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~~clouds are one of the main ~~reasons for~~causes of the biases in precipitation and atmospheric circulation simulations. In recent years, due to the rapid development of data science, ~~Machine~~machine learning (ML) parameterizations for convection and clouds have been ~~proven~~demonstrated to have the potential to perform better than conventional parameterizations. ~~At present, most of the existing~~Most previous studies ~~are~~were conducted on aqua-planet and idealized models, and the problems of ~~simulated~~simulation instability and climate drift still exist. ~~In realistic configurated models, developing a machine learning~~Developing an ML parameterization scheme remains a challenging task~~.~~ in realistically configured models. In this ~~study~~paper, a set of ~~deep~~ residual deep neural networks (ResDNNs) with a strong nonlinear fitting ability is designed to emulate a superparameterization (SP) with different ~~types of outputs. Sensitivity tests show that high~~ outputs in a hybrid ML-physical general circulation model (GCM). It can sustain stable simulations for over 10 years under real-world geographical boundary conditions. We explore the relationship between the accuracy ~~is necessary to develop a stable ML parameterization. Trial~~and stability by validating multiple deep neural network (DNN) and ResDNN sets in prognostic runs. In addition, there are significant differences in the prognostic results of the stable ResDNN sets. Therefore, trial-and-error is used to acquire the optimal ResDNN set for both high~~ performance~~ skill and long-term stability, ~~named~~ which we name the NN-Parameterization. In offline validation, the NN-Parameterization ~~emulates~~can emulate the SP ~~results far better than the conventional subgrid parameterizations. Then, in~~ in mid- to high-latitude regions with a high accuracy. However, its prediction skill over tropical ocean areas still needs improvement. In the multi-year prognostic test, ~~NN-Parameterization reproduces reasonable climate mean states but still with some biases. Most importantly, NN parameterization successfully reproduces the climate variability in a superparameterizated GCM, with an over 30-time faster running speed. Under real geographical boundary conditions, the hybrid ML-physical GCM well simulates the spatial distribution of boreal summer~~the hybrid ML-physical GCM simulates the tropical precipitation well over land and significantly improves the frequency of the precipitation extremes, which ~~is largely~~are vastly underestimated in the Community Atmospheric Model version 5 (CAM5) with ~~the~~a horizontal resolution of 1.9°×2.5°. Furthermore, the hybrid ML-physical GCM simulates ~~a stronger~~the robust signal of the Madden~~-~~–Julian oscillation with a more reasonable propagation speed than CAM5. However, there are still substantial biases with the hybrid ML-physical GCM in the mean states, including the temperature field in the tropopause and at high latitudes and the precipitation over tropical oceanic regions, which are larger

than those in CAM5. This study is a pioneer ~~to achieve~~in achieving multi-year stable climate simulations using a hybrid ML-physical GCM ~~in~~under actual land~~-~~–ocean boundary conditions~~.~~ that become sustained over 30 times faster than the target SP. It demonstrates the emerging potential ~~for~~of using ~~machine learning~~ML parameterizations in climate simulations.

1 Introduction

~~The general~~General circulation models (GCMs) have been widely used ~~for studying~~to study climate variability, prediction, and projections. Despite decades of GCM development, most GCMs ~~still~~continue to suffer from many systematic biases, especially ~~at~~in low~~ latitudes. A ~~-latitude regions. The prominent tropical bias ~~in~~of most current GCMs is referred to as 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 (Lin, 2007; Zhang et al., 2019). Convectively coupled equatorial waves and the Madden-Julian ~~Oscillation~~oscillation (MJO), ~~featured~~which are characterized by eastward propagating convective cloud clusters, are also not well simulated ~~in~~by GCMs (Ling et al., 2017; Cao and Zhang, 2017).

Many studies have attributed most of these biases to ~~the imperfection of~~deficiencies in the parameterization schemes for atmospheric moist convection and cloud processes in the current GCMs (Zhang and Song, 2010; Cao and Zhang, 2017; Song and Zhang, 2018; Zhang and Song, 2019). Cloud-related processes span a large range of spatial scales, from micron-scale cloud nucleation~~,~~ to 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 and interact with the atmospheric circulation, affecting the transport and distribution of energy (Emanual et al., 1994). Therefore, it is very important to simulate the cloud and convection ~~process~~processes in GCMs correctly. However, the ~~current~~ GCMs that are currently used for climate ~~simulation~~simulations have a horizontal resolution of ~~~100km~~100 km and a vertical hydrostatic coordinate. Thus, in most GCMs, ~~besides~~in addition to 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; Lopez-Gomez et al., 2020). ~~Those~~These 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~~which leads to large uncertainties and biases in climate simulations (Bony et al., 2015).

~~Cloud Resolving Models (CRMs), on the other hand,~~In contrast, cloud resolving models (CRMs) have long been used to simulate convection. Because CRMs have higher horizontal and vertical resolutions and can explicitly resolve the thermodynamic processes involved in convection, they simulate convection more accurately, including convective organization (Feng et al., 2018). In recent years, CRMs have been used ~~as superparameterization~~for SuperParameterization (SP) in low-resolution GCMs ~~to replace~~and have replaced conventional cumulus convection and cloud parameterization schemes. The most commonly used SP model is the superparameterized ~~version of the~~ Community Atmosphere Model

(SPCAM) developed by the National Center for Atmospheric Research (NCAR) (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 parameterization schemes, SPCAM performs better in simulating mesoscale convective systems, diurnal ~~cycles of~~ precipitation cycles, monsoons, the precipitation frequency distribution, and ~~MJOs~~the MJO (Khairoutdinov et al., 2005; Bretherton et al., 2014; Jiang et al., 2015; Jin et al., 2016; Kooperman et al., 2016). However, when using ~~2D~~a 2-D CRM ~~as~~for SP, the improvement ~~on~~of the climate mean states is not obvious (Khairoutdinov et al., 2005). ~~Also~~In addition, SPCAM requires far more computing resources (i.e., an order of magnitude or more) than a Community Atmosphere Model (CAM) with the same resolution ~~CAM in 1 to 2 orders of magnitude according to the resolution of the CRM subdomain.~~. Thus, the use of SPCAM in long-term climate simulations and ensemble ~~prediction~~predictions is restricted by the current computing ~~resource~~resources. Developing novel and computationally efficient schemes for ~~high performance~~ convection and cloud processes is ~~still an open problem~~highly desired in GCM development.

In the last 5 years, the rapid development of machine learning (ML) ~~technologies~~techniques, especially deep learning ~~technologies~~techniques such as ~~neural networks~~Neural Networks (NNs), has provided novel approaches to constructing parameterization schemes. Machine learning can identify ~~and~~, discover, and model complex nonlinear relationships that exist in large ~~data sets and model them.~~datasets. Several studies have used ~~machine learning~~ML methods to develop convection and cloud parameterization schemes (e.g., 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~~to train the ML models ~~training~~. Then, the trained ~~machine learning~~ML models are often evaluated offline against other independent reference simulations, and finally, they are implemented in a GCM to replace the conventional parameterization schemes.

