C. C. (2021). Glean: Generative latent bank for large-factor image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14245–14254. 32

[13] Charte, D., Charte, F., García, S., del Jesus, M. J., and Herrera, F. (2018). A practical tutorial on autoencoders for nonlinear feature fusion: Taxonomy, models, software and guidelines. *Information Fusion*, 44:78–96. 40

[14] Chen, M., Xie, P., Janowiak, J. E., and Arkin, P. A. (2002). Global land precipitation: A 50-yr monthly analysis based on gauge observations. *Journal of Hydrometeorology*, 3(3):249–266. 43, 83

[15] Chen, S., Zhang, Y., and Yang, Q. (2021). Multi-task learning in natural language processing: An overview. *arXiv preprint arXiv:2109.09138*. 41

[16] Cheng, J., Kuang, Q., Shen, C., Liu, J., Tan, X., and Liu, W. (2020). Reslap: Generating high-resolution climate prediction through image super-resolution. *IEEE Access*, 8:39623–39634. 14, 31, 32

[17] Cheng, R., Novak, L., and Schneider, T. (2021). Predicting the interannual variability of california's total annual precipitation. *Geophysical Research Letters*, 48(7):e2020GL091465. 39, 79, 80

[18] Chevallier, F., Chéruy, F., Scott, N., and Chédin, A. (1998). A neural network approach for a fast and accurate computation of a longwave radiative budget. *Journal of Applied Meteorology*, 37(11):1385–1397. 6, 103

[19] Chevallier, F., Morcrette, J.-J., Chéruy, F., and Scott, N. (2000). Use of a neural-network-based long-wave radiative-transfer scheme in the ecmwf atmospheric model. *Quarterly Journal of the Royal Meteorological Society*, 126(563):761–776. 6

[20] Cinquini, L., Crichton, D., Mattmann, C., Harney, J., Shipman, G., Wang, F., Ananthakrishnan, R., Miller, N., Denvil, S., Morgan, M., et al. (2014). The earth system grid federation: An open infrastructure for access to distributed geospatial data. *Future Generation Computer Systems*, 36:400–417. 15, 88

[21] Cohen, J., Pfeiffer, K., and Francis, J. (2017). Winter 2015/16: A turning point in enso-based seasonal forecasts. *Oceanography*. 38

[22] Collobert, R. and 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*, pages 160–167. 41

[23] Deng, L., Hinton, G., and 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*, pages 8599–8603. IEEE. 41

[24] Devries, T., Biswaranjan, K., and Taylor, G. W. (2014). Multi-task learning of facial landmarks and expression. In *2014 Canadian conference on computer and robot vision*, pages 98–103. IEEE. 41

[25] Dilley, M. and Heyman, B. N. (1995). Enso and disaster: droughts, floods and el niño/southern oscillation warm events. *Disasters*, 19(3):181–193. 86

[26] Dong, C., Loy, C. C., He, K., and Tang, X. (2014). Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*, pages 184–199. Springer. 13, 17, 24

[27] Dong, C., Loy, C. C., and Tang, X. (2016). Accelerating the super-resolution convolutional neural network. In *European conference on computer vision*, pages 391–407. Springer. 14, 19, 27, 30

[28] E3SM Project (2018). Energy Exascale Earth System Model (E3SM). [Computer Software] https://dx.doi.org/10.11578/E3SM/dc.20180418.36. 7, 15, 33, 42, 83, 88, 99, 104, 119

[29] Fistikoglu, O. and Okkan, U. (2011). Statistical downscaling of monthly precipitation using ncep/ncar reanalysis data for tahtali river basin in turkey. *Journal of Hydrologic Engineering*, 16(2):157–164. 3, 13

[30] Fournier, Q. and 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)*, pages 211–214. IEEE. 40

[31] Gagne, D. J., Christensen, H. M., Subramanian, A. C., and Monahan, A. H. (2020). Machine learning for stochastic parameterization: Generative adversarial networks in the lorenz'96 model. *Journal of Advances in Modeling Earth Systems*, 12(3):e2019MS001896. 103

[32] Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., and Yacalis, G. (2018). Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, 45(11):5742–5751. 103, 104

[33] Ghifary, M., Kleijn, W. B., Zhang, M., and Balduzzi, D. (2015). Domain generalization for object recognition with multi-task autoencoders. In *Proceedings of the IEEE international conference on computer vision*, pages 2551–2559. 41

