Stable climate simulations using a realistic GCM with neural network parameterizations for atmospheric moist physics and radiation processes

Xin Wang[1], Yilun Han[2], Wei Xue[1], Guangwen Yang[1], Guang J. Zhang[3]

[1]Department of Computer Science and Technology, Tsinghua University, Beijing, 100084, China
[2]Department of Earth System Science, Tsinghua University, Beijing, 100084, China
[3]Scripps Institution of Oceanography, La Jolla, CA USA

Correspondence to: Wei Xue (xuewei@tsinghua.edu.cn), Yilun Han (hanyl16@mails.tsinghua.edu.cn)

Abstract. In climate models, subgrid parameterizations of convection and cloud are one of the main reasons for the biases in precipitation and atmospheric circulation simulations. In recent years, due to the rapid development of data science, Machine learning (ML) parameterizations for convection and clouds have been proven the potential to perform better than conventional parameterizations. At present, most of the existing studies are on aqua-planet and idealized models, and the problems of simulated instability and climate drift still exist. In realistic configurated models, developing a machine learning parameterization scheme remains a challenging task. In this study, a ~~group~~set of deep residual ~~multilayer perceptrons~~neural networks (ResDNNs) with strong nonlinear fitting ability is designed to ~~learn~~emulate a superparameterization (SP) with different types of outputs. Sensitivity tests show that high accuracy is necessary to develop a stable ML parameterization ~~scheme from cloud-resolving model outputs. Multi-target training is achieved to best balance~~. Trial-and-error is used to acquire the ~~fits across diverse neural network outputs. The~~ optimal ~~machine learning parameterization, named NN-Parameterization, is further chosen among feasible candidates~~ResDNN set for both high performance and long-term ~~simulation. The results show that~~ stability, named NN-Parameterization. In offline validation, NN-Parameterization ~~performs well~~emulates the SP results far better than the conventional subgrid parameterizations. Then, in the multi-year ~~climate simulations and~~prognostic test, NN-Parameterization reproduces reasonable ~~climatology and~~climate mean states but still with some biases. Most importantly, NN parameterization successfully reproduces the climate variability in a ~~general circulation model (~~superparameterizated GCM~~),~~ with ~~a~~an over 30-time faster running speed~~ of about 30 times faster than the cloud-resolving model embedded Superparameterizated GCM~~. Under real geographical boundary conditions, the hybrid ML-physical GCM well simulates the spatial distribution of boreal summer precipitation and significantly improves the frequency of precipitation extremes, which is largely underestimated in the Community Atmospheric Model version 5 (CAM5) with the horizontal resolution of 1.9°×2.5°. Furthermore, the hybrid ML-physical GCM simulates a stronger signal of the Madden-Julian oscillation with a more reasonable propagation speed~~, which is too weak and propagates too fast in~~ than CAM5. This study is a pioneer to achieve multi-year stable climate simulations using a hybrid ML-physical GCM in actual land-ocean boundary conditions. It demonstrates the emerging potential for using machine learning parameterizations in climate simulations.

1 Introduction

The general circulation models (GCMs) have been widely used for studying climate variability, prediction and projections. Despite decades of GCM development, most GCMs still suffer from many systematic biases, especially at low latitudes. A prominent tropical bias in most current GCMs is the double intertropical convergence zone (ITCZ) syndrome, which is characterized by two parallel zonal bands of annual precipitation straddling the equator over the central and eastern Pacific (~~Mechoso et al., 1995;~~ Lin, 2007; ~~Zuidema et al., 2016;~~ Zhang et al., 2019~~; Lu et al., 2020~~). Convectively coupled equatorial waves and the Madden-Julian Oscillation (MJO), featured by eastward propagating convective cloud clusters, are also not well simulated in GCMs (Ling et al., ~~2017). The simulated MJOs in GCMs are often too weak and propagate too fast (Lin et al., 2006).~~ 2017; Cao and Zhang, 2017).

Many studies have attributed most of these biases to the imperfection of the parameterization schemes for atmospheric moist convection and cloud processes in current GCMs (~~Song and Zhang, 2009;~~ Zhang and Song, 2010; ~~Oueslati and Bellon, 2013; Crueger and Stevens, 2015; Deng et al., 2016;~~ Cao and Zhang, 2017; ~~Peters et al., 2017;~~ Song and Zhang, 2018; Zhang and Song, 2019). Cloud-related processes span a large range of spatial scales, from micron-scale cloud nucleation, meter-scale turbulence, to individual convective cells and organized convective systems, which are a few kilometers to hundreds of kilometers in size, and to tropical disturbances, which have a spatial scale of thousands of kilometers. They directly influence the radiation balance and hydrological cycle of the earth system~~. Their interaction~~ and interact with the atmospheric circulation ~~affects~~, affecting the transport and distribution of energy ~~and is the largest source of precipitation biases.~~(Emanual et al., 1994). Therefore, it is very important to simulate the cloud and convection process in GCMs correctly. However, the current GCMs used for climate simulation have a horizontal resolution of ~100km and a vertical hydrostatic coordinate. ~~Therefore~~Thus, in most GCMs, besides parameterized cloud microphysics, convection and its influence on the atmospheric circulation are represented by convective parameterization schemes, which are usually based on simplified theories, limited observations, and empirical relationships (Tiedtke, 1989; Zhang and McFarlane, 1995; ~~Zhang, 2002; Wu, 2012; Storer et al., 2015; Zhao et al., 2018; Seo et al., 2019; Xie et al., 2019; Lopez-Gomez et al., 2020; Hourdin et al.,~~Lopez-Gomez et al., 2020). Those schemes regard convective heat and moisture transport as the collective effects of idealized individual kilometer-scale convective cells. They cannot represent the effects of many complicated convective structures, including organized convective systems, leading to large uncertainties and biases in climate simulations~~.~~ (Bony et al., 2015).

Cloud Resolving Models (CRMs), on the other hand, have long been used to simulate convection. Because CRMs have higher horizontal and vertical resolutions and can explicitly resolve the thermodynamic processes in convection, they simulate convection more accurately, including convective organization (Feng et al., 2018). In recent years, CRMs have been ~~applied to~~used as superparameterization (SP) in low-resolution GCMs to replace conventional cumulus convection and cloud ~~microphysical~~ parameterization schemes~~, such as~~. The commonly used SP model is the superparameterized version of the Community Atmosphere Model (SPCAM) developed by the National Center for Atmospheric Research (Grabowski and Smolarkiewicz, 1999; Grabowski, 2001, 2004; Khairoutdinov and Randall, 2001; Randall et al., 2003; Khairoutdinov et al.,

2005). Compared with conventional cumulus convection and cloud ~~microphysical~~ parameterization schemes, SPCAM performs better in simulating mesoscale convective systems, diurnal cycles of precipitation, monsoons, precipitation frequency distribution, and MJOs (Khairoutdinov et al., 2005; Bretherton et al., 2014; Jiang et al., 2015; Jin et al., 2016; ~~Jones~~Kooperman et al., ~~2019; Hannah et al., 2020~~2016). However, when using 2D CRM as SP, the improvement on climate mean states is not obvious (Khairoutdinov et al., 2005). Also, SPCAM requires far more computing ~~resource required for SPCAM is an order~~resources than the same resolution CAM in 1 to 2 orders of magnitude ~~larger compared with that for CAM. The~~according to the resolution of the CRM subdomain. Thus, the use of SPCAM in long-term climate simulations and ensemble prediction is restricted by the current computing resource. Developing novel and computationally efficient schemes for high performance convection and cloud processes is still an open problem in GCM development.

~~The conventional theory-driven parameterization schemes are based on the limited mathematical theory and observations, guided by physical laws of atmospheric motion, while the data-driven parameterization scheme can identify and extract complex nonlinear relationships from high-resolution and high-fidelity data sets. More recently~~In the last 5 years, the rapid development of machine learning (ML) technologies, especially deep learning technologies such as neural networks (NNs), has provided novel approaches to constructing parameterization schemes. Machine learning can identify and discover complex nonlinear relationships that exist in large data sets and model them. Several studies have used machine learning methods to develop convection and cloud parameterization schemes (e.g., ~~Schneider et al., 2017; Dueben and Bauer, 2018;~~ Gentine et al., 2018; Rasp et al., 2018). These studies followed a similar approach. The first step is to derive a target dataset from a reference simulation, which is later used for machine learning ~~algorithm~~models training. Then, the trained ~~ones~~machine learning models are often evaluated offline against other independent reference simulations and finally implemented in a GCM to replace the conventional parameterization schemes ~~(Rasp, 2020).~~.

Krasnopolsky et al. (2013) first proposed a proof-of-concept for developing convection parameterization based on the NN technique. Specifically, an ensemble of shallow NNs was applied to learn convective temperature and moisture tendencies, with training data from CRM simulations forced by observations in the tropical western Pacific. The resulting convective parameterization scheme was able to simulate the main features of cloud and precipitation in the NCAR CAM4 diagnostically. However, the key issue of prognostic validation in 3-D GCMs was not addressed. Recent studies have investigated ML parameterizations in prognostic mode in simplified aqua-planet GCMs. For example, Rasp et al. (2018) developed a deep fully connected NN ~~algorithm~~(DNN) to predict convection and clouds, which was trained with the data from an aqua-planet SPCAM. The NN parameterization was then implemented in the corresponding aqua-planet CAM and produced multi-year prognostic results close to SPCAM. ~~They~~For this NN parameterization, Rasp (2020) found that minor changes, either to the training dataset or in the input/output vectors, can lead to model integration instabilities. Brenowitz and Bretherton (2019) fitted a DNN for convection and clouds to the coarse-grained data from a near-global aqua-planet cloud-resolving simulation using the System for Atmospheric Modeling (SAM). The NN scheme was then tested prognostically in a coarse-grid SAM. Their results showed that there were unphysical correlations learned by the network, and information in the upper levels from the input vector had to be removed to produce stable long-term simulations. Rather than using NNs, Yuval and O'Gorman (2020) used

random forest to develop an ML parameterization based on the training data from a high-resolution idealized 3-D model with a setup of equatorial beta plane. ~~Stable simulations that replicated the climatology of the high-resolution model were achieved after they implemented this parameterization in a coarse-resolution GCM~~They used two independent random forests to emulate different processes separately and ensured physical constraints by predicting subgrid fluxes instead of tendencies. Later, Yuval et al. (2021) completed the same task with NNs. Both works achieved stable simulations in coarse resolution aqua-planet GCMs. Brenowitz et al. (2020) proposed methods to interpret and stabilize ML parameterization of convection. In their work, a wave spectra analysis tool was introduced to explain why ML coupled GCMs blew up.

~~Although machine learning is a promising approach for developing new parameterizations, issues of instability and climate drift still prevent it from the application of machine learning parameterization in models. Rasp (2020) proposed coupled online learning to tackle instabilities and biases in NN parameterizations, which is a concept illustration using the idealized Lorenz 96 model.~~ In real-world climate models with varied underlying surfaces, convection and clouds are more diverse under different climate backgrounds, which makes the task of developing ML-based parameterizations more complicated. A few early works have shown the feasibility of using neural networks fitting cloud processes in real-world models. Han et al. (2020) used a 1-D deep residual convolutional neural network (ResNet) to emulate moist physics in SPCAM. This ResNet based parameterization fitted the targets with high accuracy and is successfully implemented in a single column model. Mooers et al. (2021) got a high-skill DNN via ~~auto-~~automated machine learning technique and forced an offline land model with DNN emulated atmospheric fields. However, neither of these studies have tested their NNs prognostically for long-term simulations ~~with ML-parameterized GCMs. This study uses a group~~. Similar to the idea of several NNs for different processes in Yuval and O'Gorman (2020), this study uses a set of NNs to emulate convection and cloud processes in SPCAM with an actual global land-ocean distribution. We ~~apply two innovative methods~~use the residual connections in ~~neural network models: multi-target training~~Han et al. (2020) to ~~achieve balanced results across diverse neural network outputs and multilayer perceptron~~ acquire super deep neural networks with ~~residual blocks (ResMLP) to enhance~~great nonlinear fitting ability. Furthermore, ~~an optimal DNN emulator is chosen among a set of well trained neural networks by using multi-target training~~we conduct systematic trial-and-error to filter out unstable NN parameterizations and ~~ResMLP, to achieve~~get the best ResDNN set with both ~~high-performance~~accuracy and long-term ~~simulations~~stability. The NN parameterization scheme is then implemented in the realistically configurated CAM to obtain long-term stable simulations. Technically, NNs are commonly implemented via high-level programming languages such as Python and deep learning libraries. However, GCMs are mainly written in Fortran, making ~~it difficult to integrate~~integrating with deep learning algorithms~~.~~ inconvenient. Therefore, we introduce ~~a DNN~~an NN-GCM coupling platform in which ~~the DNN model~~NN models and ~~the GCM~~GCMs can interact through data transmission. This coupling strategy can facilitate the development of ML-physical hybrid models with high flexibility. Under real-geography boundary conditions, our work achieves more than 10-year stable climate simulations in Atmospheric Model Intercomparison Project (AMIP)-style experiments by using a hybrid ML-physical GCM. The simulation results may show some biases in climate mean fields but successfully reproduce variability in SPCAM. To our knowledge, this is the first time a decade-long stable real-world climate simulation is achieved with a NN-based parameterization.

The remainder of this paper is organized as follows. Section 2 briefly describes the model, the experiments, the DNN algorithm, and the DNN-GCM coupling platform. Section 3 analyses the simulation stability of NNCAM. Section 4 presents the offline validation of the DNN scheme, focusing on the output temperature and moisture tendencies. Results of multi-year simulations, employing the DNN parameterization scheme, are shown in ~~section 4~~Section 5. A summary and conclusions are presented in ~~section 5~~Section 6.

2 Methods and data

In this study, we choose SPCAM as the reference model to generate target simulations. A ~~group~~set of ~~DNNs~~NNs is trained with the target simulation data using optimized hyperparameters. Then, they are organized as a subgrid physics emulator and implemented into the superparameterized version of Community Atmospheric Model (SPCAM), replacing both the CRM based SP and ~~CRM~~the radiation~~.~~ effects of the CRM. This ~~DNN~~NN-enabled GCM is referred to as NNCAM hereafter.

2.1 SPCAM setup and data generation

The GCMs used in this study are the CAM5.2 developed by the National Center for Atmospheric Research and its superparameterized version SPCAM (Khairoutdinov and Randall, 2001; Khairoutdinov et al., 2005). A complete description of CAM5 is given by Neale et al. (2012). The dynamic core of CAM5 has a horizontal resolution of 1.9°×2.5° and 30 vertical levels with a model top at about 2 hPa. To represent moist processes, CAM5 adopts a plume-based treatment of shallow convection (Park and Bretherton, 2009), a mass-flux parameterization scheme for deep convection (Zhang and McFarlane, 1995), and an advanced two-moment representation of cloud microphysical processes (Morrison and Gettelman, 2008; Gettelman et al., 2010). In the AMIP experiments we conducted, CAM5 is coupled to a land surface model Community Land Model version 4.0 (Oleson et al., 2010) and uses prescribed sea surface temperatures and sea ice concentrations.

In this study, SPCAM is used to generate the training data. In SPCAM, a two-dimensional (2-D) CRM is embedded in each grid column of the host CAM as SP. The 2-D CRM has 32 grid points in the zonal direction and 30 vertical levels that are shared with the host CAM. The CRM handles convection and cloud microphysics to replace the conventional parameterization schemes, and the radiation is calculated on the CRM subgrids to include the cloud-radiation interaction at cloud scale (Khairoutdinov et al., 2005~~), referring to CRM radiation hereafter. The other physical processes~~). Under realistic configuration, the planetary boundary layer process, orographic gravity wave drags, and the dynamic core are computed on the CAM grid ~~as usual~~. One conceptual advantage of using SPCAM as the reference simulation is that the subgrid and grid-scale processes are clearly separated, making it easy to define the parameterization task for an ML algorithm (Rasp, 2020).

2.2 NN-Parameterization

~~In the NNCAM, a DNN emulator is trained with the target data from the 2-D CRM and CRM radiation to achieve better results than the CAM5 with conventional parameterizations. In the following section, we refer to this DNN emulator as the NN-~~

~~Parameterization. As shown in Figure 1, the NN-Parameterization replaces the CRM moist physics and radiation calculations in SPCAM. It is coupled with the dynamic core and other parameterization schemes in each time integration loop in NNCAM.~~

~~NN-Parameterization is a challenging deep learning application. It integrates the NNs into a scientific computing program for continuous time integration. In the numerical model system, the prediction errors of the NNs are at the risk of being amplified by the continuous iterations with the dynamic core and other physical processes, causing model state drift and even model crashes in NNCAM. This is expected to be more difficult in the GCMs with real land-ocean distributions. Because of the energy exchange between the land and the atmosphere, the nonlinearity of the simulation system is more complicated than the idealized aqua-planet models, bringing more uncertainties and more numerical sensitivities. In this section, we propose a DNN based Parameterization that can fit the grid average of the 2-D CRM data in SPCAM, and on this basis, achieve high computational performance and stable time integration.~~

2.2.1 Data sets

The NN-Parameterization~~, as~~ is a deep learning emulator of the ~~CRM~~SP and ~~the CRM~~ its cloud-scale radiation effects in SPCAM~~, is designed to replace physics processes in the host CAM5, including deep and shallow convection, cloud microphysics and macrophysics, and radiation.~~. Therefore, the inputs of this emulator are borrowed from ~~CRM inputs~~ the SP input variables such as the grid-scale state variables and forcings ~~composed of the dynamic core and the planetary boundary diffusion. They are the~~, including specific humidity q_v, temperature T, largescale water vapor forcing $\left(\frac{\partial q_v}{\partial t}\right)_{ls}$ and temperature forcing $\left(\frac{\partial T}{\partial t}\right)_{ls}$. Additionally, we select surface pressure P_s and solar insolation (SOLIN) at the top of the model from the radiation module. The outputs of NN-Parameterization are subgrid-scale tendencies of moisture $\left(\frac{\partial q_v}{\partial t}\right)$ and of temperature $\left(\frac{\partial T}{\partial t}\right)$ at each model level as well as net shortwave and longwave radiative fluxes at both the surface and the TOA. This heating is composed of moist heating in the ~~CRM~~SP and the GCM-grid-averaged SP radiative heating ~~from the CRM radiation module. Since we use the real geography, we also~~. Also, it is important to include direct and diffuse downwelling solar radiation fluxes as output ~~the NN-Parameterization~~variables to force the coupled land surface model~~, which is critical to improve the performance of the NNCAM~~. Specifically, they are solar downward visible direct to surface (SOLS), solar downward near infrared direct to surface (SOLL), solar downward visible diffuse to surface (SOLSD), and solar downward near infrared diffuse to surface (SOLLD). ~~Those downwelling solar radiation fluxes with separation of direct versus diffusion are introduced for different land cover types and processes in the land surface model (Mooers et al., 2021). The~~In the end, the precipitation is derived from column integration of predicted moisture tendency to keep basic water conservation.

The large-scale forcings are commonly not included in previous studies with aqua-planet configuration. However, under realistic configuration, such forcings are composed of the dynamics and the planetary boundary layer diffusion, thereby carrying critical information about the complex background circulations and surface condition. Similarly, those downwelling solar radiation fluxes with separation of direct versus diffusion records the received solar energy by the coupled surface model

with different land cover types and processes (Mooers et al., 2021). If not included, the land surface is not heated up by the sun, therefore, seriously weaking the sea and land breeze and monsoon circulations.

Table 1 lists the input and output variables and their normalization factors. There are 30 model levels for each profile variables. Therefore, the input vector consists of 122 elements for 4 profile variables and 2 scalars, while the 68-element output vector is made of 2 profiles and 8 scalars. All input and output variables are normalized to ensure that they are in the same magnitude before they are put into the NN-parameterization for training, testing, model prognostic validation. The normalization factor for each variable shown in the supplemented codebase is determined by the maximum of its absolute values.

~~In this study, the target model SPCAM is run for 3 years~~The training dataset used by all considered NNs is 40% temporally random sampled from the 2-year SPCAM simulation from January 1, 1997 to December 31, ~~1999 with a time step of 30 min. The first year of SPCAM simulations~~1998. Notably, random sampling is ~~for spinup,~~only done in the ~~second~~time dimension but not in latitude and ~~third years are for training, with a total of 484 million~~longitude, including all 13,824 samples~~. In~~ from global grid points for each selected time step. To avoid any mix or temporal connection between the training ~~process, 64% of the data are randomly selected for training~~ set and ~~16% is for~~offline validation ~~in every training epoch to monitor the performance and to prevent overfitting, and the remaining 20% of the data are used for testing.~~

~~2.2.2 Multi-Target ResMLP~~

~~In this study, the NN Parameterization is an isolated column subgrid parameterization. This means that the inputs and outputs of NN-Parameterization are both 1-D vectors. Compared with machine learning tasks such as computer vision and natural language processing, the data input and output of NN-Parameterization are relatively simpler. Therefore~~set, we ~~choose the commonly used multilayer perceptron (MLP) to achieve a better generalization.~~ random sample 40% timesteps from the SPCAM simulation in the year 2000 as offline validation set in the sensitivity test.