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 the convective temperature and moisture tendencies, ~~with~~and the training data ~~from~~for the CRM simulations was forced ~~by~~using observations in the tropical western Pacific. The resulting convective parameterization scheme was able to simulate the main features of ~~cloud~~the clouds and precipitation in the NCAR CAM4 diagnostically. However, the key issue of prognostic validation in 3-D GCMs ~~was~~has not been 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 deep NN (DNN) to predict convection and clouds, which was trained ~~with the~~using data from an aqua-planet SPCAM. The ~~NN~~DNN-based parameterization was then implemented in the corresponding aqua-planet CAM and produced multi-year prognostic results that were close to the SPCAM data. For this ~~NN~~DNN-based parameterization, Rasp (2020) found that minor changes, either to the training dataset or ~~in~~to 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~~non-physical correlations were learned by the network, and the information in the upper levels obtained from the input ~~vector~~data had to be removed to produce stable long-term simulations. Rather than using NNs, Yuval and O'Gorman

(2020) used the random forest algorithm to develop an ML parameterization based on ~~the~~ training data from a high-resolution idealized 3-D model with a setup ~~of~~on the equatorial beta plane. They used two independent random forests to separately emulate the different processes ~~separately~~ and ensured the physical constraints by predicting subgrid fluxes instead of tendencies. Later, Yuval et al. (2021) completed the same task ~~with~~using NNs. Both ~~works~~methods achieved stable simulations ~~in~~for coarse resolution aqua-planet GCMs. To determine why some methods can achieve stable prognostic simulations and others cannot, Brenowitz et al. (2020) proposed methods ~~to interpret~~for interpreting and ~~stabilize~~stabilizing ML parameterization ~~of~~for convection. In their ~~work~~study, a wave spectra analysis tool was introduced to explain why the ML coupled GCMs blew up.

In real-world climate models with ~~varied~~varying 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~~earlier studies demonstrated the feasibility of using neural networks ~~fitting~~to emulate 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~~fit the targets with a high accuracy and ~~is~~was successfully implemented in a single column model. Mooers et al. (2021) ~~got~~developed a high-skill DNN ~~via~~using an automated ~~machine learning~~ML technique and forced an offline land model ~~with~~using DNN emulated atmospheric fields. However, neither of these studies ~~have~~ tested their NNs prognostically for long-term simulations. Similar to the idea of using several NNs for different processes ~~in~~proposed by Yuval and O'Gorman (2020), in this study~~ uses~~, a set of NNs was used to emulate convection and cloud processes in SPCAM with ~~an~~the actual global land-ocean distribution. We ~~use~~used the residual connections ~~in~~of Han et al. (2020) to acquire super deep neural networks with a great nonlinear fitting ability. Furthermore, we ~~conduct~~conducted systematic trial-and-error analysis to filter out unstable NN~~ parameterizations~~-Parameterizations and ~~get~~to obtain the best residual deep neural network (ResDNN) set ~~with~~in terms of both accuracy and long-term stability. The NN ~~parameterization~~-Parameterization scheme ~~is~~was then implemented in ~~the~~a realistically ~~configurated~~configured CAM to obtain long-term stable simulations. ~~Technically,~~ NNs are commonly implemented ~~via~~using high-level programming languages such as Python and deep learning libraries. However, GCMs are mainly written in Fortran, making integrating them with deep learning algorithms inconvenient. Therefore, we ~~introduce~~introduced an NN-GCM coupling platform in which NN models and GCMs can interact through data transmission. This coupling strategy ~~can facilitate~~facilitates the development of ML-physical hybrid models with a high flexibility. Under real~~-geography~~ geographic boundary conditions, ~~our work achieves~~we achieved more than 10-year-long stable climate simulations in Atmospheric Model Intercomparison Project (AMIP)-style experiments ~~by~~ using a hybrid ML-physical GCM. The simulation results ~~may show~~exhibited some biases in the mean climate ~~mean~~ fields, but they successfully ~~reproduce~~reproduced the variability in SPCAM. To our knowledge, this is the first time a decade-long stable real-world climate simulation ~~is~~has been achieved ~~with a~~using an NN-based parameterization.

The remainder of this paper is organized as follows. Section 2 briefly describes the model, the experiments, the ~~DNN~~NN algorithm, and the ~~DNN~~NN-GCM coupling platform. Section 3 ~~analyses~~analyzes the simulation stability of NNCAM. Section 4 presents the offline validation of the ~~DNN~~NN scheme, focusing on the output temperature and moisture tendencies.

~~Results~~The results of the multi-year simulations~~, employing the DNN parameterization~~ conducted using the NN-Parameterization scheme~~,~~ are ~~shown~~presented in Section 5. A summary and the conclusions are presented in Section 6.

2 Methods and ~~data~~Data

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

2.1 SPCAM setup and data generation

The GCMs used in this study ~~are~~were 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~~has been 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~~for 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~~for microphysical cloud processes (Morrison and Gettelman, 2008; Gettelman et al., 2010). In the AMIP experiments we conducted, CAM5 ~~is~~was coupled to ~~a land surface model~~ the Community Land Model version 4.0 land surface model (Oleson et al., 2010) and ~~uses~~the prescribed sea surface temperatures and sea ice concentrations~~.~~ were used.

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

2.2 NN-Parameterization

2.2.1 ~~Data sets~~Datasets

The NN-Parameterization is a deep learning emulator of the SP and its cloud-scale radiation effects in SPCAM. Therefore, the inputs of this emulator are borrowed from the SP input variables, such as the grid-scale state variables and forcings, including the 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~~selected the surface pressure P_s and solar insolation (SOLIN) at the top of the model from the radiation module. The outputs of the NN-Parameterization are subgrid-scale tendencies of the moisture $\left(\frac{\partial q_v}{\partial t}\right)$ and ~~of temperature~~ ~~$\left(\frac{\partial T}{\partial t}\right)$~~dry static energy $\left(\frac{\partial s}{\partial t}\right)$ at each model level ~~as well as~~. It should be noted that $\left(\frac{\partial s}{\partial t}\right)$ is the sum of the heating from the moist processes in the SP and the heating from the SP radiation (shortwave heating *qrs* plus longwave heating *qrl*). To complete the emulation of the cloud radiation process, apart from the commonly used net shortwave and longwave radiative fluxes at both the surface and the Top of the Atmosphere (TOA~~. This heating is composed of moist heating in the SP and the GCM-grid averaged SP radiative heating. Also,~~) (Rasp et al., 2018; Mooers et al., 2021), it is ~~important~~essential to include direct and diffuse downwelling solar radiation fluxes as output variables in order to force the coupled land surface model. Specifically, they are the 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~~).~~) fluxes. In the end, the precipitation is derived from column integration of the predicted moisture tendency to ~~keep~~ensure basic water conservation.