[34] Ghosh, S. (2010). Svm-pgsl coupled approach for statistical downscaling to predict rainfall from gcm output. *Journal of Geophysical Research: Atmospheres*, 115(D22). 3, 13

[35] Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448. 41

[36] Golaz, J.-C., Caldwell, P. M., Van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., Abeshu, G., Anantharaj, V., Asay-Davis, X. S., Bader, D. C., et al. (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. 8, 42, 88

[37] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press. 13, 43

[38] Gupta, M., Kodamana, H., and Sandeep, S. (2020). Prediction of enso beyond spring predictability barrier using deep convolutional lstm networks. *IEEE Geoscience and Remote Sensing Letters*. 54

[39] Ham, Y.-G., Kim, J.-H., and Luo, J.-J. (2019). Deep learning for multi-year enso forecasts. *Nature*, 573(7775):568–572. 81

[40] Han, Y., Zhang, G. J., Huang, X., and Wang, Y. (2020). A moist physics parameterization based on deep learning. *Journal of Advances in Modeling Earth Systems*, 12(9):e2020MS002076. 103

[41] Hannachi, A., Stephenson, D., and Sperber, K. (2003). Probability-based methods for quantifying nonlinearity in the enso. *Climate Dynamics*, 20(2-3):241–256. 40

[42] He, D., Lin, P., Liu, H., Ding, L., and Jiang, J. (2019). Dlenso: A deep learning enso forecasting model. In *Pacific Rim International Conference on Artificial Intelligence*, pages 12–23. Springer. 87

[43] He, J. and Eastman, J. R. (2020). A sequential autoencoder for teleconnection analysis. *Remote Sensing*, 12(5):851. 5, 40

[44] Hinton, G. E. and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507. 40

[45] Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780. 53, 89

[46] Hsieh, W. W., Wu, A., and Shabbar, A. (2006). Nonlinear atmospheric teleconnections. *Geophysical Research Letters*, 33(7). 39

[47] Huang, A., Vega-Westhoff, B., and Sriver, R. L. (2019). Analyzing el niño–southern oscillation predictability using long-short-term-memory models. *Earth and Space Science*, 6(2):212–221. 54

[48] Huang, X. and Ullrich, P. A. (2017). 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. 39, 79

[49] Huntingford, C., Jeffers, E. S., Bonsall, M. B., Christensen, H. M., Lees, T., and Yang, H. (2019). Machine learning and artificial intelligence to aid climate change research and preparedness. *Environmental Research Letters*, 14(12):124007. 2

[50] Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., Lamarque, J.-F., Large, W. G., Lawrence, D., Lindsay, K., et al. (2013). The community earth system model: a framework for collaborative research. *Bulletin of the American Meteorological Society*, 94(9):1339–1360. 7

[51] Jin, E. K., Kinter, J. L., Wang, B., Park, C.-K., Kang, I.-S., Kirtman, B., Kug, J.-S., Kumar, A., Luo, J.-J., Schemm, J., et al. (2008). Current status of enso prediction skill in coupled ocean–atmosphere models. *Climate Dynamics*, 31(6):647–664. 93

[52] Jong, B.-T., Ting, M., and Seager, R. (2016). El niño's impact on california precipitation: Seasonality, regionality, and el niño intensity. *Environmental Research Letters*, 11(5):054021. 39, 79

[53] Kao, H.-Y. and Yu, J.-Y. (2009). Contrasting eastern-pacific and central-pacific types of enso. *Journal of Climate*, 22(3):615–632. 4

[54] Kendall, A., Gal, Y., and 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*, pages 7482–7491. 41

[55] Kharin, V. V., Zwiers, F. W., Zhang, X., and Hegerl, G. C. (2007). Changes in temperature and precipitation extremes in the ipcc ensemble of global coupled model simulations. *Journal of Climate*, 20(8):1419–1444. 12

[56] Kim, S., Ames, S., Lee, J., Zhang, C., Wilson, A. C., and Williams, D. (2017a). Massive scale deep learning for detecting extreme climate events. *Climate Informatics*. 5

[57] Kim, S., Hori, T., and Watanabe, S. (2017b). 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)*, pages 4835–4839. IEEE. 41

[58] Krasnopolsky, V., Fox-Rabinovitz, M., Hou, Y., Lord, S., and Belochitski, A. (2010). Accurate and fast neural network emulations of model radiation for the

ncep coupled climate forecast system: climate simulations and seasonal predictions. *Monthly Weather Review*, 138(5):1822–1842. 104