2.2.2 A ResDNN Set

In the development of the NN ~~parameterization -~~Parameterization scheme, it is found that ~~the NN has a~~when different ~~fitting capability for~~variables are used as the output of the neural network, the training difficulty is quite different ~~types of output variables. It usually has higher accuracy in predicting radiative~~. Especially, the neural network's ability to fit the radiation heating and scalar fluxes ~~and temperature tendency, while the accuracy is lower in predicting moisture tendency~~is significantly stronger than the tendencies variables. This is also found in Gentine et al. (2018), in which the coefficient of determination (R^2) of radiative heating tendency is higher than that of moisture tendency at most model levels. ~~The physical processes behind the data representing convection and radiation are different, introduce high nonlinearity in the NN-Parameterization.~~

~~The distribution of different physical variables also varies greatly. Using~~We believe that using a single NN with one target to train ~~the relatively independent physical~~all variables, i.e., moisture tendency, temperature tendency, and radiation fluxes,

inevitably causes mutual interference. Since gradient descending is applied to optimize the network in training, mutual interference between different targets is expected to cause the cancel out of gradient directions used for descending (Crawshaw et al., 2020; Zhang ~~et al~~and Yang., 2021) and ultimately affect the convergence of the network.

~~As shown in Figure 2, we divide the single network prediction target, a single stacked 1-D output, into multiple targets, and~~ We use ~~multiple NNs~~different neural networks to train ~~these targets separately~~the tendency of moisture and temperature, and radiation fluxes, respectively. By doing so, we avoid the gradient cancellation between multiple targets and improve the convergence speed and fitting accuracy when training the network. As described in Section 3.1, when using the same network configuration, radiation fluxes are trained much easier with higher accuracy than tendencies of moisture and temperature. We admit that putting heating and moistening rate in two different NNs arbitrarily cut physical connections between them. But this separation is surely doing training more easily in the developing stage.

~~After dividing multiple targets, the learning target of NN-Parameterization still contains a lot of nonlinear information. This research attempts to construct a deeper neural network so that NNs can have stronger nonlinear fitting abilities. As the depth of a fully connected neural network increases, its performance becomes saturated and begins to decline when the depth increases further (He~~In this study, to mimic the column-independent SP and its radiation effects, the input and output of NN-Parameterization have to be both 1-D vectors. This means that the data input and output of NN-Parameterization are much simpler than those in the existing mainstream machine learning problems, such as image recognition and text-speech recognition, so it is impossible to apply most of the existing complex neural networks directly. Taking the convolutional neural network CNN as an example, the study of Albawi et al. (2017) shows that CNN has more advantages than DNN in the learning of large-scale images. The problem we face is that the input is a 122-dimensional vector stitched by multiple different physical quantities with only 4 30-element 1D profiles plus 2 scalars, which cannot meet the requirements of "large-scale" (generally at least 32×32 two-dimensional images). So, there is no need to use CNN. Hornik et al. (1989) proved that a single-layer neural network can approximate any function. Although the problem that NN-Parametrization needs to deal with is highly nonlinear, from the point of view of machine learning, it is essentially a mapping problem from a 122-element 1D-vector to a 1D-vector with a length of 68. According to the universal approximation theorem, DNN is feasible. Therefore, when constructing NN-Parametrization, we first tried to use DNN for fitting, and introduced residual connections to extend DNN to ResDNN.

After numerous experiments, we got the best hyperparameters of DNN and ResDNN. When training a Fully connected DNN, the hidden layer width of the network should be set to 512, and the network depth should not exceed 7, otherwise it will affect the convergence of the DNN. In order to make the neural network capture more non-linear information, enhance the fitting ability. We introduce skip connections to extend the 7-layer DNN to 14-layer ResDNN. The network structure of ResDNN is shown in Figure 1. In the training process, both DNN and ResDNN use an initial learning rate of 0.001 and a learning rate decaying strategy as cosine annealing (Loshchilov et al., 2016) without dropout and L2 regularization. Adam (Kingma and Ba, 2014) is chosen as the optimizer to minimize the mean squared errors (MSEs). The results in Figure 2 show that ResDNN can fit data is significantly better than DNN, with details described in Section 3.1. At the same time, the sensitivity tests in section 3 also prove that no DNN model can ensure the stable simulation of NNCAM. So, we chose

ResDNNs sets as stable candidates to build NN-Parameterization. After obtaining all well-fit ResDNN sets, the next step is to couple the candidates into NNCAM one by one for prognostic tests and find sets that can support stable simulation. To complete this extremely challenging task, we have more than 50 prognostic tests. All experiments and analyses on stability will be introduced in section 3 as well.

~~et al., 2016).~~ ~~For this reason, we use residual blocks to solve the problem of network degradation. By testing different depths and widths, we finally determined the network architecture for the optimal performance. As shown in Figure 3, a total of 7 residual blocks are used to form a deep neural network, and each residual block includes two 512 node-wide dense layers connected with a layer jump. Therefore, the deep neural network that has 14 layers with 3.5 million parameters is named ResMLP hereafter. As shown in Figure 4, under multi-target training with the same hyperparameters, ResMLP's fitting accuracy (R^2) for different output variables is better than MLP, achieving a balanced prediction result, that is, the R^2 of each variable is as close to 1 as possible, especially tendencies of moisture and temperature.~~

~~Ott et al. (2020) found that on aqua-planet, the fitting accuracy of an NN-Parameterization is positively correlated with the stability of prognostic validation. In the real-world prognostic validation, we also find that the fitting accuracy of NN-Parameterization has an important impact on stability. When NN-Parameterization is under-fitting, NNCAM can only run for a few days before it crashes. To obtain a feasible NN-Parameterization, a well-fit is necessary. As shown in Table 2, after multiple trials, we determine the hyperparameter configuration of ResMLP for high fitting accuracy. It can support NNCAM to run stably for at least several months.~~

~~However, we find that a high R^2 may not guarantee the stability of real-world prognostic validation. We have trained several groups of ResMLPs with doubled samples and epochs, having their R^2 increased by 1% to 2%. Surprisingly, NNCAM using these ResMLPs crashed even earlier than before. As a result, high fitting accuracy is necessary but not sufficient for both high performance and long-term simulations. In this work, we proposed a trial-and-error method to effectively find the optimal neural network that can guarantee both multi-year and high performance simulations. Firstly, we used the hyperparameter configuration of Table 2 as a baseline and prepared 50 groups of ResMLPs with similar R^2 as candidate models using different train samples and epochs. Secondly, we conducted comprehensive prognostic tests on these candidate neural networks and obtained the feasible NN-Parameterization schemes that can support NNCAM's stable simulation for multiple years. To our knowledge, running the relatively shorter NNCAM simulation is enough to screen out the feasible networks for stable simulations since we found that the ResMLP groups that cannot support long-term integration made NNCAM to collapse within half a year of simulation; and the feasible groups, on the other hand, coupled stably in NNCAM for the first half a year and showed no sign of crashes in the 10-year prognostic simulation. Finally, the optimal NN-Parameterization was selected for the best performance ResMLP group among the feasible candidats.~~

2.2.3 Implementation of NN-Parameterization

~~After being fully trained with CRM data, the~~The NN-Parameterization is implemented into SPCAM to replace both the CRM ~~and CRM~~based superparameterization and its radiation effects on the basis of coarse grid average. In the beginning of each

timestep, NNCAM calls the NN-Parameterization and predict the moisture tendency $\left(\frac{\partial q_v}{\partial t}\right)$, the temperature tendency $\left(\frac{\partial T}{\partial t}\right)$ and radiation fluxes. Then the DNN predictions are returned to NNCAM, updating the model states and fluxes. Additionally, the surface total precipitation is derived from column integration of the predicted moisture tendency. The near-surface conditions of the atmosphere and downwelling radiation fluxes are transferred to the land surface model. After the coupling of the land surface model and the prescribed SST, the host CAM5 performs the planetary boundary layer diffusion and let its dynamic core complete a timestep integration (Figure 1). In the next timestep, the dynamic core returns the new model states to the NN-Parameterization as inputs again. During the whole process, NN-Parameterization and GCM will constantly update each other's status. How to couple the NN Parameterization ~~(DNN)~~ with GCM and run efficiently and effectively is the key ~~of~~to the implementation of NNCAM. To solve these problems, we develop the ~~DNN~~NN-GCM coupler that integrates ~~DNN~~NNs into NNCAM, which will be introduced in the following section.

2.3 The ~~DNN~~NN-GCM Coupler

Deep learning research mainly uses machine learning frameworks based on Python interfaces to train neural network models and deploy them through C++ or Python programs. While GCM is mainly developed in Fortran, it is a very challenging work to call a neural network model based on Python/C++ interface in GCM codes written in Fortran. Solving the problem of code compatibility between ~~DNN~~NN and GCM can significantly help develop ~~DNN~~NN based Parameterizations for climate models.

To implement a ~~DNN~~NN based Parameterization in current climate model which is mostly developed in Fortran, many researchers try to get the network parameters (e.g., weight, bias) from the machine learning models and implement the ~~DNN~~NN models (e.g., DNNs) with hard coding in Fortran. At runtime, NNCAM will call ~~DNN~~a NN parametrization as a function (Rasp et al., 2018; Brenowitz and Bretherton, 2019). Recently, some researchers have developed a Fortran-neural network interface that can be used to deploy DNNs into GCMs (Ott et al., 2020). This interface can import neural network parameters from outside of Fortran program, and the Fortran-based implementation ensures that it can be flexibly deployed in GCMs. However, embedding ~~DNN~~a NN parameterization in NNCAM is still a troublesome task with no existing coupling framework to support many of the latest network structures. This problem will restrict developers from building more powerful ~~DNNs~~NNs and deploying them in NNCAM.

~~In this research, we regard NN-Parameterization as a component model coupled to NNCAM.~~ We develop the coupler to bridge NN-Parameterization with the host CAM5. Through this coupler, the neural network can communicate with the dynamic core and other physical schemes in NNCAM in each time step. When NNCAM is running, as shown in ① in Figure 5, the coupler receives the state and forcing output from dynamic core in Fortran based CAM5. For each input variable, we use the native MPI interface in CAM5 to gather the data of all processes to the master process into a tensor. Then, as shown in ② of Figure 5, the coupler will transmit the gathered tensor through the data buffer to the NN-Parameterization running on the same node as the master process. The NN-Parameterization gets the input, infers the outputs, and transmits them back to the coupler. As shown in ③ of Figure 5, the coupler will first write these tendencies and radiation fluxes back to the master process and

then broadcast the data to CAM5 processes running on the computing nodes through the MPI transmission interface. Therefore, other parameterizations get the predictions from NN-Parameterization to complete the follow-up procedures (④ in Figure 5).

In practice, the ~~DNN~~NN-GCM Coupler introduces a data buffer that supports system-level interface, which is accessible by both Fortran based GCM and Python based ~~DNN~~NN without supplementary foreign codes. This can avoid code compatibility issues when building Machine Learning coupled numerical models. It supports all mainstream machine learning frameworks, including native PyTorch and ~~Tensorflow~~TensorFlow. Based on the coupler, one can efficiently and flexibly deploy the Deep Learning Model in NNCAM, and can even take advantage of the latest developed neural networks.

All neural network models deployed through ~~DNN~~NN-GCM Coupler can support GPU accelerated inference to achieve excellent computing performance. In this study, we ran SPCAM and NNCAM on ~~196~~192 CPU cores. NNCAM also used 2 GPUs for acceleration. During the NNCAM runtime, each time step of NNCAM requires NN-Parameterization to complete an inference and conduct data communication with NNCAM. This is a typical high-frequency communication scenario. We evaluated the amount of data (about 20MB for CAM5 with the horizontal resolution of 1.9°×2.5°) that needs to be transmitted for each communication, and determined to establish a data buffer on a high-speed solid-state drive to ensure a balance of performance and compatibility. It takes about $1x10^{-2}$ seconds to access the data buffer in each time step, which is enough to support the efficient simulation of NNCAM. The Simulation Years per Day (SYPD) of NNCAM based on ~~DNN~~NN-GCM Coupler has ~~a~~an impressive performance improvement, when using 192 Intel CPU cores, the SYPD of ~~nearly 30x compared to~~ SPCAM~~, and reaches half~~ is 0.3, the ~~speed~~SYPD of CAM5 is 20, and the SYPD of NNCAM is 10. It is worth noting that, NNCAM based on NN-GCM coupler uses an additional GPU to accelerate NN-Parameterization. When NN-GCM Coupler is not used, NN-Parameterization is implemented by Fortran and accelerated by Fortran-based Math Kernel Library, the SYPD is 1.5.

3 A Road to Stability

3.1 Sensitivity Tests and Trial-and-error

To develop a stable NN parameterization, we propose a ResDNN set, where each neural network is responsible for predicting a class of variables (see section 2.2.2). One may wonder whether the ResDNN architecture is necessary and whether offline accuracy of NNs matters in online stability. This section tries to deal with the questions via a series of sensitivity tests.

To prove the necessity of the ResDNN architecture, we use the 7-layer DNN as the control group. We do not include other types of ML architecture, since random forest is less likely to perform as accurately as neural networks and cannot be implemented in GPUs (Yuval et al., 2021), and 1D CNN is not widely used in other studies except Han et al. (2020) with unknown prognostic performance.

The prognostic tests of NN parameterization begin at 1998-01-01 as a startup. As initialization, calling the SP in SPCAM at the first step is required to generate the correct largescale forcings as the input for NN parameterizations. In the sensitivity test, we freeze the ResDNN for the 8 radiation fluxes to simplify the neural network choices, since their offline validation is

extremely accurate with R^2 above 0.98 over 50 training epochs (Figure 2b). Different from the accurately trained radiation fluxes, the tendencies of temperature and moisture are less accurate and can hypothetically affect the prognostic performance. To evaluate the tendency of moistening and heating in one metric, we introduce the MSE of moist static energy changing rate ($dh = C_p dT + L_v dq_v$) as:

$$MSE_h = \left\| \frac{1}{g}(dh_{NN} - dh_{SPCAM})\Delta p \right\|_2, \quad (1)$$

where g is the gravity constant, C_p refers to the heat capacity of air, L_v is the latent heat of water vapor, and Δp is the layer thickness. Multiple ResDNN pairs for dq_v and dT and DNN pairs are trained from 5 epochs to 50 epochs, carrying different offline validation accuracy.

Figure 4 shows the offline validation MSE_h versus the prognostic steps. First, DNN parameterizations (blue triangles) are systematic less accurate than ResDNN ones (blue dots and black inverted triangles), which is consistent with Figure 2a. They cannot run stably in prognostic tests with the best DNN parameterization to sustain half a year of simulation. For the ResDNNs, the less well-trained ones with high MSE crash for a shorter simulation period than DNNs. However, when the offline MSE decreases to a certain level (e.g., 290 W^2/m^4), some of the ResDNN parameterizations are stable for extreme long-term simulations, while others remain unstable.

Generally, a NN parameterization that can support long-term integration should have both good generalization abilities and high accuracy for training and validation. Above all, sufficient accuracy is necessary for all neural networks. From Figure 4, it can be interpreted that a vague threshold exists in the validation MSE. ResDNNs can be trained for higher accuracy since they are much deeper than DNNs with much higher model capacity. So, they are more competent than DNNs in this job. On the other hand, studies showed that high-capacity models are harder to train and more likely to overfit (Goodfellow et al., 2016). ~~3~~Thus, the prognostic stability differences between less well-trained ResDNNs and the well-trained ones are drastic compared with DNNs. Also, some overly trained ResDNNs with lowest validation loss are speculated to overfit. Those overfitting models are less likely to generalize to unknown backgrounds caused by accumulated errors in the ML-GCM system, ending up model crashes. However, those are just intuitive experiences but not guarantee ways for stability.

In the time evolution of the global averaged total energy (Figure 5). The system energy grows exponentially and then blows up for unstable ResDNN parameterizations (the red and orange lines). In contrast, the stable ones can keep the total energy at a certain level and reproduce the annual cycle fluctuations in SPCAM. Among the stable ResDNN schemes, some can get nearly a perfect reproduction of the total energy evolution of SPCAM (the blue line), while some inaccurately simulate the climate state with a large deviation (green line). Therefore, among the accurate ResDNN parameterizations (e.g., offline validation $MSE_h < 290\ W^2/m^4$), we still have to use the trial-and-error to filter out unstable ones and then select the best ResDNN pair for moistening and heating rate that can reduplicate the total energy time evolution of SPCAM with the least deviation. We name this best ResDNN pair together with the ResDNN in charge of radiation fluxes the NN-Parameterization. This NN-GCM coupled model is called NNCAM and is later evaluated for climate mean states and variability.

3.2 Gravity Wave Diagnosis

It is still a question of why unstable NN parameterizations blow up models. The fast-growing energy of the unstable runs indicates a possible underlying unrealistic energy amplifying mechanism in the NN-GCM coupled system. Brenowitz et al. (2020) offered interpretations. When an unstable NN parameterization is coupled with dynamics, it tends to amplify any unrealistic perturbation caused by emulation errors and pass it to the entire system through gravity waves. In contrast, the stable NN parameterizations tend to dump all the perturbs quickly. This is true in our study with realistic configuration. Such unstable gravity waves are observed in the prognostic simulation of an unstable ResDNN (the red line in Figure 5). The animation in Movie S1 records the first unrealistic wave and Movie S2 documents more intense waves afterward with a perfectly round shape. Brenowitz et al. (2020) also introduced an analysis tool that calculates wave energy spectra of a hierarchy model that couples the linear response functions (LRF) of a NN parameterization to a simplified two-dimensional linear dynamic system, where perturbations can propagate in 2D gravity waves. We apply the tool in our study and detect similar results of unstable mode for the unstable ResDNN with positive energy growth rate across all wave numbers at phase speed between 5 m/s to 20 m/s (Figure S1b). While the stable ResDNN shows a stable mode with the growth rate of nearly all wave numbers and phases below zero (Figure S1a).

4 Offline Validation of NN-Parameterization

~~To assess~~Before evaluating the prognostic results, demonstration of offline performance with geographic information is needed for the following purposes: 1) To show how well ~~the NN-Parameterization learns the subgrid tendencies from the CRM and its effects on radiation in SPCAM, we~~our NN-Parameterization emulates the SP in realistic configuration compared with baseline CAM5 physics and with previous studies. 2) To reveal the strength and weakness of NN emulations with correct input, give clues to the analysis of prognostic results in the following section. We performed offline testing with a realistically configurated SPCAM from January ~~1997~~1st 1999 to December ~~1998~~31st 2000, where NN-Parameterization is diagnostically run paralleled to the ~~CRM in SPCAM.~~SP, and so does the CAM5 physics. The results over the entire second year of ~~1998~~the period are chosen for evaluation~~. As suggested~~, completely independent from the training dataset. Following the conventions in Han et al. (2020) and Mooers et al. ~~(2021), faced with the diverse realistic boundary condition, it is necessary to conduct such evaluation before prognostic experiments. We~~(2021), we choose mean fields and coefficient of determination (R^2) as the two metrics ~~in the offline testing~~for evaluations.

The mean diabatic heating and drying rates produced by convection and large-scale condensation in SPCAM and NN-Parameterization are in close agreement. Figure 6 shows the latitude-height cross-sections of the annual mean heating and moistening rates in SPCAM and the corresponding NN-Parameterization. At 5 °N, SPCAM shows maximum latent heating in the deep troposphere, corresponding to deep convection at the ITCZ. In the subtropics, there is heating and moistening in the lower troposphere, corresponding to stratocumulus and shallow convection in the subtropics. In the midlatitudes, there is a secondary heating maximum below 400 hPa due to midlatitude storm tracks. All these features are well reproduced by NN-

Parameterization. Note that in the ~~midtroposphre~~midtroposphere, the ITCZ peak in the drying rates is slightly weaker in NN-Parameterization compared with that of SPCAM (Figure 6c and 6d).

In addition to the mean fields, the high prediction skill of NN-Parameterization is also shown in the spatial distribution of R^2. ~~We choose pressure-latitude cross-sections to better~~To demonstrate R^2 for the 3D variables such as diabatic heating and moistening~~. Therefore,~~, same as Mooers et al. (2020), zonal averages are calculated in advance~~.~~ before R^2 calculation for each location in the pressure-latitude cross-section. For diabatic heating, R^2 is above 0.7 over the entire mid to low troposphere~~,~~ and the high skill regions with R^2 greater than 0.9 concentrates in low levels but are extended to mid-troposphere in storm tracks (Figure 7a). As for the moistening rate, the high skill zones concentrate in the mid to upper troposphere (Figure 7b), leaving low skill areas below. Those regions with low accuracy are generally ~~locate~~located in the mid to low troposphere in tropics and subtropics, corresponding to deep convection at ITCZ and shallow convection in subtropics. Nonetheless, the tendencies from diagnostic CAM5 parameterization hardly draw any similarity to those simulated by the SP except for a few locations in the mid to upper troposphere in tropics and polar regions (Figure 7c & 7d).

The global distribution of R^2 for the ~~derived~~ precipitation predictions is shown in Figure 8. Our NN-Parameterization shows a great prediction skill globally, especially in the midlatitude storm tracks. The prediction skill is relatively low in many areas between 30°S to 30°N and some midlatitude continents~~. In~~ (Figure 8a), in particular,~~ the prediction skill of precipitation is~~ not ideal in the ITCZ deep convection regions. Moreover, for shallow convection in Subtropical Eastern Pacific and Subtropical Eastern Atlantic, the precipitation prediction skill hits bottom, corresponding to the subtropical low skill zones for moistening rate (Figure ~~7b).~~6b). On the other hand, the total precipitation simulated by CAM5 parameterizations is much less analogous to the SP than NN-Parameterization with a systematically lower accuracy globally. CAM5 precipitation can reach a relatively high accuracy along the mid-latitude storm tracks but fail most regions in the tropics.