The large-scale forcings ~~are commonly~~were often not included in previous studies ~~with~~that used an aqua-planet configuration. However, under a realistic configuration, such forcings are composed of the dynamics and the planetary boundary layer diffusion, ~~thereby carrying~~and thus, they carry critical information about the complex background circulations and surface ~~condition~~conditions. Similarly, ~~those~~the downwelling solar radiation fluxes with direct separation ~~of direct~~ versus diffusion ~~records~~record the ~~received~~ solar energy received by the coupled surface model ~~with~~for different land cover types and processes (Mooers et al., 2021). If they are not included, the land surface is not heated ~~up~~ by the sun, ~~therefore,~~which seriously ~~weaking~~weakens the sea and land breeze and monsoon circulations. In this study, we used the vertical integration of the NN predicted moisture tendency as an approximation of the surface precipitation, which has also been used in previous studies (e.g., O'Gorman et al., 2018; and Han et al., 2020). In the offline validation test, we observed negative precipitation events (27% occurrence in 1 year of results). Nonetheless, 93% of the negative precipitation events had a magnitude of less than 1 mm/day. In the online prognostic runs, reasonable rainfall results (more details will be provided in Section 5) were achieved using this approximation scheme.

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

The training dataset used by all of the considered NNs ~~is~~consisted of 40% of the temporally ~~random~~randomly sampled data from the 2-year SPCAM simulation from January 1, 1997, to December 31, 1998. ~~Notably,~~ It should be noted that random sampling ~~is~~was only done in the time dimension, but not in the latitude and longitude dimensions, including all 13,824 samples from the global grid points for each selected time step. To avoid any ~~mix~~mixing or temporal connection between the training set and the offline validation set, we ~~random sample~~randomly sampled 40% ~~timesteps~~of the time steps from the SPCAM simulation in ~~the year~~ 2000 ~~as~~ to produce the offline validation set ~~in~~used for the sensitivity test.

2.2.2 A ResDNN ~~Set~~set

~~In~~During the development of the NN-Parameterization scheme, it ~~is~~was found that when different variables are used as the output of the neural network, the ~~training~~ difficulty of the training is quite different. ~~Especially~~In particular, the neural network's ability to fit the radiation heating and scalar fluxes is significantly stronger than the tendencies variables. ~~This is also found in~~ Gentine et al. (2018~~), in which~~) also reported this, and they found that the coefficient of determination (R^2) of the radiative heating tendency ~~is~~was higher than that of the moisture tendency at most model levels. We ~~believe~~think that using a single NN with one ~~target~~output to train all of the variables~~,~~ (i.e., the moisture tendency, ~~temperature~~dry static energy tendency, and radiation fluxes~~, inevitably causes~~) is possible to cause mutual interference. Since gradient descending is applied to optimize the network ~~in~~during the training, mutual interference between different ~~targets is expected to~~outputs will cause the ~~cancel out of~~ gradient directions used for the descending ~~(Crawshaw~~to cancel out (Yu et al., 2020~~; Zhang and Yang., 2021) and~~), which will ultimately affect the convergence of the network. ~~We use~~Thus, we used three different neural networks with the same hyperparameters to train

(1) the tendency of the moisture~~ and temperature, and~~;

(2) the tendency of the dry static energy; and

(3) the radiation fluxes at the surface and TOA.

It should be noted that the radiation fluxes include the net shortwave and longwave radiative fluxes at the surface (FSNS and FLNS, respectively~~.~~) and at the TOA (FSNT and FLNT, respectively) and four solar radiation fluxes (SOLS, SOLL, SOLSD, and SOLLD). By doing so, we ~~avoid~~avoided the gradient cancellation ~~between multiple targets and improve~~and improved the convergence speed and fitting accuracy when training the network. As will be described in Section 3.1, when using the same network configuration, the radiation fluxes are trained ~~much easier with~~ more easily and have a higher accuracy than the tendencies of the moisture and temperature. We admit that putting the heating and moistening ~~rate~~rates in two different NNs arbitrarily ~~cut~~cuts the physical connections between them. ~~But~~However, this separation ~~is surely doing~~makes the training ~~more easily~~easier in the ~~developing~~development stage.

In this study, to mimic the column-independent SP and its radiation effects, the input and output of the NN-Parameterization ~~have to be~~ both had to be 1-D vectors. This means that the ~~data~~ input and output of the NN-Parameterization are much simpler than those in ~~the~~ existing mainstream ~~machine learning~~ML problems, such as image recognition and text-speech recognition~~, so~~. Thus, it is impossible to directly 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~~demonstrated 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~~it is feasible~~.~~ for a DNN to map from a 122-element 1-D vector to a 1-D vector with a length of 68, which is what the NN-Parameterization does. Therefore, when constructing the NN-Parametrization, we first tried to use a DNN for the fitting~~,~~ and introduced residual connections to extend the DNN in to a ResDNN.

After numerous experiments, we ~~got~~obtained the best hyperparameters ~~of~~for the DNN and ResDNN. When training a ~~Fully~~fully connected DNN, the hidden layer width of the network should be set to 512, and the network's depth should not exceed 7~~,~~; otherwise~~ it will affect~~, the convergence of the DNN~~.~~ will be affected. In order to make the neural network capture more non-linear information, ~~enhance~~ the fitting ability~~.~~ was enhanced. We ~~introduce~~introduces skip connections to extend the 7-layer DNN to a 14-layer ResDNN. The network structure of the ResDNN is shown in Figure 1. In the training process, both the DNN and ResDNN use an initial learning rate of 0.001 and a learning rate decaying strategy ~~as~~for the cosine annealing (Loshchilov et al., 2016) without dropout and L2 regularization. Adam (Kingma and Ba, 2014) ~~is~~was chosen as the optimizer to minimize the mean ~~squared~~square errors (MSEs). ~~The results in~~ The specific hyperparameter searching space of the DNN and ResDNN is documented in Table S1.

Figure 2 ~~show~~shows that the ResDNN ~~can fit~~fits the data ~~is~~ significantly better than the 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~~. We chose ~~ResDNNs~~ResDNN sets as stable candidates to build the NN-Parameterization. After obtaining ~~all~~ well-fit ResDNN sets, the next step is to couple the candidates into NNCAM one by one for the prognostic tests and to find the sets that can support a stable simulation.~~ To complete this extremely challenging task, we have more than 50 prognostic tests.~~ All of the experiments and analyses ~~on~~related to the stability will be introduced in ~~section~~Section 3~~ as well~~.

2.2.3 Implementation of NN-Parameterization

The NN-Parameterization is implemented into SPCAM to replace both the CRM~~-~~-based superparameterization and its radiation effects based on the ~~basis~~average of the coarse grid~~ average. In~~. At the beginning of each ~~timestep~~time step, NNCAM calls the

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

2.3 ~~The~~ NN-GCM ~~Coupler~~coupler

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

To implement ~~a~~an NN-~~-~~based ~~Parameterization~~parameterization in the current climate ~~model~~models, which ~~is~~are mostly developed in Fortran, many researchers ~~try~~have attempted to ~~get~~obtain the network parameters (e.g., the weight~~,~~ and bias) from the ~~machine learning~~ML models and implement the NN models (e.g., DNNs) ~~with~~using hard coding in Fortran. At the runtime, NNCAM will call ~~a~~an 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~~in GCMs (Ott et al., 2020). This interface can import neural network parameters from outside of the Fortran program, and the Fortran-based implementation ensures that it can be flexibly deployed in GCMs. However, embedding ~~a~~an NN~~-parameterization~~-Parameterization in NNCAM is still a troublesome task ~~with~~, and there is no existing coupling framework to support many of the latest network structures. This problem ~~will restrict developers~~prevents researchers from building more powerful NNs and deploying them in NNCAM.