[59] Krasnopolsky, V. M., Chalikov, D. V., and Tolman, H. L. (2002). A neural network technique to improve computational efficiency of numerical oceanic models. *Ocean Modelling*, 4(3-4):363–383. 6

[60] Krasnopolsky, V. M., Fox-Rabinovitz, M. S., and Belochitski, A. A. (2013). Using ensemble of neural networks to learn stochastic convection parameterizations for climate and numerical weather prediction models from data simulated by a cloud resolving model. *Advances in Artificial Neural Systems*, 2013. 6

[61] Krasnopolsky, V. M., Fox-Rabinovitz, M. S., and Chalikov, D. V. (2005). New approach to calculation of atmospheric model physics: Accurate and fast neural network emulation of longwave radiation in a climate model. *Monthly Weather Review*, 133(5):1370–1383. 6, 102, 103

[62] Kug, J.-S., Jin, F.-F., and An, S.-I. (2009). Two types of el niño events: cold tongue el niño and warm pool el niño. *Journal of Climate*, 22(6):1499–1515. 4, 40

[63] Lacis, A. A. and Oinas, V. (1991). A description of the correlated k distribution method for modeling nongray gaseous absorption, thermal emission, and multiple scattering in vertically inhomogeneous atmospheres. *Journal of Geophysical Research: Atmospheres*, 96(D5):9027–9063. 102

[64] Larkin, N. K. and Harrison, D. (2005). Global seasonal temperature and precipitation anomalies during el niño autumn and winter. *Geophysical Research Letters*, 32(16). 4

[65] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444. 13

[66] LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., and Jackel, L. (1989). Handwritten digit recognition with a back-propagation network. *Advances in neural information processing systems*, 2. 13

[67] Lee, S.-K., Lopez, H., Chung, E.-S., DiNezio, P., Yeh, S.-W., and Wittenberg, A. T. (2018). On the fragile relationship between el nio and california rainfall. *Geophysical Research Letters*, 45(2):907–915. 38

[68] Li, G., Ren, B., Yang, C., and Zheng, J. (2010). Indices of el niño and el niño modoki: an improved el niño modoki index. *Advances in Atmospheric Sciences*, 27(5):1210–1220. 39

[69] Liu, P., Qiu, X., and Huang, X. (2017). Adversarial multi-task learning for text classification. *arXiv preprint arXiv:1704.05742*. 41

[70] Liu, Y., Ganguly, A. R., and Dy, J. (2020). Climate downscaling using ynet: A deep convolutional network with skip connections and fusion. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 3145–3153. 17

[71] Liu, Y., Racah, E., Correa, J., Khosrowshahi, A., Lavers, D., Kunkel, K., Wehner, M., Collins, W., et al. (2016). Application of deep convolutional neural networks for detecting extreme weather in climate datasets. *arXiv preprint arXiv:1605.01156*. 5

[72] Livezey, R. E. and Chen, W. (1983). Statistical field significance and its determination by monte carlo techniques. *Mon. Wea. Rev*, 111(1):46–59. 55

[73] Luo, Z., Yu, J., and Liu, Z. (2019). The super-resolution reconstruction of sar image based on the improved fsrcnn. *The Journal of Engineering*, 2019(19):5975–5978. 18

[74] LHeureux, M. L., Takahashi, K., Watkins, A. B., Barnston, A. G., Becker, E. J., Di Liberto, T. E., Gamble, F., Gottschalck, J., Halpert, M. S., Huang, B., et al. (2017). Observing and predicting the 2015/16 el niño. *Bulletin of the American Meteorological Society*, 98(7):1363–1382. 4, 38

[75] LHeureux, M. L., Tippett, M. K., and Becker, E. J. (2021). Sources of subseasonal skill and predictability in wintertime california precipitation forecasts. *Weather and Forecasting*, 36(5):1815 – 1826. 38, 79, 80

[76] Mahajan, S., Evans, K. J., Branstetter, M., Anantharaj, V., and Leifeld, J. K. (2015). Fidelity of precipitation extremes in high resolution global climate simulations. *Procedia Computer Science*, 51:2178–2187. 12, 32

[77] Mahajan, S., North, G. R., Saravanan, R., and Genton, M. G. (2012). Statistical significance of trends in monthly heavy precipitation over the us. *Climate dynamics*, 38(7):1375–1387. 57, 69

[78] Mahajan, S., Tang, Q., Keen, N. D., Golaz, J.-C., and 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*, pages 1 – 62. 42