Generally, NN-Parameterization ~~shows high performance~~ performs far better than CAM5 parameterization in the 1-year offline testing ~~regarding mean fields and fitting R^2. As suggested in previous studies, manually tunned fully-connected neural networks often fail~~ and shows similar accuracy as the ~~mission~~DNN in Mooers et al. (2020). The real-geography data can significantly decrease the emulation skill of ~~fitting variables in realistically configurated simulations~~a deep learning model (Mooers et al., 2021~~). NN-Parameterization succeeds in fitting diabatic heating and moistening rate, which suggests that, with a multi-target framework and the implementation of residual blocks, a well-designed DNN has a great potential~~), where the convection backgrounds are much more complex with meridional and zonal asymmetric and seasonal variated circulations, not to mention the orograph and various types of underlying land surface. In that case, the ResDNN is a valuable NN architecture that can bring good performance as the automated hyperparameter tuning algorithm without searching for hundreds of ~~serving as a replacement of convection and cloud parameterization. Low accuracy of DNN prediction~~NN candidates. Still, our NN-Parameterization is exposed to low accuracy predictions in subtropical shallow convection areas~~-is~~, a great challenge for machine learning emulation of moistening rate and precipitation. ~~Similar results are noted in previous studies (Gentine et al., 2018 & Mooers et al., 2021). However, even with this shortcoming, NN-Parameterization still manages to carry in multi-year prognostic simulations with reasonable results shown in the coming Section 4.~~ In those regions, the local

variance/std is close to zero. But the NNs in our study are trained in the loss function of mean squared error, which is not sensitive to small values.

45 Long-term Prognostic Validation

As described in Section 2.2, NN-Parameterization is selected for best prognostic performance in Section 3.1. It is coupled in NNCAMthe realistic configurated SPCAM to replace the conventional deep and shallow convection, microphysics and macrophysics, and SP and its cloud-scale radiation parameterizations. At the same time, the planetary boundary layer diffusion, as well as the dynamic core are still kept in the host GCM. Starting with the SPCAM checkpoint on July 1,effects. This coupled model is called NNCAM afterwards and is compared with SPCAM and CAM5. All three model starts at January 1st 1998, NNCAM is as start up. They are all run for 96 years with the first year for spin up and a half to December 31, 2007. After a spinup for half a year, we choose the next four5 years from January 1,1st 1999 to December 31, 200231st 2003 for evaluation. Although long-term stable run of NNCAM is achieved, a very slow climate drift is still inevitable. Thus, we do not put the last and comparison. Later, the simulation of NNCAM is extended for another 5 years into analysis. In addition, as the referencing target model, SPCAM is only run from January 1, 1997 to December 31, 1999 dueto December 31st 2008 to show its stability. Due to excessive computing resources consumption. We gather the results of the last two years to compare with NNCAM simulations. CAM5, as, the coarse-grided conventional parameterized control model, is run from January 1, 1998 to December 31, 2002. Similarly, the first year is for spinup and the last four years for analysis.simulation of SPCAM is not get extended. In analysis of prognostic results, the following are selected for demonstration of climatology and variability: multi-year mean fields of temperature and humidity, precipitation, precipitation frequency distribution, and the Madden Julian Oscillation.

45.1 Climatology

5.1.1 Vertical profiles of temperature and humidity

In this section, we first evaluate the vertical structure of the mean temperature and humidity. Figure 9 presentsshows the zonally averaged vertical profiles of air temperature and specific humidity as simulated by the NNCAM and the CAM5, in contrast to the SPCAM simulations. Overall, the NNCAM does an excellent job in reproducing thesimulate reasonable thermal and moisture structure. The resulting However, it is shown that NNCAM has some biases in mean state infields of temperature and specific humidity of NNCAM closely resembles SPCAM throughout the troposphere., which is shown as larger root mean squared errors (RMSEs) or larger differences than CAM5 (Figure S2). The larger deviations are temperature biases in the tropopause, where the cold-point region is thinner and warmer in NNCAM than in SPCAM and CAM5. In addition, there are cold biases above 200 hPa and warm biases blow over polar regions and in NNCAM. For the humidity field, there are slight dry biases over the equator in NNCAMand wet biases elsewhere in NNCAM. Even with the biases, the climate mean states are consistent with those in the last 5-year simulation for NNCAM (Figure S3), which indicates almost no climate drift in the long-term simulation.

45.1.2 Precipitation

Figure 10 shows the spatial distributions of winter (December-January-February) and summer (June-July-August) mean precipitation simulated by SPCAM, NNCAM, and CAM5~~, with~~. The SPCAM simulation results are regarded as reference precipitation. In SPCAM (Figure 10a and 10b), massive precipitation can be found in regions of Asian monsoon and midlatitude storm tracks over the northwest Pacific and Atlantic oceans. In the tropics, the primary peaks of rainfall are in the eastern Indian Ocean and Maritime Continent regions. Furthermore, two zonal precipitation bands are located at 0°–10°N in the equatorial Pacific and Atlantic oceans, constituting the northern ITCZ. The southern South Pacific Convergence Zone (SPCZ) is mainly located around 5°S–10°S near the western Pacific warm pool region and experiences a southeast tilt as it extends eastward into the central Pacific. The main spatial patterns of SPCAM precipitation climatology are properly reproduced by both NNCAM and CAM5. In NNCAM, strong rainfall centers are well simulated over the tropical land regions over Maritime Continent, the Asian monsoon region, and South America and Africa (Figure 10c and 10d). In addition, the heavy summertime precipitation over the ~~northwestern~~Northwestern Pacific simulated by SPCAM is well represented in ~~both SPCAM and~~ NNCAM (Figure ~~10b~~10a and ~~10d~~10c). In CAM5, there is too little precipitation over that area (Figure ~~10f), which is a common bias in many GCMs. However, the equatorial region is too dry in~~10e). Moreover, NNCAM~~, especially over sea surface area in boreal winter~~ can maintain the spatial pattern and global average of precipitation in the next 5-year simulation, reassuring its long-term stability (Figure ~~10e).~~ S4).

~~Figure 11 shows the annually averaged zonal mean precipitation from SPCAM, NNCAM, and CAM5. NNCAM generally reproduces very similar latitudinal variations to that of SPCAM, but with weaker precipitation intensity near the equator. Precipitation of both NNCAM and CAM5 agrees with SPCAM in global annual averages (Figure 11a). NNCAM results even come closer with SPCAM than CAM5 in boreal summer (Figure 11f). In ANN and DJF average, precipitation of NNCAM is underestimated over tropical continents but overestimated in subtropical land regions compared with SPCAM targets (Figure 11d, e). When the ocean areas are included, the SPCZ in NNCAM is excessively separated from the ITCZ in boreal winter. Its precipitation center south shifted, resulting in a minimum zone of equatorial precipitation (Figure 11b). NNCAM shows moderate precipitation prediction biases in some latitudes in the tropics which are corresponding to the low skill regions of tropical and subtropical moistening rate and rainfall in Section 3. In our speculation, the weaker drying tendencies of the ITCZ midtroposphre from the NN parameterization causes separated convergence zones, leading to underestimation of equatorial rainfall rate.~~

Generally, NNCAM draws more similarity to SPCAM than CAM5 in spatial distribution of summertime multiyear precipitation with smaller RMSE and global averaged biases. However, in the difference plot (Figure 11), NNCAM moderately underestimates precipitation along the equator, Indian monsoon region, and maritime continent in summer (Figure 11a). In boreal winter, NNCAM simulates a weak and excessively separated SPCZ from ITCZ, with both precipitation centers shifting away from each other. As a result, we detect underestimation in the equatorial regions of the maritime continent as well as the SPCZ but overestimation on the north of the equator in the West Pacific (Figure 11b), which makes NNCAM less resemble

SPCAM than CAM5 in this season. This simulation biases in NNCAM are speculated linked to the weaker drying tendencies of the ITCZ midtroposphere from the NN parameterization and low accuracy of NNCAM predictions in tropics.

5.2 Variability

5.2.1 Frequency Distribution of Precipitation

Moreover, NNCAM shows better performance in simulating precipitation extremes. Figure 12 shows the probability densities function of simulated daily precipitation in the tropics (30°S−30°N) with a precipitation intensity interval of 1 mm day^{-1}. In CAM5, ~~the frequency of light rainfall events with values smaller than 1 mm day^{-1} is lower than that in SPCAM, and~~ heavy precipitation events exceeding 20 mm day^{-1} are ~~insufficient.~~greatly underestimated. In addition, light to moderate precipitation events between 2−20 mm day^{-1} are overestimated with an unreal probability peak around 10 mm day^{-1} in CAM5, which is a typical simulation bias found in simulations with parameterized convection but not in explicitly resolved convections (Holloway et al., 2012). Compared with CAM5, the spectral distribution of precipitation in NNCAM is much closer to SPCAM. ~~Both light and~~ heavy rainfall events are substantially enhanced, and the overestimated precipitation occurrence between 2−20 mm day^{-1} is reduced~~. In addition, NNCAM avoids the unreal probability peak around 10 mm day^{-1} appeared in CAM5, which is a common simulation bias found in simulations with parameterized convection but not in explicitly resolved convections (Holloway et al., 2012).~~ with no spurious peak around 10 mm day^{-1}.

~~4.3~~5.2.2 The MJO

The MJO is a crucial tropical intraseasonal variability at the time scale of 20–100 days (Wheeler and Kiladis, 1999). Figure 13 presents the wavenumber and frequency spectra for equatorial precipitation daily anomalies from SPCAM, NNCAM, and CAM5 in 4 consecutive boreal winter from 1999 to 2003. SPCAM shows ~~a concentration of~~widespread power signals over zonal number of 1-4 and periods between 20-100 plus a peak around at zonal numbers of 1−3 and periods of ~~30−40~~70-100-day for eastward propagation (Figure 13a). ~~In~~Similarly, in NNCAM, there is a spectral peak at the wavenumbers of 1−2 and ~~longer~~ periods of ~~greater than~~ 50-80 ~~days (Figure 13b). There is also a second spectral peak with comparable wavenumbers at 50−80~~ day ~~periods~~ for ~~eastward~~east propagation~~,~~ (Figure 13b), exhibiting intense intraseasonal signals. For CAM5 (Figure 13c), the spectral power is concentrated around 30-day and more extended periods (greater than 80 days) at wavenumber 1 for eastward propagation. In addition, CAM5 also shows signals of westward propagation around 30-day period. Compared with CAM5, NNCAM shows stronger intraseasonal power and resembles SPCAM better. To quantify this similarity, we calculate the coefficient of determination R^2 of the precipitation spectrum in NNCAM and CAM5, using the spectrum in SPCAM as the target value. The precipitation spectrum R^2 in NNCAM (0.51) is much higher than that in CAM5 (0.40).

The MJO is characterized by the eastward propagation of deep convective structures ~~with an average phase speed of around 5 m s^{-1}~~ along the equator. Generally, it generally forms over the Indian Ocean, strengthens over the Pacific, and weakens in the eastern Pacific due to interaction with cooler SSTs (Madden and Julian, ~~1971~~1972). Figure 14 presents the longitude-

time lag evolution of 10°S−10°N meridional averaged daily anomalies of intraseasonal (filtered with 20-100 day bandpass) precipitation ~~anomalies.~~and 200 hPa zonal wind (U200) in boreal winter. The results show that both SPCAM and NNCAM ~~and CAM5~~ reasonably reproduce the eastward propagating convection from the Indian Ocean across the Maritime Continent to the Pacific (Figure ~~14b~~14a and ~~14c).~~14b), confirmed by both precipitation field and U200 field. Therefore, we conclude that NNCAM captures the key MJO propagation simulated in SPCAM. ~~The average phase speed~~In contrast, the time lag plot of ~~eastward~~CAM5 depicts an unpleasant west propagation. Same as the precipitation spectrum, R^2 of ~~deep convection~~ the time lag coefficient is shown to quantify the resemblance. The time lag coefficient of U200 in NNCAM is much closer to ~~5 m s^{-1}~~SPCAM than ~~the overly fast propagation speed~~CAM5, with a way higher R^2, indicating that the NN-Parameterization successfully emulates the convection variability of the SP and reflects it in ~~CAM5, denoted by the dashed line in Figure 14b and 14c.~~the dynamic fields.

~~5~~ 6 Summary and Conclusions ~~and Discussion~~

This study investigates the potential of ~~DNN~~deep neural network based parameterizations ~~embedded into~~ in SPCAM ~~in reproducing~~to reproduce long-term climatology and climate variability. We present NN-Parameterization, a ~~group of organized ResMLPs~~ResDNN set, to emulate ~~convection,~~the SP with a 2D CRM and its cloud ~~processes and~~ scale radiation in ~~a~~effects in a realistic configurated SPCAM with ~~a realistic global~~true land-ocean distribution and orography. The input variables to the NN-Parameterization include specific humidity, temperature, largescale water vapor and temperature forcings, surface pressure and solar insolation. The output variables of the NN-Parameterization consist of the subgrid tendencies of moisture and temperature ~~as well as radiation fluxes. The output variables are divided into multiple groups, and each group is trained as one target through independent neural network following the same architecture. Both long-term stable and high performance climate simulations are finally obtained in this work. To effectively bridge the host CAM and NN-Parameterization, we have expanded the coupler idea in a earth system model to ensure the flexibility of embedding DNN into GCM. As a result, we can efficiently use DNNs to construct NN-Parameterization that can support NNCAM stable simulation for multi-years. This study is the first attempt to achieve stable climate simulations using a hybrid ML-physical GCM, which is configurated with real land-ocean distributions.~~, net radiation fluxes at the top of the model and surface, and solar radiation fluxes down to the surface. We proposed a set of 14-layer deep residual neural networks in which each NN is in charge of one type of output variable. With such a design, we gain the best emulation accuracy for each predictor. Via a systematic trial-and-error searching procedure, we are able to firstly select sets of ResDNNs that support stable prognostic climate simulations and then choose the best set with lowest climate errors as the formal NN-Parameterization. Moreover, a mechanism of unreal perturbation amplification is found in GCM simulations with unstable NN parameterizations with the spectrum diagnostic tool invented in Brenowitz et al. (2020).

The offline test shows the great skills of the NN-Parameterization in emulating the ~~CRM~~SP outputs and ~~CRM~~its cloud scale radiation effects in SPCAM. The overall diabatic heating and drying rates in the NN-Parameterization and SPCAM are

in close agreement. When implemented in the host ~~coarse-grided CAM5~~ SPCAM to replace ~~the conventional schemes of moist physics~~its time-consuming SP and its radiation effects, the NN-Parameterization successes in an extensive long-term stable prognostic simulation and ~~performs well in reproducing the~~predicts reasonable mean vertical structures in temperature and humidity, and the precipitation distributions. ~~The prominent~~Compared with the SPCAM target simulation, NNCAM still produces some biases in mean fields, such as a warmer troposphere over polar regions and tropopause and strong precipitation underestimation in equatorial regions. On the other hand, the better climate variability in SPCAM over CAM5 is well learned by our NN-Parameterization and reproduced in NNCAM with better frequency in extreme rainfall, similar MJO ~~signal~~spectrum and ~~its phase~~propagation direction and speed ~~are also well captured using our NN~~. Although with the biases in climate states so far, NNCAM can still be regarded as the first attempt to prognostically couple a NN-based parameterization in realistic configurated 3D GCM.

~~Machine~~Many previous studies have well-studied machine learning parameterizations implemented in aqua-planet configurated 3D GCM ~~have been well studied in many previous research.~~. Some faced instability in coupled simulations (Brenowitz ~~&~~and Bretherton, 2019), ~~and some tried to solve such instability through online learning (Rasp, 2020). Some others achieved~~while some succeeded in long-term stable prognostic simulations with deep fully-connected neural networks (Rasp et al., 2018; Yuval et al., 2021) as well as random forest (Yuval ~~&~~and O'Gorman, 2020). In contrast to aqua-planet simulations, the spatial heterogeneity is prominent over land in GCMs which are configurated with real-geography boundary conditions. In this case, a plain fully connected neural networks the ~~conventional MLPs have been unable to fit CRM outputs~~SP output (Mooers et al., 2021). The convection, clouds, and the interacted radiation of the CRM together with real-geography boundary conditions are ~~no doubtfully~~without doubt far more complicated ~~and nonlinear~~ than in idealized models. ~~We first propose a multi-target DNN method to control nonlinearity, reduce gradient directions cancellation during training, and make each output variable reach~~To meet the ~~best fit state. Moreover~~new demand under realistic configuration, we design ~~ResMLP~~ResDNN with sufficient depth to further improve the nonlinear fitting ability of NN-Parameterization. ~~A trial-and-error method is further presented to effectively find~~With the ~~optimal neural network for NN-Parameterization. By doing so~~skip connections, the ~~NN-Parameterization we constructed achieved good fitting accuracy in~~7-layer DNN models can be extended to 14 layers, therefore, significant improving offline ~~testing and successfully carried out~~accuracy. In the prognostic tests, a few ResDNN parameterizations can support long term stable ~~online simulations for multiple years~~run, while all DNN parameterizations are so far found unstable.

Trial-and-error is still the only way to find stable NN parameterizations. So far, we have not come up with an a priori mothed that guaranteed stability. However, we do find some clues in the sensitivity tests. We believe sufficient offline accuracy is essential for online stability by confirming all inaccurate NN parameterizations unstable. On the other hand, some highly accurate ones still crash the prognostic simulation, where we find rapid increasing total energy. This mechanism is that unstable NNs cannot damp neural network emulation errors but amplify and propagate them to the entire system through gravity waves.

The prognostic biases in mean fields in speculated as a result of by the combined effect of the emulation errors of all the NN-Parameterization prediction fields. Further study is required. Still, it can be related to the spatially non-uniform accuracy

of NN-Parameterization, such as relatively low fitting accuracy in tropical deep convective regions and shallow subtropical convection and stratiform cloud regions. Such problems have also been reported in previous studies (Gentine et al., 2018; Mooers et al., 2021). We believe that a NN parameterization with heterogeneous characteristics across different regions, rather than a globally uniform scheme, can further improve the fitting accuracy in this tropical and subtropical region.

Embedding deep neural networks into Fortran based atmospheric models is still a handicap. Before this study, researchers mainly used hard coding to build neural networks (Rasp et al., 2018; Brenowitz and Bretherton, 2019). An easier way is to use Fortran based neural network libraries that can flexibly import network parameters (Ott et al., 2020). These methods have successfully implemented ~~DNN~~NN in GCM, but they can only support dense layer based ~~DNN~~NN. As a result, developers cannot take advantage of the most advanced neural network structures such as convolution, shortcut, self-attention, variational autoencoder, etc., to build powerful DNN based Parameterizations. In this research, through ~~DNN~~NN-GCM Coupler, NN-Parameterization can support the mainstream GPU-enabled machine learning frameworks. Thanks to the simple and effective implementation of the DNN-GCM Coupler, our NNCAM achieves 30 times SYPD compared to SPCAM by using ~~four 14-layer deep ResMLPs~~a ResDNN set in NN-Parameterization, although these DNNs are much deeper than the previous state-of-the-art fully-connected NNs in this field.

~~Different from the existing works, we find that high fitting accuracy of NN-Parameterization is necessary but not sufficient for both high performance and long-term simulations since neural networks with high fitting accuracy may crash before achieving years of stable simulation. The mechanism of the observed early-crashed simulations needs to be further investigated. We obtained the optimal NN-Parameterization that can perform both multi-year and high performance simulations by trial and error as an initial attempt. In addition, we observe that NN-Parameterization has spatially non-uniform accuracy, such as relatively low fitting accuracy in tropical deep convective regions and subtropical shallow convection and stratiform cloud regions. Such problems have also been reported in previous studies (Gentine et al., 2018; Mooers et al., 2021). We believe that a NN parameterization with heterogeneous characteristics across different regions, rather than a globally uniform scheme, can further improve the fitting accuracy in this tropical and subtropical region.~~

Code and data availability. The original training and testing data can be accessed at https://doi.org/10.5281/zenodo.5625616. The source codes of SPCAM version 2 ~~can be accessed at https://svn-ccsm-release.cgd.ucar.edu/model_development_releases/spcam2_0-cesm1_1_1. The training and testing data and the associated codes of the NN parameterization scheme~~and NNCAM have been archived, and made publicly available for downloading from ~~http~~https://doi.org/10.5281/zenodo.~~5153501~~5596273.

Author contributions. XW trained the deep learning model, constructed the DNN-GCM Coupler, performed the NNCAM and CAM5 experiments, and wrote the main part of the paper. YH conducted the SPCAM simulations, offered valuable suggestions on the development of the NN parameterization, and participated in the writing of the paper~~.~~ and revision. WX supervised this work, provided critical comments on this work and participated in the writing of the paper. GJZ provided key points for this

research and participated in the revision of the paper. GWY supported this research and gave important opinions. All the authors discussed the model development and the results.

Competing interests. The authors declare no conflict of interest.

Acknowledgements. This work is partially supported by the National Key R&D Program of China (grant no. 2017YFA0604500)), and the National Natural Science Foundation of China (grant no. 42130603). Dr. Yilun Han is supported by National Key R&D Program of China (grant no. 2017YFA0604000). We thank Prof. Yong Wang for his guidance on SPCAM simulations and valuable discussions on this work. We also thank Prof. Yixiong Lu for providing professional advice on the evaluation of the simulation results of NNCAM.