We ~~develop the~~developed a coupler to bridge the 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,~~3), the coupler receives the state and forcing output from the dynamic core in the Fortran-~~-~~based CAM5. For each input variable, we ~~use~~used the native Message Passing Interface (MPI) interface in

CAM5 to gather the data ~~of~~for all of the processes ~~to~~into the master process into a tensor. Then, ~~as shown in ② of Figure 5,~~ the coupler ~~will transmit~~transmits the gathered tensor through the data buffer to the NN-Parameterization running on the same node as the master process~~.~~ (② in Figure 3). The NN-Parameterization ~~gets~~obtains the input, infers the outputs, and transmits them back to the coupler. As shown in ③ ~~of~~in Figure ~~5~~3, the coupler ~~will first write~~writes these tendencies and radiation fluxes back to the master process, and then~~ broadcast~~, it broadcasts the data to the CAM5 processes running on the computing nodes through the MPI transmission interface. Therefore, other parameterizations ~~get~~obtain the predictions from the NN-Parameterization to complete the follow-up procedures (④ in Figure ~~5).~~ 3).

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

All neural network models deployed ~~through~~using the NN-GCM ~~Coupler~~coupler can support a Graphics Processing Unit (GPU) accelerated inference to achieve excellent computing performance. In this study, we ran SPCAM and NNCAM on 192 CPU cores. NNCAM also used ~~2~~two GPUs for acceleration. During the NNCAM runtime, each time step of NNCAM requires the 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~~20 MB for CAM5 with ~~the~~a horizontal resolution of 1.9°×2.5°) that needs to be transmitted for each communication~~,~~ and ~~determined~~decided to establish a data buffer on a high-speed solid-state drive to ensure a balance ~~of~~between 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~~simulation years per ~~Day~~day (SYPD) of NNCAM based on the NN-GCM ~~Coupler has~~coupler represents an impressive performance improvement~~, when~~. When using 192 Intel CPU cores, the SYPD of SPCAM is 0.3, the SYPD of CAM5 is 20, and the SYPD of NNCAM is 10. It ~~is worth noting~~should be noted that~~,~~ NNCAM based on the NN-GCM coupler uses an additional GPU to accelerate the NN-Parameterization. When the NN-GCM ~~Coupler~~coupler is not used, the NN-Parameterization is implemented ~~by~~using Fortran and is accelerated by the Fortran-based Math Kernel Library, and the SYPD is 1.5.

3 A Road to Stability

3.1 ~~Sensitivity Tests and~~ Trial-and-error

To develop a stable NN~~ parameterization~~-Parameterization, we propose ~~a ResDNN set, where~~the use of a set of three ResDNNs, in which each neural network is responsible for predicting a class of variables (see ~~section~~Section 2.2.2). ~~One may wonder~~

~~whether the ResDNN architecture is necessary and whether offline accuracy of NNs matters in~~ Ott et al. (2020) demonstrated that there is a negative correlation between the offline MSE and 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~~ when using tendencies 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~~outputs in ~~GPUs (Yuval et al., 2021), and 1D CNN is not widely~~aqua-planet simulations. Since we also used tendencies as outputs in ~~other studies except Han et al. (2020) with unknown prognostic performance.~~ ~~The prognostic tests of~~the real-world simulations, we conclude that an NN-based parameterization ~~begin at 1998-01-01 as~~that can support long-term integration should have a ~~startup.~~high accuracy regarding training and validation. As ~~initialization, calling the SP~~was described in ~~SPCAM at the~~Section 2.2.2, we tried DNNs first~~ step is required to generate the correct largescale forcings as the input for NN~~, and then, we extended the DNNs to ResDNNs to achieve a high offline accuracy (Figure 2). Even through more accurate ResDNNs have a higher probability of becoming stable parameterizations (Figure 4), we still do not have a way to determine the stability a priori. Therefore, we still used the trial-and-error method to filter out unstable ones and then selected the best ResDNN set that could reduplicate the total energy time evolution of SPCAM with the least deviation, i.e., the NN-Parameterization.

3.2 Sensitivity tests

We conducted prognostic runs of all three neural networks in each NN set using the NN-GCM coupler. To demonstrate the reality behind the relationships between the offline accuracy and online stability under a real-world configuration, we conducted sensitivity tests using 10 DNN sets and dozens of ResDNN sets and conducted the training and evaluation using the settings described in Section 2.2.2. In the sensitivity ~~test~~tests, we ~~freeze~~conducted prognostic runs (see details in Section 3.2) using all three neural networks in each NN set using the NN-GCM coupler.

First, we selected the best ResDNN for the ~~8~~ radiation fluxes ~~to simplify the neural network choices,~~at the surface and the TOA that was shared in every NN set since their offline validation ~~is extremely~~was exceptionally accurate with R^2 ~~above~~≥ 0.98 over 50 training epochs (Figure 2b). ~~Different from~~In contrast to the accurately trained radiation fluxes, the tendencies of ~~temperature~~the dry static energy and moisture are less accurate and can ~~hypothetically~~ affect the prognostic performance. To evaluate ~~the tendency of moistening and heating in~~ those two tendencies using one metric, we ~~introduce~~introduced the MSE of the rate of change of the moist static energy ~~changing rate~~ $(dh = \cancel{C_p dT} ds + L_v dq_v$ ~~) as:~~):

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

where g is the acceleration due to gravity~~ constant~~, C_p ~~refers to~~is 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~~for dq_v and ds were trained from ~~5~~five epochs to 50 epochs, ~~carrying~~resulting in different offline validation ~~accuracy.~~ accuracies. We used the maximum number of steps until the model crashed to measure the prognostic performance.

Figure 4 shows the offline validation MSE_h versus the maximum prognostic steps. ~~First,~~The DNN~~-~~-parameterizations (blue triangles) are ~~systematic~~systematically less accurate than the ResDNN~~ ones ~~-parameterizations (blue dots and black inverted triangles), which is consistent with Figure 2a. They ~~cannot run stably~~could not sustain half a year of simulation in the 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~~MSEs also crashed after short simulation ~~period than DNNs.~~periods. However, when the offline MSE ~~decreases~~decreased to a certain level (e.g., 290 W^2/m^4), ~~some~~10 of the ResDNN~~-~~-parameterizations ~~are~~were stable ~~for extreme~~in long-term simulations~~, while others remain unstable.~~

~~Generally, a NN parameterization~~ of over 10 years (black inverted triangles). We speculate 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 ~~the more ~~competent~~accurate ResDNN sets have a higher probability of becoming stable NN-Parameterizations since all of the stable NN-Parameterizations are ResDNNs.