[79] Majewski, D., Liermann, D., Prohl, P., Ritter, B., Buchhold, M., Hanisch, T., Paul, G., Wergen, W., and Baumgardner, J. (1998). The new global icosahedral–hexagonal grid point model gme of the deutscher wetterdienst. In *Proc. ECMWF Seminar on Recent Developments in Numerical Methods for Atmospheric Modelling, ECMWF, Reading, United Kingdom*. Citeseer. 102

[80] Majewski, D., Liermann, D., Prohl, P., Ritter, B., Buchhold, M., Hanisch, T., Paul, G., Wergen, W., and Baumgardner, J. (2002). The operational global icosahedral–hexagonal gridpoint model gme: description and high-resolution tests. *Monthly Weather Review*, 130(2):319–338. 102

[81] Masti, D. and Bemporad, A. (2018). Learning nonlinear state-space models using deep autoencoders. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 3862–3867. IEEE. 40

[82] Moncet, J. and Clough, S. (1997). Accelerated monochromatic radiative transfer for scattering atmospheres: Application of a new model to spectral radiance

observations. *Journal of Geophysical Research: Atmospheres*, 102(D18):21853–21866. 102

[83] Morcrette, J., Barker, H. W., Cole, J., Iacono, M. J., and Pincus, R. (2008a). Impact of a new radiation package, mcrad, in the ecmwf integrated forecasting system. *Monthly weather review*, 136(12):4773–4798. 102

[84] Morcrette, J.-J., Mozdzynski, G., and Leutbecher, M. (2008b). A reduced radiation grid for the ecmwf integrated forecasting system. *Monthly weather review*, 136(12):4760–4772. 103

[85] Mu, B., Ma, S., Yuan, S., and 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)*, pages 215–219. IEEE. 54

[86] Nicholls, R. J. and Cazenave, A. (2010). Sea-level rise and its impact on coastal zones. *science*, 328(5985):1517–1520. 12

[87] Nowack, P., Braesicke, P., Haigh, J., Abraham, N. L., Pyle, J., and Voulgarakis, A. (2018). Using machine learning to build temperature-based ozone parameterizations for climate sensitivity simulations. *Environmental Research Letters*, 13(10):104016. 6

[88] O'Gorman, P. A. and Dwyer, J. G. (2018). Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change,

and extreme events. *Journal of Advances in Modeling Earth Systems*, 10(10):2548–2563. 6

[89] Pal, A., Mahajan, S., and Norman, M. R. (2019). Using deep neural networks as cost-effective surrogate models for super-parameterized e3sm radiative transfer. *Geophysical Research Letters*, 46(11):6069–6079. 6, 101, 103, 104, 107

[90] Pal, M., Maity, R., Ratnam, J., Nonaka, M., and Behera, S. K. (2020). Long-lead prediction of enso modoki index using machine learning algorithms. *Scientific reports*, 10. 87

[91] Pan, B., Hsu, K., AghaKouchak, A., Sorooshian, S., and Higgins, W. (2019a). Precipitation prediction skill for the west coast united states: From short to extended range. *Journal of Climate*, 32(1):161 – 182. 79

[92] Pan, Z., Yu, W., Yi, X., Khan, A., Yuan, F., and Zheng, Y. (2019b). Recent progress on generative adversarial networks (gans): A survey. *IEEE Access*, 7:36322–36333. 117

[93] Philander, S. G. H. (1983). El niño southern oscillation phenomena. *Nature*, 302(5906):295–301. 4, 86

[94] Plaut, E. (2018). From principal subspaces to principal components with linear autoencoders. *arXiv preprint arXiv:1804.10253*. 45

[95] Rasp, S., Pritchard, M. S., and Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, 115(39):9684–9689. 103

[96] Rayner, N., Parker, D., Folland, C., Horton, E., Alexander, L., and Rowell, D. (2003). The global sea-ice and sea surface temperature (hadisst) data sets. *J. Geophys. Res.* 42, 83

[97] Raza, S. E. A., Cheung, L., Epstein, D., Pelengaris, S., Khan, M., and 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)*, pages 337–340. IEEE. 41

[98] Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., et al. (2019). Deep learning and process understanding for data-driven earth system science. *Nature*, 566(7743):195–204. 2

[Renard et al.] Renard, B., Lang, M., Bois, P., Dupeyrat, A., Mestre, O., Niel, H., Sauquet, E., Prudhomme, C., Parey, S., Paquet, E., Neppel, L., and Gailhard, J. Regional methods for trend detection: Assessing field significance and regional consistency. *Water Resources Research*, 44(8). 67