References

Albawi, S., Mohammed, T. A., and Al-Zawi, S.: Understanding of a convolutional neural network, 2017 International Conference on Engineering and Technology (ICET), 21-23 Aug. 1-6, 10.1109/ICEngTechnol.2017.8308186, 2017.

Bony, S., Stevens, B., Frierson, D. M. W., Jakob, C., Kageyama, M., Pincus, R., Shepherd, T. G., Sherwood, S. C., Siebesma, A. P., Sobel, A. H., Watanabe, M., and Webb, M. J.: Clouds, circulation and climate sensitivity, Nature Geoscience, 8, 261-268, 10.1038/ngeo2398, 2015.

Brenowitz, N. D. and Bretherton, C. S.: Spatially extended tests of a neural network Parameterization trainedExtended Tests of a Neural Network Parametrization Trained by coarse-graining, J. Adv. Model.Coarse-Graining, Journal of Advances in Modeling Earth Syst.,Systems, 11, 2728–2744, doi:10.1029/2019MS001711 2019ms001711, 2019.

Brenowitz, N. D., Beucler, T., Pritchard, M., and Bretherton, C. S.: Interpreting and Stabilizing Machine-Learning Parametrizations of Convection, Journal of the Atmospheric Sciences, 77, 4357-4375, 10.1175/jas-d-20-0082.1, 2020.

Bretherton, C. S., Blossey, P. N., and Stan, C.: Cloud feedbacks on greenhouse warming in the superparameterized climate model SP-CCSM4, J. Adv. Model.Journal of Advances in Modeling Earth Syst.,Systems, 6, 4, 1185–1204, https://doi.org/10.1002/2014MS000355, 2014.

Bony, S., B. Stevens, D. M. W. Frierson, et al.: Clouds, circulation and climate sensitivity, Nature Geoscience, 8, 261-268, 2015.

Cao, G. and Zhang, G. J.: Role of vertical structure Vertical Structure of convective heatingConvective Heating in MJO simulationSimulation in NCAR CAM5, J. .3, Journal of Climate, 30, 7423–7439, doi:10.1175/JCLI-Djcli-d-16-0913.1, 2017.

Crawshaw, Michael. "M.: Multi-task learning with deep neural networks: A survey.", arXiv preprint arXiv:2009.09796 (, 2020).

Emanuel, K. A., David Neelin, J., and Bretherton, C. S.: ~~Crueger, T. and Stevens, B.: The effect of atmospheric radiative heating by clouds on the Madden-Julian Oscillation, J. Adv. Model. Earth Syst., 7, 854–864, doi:10.1002/2015MS000434, 2015.~~

~~Deng, Q., Khouider, B., Majda, A.~~ On large-scale circulations in convecting atmospheres, Quarterly Journal of the Royal Meteorological Society, 120, 1111-1143, 10.1002/qj.49712051902, 1994.

~~J., and Ravindran, A.: Effect of stratiform heating on the planetary scale organization of tropical convection, J. Atmos. Sci., 73, 371–392, doi:10.1175/JAS-D-15-0178.1, 2016.~~

~~Dueben, P. D. and Bauer, P.: Challenges and design choices for global weather and climate models based on machine learning, Geosci. Model Dev., 11, 3999–4009, doi:10.5194/gmd-11-3999-2018, 2018.~~

~~Feng, Z., Leung, L. R., Houze Jr, R. A., Hagos, S., Hardin, J., Yang, Q., ... & Fan, J.: Structure and evolution of mesoscale convective systems: Sensitivity to cloud microphysics in convection-permitting simulations over the United States. Journal of Advances in Modeling Earth Systems, 10(7), 1470-1494., 2018~~

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., and Yacalis, G.: Could machine learning break the convection parameterization deadlock~~?~~?, Geophysical Research Letters, 45, 5742–5751~~, doi:10.1029/2018GL078202~~, 2018.

Gettelman, A., Liu, X., Ghan, S. J., Morrison, H., Park, S., Conley, A. J., Klein, S. A., Boyle, J., Mitchell, D. L., and Li, J.-L. F.: Global simulations of ice nucleation and ice supersaturation with an improved cloud scheme in the Community Atmosphere Model, ~~J. Geophys. Res., 115, D18216, doi:~~Journal of Geophysical Research: Atmospheres, 115, https://doi.org/10.1029/2009JD013797, 2010.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.

Grabowski, W. W.: Coupling ~~cloud processes~~Cloud Processes with the ~~large scale dynamics using~~Large-Scale Dynamics Using the Cloud-Resolving Convection Parameterization (CRCP), ~~J. Atmos. Sci.,~~Journal of the Atmospheric Sciences, 58, 978–997, 10.1175/1520-0469(2001~~.~~)058<0978:Ccpwtl>2.0.Co;2, 2001.

Grabowski, W. W.: An ~~improved framework~~Improved Framework for ~~superparameterization, J. Atmos. Sci.,~~Superparameterization, Journal of the Atmospheric Sciences, 61, 1940-1952, 10.1175/1520-0469(2004~~.~~)061<1940:Aiffs>2.0.Co;2, 2004.

Grabowski, W. W. and Smolarkiewicz, P. K.: CRCP: ~~A~~a Cloud Resolving Convection Parameterization for modeling the tropical convecting atmosphere, Physica D: Nonlinear Phenomena, 133, 171-178, ~~1999.~~https://doi.org/10.1016/S0167-2789(99)00104-9, 1999.

~~He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.~~

Han, Y., Zhang, G. J., Huang, X., and Wang, Y.: A ~~moist physics parameterization based on deep learning, J. Adv. Model.~~Moist Physics Parameterization Based on Deep Learning, Journal of Advances in Modeling Earth ~~Syst.,~~Systems, 12, e2020MS002076, 10.1029/2020ms002076, 2020.

Hannah, W. M., Jones, C. R., Hillman, B. R., et al.: Initial results from the super-parameterized E3SM, J. Adv. Model. Earth Syst., 12, 1, 2020.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep residual learning for image recognition, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, Nevada, June 26 - July 1, 770-778, 2016.

Holloway, C. E., Woolnough, S. J., and Lister, G. M. S.: Precipitation distributions for explicit versus parametrized convection in a large-domain high-resolution tropical case study, Q. J. R. Meteorol. Soc.,Quarterly Journal of the Royal Meteorological Society, 138, 1692-1708, https://doi.org/10.1002/qj.1903, 2012.

Hornik, K., Stinchcombe, M., and White, H.: Multilayer feedforward networks are universal approximators, Neural Networks, 2, 359-366, https://doi.org/10.1016/0893-6080(89)90020-8, 1989.

Jiang, X., Waliser, D. E., Xavier, P. K., Petch, J., Klingaman, N. P., Woolnough, S. J., Guan, B., Bellon, G., Crueger, T., DeMott, C., Hannay, C., Lin, H., Hu, W., Kim, D., Lappen, C.-L., Lu, M.-M., Ma, H.-Y., Miyakawa, T., Ridout, J. A., Schubert, S. D., Scinocca, J., Seo, K.-H., Shindo, E., Song, X., Stan, C., Tseng, W.-L., Hourdin, F., Rio, Wang, W., Wu, T., Wu, X., Wyser, K., Zhang, G. J., and Zhu, H.: Vertical structure and physical processes of the Madden-Julian oscillation: Exploring key model physics in climate simulations, Journal of Geophysical Research: Atmospheres, 120, 4718-4748, 10.1002/2014jd022375, 2015.

Khairoutdinov, M., Randall, D., and DeMott, C.: Simulations of the Atmospheric General Circulation Using a Cloud-Resolving Model as a Superparameterization of Physical Processes, Journal of the Atmospheric Sciences, 62, 2136-2154, 10.1175/jas3453.1, 2005.

Khairoutdinov, M. F.C., Grandpeix, J.-Y., Madeleine, J. B., Cheruy, F., Rochetin, N., et al.: LMDZ6A: The atmospheric component of the IPSL climate model with improved and better tuned physics, J. Adv. Model. Earth Syst., 12, e2019MS001892, doi:10.1029/2019MS001892, 2020.

Jin, Y. and Stan, C.: Simulation of East Asian Summer Monsoon (EASM) in SP-CCSM4: Part I–Seasonal mean state and intraseasonal variability, Journal of Geophysical Research: Atmospheres, 121, 13, 7801–7818, 2016.

Jones, T. R., Randall, D. A., and Branson, M. D.: Multiple-instance superparameterization: 1. Concept, and predictability of precipitation. J. Adv. Model. Earth Syst., 11, 11, 3497–3520, 2019.

Khairoutdinov, M. F. and Randall, D. A.: A cloud resolving model as a cloud parameterization in the NCAR Community Climate System Model: Preliminary results, Geophysical Research Letters, 28, 3617–3620, 10.1029/2001gl013552, 2001.

Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.

Kooperman, G. J., Pritchard, M. S., Burt, M. A., Khairoutdinov, M. F., Randall, D. Branson, M. D., and Randall, D. A.: Robust effects of cloud superparameterization on simulated daily rainfall intensity statistics across multiple versions of the Community Earth System Model, Journal of Advances in Modeling Earth Systems, 8, 140-165, 10.1002/2015ms000574, 2016.

A., and DeMott, C.: Simulations of the atmospheric general circulation using a cloud-resolving model as a superparameterization of physical processes, J. Atmos. Sci., 62, 2136–2154, 2005.

Krasnopolsky, V. M., Fox-Rabinovitz, M. S., and Belochitski, A. A.: 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, 1–13, doi:10.1155/2013/485913, 5, 2013.

Lin, J.-L.: The doubleDouble-ITCZ problemProblem in IPCC AR4 coupledCoupled GCMs: Ocean-atmosphere feedback analysis, J. Atmosphere Feedback Analysis, Journal of Climate, 20, 4497-4525, doi:10.1175/JCLI4272jcli4272.1, 2007.

Lin, J. L., and Coauthors: Tropical intraseasonal variability in 14 IPCC AR4 climate models. Part I: Convective signals, J. Climate, 19, 2665–2690.

Ling, J., Li, C., Li, T., Jia, X., Khouider, B., Maloney, E., Vitart, F., Xiao, Z., and Zhang, C.: Challenges and opportunitiesOpportunities in MJO studiesStudies, Bulletin of the American Meteorological Society, 98, ES53–-ES56, 10.1175/bams-d-16-0283.1, 2017.

Lopez-Gomez, I., Cohen, Y., He, J., Jaruga, A., and Schneider, T.: A generalized mixing length closureGeneralized Mixing Length Closure for eddy-diffusivity mass-flux schemesEddy-Diffusivity Mass-Flux Schemes of turbulenceTurbulence and convection, J. Adv. Model.Convection, Journal of Advances in Modeling Earth Syst.,Systems, 12, e2020MS002161, https://doi.org/10.1029/2020MS002161, 2020.

Lu, Y., Wu, T., Li, Y., and Yang, B.: Mitigation of the double ITCZ syndrome in BCC-CSM2-MR through improving parameterizations of boundary layer turbulence and shallow convection, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-40, in review, 2020.

Loshchilov, I. and Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts, arXiv preprint arXiv:1608.03983, 2016.

Madden, R. A. and Julian, P. R.: Detection of : Description of Global-Scale Circulation Cells in the Tropics with a 40–50 day oscillation in the zonal wind in the tropical Pacific, J. Atmos. Sci., 28, 702–708, 1971.

Mechoso, C. R., Robertson, A. W., Barth, N., Davey, M. K., Delecluse,Day Period, Journal of Atmospheric Sciences, 29, 1109-1123, P., Gent, P. R., Ineson, S., Kirtman, B., Latif, M., Le Treut,H., Nagai, T., Neelin, J. D., Philander, S. G. H., Polcher, J., Schopf, P. S., Stockdale, T., Suarez, M. J., Terray,L., Thual, O., and Tribbia, J. J.: The seasonal cycle over the tropical Pacific in coupled ocean atmosphere general circulation models, Mon. Weather Rev., 123(9), 2825-2838, doi:10.1175/1520-0493(1995)123<2825:TSCOTT0469(1972)029<1109:Dogscc>2.0.COCo;2, 19951972.

Mooers, G., Pritchard, M., Beucler, T., Ott, J., Yacalis, G., Baldi, P., &and Gentine, P. (2021).: Assessing the Potential of Deep Learning for Emulating Cloud Superparameterization in Climate Models With Real--Geography Boundary Conditions., Journal of Advances in Modeling Earth Systems, 13(5), e2020MS002385, https://doi.org/10.1029/2020MS002385, 2021.

Morrison, H. and Gettelman, A.: A new two-moment bulk stratiform cloud microphysics schemeNew Two-Moment Bulk Stratiform Cloud Microphysics Scheme in the Community Atmosphere Model, versionVersion 3 (CAM3). Part I:

Description and numerical tests, J. Clim.,Numerical Tests, Journal of Climate, 21, 3642–-3659, 10.1175/2008jcli2105.1, 2008.

Neale, R. B., Chen, C.-C., Gettelman, A., Lauritzen, P. H., Park, S., Williamson, D. L., et al.Conley, A. J., Garcia, R., Kinnison, D., and Lamarque, J.-F.: Description of the NCAR Community Atmosphere Model (CAM5community atmosphere model (CAM 5.0), NCAR Tech.Technical Note NCAR/TN-486+STR, 268pp., Natl. Cent. For Atmos. Res., Boulder, Colo, 1, 1-12, 2012.

Oleson, K. W., Lawrence, D. M., Gordon, B., Flanner, M. FG., Kluzek, E., Peter, J., et al.Levis, S., Swenson, S. C., Thornton, E., and Feddema, J.: Technical description of version 4.0 of the Community Land Model (CLM), NCAR Technical Note, NCAR/TN-486+STR, 2010.

Ott, J., Pritchard, M., Best, N., Linstead, E., Curcic, M., and Baldi, P.: A Fortran-Keras deep learning bridgeDeep Learning Bridge for scientific computing, arXiv:2004.10652Scientific Computing, Scientific Programming, 2020., 8888811, 10.1155/2020/8888811, 2020.

Park, S. and Bretherton, C. S.: The University of Washington shallow convectionShallow Convection and moist turbulence schemesMoist Turbulence Schemes and their impactTheir Impact on climate simulationsClimate Simulations with the Community Atmosphere Model, J. Clim.,Journal of Climate, 22, 3449–-3469, 10.1175/2008jcli2557.1, 2009.

Randall, D., Khairoutdinov, M., Arakawa, A., et al.and Grabowski, W.: Breaking the cloud parameterization deadlockCloud Parameterization Deadlock, Bulletin of the American Meteorological Society, 84, 1547–-1564, 10.1175/bams-84-11-1547, 2003.

Rasp, S.: Coupled online learning as a way to tackle instabilities and biases in neural network parameterizations: general algorithms and Lorenz 96 case study (v1.0), Geosci. Rasp,Model Dev., 13, 2185-2196, S.:Coupled online learning as a way to tackle instabilities and biases in neural network parameterizations: general algorithms and Lorenz 96 case study (v1.0), Geosci. Model Dev., 13, 2185-2196, doi:10.5194/gmd-13-2185-2020, 2020.

Rasp, S., Pritchard, M. S., and Gentine, P.: Deep learning to represent subgrid processes in climate models, Proceedings of the National Academy of Sciences, 115, 9684–-9689, 10.1073/pnas.1810286115, 2018.

Schneider, T., Lan, S., Stuart, A., and Teixeira, J.: Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations, Geophys. Res. Lett., 44, 12396–12417, doi:10.1002/2017GL076101, 2017.

Seo, E., Lee, M. I., Kim, D., Lim, Y. K., Schubert, S. D., and Kim, K. M.: Inter-annual variation of tropical cyclones simulated by GEOS-5 AGCM with modified convection scheme, Journal of Climatology, 39, 4041–4057.

Song, X. and Zhang, G. J.: The Roles of Convection parameterization, tropical Pacific double ITCZ, and upper-ocean biasesParameterization in the NCAR CCSM3. Part I: Climatology and atmospheric feedback, J. Climate, 22, 4299–4315, doi:10.1175/2009JCLI2642.1, 2009.

Storer, R. L., Griffin, B. M., Höft, J., Weber, J. K., Raut, E., Larson, V. E., Wang, M., and Rasch, P. J.: Parameterizing deep convection using the assumed probability density function method, Geosci, Model Dev., 8, 1–19, doi:10.5194/gmd-8-1-2015, 2015.

Oueslati, B. and Bellon, G.: Convective entrainment and large scale organization of tropical precipitation: Sensitivity of the CNRM-CM5 hierarchy of models, J. Climate, 26, 2931–2946, doi:10.1175/JCLI-D-12-00314.1, 2013.

Peters, K., Crueger, T., Jakob, C., and Möbis, B.: Improved MJO-simulation in ECHAM6.3 by coupling a Stochastic Multicloud Model to the convection scheme, J. Adv. Model. Earth Syst., 9, 193–219, doi:10.1002/2016MS000809, 2017.

Rasp, S.: Coupled online learning as a way to tackle instabilities and biases in neural network parameterizations: general algorithms and Lorenz 96 case study (v1.0), Geosci. Model Dev., 13, 2185–2196, doi:10.5194/gmd-13-2185-2020, 2020.

Song, X. and Zhang, G. J.: The roles of convection parameterization in the formation of double ITCZ syndromeFormation of Double ITCZ Syndrome in the NCAR CESM: I. Atmospheric processes, J. Adv. Model.Processes, Journal of Advances in Modeling Earth Syst.,Systems, 10, 842-866, https://doi.org/10.1002/2017MS001191, 2018.

Tiedtke, M.: A comprehensive mass flux scheme for cumulus parameterization in large-scale models, Mon.: A Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale Models, Monthly Weather Rev.,Review, 117, 1779–1800, 10.1175/1520-0493(1989)117<1779:Acmfsf>2.0.Co;2, 1989.

Wheeler, M. and Kiladis, G. N.: Convectively coupled equatorial wavesCoupled Equatorial Waves: Analysis of clouds and temperature in the wavenumber-frequency domain, J. Atmos. Sci.,Clouds and Temperature in the Wavenumber–Frequency Domain, Journal of the Atmospheric Sciences, 56, 374–399, 10.1175/1520-0469(1999)056<0374:Ccewao>2.0.Co;2, 1999.

Wu, T.: A mass-flux cumulus parameterization scheme for large-scale models: Description and test with observations, Clim. Dyn., 38, 725–744, 2012.

Xie, S., Wang, Y.-C., Lin, W., Ma, H.-Y., Tang, Q., Tang, S. et al.: Improved diurnal cycle of precipitation in E3SM with a revised convective triggering function, J. Adv. Model. Earth Syst., 11, 2290–2310, doi:10.1029/2019MS001702, 2019.

Yuval, J. and O'Gorman, P. A.: Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions, Nature Communications, 11, 3295, doi:10.1038/s41467-020-17142-3, 2020.

Yuval, J., O'Gorman, P. A., and Hill, C. N.: Use of Neural Networks for Stable, Accurate and Physically Consistent Parameterization of Subgrid Atmospheric Processes With Good Performance at Reduced Precision, Geophysical Research Letters, 48, e2020GL091363, https://doi.org/10.1029/2020GL091363, 2021.

Zhang, G. J. and McFarlane, N. A.: Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian Climate Centerclimate centre general circulation model, Atmos. Atmosphere-Ocean., 33, 407–446, 10.1080/07055900.1995.9649539, 1995.

Zhang, G. J.: Convective quasi-equilibrium in midlatitude continental environment and its effect on convective parameterization. Journal of Geophysical Research: Atmospheres, **107(D14),** 4220, doi:10.1029/2001JD001005., 2002

Zhang, G. J. and Song, X.: Convection ~~parameterization, tropical~~Parameterization, Tropical Pacific ~~double~~Double ITCZ, and ~~upper-ocean biases~~Upper-Ocean Biases in the NCAR CCSM3. Part II: Coupled ~~feedback~~Feedback and the ~~role of ocean heat transport, J.~~ Role of Ocean Heat Transport, Journal of Climate, 23, 800~~–~~-812, ~~doi:~~10.1175/~~2009JCLI3109~~2009jcli3109.1, 2010.

Zhang, G. J., Song, X., and Wang, Y.: The double ITCZ syndrome in GCMs: A coupled feedback problem among convection, clouds, atmospheric and ocean circulations, ~~Atmos. Res.,~~Atmospheric Research, 229, 255-268, https://doi~~:~~.org/10.1016/j.atmosres.2019.06.023, 2019.

Zhang, ~~Yu,~~Y. and ~~Qiang~~ Yang~~. "~~, Q.: A ~~survey~~Survey on ~~multi-task learning."~~Multi-Task Learning, IEEE Transactions on Knowledge and Data Engineering ~~(~~, 1-1, 10.1109/TKDE.2021~~).~~

~~Zhao, M., Golaz, J.-C., Held, I. M., Guo, H., Balaji, V., Benson, R., et al.: The GFDL global atmosphere and land model AM4.0/LM4.0: 2. Model description, sensitivity studies, and tuning strategies, J. Adv. Model. Earth Syst., 10, 735–769, doi:10.1002/2017MS001209, 2018.~~

~~Zuidema, P., Chang, P., Medeiros, B., Kirtman, B. P., Mechoso, R., Schneider, E. K., Toniazzo, T., Richter, I., Small, R. J., Bellomo, K., Brandt, P., de Szoeke, S., Farrar, J. T., Jung, E., Kato, S., Li, M., Patricola, C. M., Wang, Z., Wood, R., and Xu, Z.: Challenges and prospects for reducing coupled climate model SST biases in the eastern tropical Atlantic and Pacific oceans: The U.S. CLIVAR eastern tropical oceans synthesis working group, B. Am. Meterol. Soc., 97, 2305-2328, doi:10.1175/BAMS-D-15-00274.1, 2016~~.3070203, 2021.