A few unstable ResDNN sets are equally or more accurate than ~~DNNs in this job. On the other hand,~~the stable ones. Previous studies ~~showed~~have shown that high-capacity (more hidden layers and more weights and biases) models are harder to train and are more likely to ~~overfit~~ produce overfitting (Goodfellow et al., 2016). ~~Thus, the prognostic stability differences between less well-trained ResDNNs and the well-trained ones are drastic compared with DNNs. Also, some~~Some overly trained ResDNNs with lowest validation loss are speculated to ~~overfit. Those~~produce overfitting~~ models~~, and therefore, they 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~~causing the model to crash.

In the time evolution of the ~~global~~globally averaged total energy (Figure 5~~). The~~), 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 of fluctuations in SPCAM. Among the stable ResDNN ~~schemes~~sets, some can ~~get nearly a perfect reproduction of~~almost perfectly reproduce the total energy evolution of SPCAM (the blue line~~), while~~). However, some inaccurately simulate the climate state with a ~~large~~significant 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~~Apart from global averages, the prognostic results of the 10 stable 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~~sets vary from each other in terms of the global distribution. Figure S1 shows the precipitation spread across all of the stable NN sets for the prognostic simulation from 1999 to 2003. The obvious standard deviation centers coincide with the heavy tropical precipitation areas.

3.~~2~~3 Gravity ~~Wave Diagnosis~~wave diagnosis

It is still ~~a question of~~unclear why unstable NN~~ parameterizations ~~-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 NN-GCM system. Brenowitz et al. (2020) offered several interpretations. When an unstable NN~~ parameterization~~-Parameterization is coupled with dynamics, it tends to amplify any unrealistic ~~perturbation ~~perturbations caused by emulation errors and to pass it to the entire system through gravity waves. In contrast, the stable NN~~ parameterizations~~-Parameterizations tend to dump all of the ~~perturbs~~perturbations quickly. This ~~is~~was found to be true in our study ~~with~~for the realistic configuration. Such unstable gravity waves ~~are~~were 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 the more intense waves ~~afterward ~~with a perfectly round shape~~.~~ after this point in time. Additionally, we found that our instable waves mostly occurred in the tropics, which is different from the mid-latitude instability that occurs when using ML parameterizations in aqua-planet simulations (Brenowitz et al~~.~~., 2020).

Brenowitz et al. (2020) also introduced an analysis tool that calculates the wave energy spectra of a hierarchy model that couples the ~~linear response functions~~Linear Response Functions (LRF) of ~~a~~an NN-based parameterization to a simplified two-dimensional linear dynamic system, ~~where~~in which perturbations can propagate in ~~2D~~2-D gravity waves. We ~~apply~~applied the tool in ~~our~~this study and ~~detect~~detected similar results ~~of~~in the unstable mode for the unstable ResDNN with a positive energy growth rate across all wave numbers at phase ~~speed between~~speeds of 5~~ m/s to ~~-20 m/s (Figure ~~S1b). While~~S2b). In contrast, the stable ResDNN ~~shows~~exhibited a stable mode ~~with~~for the growth rate of nearly all wave numbers and phases below zero (Figure ~~S1a~~S2a).

4 Offline Validation of NN-Parameterization

Before evaluating the prognostic results, ~~demonstration of~~the offline performance with geographic information ~~is needed~~needs to be demonstrated for the following purposes: 1) ~~To~~to show how well our NN-Parameterization emulates the SP ~~in~~for a realistic configuration compared with the baseline CAM5 physics and ~~with ~~previous studies~~.~~; and 2) ~~To~~to reveal the ~~strength~~strengths and ~~weakness~~weaknesses of the NN emulations with the correct input~~, give~~ and to provide clues to the analysis of the prognostic results in the following section. We performed offline testing ~~with~~using a realistically ~~configurated~~configured SPCAM from January ~~1st~~1, 1999, to December ~~31st~~31, 2000, ~~where~~in which the NN-Parameterization ~~is~~was diagnostically run ~~paralleled~~parallel to the SP~~,~~ and so ~~does~~was the CAM5 physics. The results ~~over~~for the entire second year of the simulation period ~~are~~were chosen for evaluation, which was completely independent from the training dataset. Following the conventions ~~in~~of Han et al. (2020) and Mooers et al. (2021), we ~~choose~~used the mean fields and the coefficient of determination (R^2) as the ~~two~~evaluation metrics~~ for evaluations.~~. It should be noted that the NN-Parameterization was tuned to emulate the SP, and the CAM's parameterization was tuned to obtain close results to the observations. The latter is merely introduced as a baseline.

The mean diabatic heating and drying rates produced by convection~~ and~~, large-scale condensation, and cloud radiation effects in SPCAM and the 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~~exhibits the maximum latent heating in the deep troposphere, corresponding to the deep convection ~~at~~in the ITCZ. In the subtropics, ~~there is~~ heating and moistening occur in the lower troposphere, corresponding to the stratocumulus and shallow convection in the subtropics. In the ~~midlatitudes~~mid-latitudes, there is a secondary heating maximum below 400 hPa due to ~~midlatitude~~the mid-latitude storm tracks. All of these features are well reproduced by the NN-Parameterization. ~~Note~~It should be noted that in the ~~midtroposphere~~mid-troposphere, the ~~ITCZ~~ peak in the drying ~~rates~~ rate in the ITCZ is slightly weaker in the NN-Parameterization ~~compared with that of~~than in SPCAM (~~Figure~~Figures 6c and 6d).

In addition to the mean fields, the high prediction skill of the NN-Parameterization is also ~~shown in~~demonstrated by the spatial distribution of the R^2~~.~~ values. To ~~demonstrate~~ illustrate the R^2 ~~for~~values of the ~~3D~~3-D variables such as the diabatic heating and moistening, ~~same as~~following Mooers et al. (~~2020),~~2021), the zonal averages ~~are~~were calculated in advance before the R^2 calculation for each location in the pressure-latitude cross-section. For the diabatic heating, the R^2 value is ~~above~~ >0.7 ~~over~~throughout the ~~entire mid to low~~middle and lower troposphere, and the high skill regions with R^2 values of greater than 0.9 ~~concentrates~~are concentrated in the low levels but ~~are extended to~~ extend into the mid-troposphere in the storm tracks (Figure 7a). ~~As for~~For the moistening rate, the high skill zones ~~concentrate~~are concentrated in the ~~mid to~~middle and upper troposphere (Figure 7b), ~~leaving~~with low skill areas below. ~~Those~~The regions with ~~low accuracy~~lower accuracies are generally located in the ~~mid to low~~ middle and lower troposphere in the tropics and subtropics, ~~corresponding~~which correspond to the deep convection ~~at~~in the ITCZ and the shallow convection in the subtropics. Nonetheless, the tendencies ~~from~~of the diagnostic CAM5 parameterization ~~hardly draw any similarity~~are not similar to those simulated by the SP, except for a few locations in the ~~mid to~~middle and upper troposphere in the tropics and polar regions (~~Figure~~Figures 7c ~~&~~and 7d).