[100] Saunders, M., Chandler, R., Merchant, C., and Roberts, F. (2000). Atlantic hurricanes and nw pacific typhoons: Enso spatial impacts on occurrence and

landfall. *Geophysical research letters*, 27(8):1147–1150. 86

[101] Schmidt, G. (2010). The real holes in climate science. *Nature*, 463:21. 12

[102] Shinohara, Y. (2016). Adversarial multi-task learning of deep neural networks for robust speech recognition. In *Interspeech*, pages 2369–2372. San Francisco, CA, USA. 41

[103] Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. 14

[104] Stengel, K., Glaws, A., Hettinger, D., and King, R. N. (2020). Adversarial super-resolution of climatological wind and solar data. *Proceedings of the National Academy of Sciences*, 117(29):16805–16815. 13, 15, 31, 32

[105] Takahashi, K., Montecinos, A., Goubanova, K., and Dewitte, B. (2011). Enso regimes: Reinterpreting the canonical and modoki el niño. *Geophysical Research Letters*, 38(10). 40

[106] Tang, Y. and 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). 40, 45

[107] Thomas, E. E. and Vimont, D. J. (2016). Modeling the mechanisms of linear and nonlinear enso responses to the pacific meridional mode. *Journal of Climate*, 29(24):8745–8761. 40

[108] Tolman, H. L., Krasnopolsky, V. M., and Chalikov, D. V. (2005). Neural network approximations for nonlinear interactions in wind wave spectra: direct mapping for wind seas in deep water. *Ocean Modelling*, 8(3):253–278. 6

[109] Toshniwal, S., Tang, H., Lu, L., and Livescu, K. (2017). Multitask learning with low-level auxiliary tasks for encoder-decoder based speech recognition. *arXiv preprint arXiv:1704.01631*. 41

[110] Trafalis, T. B., Adrianto, I., Richman, M. B., and Lakshmivarahan, S. (2014). Machine-learning classifiers for imbalanced tornado data. *Computational Management Science*, 11(4):403–418. 5

[111] Trenberth, K. E. (2012). Framing the way to relate climate extremes to climate change. *Climatic change*, 115(2):283–290. 12

[112] Trenberth, K. E. and Stepaniak, D. P. (2001). Indices of el niño evolution. *Journal of climate*, 14(8):1697–1701. 4, 39

[113] Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R., and Ganguly, A. R. (2017). Deepsd: Generating high resolution climate change projections through single image super-resolution. In *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining*, pages 1663–1672. 12, 13, 17, 18

[114] Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R., and Ganguly, A. R. (2018). Generating high resolution climate change projections through single

image super-resolution: An abridged version. In *International Joint Conferences on Artificial Intelligence Organization*. 13, 31

[115] Ventura, V., Paciorek, C. J., and Risbey, J. S. (2004). Controlling the proportion of falsely rejected hypotheses when conducting multiple tests with climatological data. *Journal of Climate*, 17(22):4343–4356. 67

[116] Von Storch, H. and Zwiers, F. W. (2001). *Statistical analysis in climate research*. Cambridge university press. 40

[117] Vu, M. T., Aribarg, T., Supratid, S., Raghavan, S. V., and Liong, S.-Y. (2016). Statistical downscaling rainfall using artificial neural network: significantly wetter bangkok? *Theoretical and applied climatology*, 126(3-4):453–467. 3, 13

[118] Wang, C. (2018). A review of enso theories. *National Science Review*, 5(6):813–825. 86

[119] Wang, G., Zhuang, Y., Fu, R., Zhao, S., and Wang, H. (2021a). Improving seasonal prediction of california winter precipitation using canonical correlation analysis. *Journal of Geophysical Research: Atmospheres*, 126(17):e2021JD034848. 39, 79, 80

[120] Wang, L. and Yoon, K.-J. (2022). Semi-supervised student-teacher learning for single image super-resolution. *Pattern Recognition*, 121:108206. 32

[121] Wang, Q., Gao, Q., Wu, L., Sun, G., and Jiao, L. (2021b). Adversarial multi-path residual network for image super-resolution. *IEEE Transactions on Image Processing*, 30:6648–6658. 32

[122] Wang, S., Anichowski, A., Tippett, M. K., and 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. 38, 79, 80

[123] Wang, W., Huang, Y., Wang, Y., and Wang, L. (2014). Generalized autoencoder: A neural network framework for dimensionality reduction. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 490–497. 40