Table 1. Input and output variables. For inputs, $\boldsymbol{q_v(z)}$ denotes the vertical profile of water vapor. $\boldsymbol{T(z)}$is the profile of temperature, and $\boldsymbol{dq_{v\ l.s.}(z)}$and $\boldsymbol{dT_{l.s.}}$ are the large scale forcing of water vapor and temperature, respectively. $\boldsymbol{P_s}$ is the surface pressure and *Solin* is the TOA solar insolation. For outputs, $\boldsymbol{dq_v(z)}$ and $\boldsymbol{dT(z)}$ are the tendencies of water vapor and temperature due to moist physics and radiative processes calculated by the NN-Parameterization. The net longwave and shortwave fluxes at the surface and the TOA are surface net longwave flux (FLNS), surface net shortwave flux (FLNT), TOA net longwave flux (FLNT), and TOA net shortwave fluxes (FSNT). The 4 downwelling solar radiation including solar downward visible direct to surface (SOLS), solar downward near infrared direct to surface (SOLL), solar downward visible diffuse to surface (SOLSD), and solar downward near infrared diffuse to surface (SOLLD) are shortwave radiation fluxes reaching the surface.

Inputs	Outputs
$q_v(z)$, $T(z)$, $dq_{vls}(z)$, $dT_{ls}(z)$, P_s, *Solin*	$dq_v(z)$, $dT(z)$, *FLNS, FSNS, FLNT, FSNT, SOLS*, *SOLL*, *SOLSD*, *SOLLD*

Table 2. Training configuration for the baseline ResMLP

Description	Value
Number of samples trained per iteration	1024
Strategies for Declining Learning Rate	coslr
Initial learning rate	0.001
Number of rounds to traverse the data set	50
Probability of randomly setting neurons to 0	0
L2 regularization	0

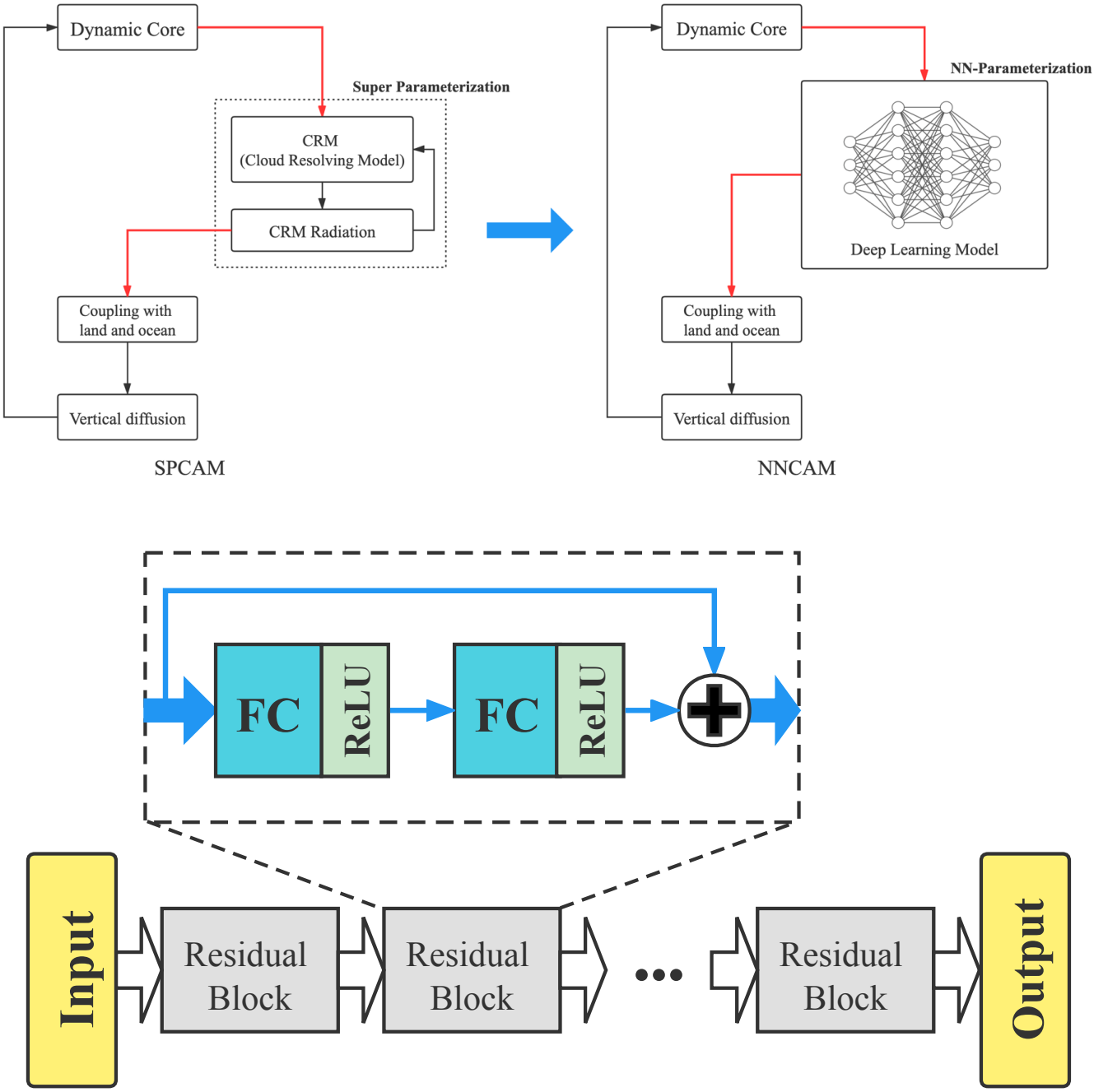


Figure 1. ~~Workflow diagram of the NNCAM. In the NNCAM, both the CRM and CRM radiation in SPCAM are replaced with the NN-Parameterization. All other model components, including the dynamic core, coupled land surface model and prescribed ocean, and planetary boundary layer, remain the same as in SPCAM.~~

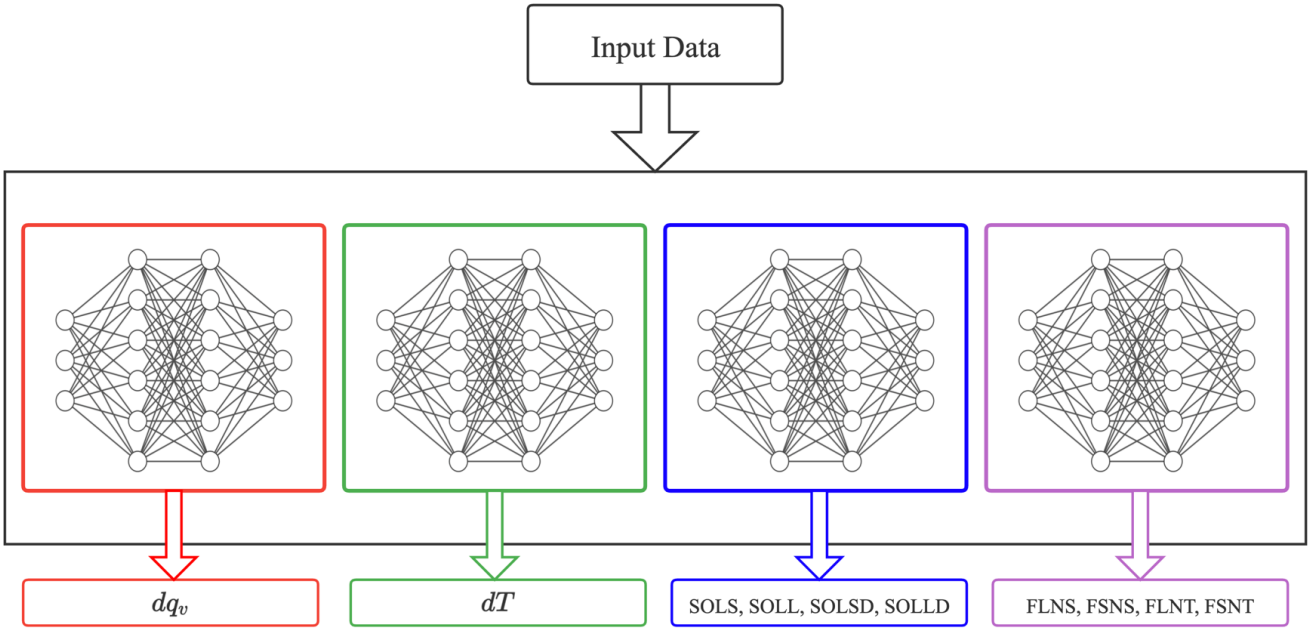
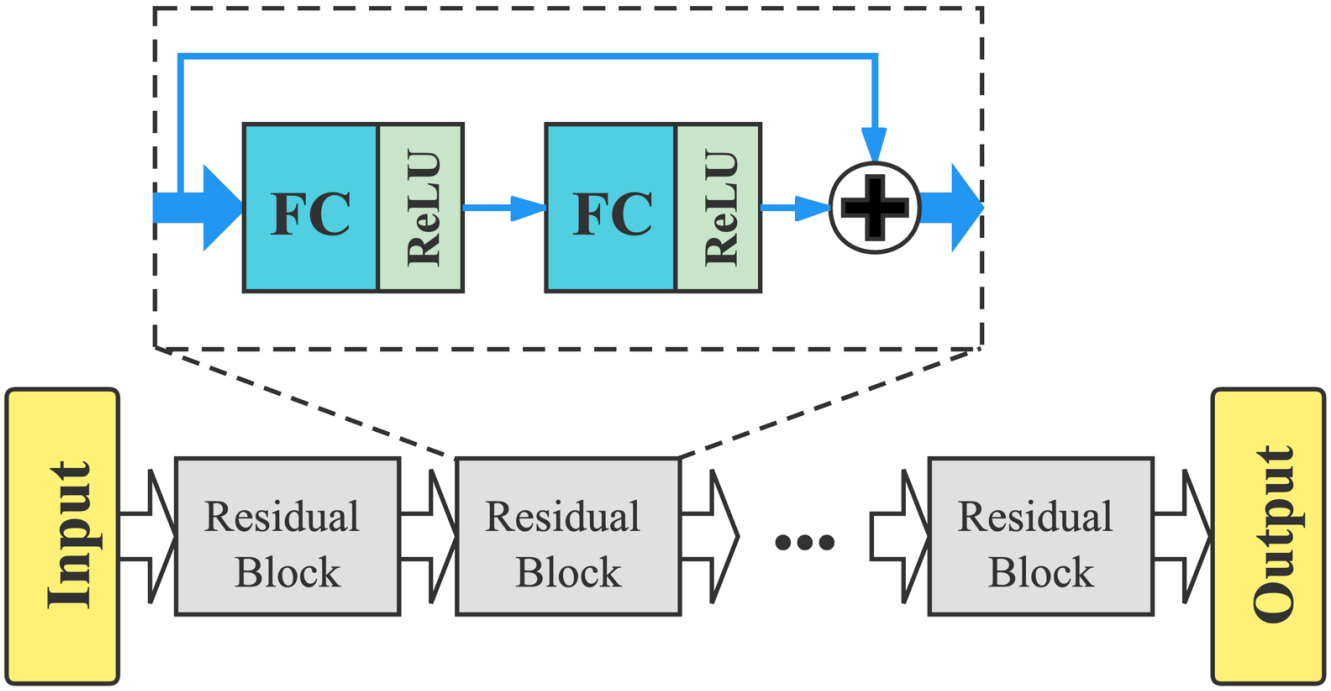
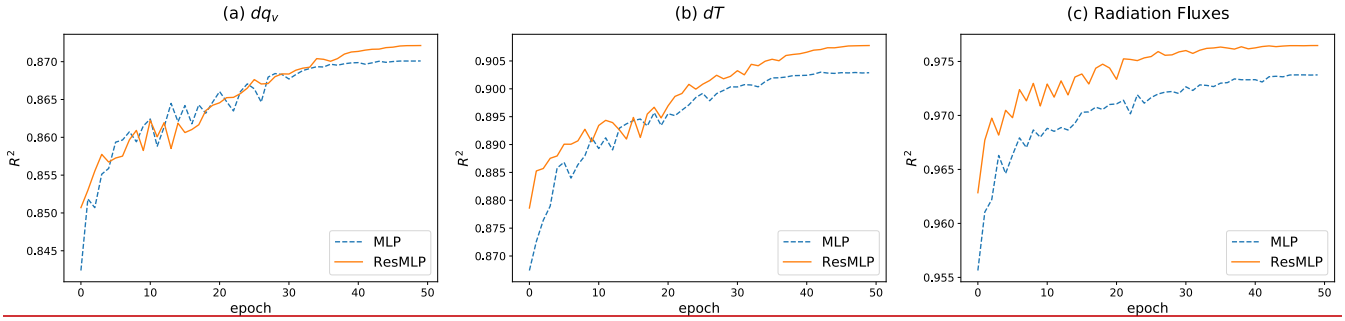


Figure 2. Schematic diagram of multi-target NN Parameterization. It is composed of three neural networks, sharing unified inputs, to predict $dq_v(z)$, $dT(z)$, the downwelling solar radiation fluxes at the surface, and net radiation fluxes at the surface and the TOA respectively.



~~**Figure 3.**~~ Schematic showing the structure of ~~ResMLP~~ResDNN. It consists of 7 residual blocks, each of which (shown in dashed box) contains two 512 node-wide dense (fully-connected) layers with a ReLU as activation, and a layer jump. The input and output are discussed in section 2.2.~~2~~1.



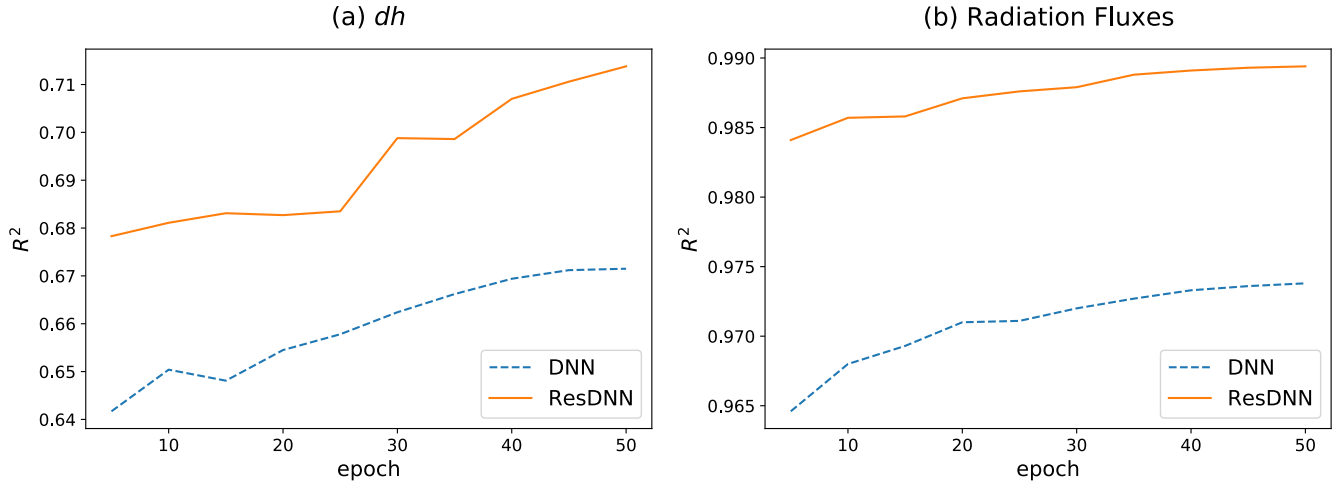
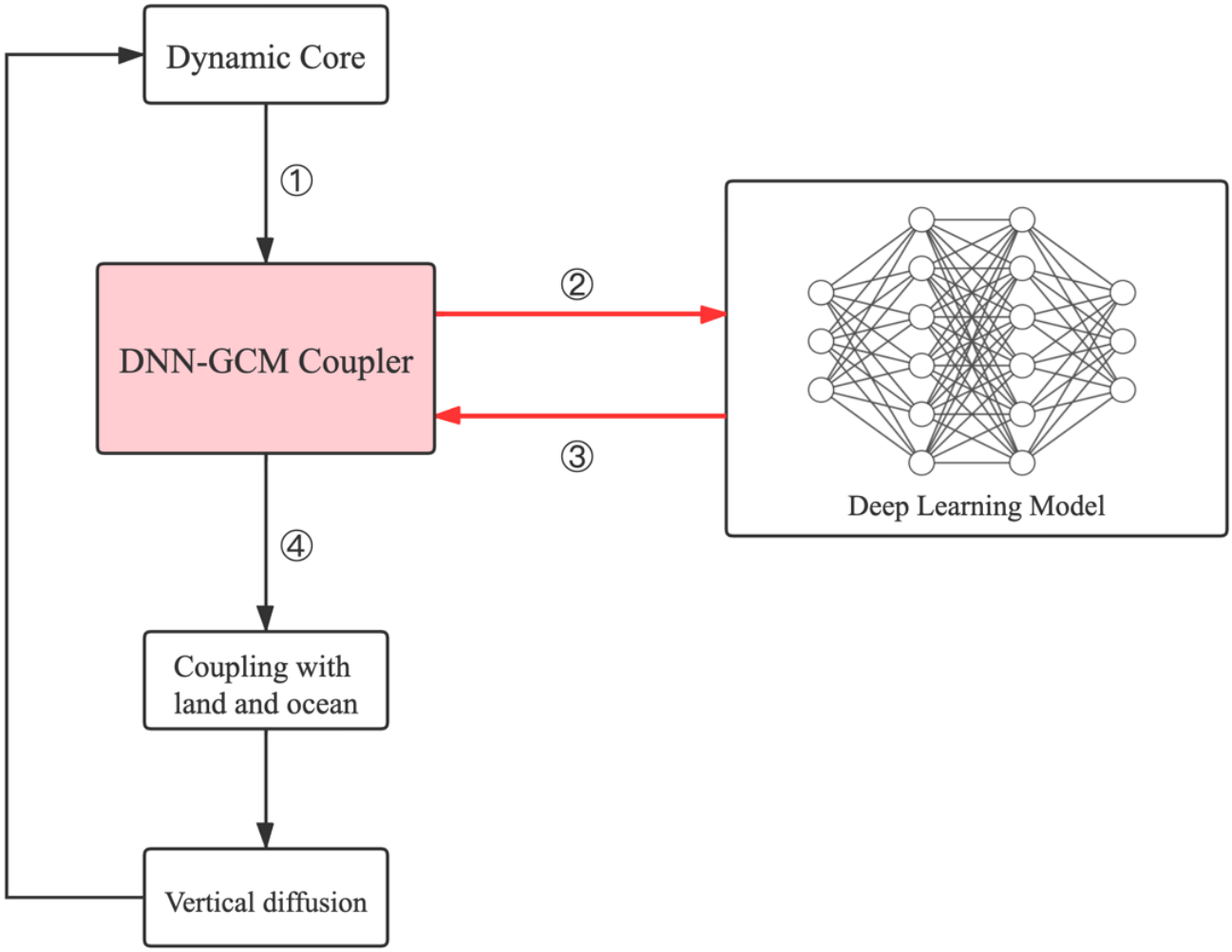


Figure 42. Fitting accuracies (R^2) of both the proposed ~~ResMLP~~ResDNN (orange solid lines) and ~~MLP~~DNN (blue dashed lines) for different targets. (a) ~~represents~~shows the ~~fitting accuracy~~R^2 of ~~moistening, (b) is the fitting accuracy of heating,~~moist static energy changing rate (dh) versus training epochs and (~~e~~b) shows the fitting accuracy of the average R^2 over the 8 radiation fluxes. Note: Spatial averaging of MSE is performed before calculating R^2.