The global distribution of the R^2 ~~for~~values of the precipitation predictions is shown in Figure 8. Our NN-Parameterization ~~shows a great prediction skill globally~~produced excellent predictions in most of the in mid- and high-latitude regions, especially in the ~~midlatitude~~ storm tracks. ~~The~~However, the prediction skill is relatively low in many of the ocean areas between 30°S ~~to~~and 30°N and in some ~~midlatitude~~ mid-latitude areas over continents (Figure 8a~~), in~~). In particular, the results are not ideal ~~in the ITCZ deep convection regions. Moreover, for shallow convection in Subtropical~~ along the equatorial regions, in the subtropical Eastern Pacific, and ~~Subtropical~~in the subtropical Eastern Atlantic~~, the precipitation prediction skill hits bottom, corresponding to the subtropical~~. These areas correspond to the low skill zones ~~for~~of the moistening rate ~~(Figure 6b). On the other hand~~in the middle and lower troposphere from the equator to the subtropics (Figure 7b). As a baseline, the total precipitation simulated ~~by~~using the CAM5 parameterizations is much less analogous to the SP than the NN-Parameterization

~~with~~and has a systematically lower accuracy globally. The CAM5 precipitation can ~~reach~~achieve a relatively high accuracy along the mid-latitude storm tracks, but ~~fail~~it fails in most regions in the tropics~~.~~ (Figure 8b).

Generally, the NN-Parameterization ~~performs~~performed far better than the CAM5 parameterization in the 1-year period in the offline testing, and ~~shows similar~~ it had an accuracy ~~as~~similar to that of the DNN ~~in~~used by Mooers et al. (~~2020~~2021). The use of real~~-geography~~ geographic data can significantly decrease the emulation skill of a deep learning model (Mooers et al., 2021~~), where~~). This is because the convection backgrounds of real geographic data are much more complex with meridional and zonal asymmetric and ~~seasonal variated~~seasonally varying circulations~~, not to mention~~. In addition, the ~~orograph~~orography and various types of underlying land ~~surface. In that~~surfaces also add complexity. In this case, the ResDNN is a valuable NN architecture that ~~can bring good performance~~performs well as ~~the~~an automated hyperparameter tuning algorithm ~~without searching~~that does not need to search for hundreds of NN candidates. Still, our NN-Parameterization ~~is exposed to~~produced low accuracy predictions ~~in subtropical shallow~~along the equator over the oceans where the convection ~~areas, a great challenge for machine learning emulation of moistening rate and precipitation. 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~~ complex and vigorous and in subtropical ocean areas where the convection is ~~not sensitive to small values.~~

~~**5 Long-term Prognostic Validation**~~

weak and concentrated at low levels. This indicates that the NN-Parameterization is ~~selected for~~still inadequate in rems of its emulation skill when simulating various types of deep and shallow convection in the tropics.

5 Long-Term Prognostic Validation

The NN-Parameterization produced the best prognostic performance in Section 3.1. It ~~is~~was coupled in the ~~realistic configurated~~realistically configured SPCAM to replace the SP and its cloud-scale radiation effects. This coupled model is called referred to as NNCAM ~~afterwards~~hereinafter and is compared with SPCAM and CAM5. ~~All~~ The start time of all three model ~~starts at~~was January ~~1st~~1, 1998~~ as start up~~. They ~~are~~were all run for 6 years with the first year for spin up and the next 5 years ~~from~~ (January ~~1st~~1, 1999, to December ~~31st~~31, 2003) for evaluation and comparison. Later, the simulation of NNCAM ~~is~~was extended for another 5 years to December ~~31st~~31, 2008, to ~~show~~demonstrate its stability. Due to the excessive computing resources ~~consumption~~required, the SPCAM simulation ~~of SPCAM is~~was not ~~get~~ extended. In the analysis of the prognostic results, the following ~~are~~were selected ~~for demonstration of~~to demonstrate the multi-year climatology and variability:~~ multi-year~~

(1) The mean ~~fields of~~ temperature and humidity~~,~~ fields;

(2) The mean precipitation~~,~~ field;

(3) The precipitation frequency distribution~~,~~; and~~ the~~

(4) The Madden–Julian Oscillation.

As was mentioned in the introduction section, SPCAM, which uses the 2-D SAM as the SP, does not simulate better mean climate states than its host coarse-grid model CAM5, but it excels in climate variability. What is remarkable about NNCAM is not its performance in simulating the mean climate, but its ability to achieve a stable multi-year prognostic simulation under a real-world global land–ocean distribution. The advantages and problems of this study will provide important references for future research on NN-based stable long-term model integrations.

5.1 Climatology

5.1.1 Vertical profiles of temperature and humidity

In this section, we ~~first~~ evaluate the vertical ~~structure~~structures of the mean temperature and humidity fields. Figure 9 shows the zonally averaged vertical profiles of the air temperature and specific humidity ~~as~~ simulated ~~by the~~using NNCAM and ~~the~~ CAM5~~, in contrast~~ compared to the SPCAM simulations. Overall, ~~the~~ NNCAM ~~simulate~~simulated reasonable thermal and moisture ~~structure~~structures. However, ~~it is shown that NNCAM has some biases in~~the multi-year mean ~~fields of~~ temperature and ~~humidity~~moisture fields produced by NNCAM are more biased than those produced by CAM5, which is ~~shown as~~ reflected by the larger root mean ~~squared~~square errors (RMSEs) ~~or~~(Figure 9) and larger differences ~~than~~compared to those of CAM5 (Figure ~~S2).~~ 10). The larger deviations are temperature biases in the tropopause~~, where~~. In this region, 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 the polar regions in NNCAM. For the humidity field, there are slight dry biases over the equator and wet biases elsewhere in NNCAM. Even with ~~the~~these biases, the mean climate ~~mean~~ states are consistent with those in the last 5~~-year~~ years of the simulation for NNCAM (Figure S3), which indicates ~~almost no~~that the climate ~~drift~~states simulated by NNCAM are constant in the long-term simulation.

5.1.2 Precipitation

Figure ~~10~~11 shows the spatial distributions of the winter (December-January-February) and summer (June-July-August) mean precipitation simulated ~~by~~using SPCAM, NNCAM, and CAM5. The SPCAM simulation results are regarded as the reference precipitation. In SPCAM (~~Figure 10a~~Figures 11a and ~~10b~~11b), massive precipitation can be ~~found~~observed in ~~regions of~~the Asian monsoon region and ~~midlatitude~~the mid-latitude storm tracks over the northwest Pacific and Atlantic oceans. In the tropics, the primary peaks ~~of~~in the rainfall ~~are~~occur in the eastern Indian Ocean and Maritime Continent regions. ~~Furthermore~~In addition, 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 at around 5~~°S –~~-10°S near the western Pacific warm pool region and ~~experiences a southeast tilt~~tilts southeastward as it extends eastward into the central Pacific. The main spatial patterns of the SPCAM precipitation ~~climatology~~ are properly reproduced by both NNCAM and CAM5. ~~In~~For NNCAM, the strong rainfall centers are well simulated over the tropical land regions ~~over~~of the Maritime Continent, the Asian monsoon region, ~~and~~ South America, and Africa (~~Figure 10c~~Figures 11c and ~~10d~~11d). In addition, the heavy summertime precipitation over the Northwestern Pacific simulated by SPCAM is well represented ~~in~~by NNCAM (~~Figure 10a~~Figures 11a and ~~10c). In~~11c). For CAM5, there is too little precipitation over ~~that~~this area (Figure ~~10e~~11e). Moreover, NNCAM ~~can maintain~~maintained the spatial pattern and global average of the precipitation in the next 5~~-year~~ years of the simulation, ~~reassuring~~demonstrating its long-term stability (Figure S4).