[124] Wang, Y., Yao, H., and Zhao, S. (2016). Auto-encoder based dimensionality reduction. *Neurocomputing*, 184:232–242. 40

[125] Wehner, M. F., Reed, K. A., Li, F., Prabhat, Bacmeister, J., Chen, C.-T., Paciorek, C., Gleckler, P. J., Sperber, K. R., Collins, W. D., Gettelman, A., and Jablonowski, C. (2014). The effect of horizontal resolution on simulation quality in the Community Atmospheric Model, CAM5.1. *Journal of Advances in Modeling Earth Systems*, 6(4):980–997. 32

[126] Wehner, M. F., Smith, R. L., Bala, G., and Duffy, P. (2010a). The effect of horizontal resolution on simulation of very extreme us precipitation events in a

global atmosphere model. *Climate dynamics*, 34(2-3):241–247. 12

[127] Wehner, M. F., Smith, R. L., Bala, G., and Duffy, P. (2010b). The effect of horizontal resolution on simulation of very extreme us precipitation events in a global atmosphere model. *Climate Dynamics*, 34(2-3). 32

[128] Wilks, D. S. (2006). On field significance and the false discovery rate. *Journal of Applied Meteorology and Climatology*, 45(9):1181–1189. 67

[129] Williams, I. N. and 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. 39, 40, 67

[130] Wu, A., Hsieh, W. W., and Shabbar, A. (2005). The nonlinear patterns of north american winter temperature and precipitation associated with enso. *Journal of Climate*, 18(11):1736 – 1752. 39

[131] Xiaoqun, C., Yanan, G., Bainian, L., Kecheng, P., Guangjie, W., and Mei, G. (2020). Enso prediction based on long short-term memory (lstm). In *IOP Conference Series: Materials Science and Engineering*, volume 799, page 012035. IOP Publishing. 87

[132] Yaguchi, K., Ikarigawa, K., Kawasaki, R., Miyazaki, W., Morikawa, Y., Ito, C., Shuzo, M., and 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, pages 364–367. 41

[133] Yan, J., Mu, L., Wang, L., Ranjan, R., and Zomaya, A. Y. (2020). temporal convolutional networks for the advance prediction of enso. *Scientific Reports*, 10(1):1–15. 87

[134] Yang, H., Piao, S., Huntingford, C., Peng, S., Ciais, P., Chen, A., Zhou, G., Wang, X., Gao, M., and Zscheischler, J. (2019). Strong but intermittent spatial covariations in tropical land temperature. *Geophysical Research Letters*, 46(1):356–364. 5

[135] Yu, J.-Y. and Kim, S. T. (2010). Identification of central-pacific and eastern-pacific types of enso in cmip3 models. *Geophysical Research Letters*, 37(15). 40

[136] Yu, L., Heilman, W. E., Zhong, S., and Bian, X. (2017a). Relationships of the symmetric and asymmetric components of enso to us extreme precipitation. *Atmospheric Science Letters*, 18(12):466–474. 67

[137] Yu, L., Zhong, S., Heilman, W. E., and Bian, X. (2017b). A comparison of the effects of el niño and el niño modoki on subdaily extreme precipitation occurrences across the contiguous united states. *Journal of Geophysical Research: Atmospheres*, 122(14):7401–7415. 67

[138] Yuval, J. and OGorman, P. A. (2020). Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions. *Nature communications*, 11(1):1–10. 103

[139] Zamparo, L. and Zhang, Z. (2015). Deep autoencoders for dimensionality reduction of high-content screening data. *arXiv preprint arXiv:1501.01348*. 40

[140] Zou, Y., Zhang, L., Liu, C., Wang, B., Hu, Y., and Chen, Q. (2021). Super-resolution reconstruction of infrared images based on a convolutional neural network with skip connections. *Optics and Lasers in Engineering*, 146:106717. 32

Vita

Linsey Passarella was born in Fort Myers, Florida in 1996 to Jill and Ken Passarella. She graduated from Cypress Lake High School in 2014 and from College of Charleston in 2018 with a double bachelors degree in Physics and Meteorology. She began an internship at Oak Ridge National Lab in the summer of 2018 and joined the University of Tennessee's Bredesen Center that fall. She joined ORNL as an RD Assitant in Machine Learning Research for the Augmented Analyst Intelligence group in October 2021 and graduated with her PhD in Data Science and Engineering in 2022.