Dynamic Core
①
DNN-GCM Coupler
②
③
Deep Learning Model
④
Coupling with
land and ocean
Vertical diffusion

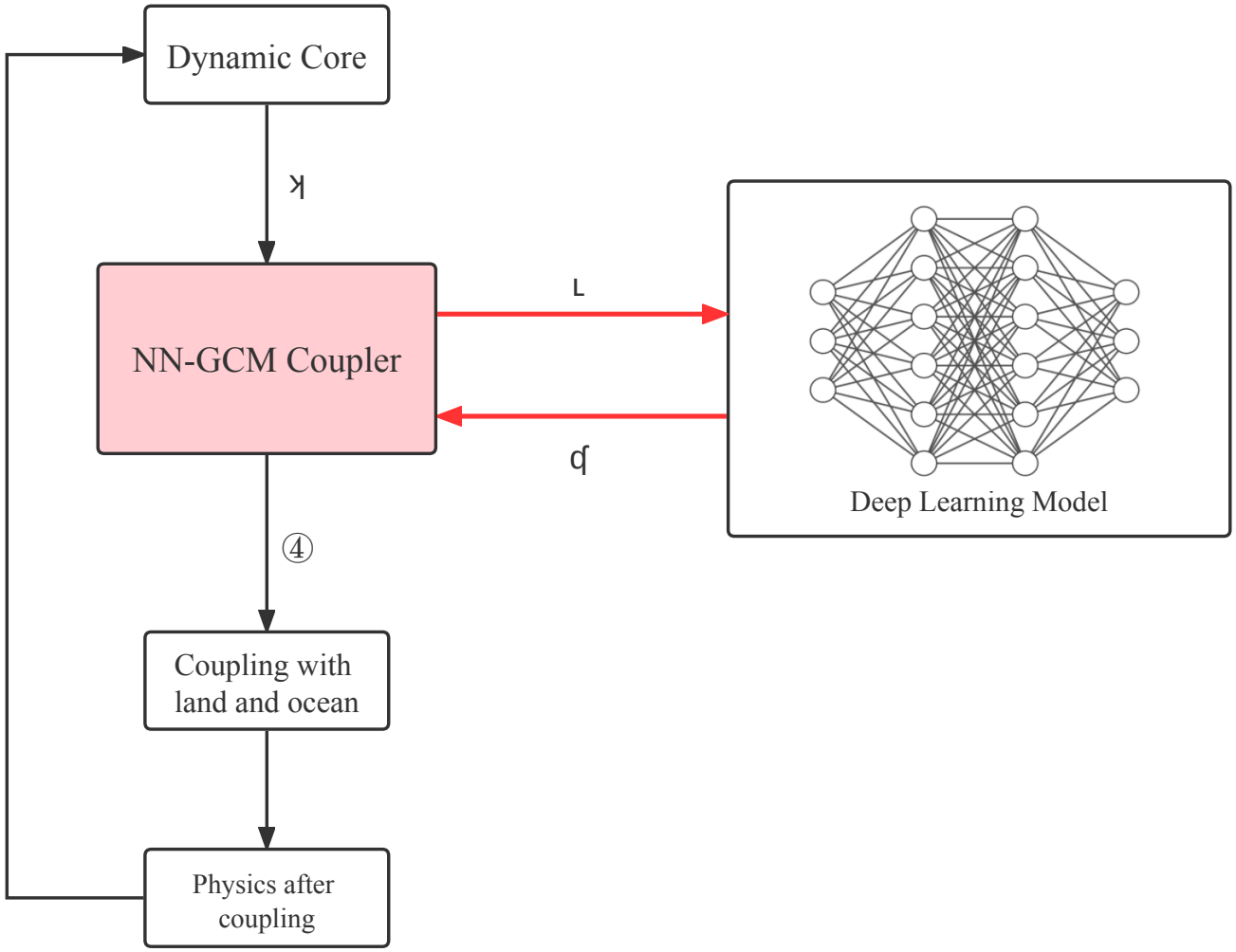
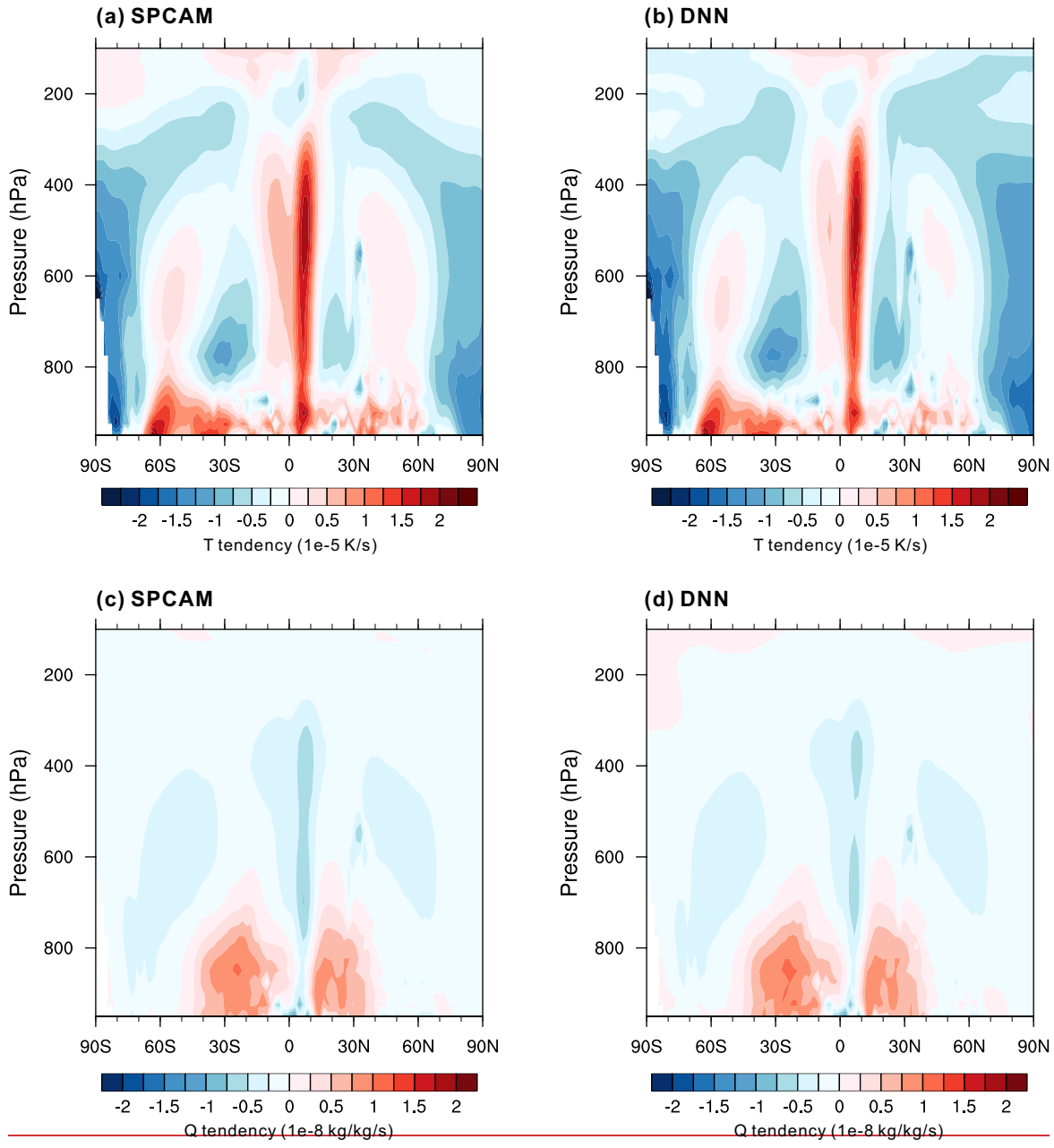


Figure ~~5~~3. A flow chart of NNCAM including ~~DNN~~NN-GCM Coupler. NNCAM runs in the direction of the arrow, and each box represents a module. Among them, DNN-GCM Coupler is indicated by light red. NN-Parameterization is shown in the sub-figure on the right. Note: ① represents the dynamic core transmits data to ~~DNN~~NN-GCM Coupler; ② and ③ represent the data communication between ~~DNN~~NN-GCM Coupler and NN-Parameterization; ④ represents the host GCM accepts the result from NN-Parameterization.


(a) SPCAM
(b) DNN
(c) SPCAM
(d) DNN
Pressure (hPa)
200
400
600
800
90S
60S
30S
0
30N
60N
90N
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
T tendency (1e-5 K/s)
Q tendency (1e-8 kg/kg/s)

910

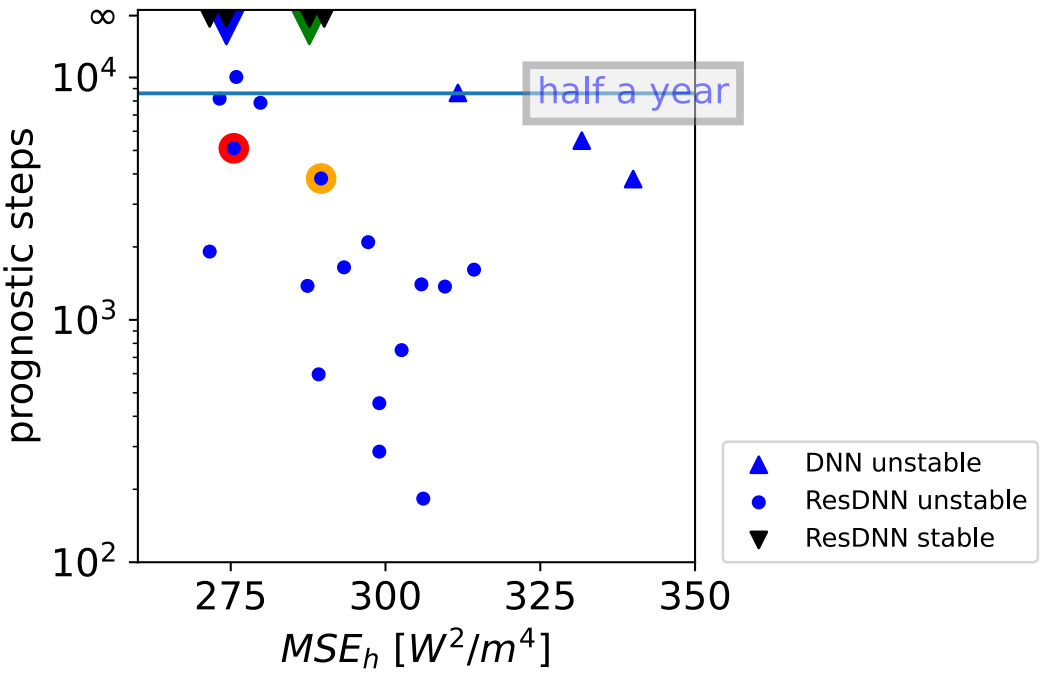


Figure 4. The offline moist static energy mean square error vs. prognostic steps. The black reversed triangles are stable NN coupled prognostic simulations lasting more than 10 years, blue ones are unstable simulations, and the blue triangles are for DNNs. The marked dots with colored outline are later exhibited in Figure 5 for time evolution of global averaged energy.

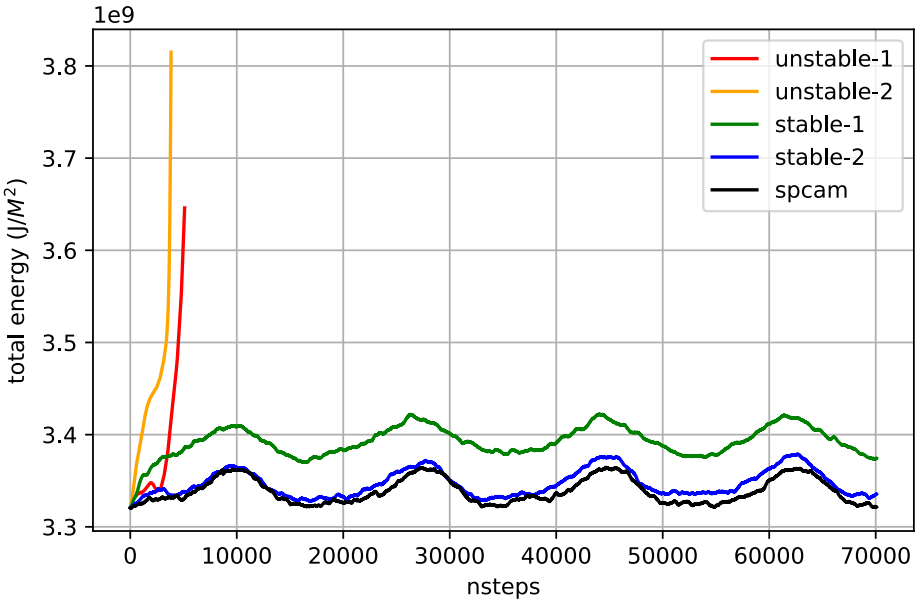
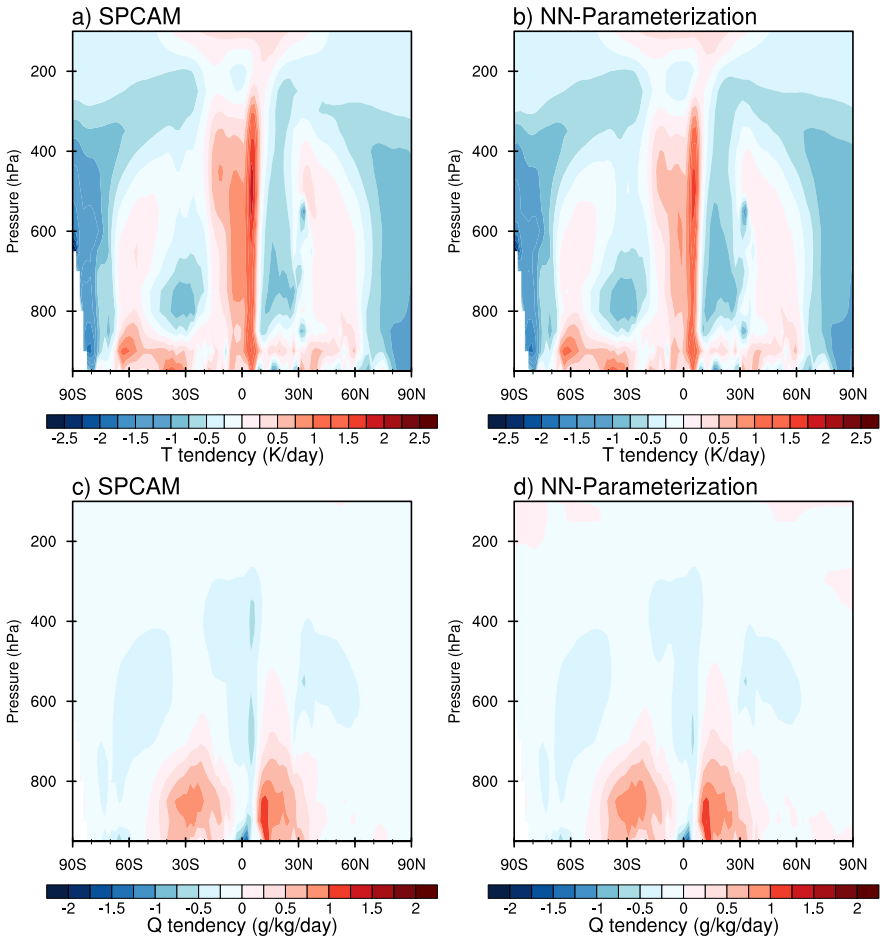
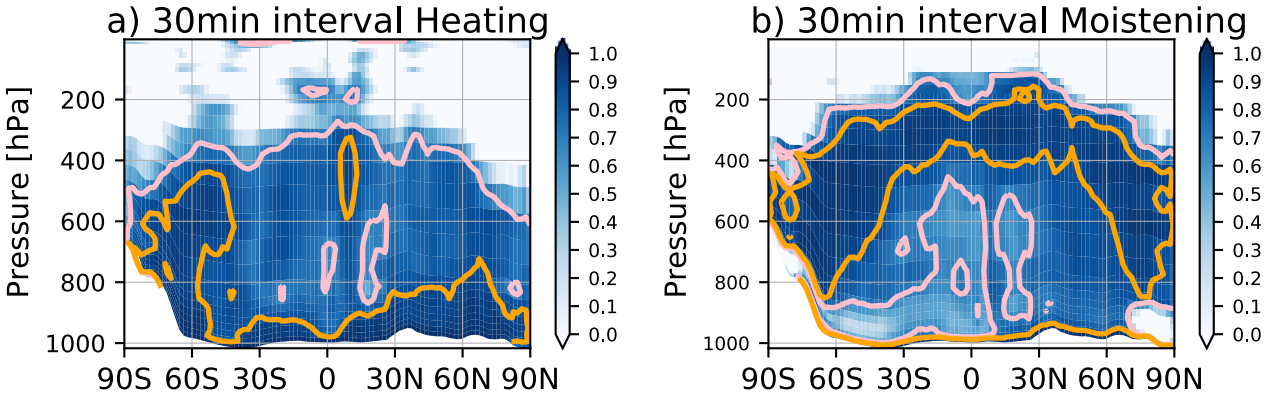


Figure 5. Time evolution of global averaged column integral total energy of NNCAM with different ResDNN parameterizations (marked with the same colors in Figure 4) and SPCAM target (the black line): Blue for stable and accurate ResDNN, green for a stable but deviated ResDNN, orange and red lines for unstable ResDNN.



925 **Figure 6.** Latitude-pressure cross sections of annual and zonal mean heating (top) and moistening (bottom) from moist physics during ~~1997–1998~~the year 2000 for (a, c) SPCAM simulations, and (b, d) offline test by the ~~DNN~~NN-Parameterization.


a) 30min interval Heating
b) 30min interval Moistening
Pressure [hPa]
200
400
600
800
1000
90S 60S 30S 0 30N 60N 90N
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

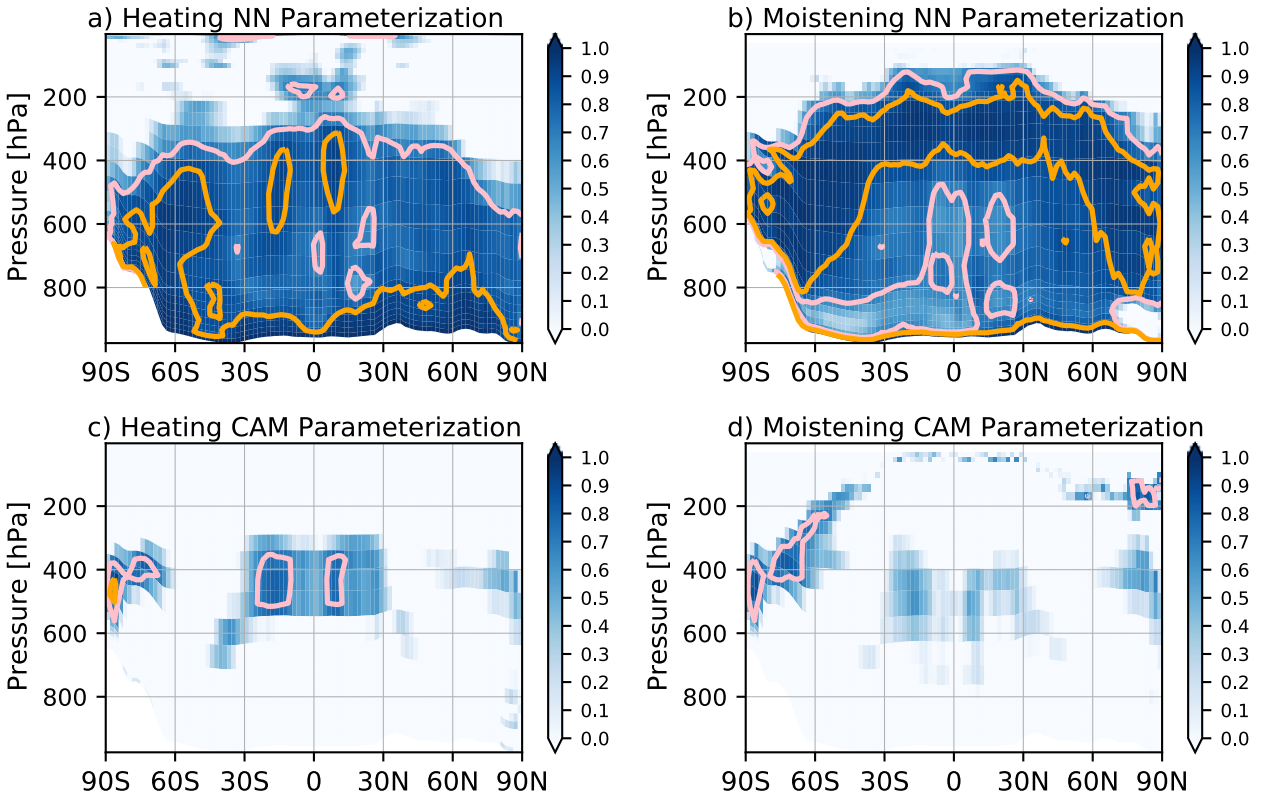
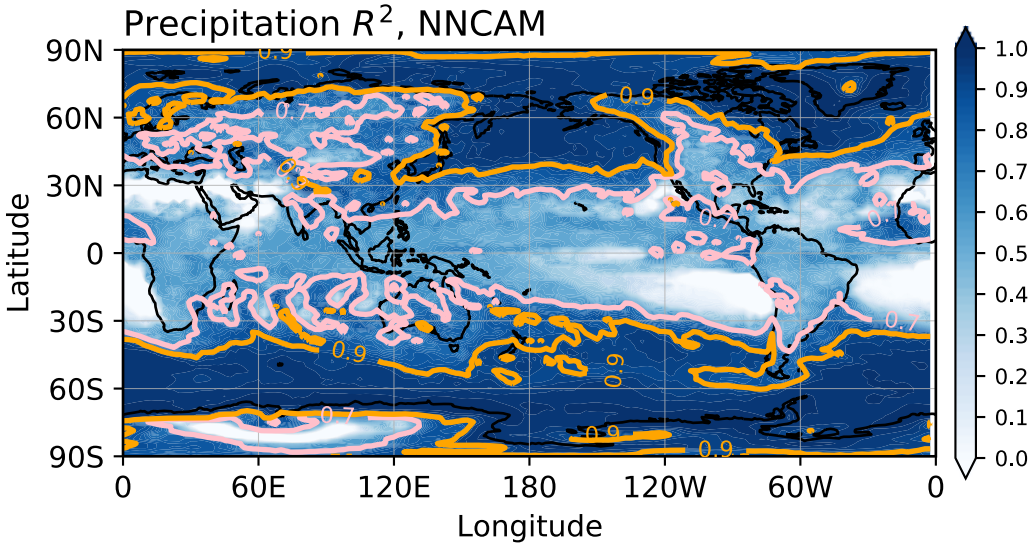


Figure 7. Latitude-pressure cross sections of coefficient of determination (R^2) for zonal averaged heating (~~a~~left panels) and moistening (~~b~~right panels). They are predicted by (a & b) NN-~~parameterization~~Parameterization in the offline one-year SPCAM run, and (c & d) by offline CAM5 parameterizations. Both are evaluated at 30-min timestep interval. Note: areas where R^2 is greater than 0.7 are contoured in pink and those greater than 0.9 are contoured in orange.