Generally, the NNCAM ~~draws~~results are more ~~similarity~~similar to SPCAM than the CAM5 results in terms of the spatial distribution of the summertime ~~multiyear~~multi-year precipitation, with smaller ~~RMSE~~RMSEs and ~~global~~globally averaged biases. However, ~~in the~~on a difference plot (Figure ~~11~~S5), NNCAM moderately underestimates the precipitation along the equator, in the Indian monsoon region, and ~~maritime continent~~over the Maritime Continent in the summer (Figure ~~11a~~S5a). In the boreal winter, NNCAM simulates a weak ~~and~~SPCZ that is excessively separated ~~SPCZ~~ from the ITCZ, with both precipitation centers ~~shifting~~shifted away from each other. As a result, ~~we detect~~ underestimation occurs in the equatorial regions of the ~~maritime continent as well as~~Maritime Continent and in the SPCZ~~ but~~, while overestimation ~~on~~occurs to the north of the equator in the ~~West~~Western Pacific (Figure ~~11b), which makes~~S5b), and thus, NNCAM ~~less resemble~~resembles SPCAM less 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~~underestimation of ~~NNCAM predictions~~the precipitation along the equator can also be observed in the zonal mean multi-year precipitation plots (Figure 12). There is a more significant minimum zone in the equatorial precipitation near the equator compared with in SPCAM and CAM5 for the annual average (Figure 12a) and the boreal winter average (Figure 12c).

In contrast to the oceanic rainfall, NNCAM predicts the precipitation over the land surfaces with good skill in the tropics (land fraction equal to 1), which resembles the tropical land rainfall intensity of SPCAM and Tropical Rainfall Measuring Mission (TRMM) observations of the annual and boreal summer averages (Figures 12d and 12e). According to Kooperman et al. (2016), SPCAM predicts the Asian and African Monsoon activity better, which leads to the more accurate land rainfall in such areas. This is related to the stronger convective variability in the SP than the conventional parameterizations. As an emulator of SPCAM, NNCAM inherits this strength.

5.2 Variability

5.2.1 Frequency ~~Distribution~~distribution of ~~Precipitation~~precipitation

Moreover, NNCAM ~~shows~~exhibited a better performance in simulating the precipitation extremes. Figure ~~12~~13 shows the probability ~~densities~~density function of the simulated daily precipitation in the tropics (30°S−30°N) with a precipitation intensity interval of 1 mm day^{-1}. ~~In~~For CAM5, the heavy precipitation events exceeding 20 mm day^{-1} are greatly underestimated. In addition, for CAM5, the 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~~and no explicitly resolved ~~convections~~convection (Holloway et al., 2012). Compared with CAM5, the spectral distribution of the precipitation ~~in~~for NNCAM is much closer to that of SPCAM. The heavy rainfall events are substantially enhanced, and the overestimated moderate precipitation ~~occurrence between~~ (2−20 mm day^{-1}) is reduced, with no spurious peak at around 10 mm day^{-1}.

5.2.2 The MJO

The MJO is a crucial tropical intraseasonal variability ~~at the~~that occurs on a time scale of 20–100 days (Wheeler and Kiladis, 1999). Figure ~~13~~14 presents the wavenumber and frequency spectra for the daily equatorial precipitation ~~daily~~ anomalies ~~from~~for SPCAM, NNCAM, and CAM5 in ~~4~~four consecutive boreal ~~winter~~winters from 1999 to 2003. SPCAM shows widespread power signals over ~~zonal number of~~ zones 1–4 and periods ~~between~~of 20–100 ~~plus~~days, as well as a peak around ~~at zonal~~zone numbers ~~of~~ 1−3 and periods of 70–100 ~~day~~ days for the eastward propagation (Figure ~~13a~~14a). Similarly, ~~in~~for NNCAM, there is a spectral peak at ~~the~~ wavenumbers of 1−2 and periods of 50–80 ~~day~~days for ~~east~~the eastward propagation (Figure ~~13b~~14b), exhibiting intense intraseasonal signals. For CAM5 (Figure ~~13c~~14c), the spectral power is concentrated around 30 ~~day~~ days and exhibits more extended periods (greater than 80 days) at a wavenumber of 1 for the eastward propagation. In addition, CAM5 also shows signals of westward propagation ~~around~~with a 30-day period. Compared with CAM5, NNCAM ~~shows~~exhibits stronger intraseasonal power and resembles SPCAM better. To quantify this similarity, we ~~calculate~~calculated the ~~coefficient~~coefficients of determination R^2 ~~of~~for the precipitation ~~spectrum in~~spectra of NNCAM and CAM5~~,~~ using the spectrum ~~in~~of SPCAM as the target value. The R^2 value of the precipitation spectrum ~~R^2 in~~ NNCAM (0.51) is much higher than that ~~in~~for CAM5 (0.40).

The MJO is characterized by the eastward propagation of deep convective structures along the equator. Generally, it ~~generally~~ forms over the Indian Ocean, strengthens over the Pacific, and weakens ~~in~~over the eastern Pacific due to ~~interaction~~interactions with cooler SSTs (Madden and Julian, 1972). Figure ~~14~~15 presents the longitude-time lag evolution ~~of~~for the 10°S−10°N meridional averaged daily anomalies of the intraseasonal (filtered ~~with~~using a 20–100-day bandpass) precipitation and 200 hPa zonal wind (U200) in the boreal winter. The results show that both SPCAM and NNCAM reasonably reproduce the eastward ~~propagating~~propagation of the convection from the Indian Ocean across the Maritime Continent ~~to~~and into the Pacific (~~Figure 14a~~Figures 15a and ~~14b),~~15b). This is confirmed by both the precipitation field and U200 field. Therefore, we conclude that NNCAM captures the key MJO propagation simulated ~~in~~by SPCAM. In contrast, the time lag plot ~~of~~for CAM5 depicts an ~~unpleasant west~~inaccurate westward propagation. ~~Same as~~Similar to the precipitation spectrum, the R^2 value of the time lag coefficient is shown to quantify the ~~resemblance.~~similarities between the simulations. The time lag coefficient of the U200 ~~in~~field for NNCAM is much closer to that for SPCAM than CAM5, with a ~~way~~much higher R^2~~,~~ value, indicating that the NN-Parameterization successfully emulates the convection variability of the SP~~ and reflects it~~, which is reflected in the dynamic fields.