Precipitation R^2, NNCAM
90N
60N
30N
0
30S
60S
90S
Latitude
0
60E
120E
180
120W
60W
0
Longitude
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

945

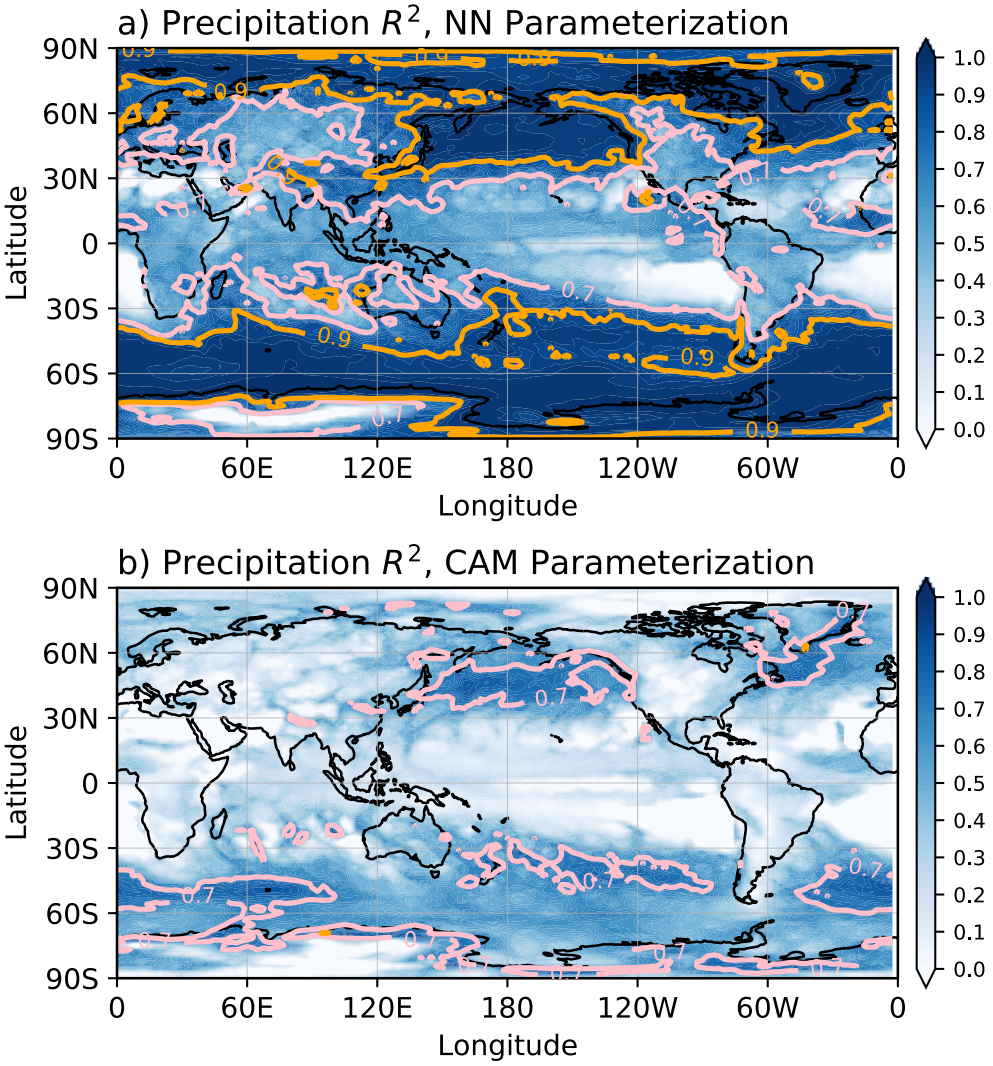
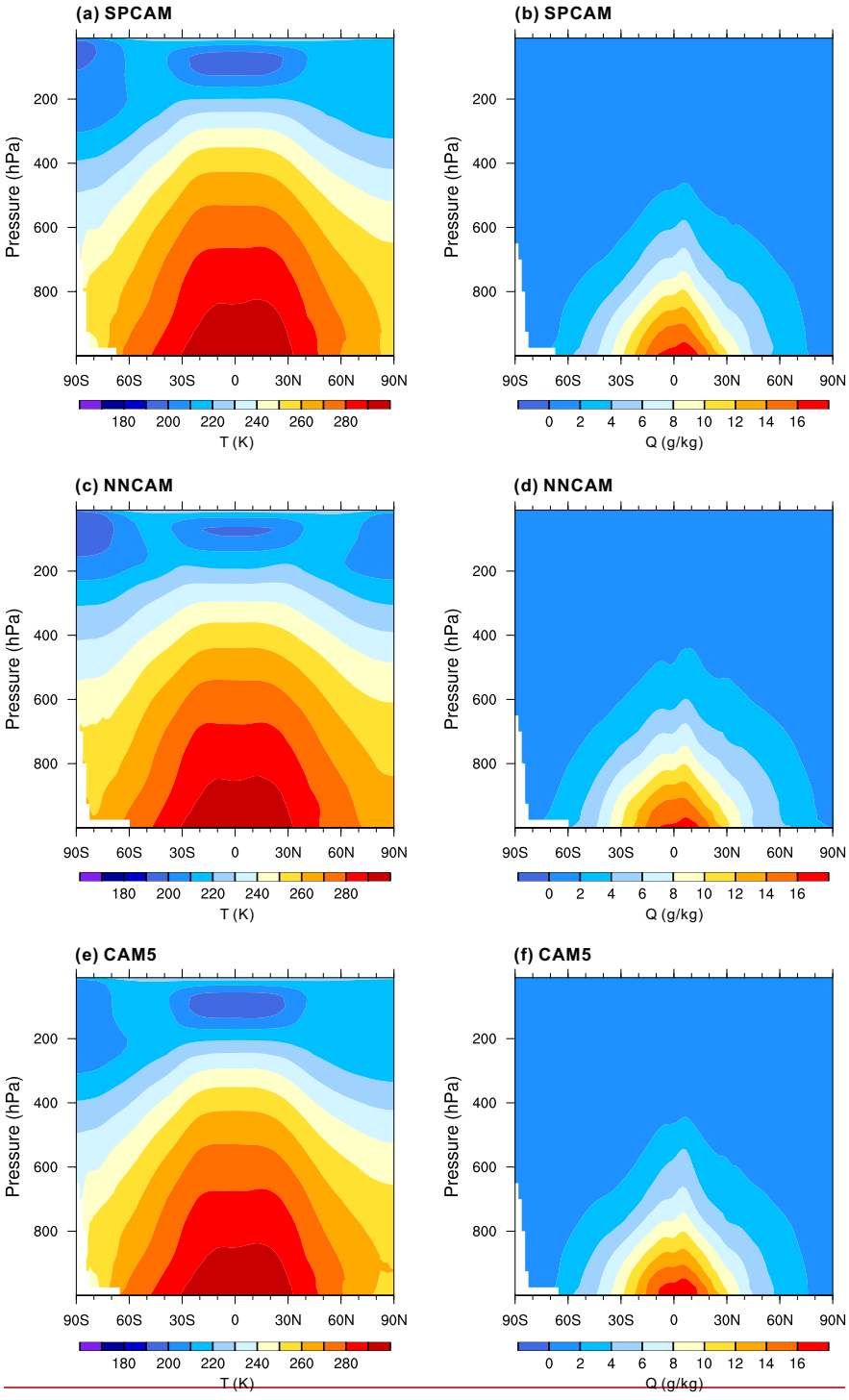


Figure 8. Latitude-pressure cross sections of coefficient of determination (R^2) for the derived precipitation predicted by NN-parameterization (a) and total precipitation from CAM5 parameterization (b) in the offline one-year SPCAM run. The predictions and SPCAM targets are in 30min timestep interval. Note: areas where R^2 is greater than 0.7 are contoured in pink and those greater than 0.9 are contoured in orange.


(a) SPCAM
(b) SPCAM
(c) NNCAM
(d) NNCAM
(e) CAM5
(f) CAM5
Pressure (hPa)
200
400
600
800
90S
60S
30S
0
30N
60N
90N
180
200
220
240
260
280
T (K)
0
2
4
6
8
10
12
14
16
Q (g/kg)

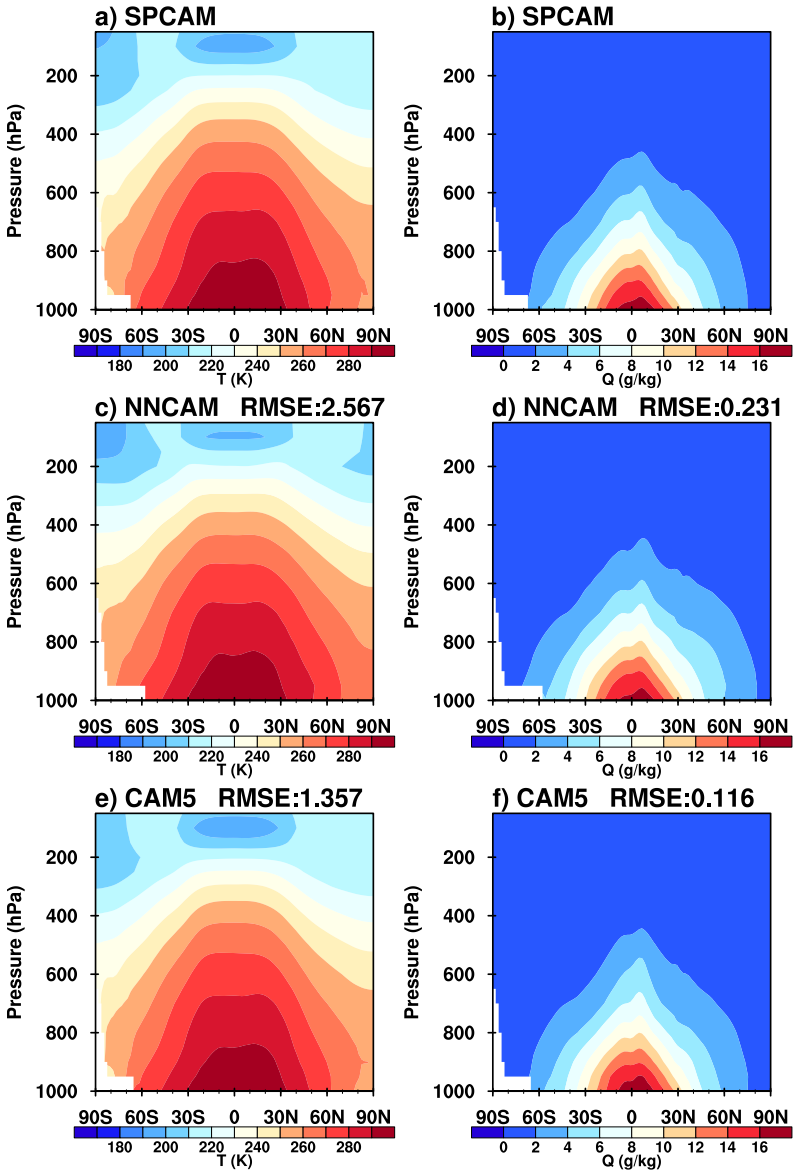
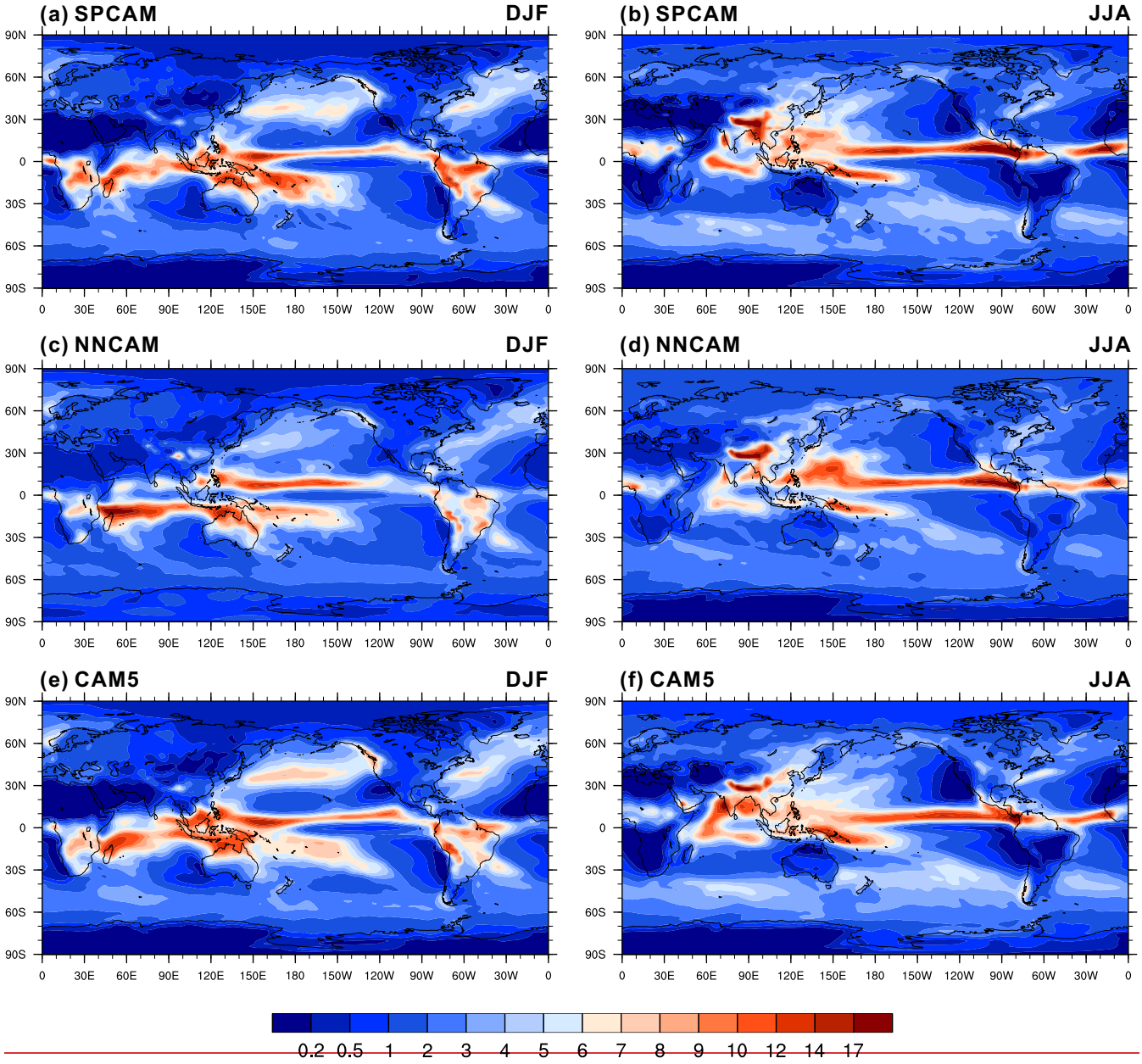


Figure 9. Latitude-pressure cross sections of annual and zonal mean temperature (left panels) and specific humidity (right panels) from (a, b) SPCAM (~~1998–2001~~1999–2003), (c, d) NNCAM (1999–2003), and (e, f) CAM5 (1999–2003).


(a) SPCAM
DJF
(b) SPCAM
JJA
(c) NNCAM
DJF
(d) NNCAM
JJA
(e) CAM5
DJF
(f) CAM5
JJA
90N
60N
30N
0
30S
60S
90S
0 30E 60E 90E 120E 150E 180 150W 120W 90W 60W 30W 0
0.2 0.5 1 2 3 4 5 6 7 8 9 10 12 14 17

960

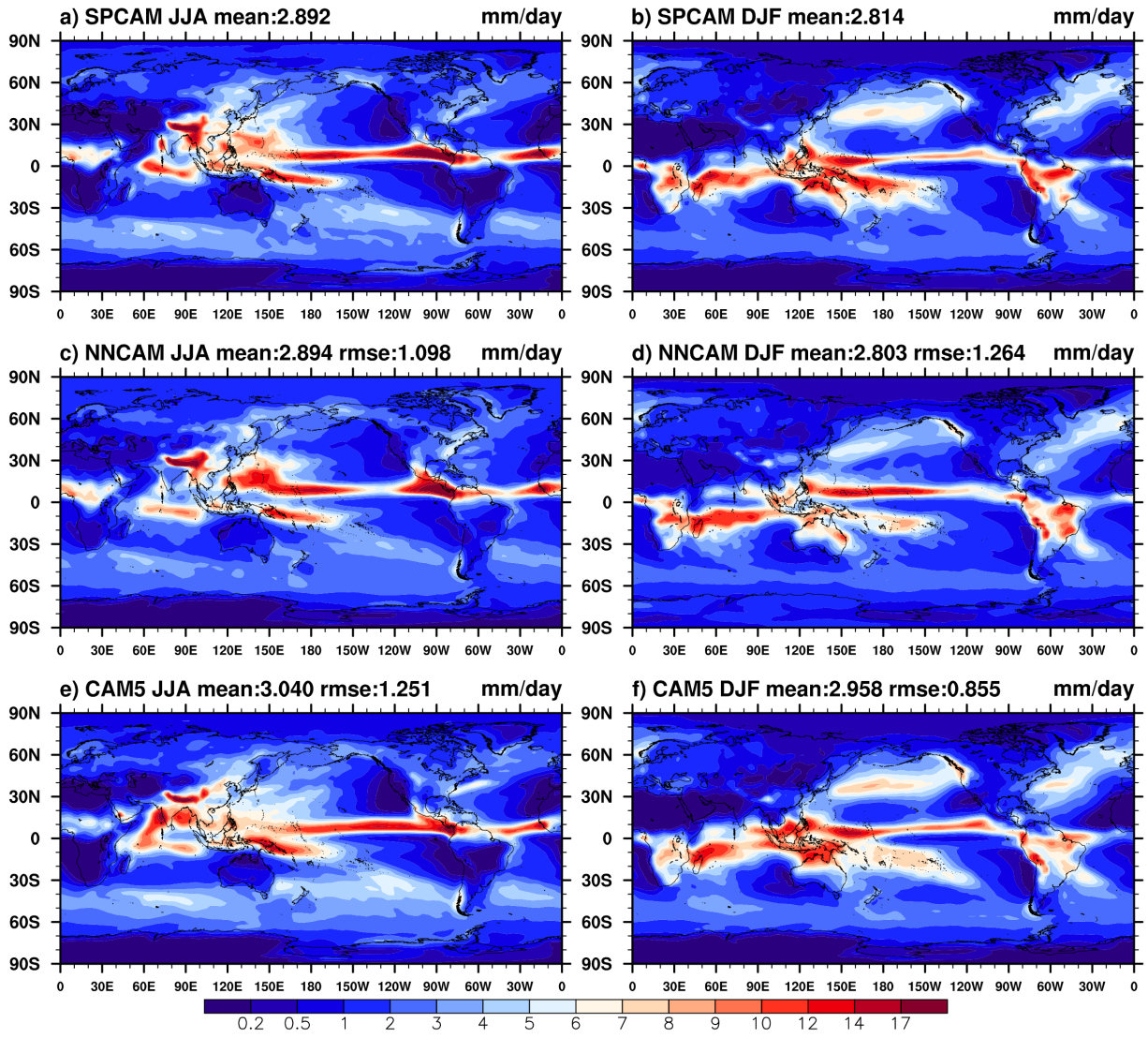
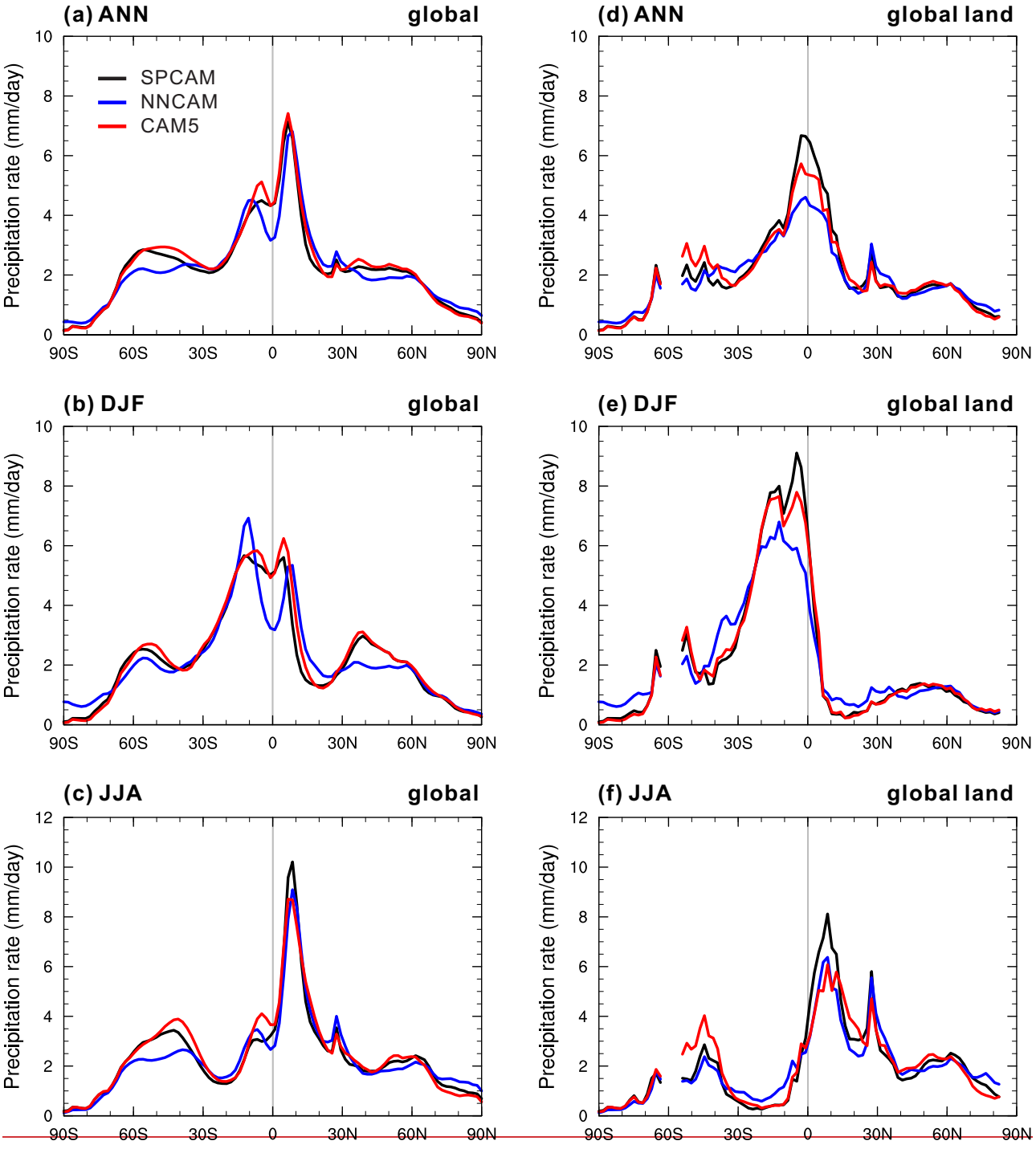


Figure 10. The mean precipitation rate (mm day^{-1}) of June-July-August (left panels) and December-January-February (~~left panels) and June-July-August (~~right panels) for (a, b) SPCAM (~~1998−2001~~1999−2003), (c, d) NNCAM (1999−2003), and (e, f) CAM5 (1999−2003).


(a) ANN
global
(d) ANN
global land
(b) DJF
global
(e) DJF
global land
(c) JJA
global
(f) JJA
global land
SPCAM
NNCAM
CAM5
Precipitation rate (mm/day)
90S
60S
30S
0
30N
60N
90N

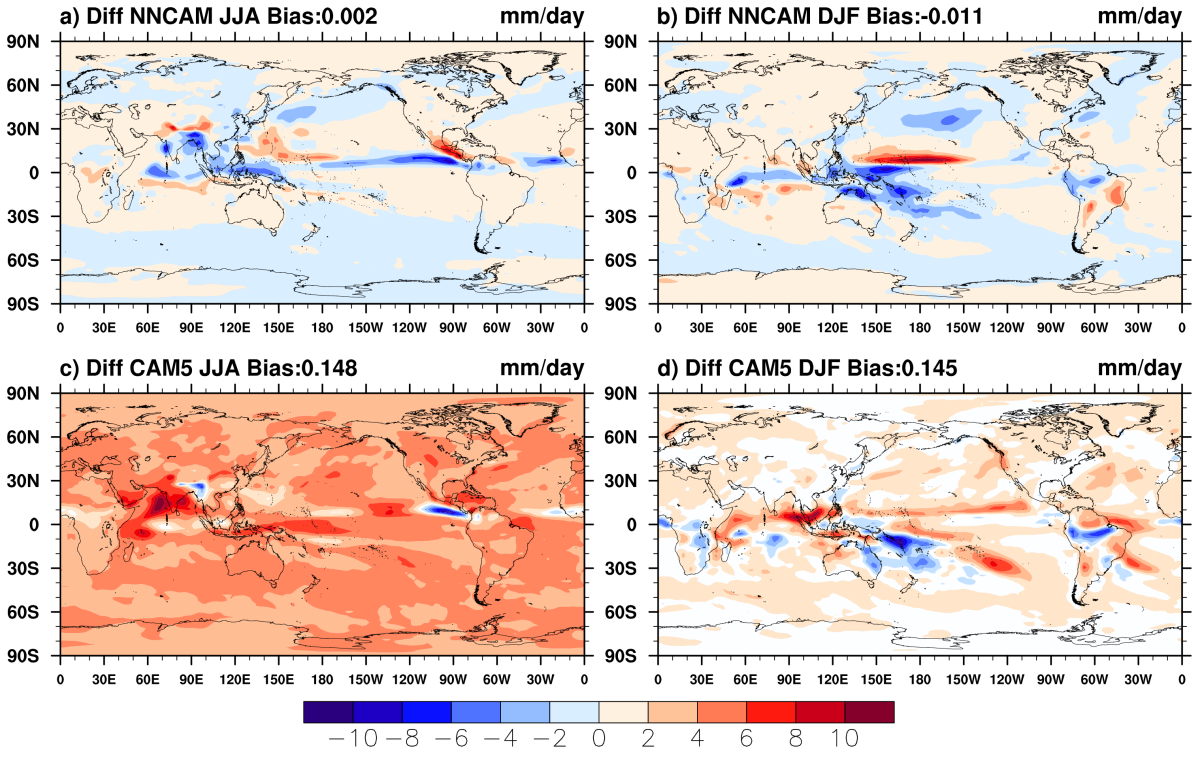
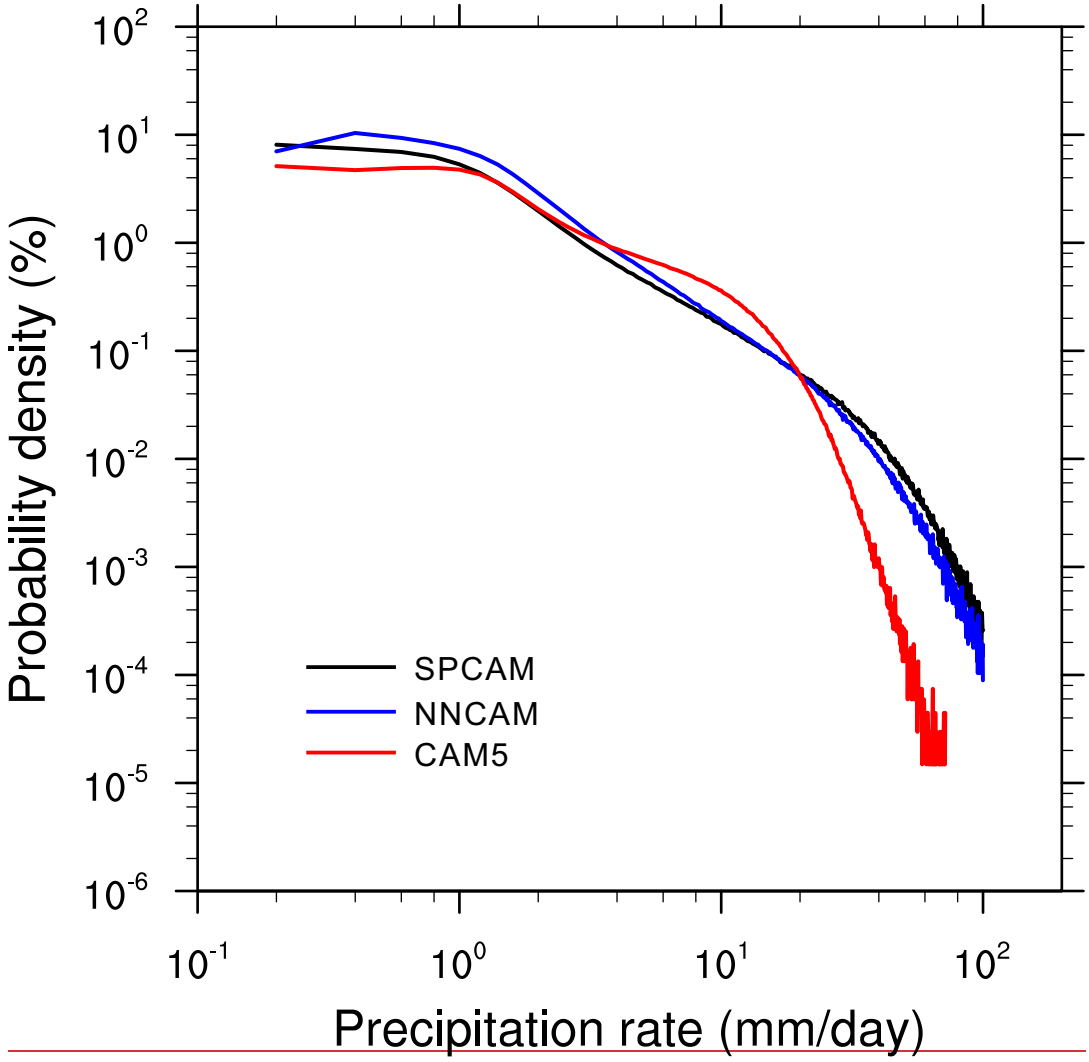


Figure 11. ~~The zonal mean~~Global distribution of precipitation ~~rate (mm day^{-1})~~difference averaged ~~for (a, d) the annual mean, (b, e) December-January-February,~~over boreal summer (left panels) and ~~(c, f) June-July-August. Black, blue and red solid lines denotes SPCAM,~~winter (right panels) between NNCAM and SPCAM (a & b) and between CAM5~~, respectively.~~ and SPCAM (c & d).


Probability density (%)
Precipitation rate (mm/day)
SPCAM
NNCAM
CAM5

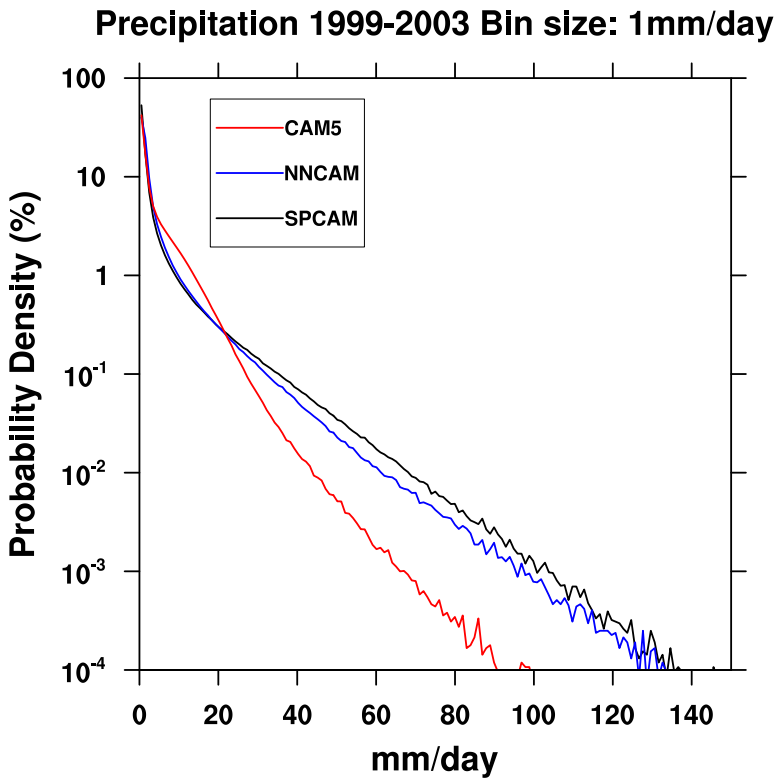
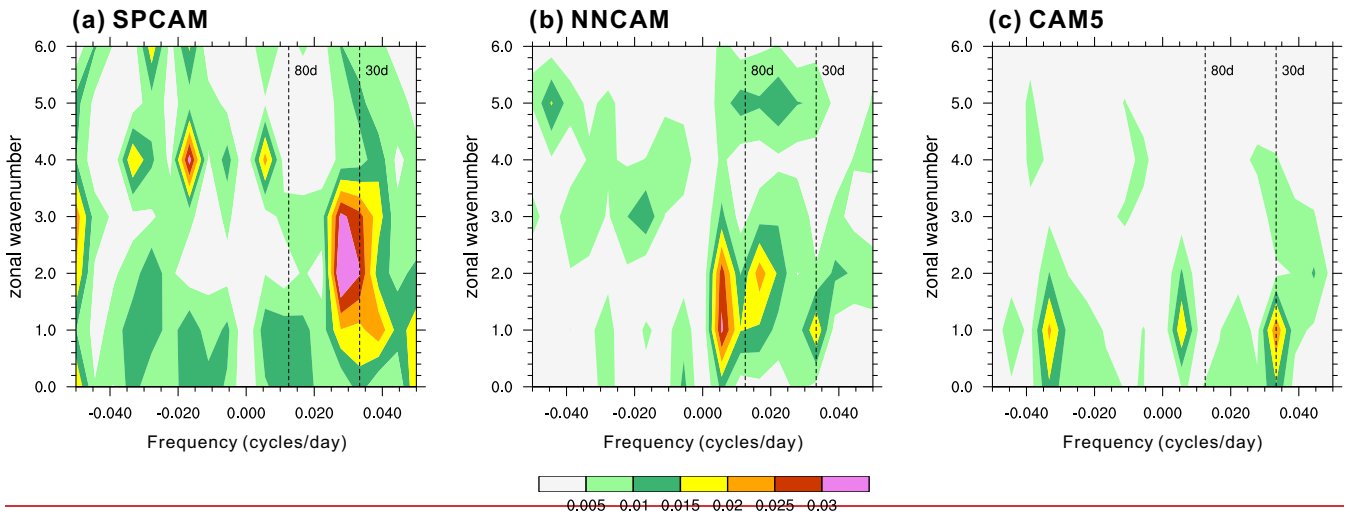


Figure 12. Probability densities of daily mean precipitation in the tropics (30°S−30°N) from the three model simulations. Black, blue and red solid lines ~~denotes~~denote SPCAM, NNCAM and CAM5, respectively.



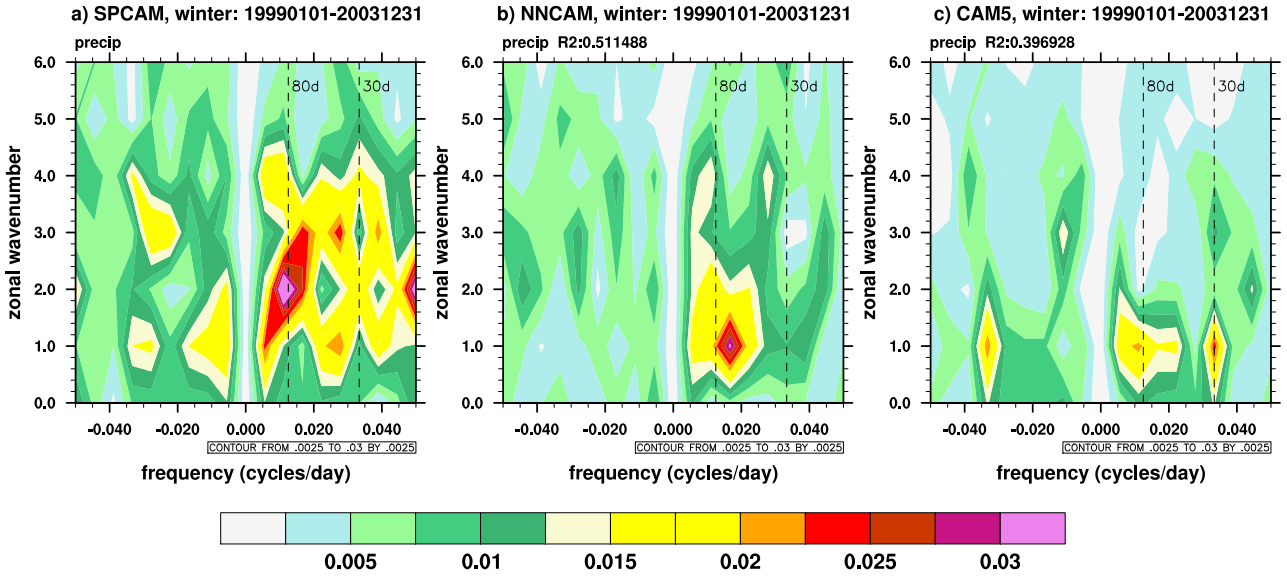


Figure 13. The wavenumber-frequency spectra of 10°S−10°N daily precipitation anomalies for (a, b) SPCAM, (c, d) NNCAM, and (e, f) CAM5 simulations for boreal winter.

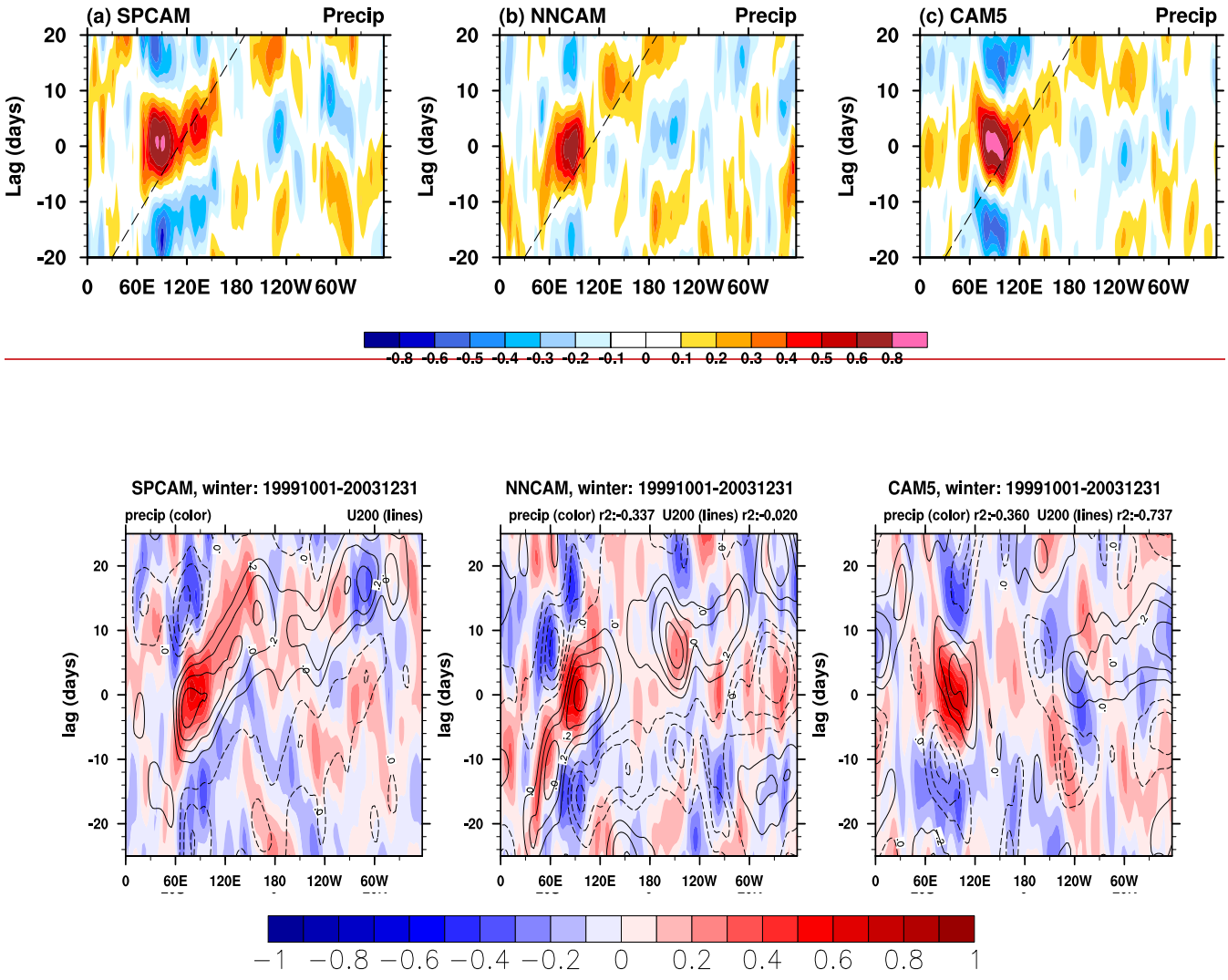


Figure 14. Longitude-time evolution of lagged correlation coefficient for the 20-100 ~~d~~day band-pass-filtered precipitation anomaly (averaged over 10°S−10°N) against regionally averaged precipitation (shaded) and zonal wind at 200hPa (contoured) over the equatorial eastern Indian Ocean (80E−100°E, 10°S−10°N). ~~Dashed lines in each panel denote the 5 m s^{-1} eastward propagation speed.~~