6 Summary and Conclusions

~~This~~In this study ~~investigates~~, the potential of deep neural network~~-~~-based parameterizations in SPCAM to reproduce long-term climatology and climate variability~~.~~ was investigated. We ~~present~~developed an NN-Parameterization~~,~~ via a ResDNN set~~,~~ to emulate the SP with a ~~2D~~2-D CRM and its cloud scale radiation ~~in effects in~~ for a ~~realistic configurated~~realistically configured SPCAM with a true land-ocean distribution and orography. The input variables ~~to~~of the NN-Parameterization include the specific humidity, temperature, largescale water vapor and temperature forcings, surface pressure, and solar insolation. The output variables of the NN-Parameterization ~~consist of~~include the subgrid tendencies of the moisture and ~~temperature, net~~ dry static energy and the radiation fluxes ~~at the top of the model and surface, and solar radiation fluxes down to the surface.~~. We ~~proposed~~propose a set of 14-layer deep residual neural networks, in which each NN is in charge of one ~~type~~group of output ~~variable~~variables. With such a design, we ~~gain the~~gained a best emulation accuracy for each predictor. ~~Via a~~Through systematic trial-and-error searching ~~procedure~~, we ~~are~~were able to ~~firstly~~ select sets of ResDNNs that support stable prognostic climate simulations, and then ~~choose~~, we chose the best set with the lowest climate errors as the formal NN-Parameterization. Moreover, ~~a~~the mechanism of the unreal perturbation amplification ~~is found~~was identified in the GCM simulations with unstable NN ~~parameterizations with~~ Parameterizations using the spectrum diagnostic tool invented ~~in~~by Brenowitz et al. (2020).

The offline ~~test shows~~tests demonstrated the ~~great~~good skills of the NN-Parameterization in emulating the SP outputs and ~~its~~the cloud scale radiation effects ~~in~~of SPCAM. The overall diabatic heating and drying rates in the NN-Parameterization and SPCAM are in close agreement. When implemented in the host SPCAM to replace its time-consuming SP and its radiation effects, the NN-Parameterization ~~successes in~~successfully produced an extensive stable long-term ~~stable~~ prognostic simulation and ~~predicts~~predicted reasonable mean vertical ~~structures in~~ temperature and humidity~~,~~ structures and ~~the~~ precipitation distributions. Compared with the SPCAM target simulation, NNCAM still produces some biases in the mean fields, such as a warmer troposphere over the polar regions and in the tropopause and underestimation of strong precipitation ~~underestimation~~ in the equatorial regions. ~~On the other hand~~In addition, the better climate variability ~~in~~of SPCAM ~~over~~compared to CAM5 ~~is well~~was learned well by our NN-Parameterization and was reproduced ~~in~~by NNCAM, with better frequency ~~in~~for extreme rainfall~~,~~ and a similar MJO spectrum ~~and~~, propagation direction, and speed. ~~Although with~~Despite the current biases in the climate states ~~so far~~, NNCAM can still be regarded as ~~the~~a first attempt to ~~prognostically~~ couple a NN-based parameterization ~~in realistic configurated 3D~~and a realistically configured 3-D GCM.

Many previous studies have ~~well-studied machine learning~~investigated ML parameterizations implemented in aqua-planet ~~configurated 3D GCM.~~configured 3-D GCMs. Some faced instability problems in coupled simulations (Brenowitz and Bretherton, 2019), while ~~some~~others succeeded in producing stable long-term ~~stable~~ prognostic simulations with deep fully-connected neural networks (Rasp et al., 2018; Yuval et al., 2021~~)~~), as well as random forest algorithms (Yuval and O'Gorman, 2020). In contrast to aqua-planet simulations, the spatial heterogeneity is prominent over the land in GCMs, which are ~~configurated with~~ configured using real ~~geography~~ geographic boundary conditions. ~~In this case, a plain fully connected neural~~

~~networks the SP output (Mooers et al., 2021).~~ The convection, clouds, and interactions with the ~~interacted~~ radiation ~~of~~in the CRM ~~together with~~ and the real~~-geography~~ geographic boundary conditions are without a doubt far more complicated than in idealized models. To meet the new demand ~~under~~for realistic ~~configuration~~configurations, we ~~design~~designed a ResDNN with sufficient depth to further improve the nonlinear fitting ability of the NN-Parameterization. With the skip connections, the 7-layer DNN models can be extended to 14 layers, ~~therefore, significant~~thereby significantly improving the offline accuracy. In the prognostic tests, a ~~few~~dozen ResDNN~~-~~-parameterizations ~~can support~~ supported a stable long~~-~~-term ~~stable~~ run, while all of the DNN~~-~~-parameterizations ~~are so far~~tested were found to be unstable.

Trial-and-error is still ~~the~~our only way to find stable NN-based parameterizations. ~~So~~Thus far, we have not ~~come up with~~developed an a priori ~~mothed~~method that ~~guaranteed~~guarantees stability. However, we ~~do~~did find some clues in the sensitivity tests. We believe sufficient offline accuracy is essential for online stability and can be achieved by confirming all of the inaccurate NN~~-parameterizations~~ -Parameterizations as unstable. ~~On the other hand~~In addition, some of the highly accurate ones still crash the prognostic simulation~~, where we find rapid increasing~~. In this case, the total energy~~.~~ was found to increase rapidly. This mechanism is that unstable NNs cannot damp the neural network emulation errors~~ but~~, and they amplify and propagate them to the entire system through gravity waves.

The prognostic biases ~~in~~of the mean fields ~~in~~are speculated ~~as~~to be a result of ~~by~~ the combined effect of the emulation errors of all of the NN-Parameterization prediction fields. Further study is required. Still, it ~~can~~may be related to the spatially non-uniform accuracy of the NN-Parameterization, such as the relatively low fitting accuracy in the tropical deep convective regions and the 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~~an NN~~-parameterization~~-Parameterization with heterogeneous characteristics across different regions, rather than a globally uniform scheme, can further improve the fitting accuracy in ~~this~~these tropical and subtropical ~~region~~regions.

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~~method is to use Fortran~~-~~-based neural network libraries that can flexibly import network parameters (Ott et al., 2020). These methods have been used to successfully ~~implemented NN~~implement NNs in ~~GCM~~GCMs, but they can only support dense, layer~~-~~-based ~~NN~~NNs. As a result, developers cannot take advantage of the most advanced neural network structures, such as convolution, shortcut, self-attention, and variational autoencoder~~, etc.,~~ structures, to build powerful ~~DNN~~ ML-based ~~Parameterizations.~~parameterizations. In this ~~research, through~~study, using an NN-GCM ~~Coupler,~~coupler, the NN-Parameterization ~~can~~could support the mainstream GPU-enabled ~~machine learning~~ML frameworks. Thanks to the simple and effective implementation of the ~~DNN~~NN-GCM ~~Coupler~~coupler, our NNCAM ~~achieves~~ achieved an SYPD 30 times ~~SYPD compared to~~that of SPCAM by using a ResDNN set ~~in~~and NN-Parameterization, ~~although~~even though these DNNs are much deeper than the previous state-of-the-art fully~~-~~ connected NNs in this field.

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 and NNCAM have been archived~~,~~ and made publicly available for downloading from https://doi.org/10.5281/zenodo.5596273.

Competing interests. The authors declare no conflict of interest.

Acknowledgements. This work is partially supported by 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 Parametrization Trained by Coarse-Graining, Journal of Advances in Modeling Earth Systems, 11, 2728-2744, 10.1029/2019ms001711, 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, Journal of Advances in Modeling Earth Systems, 6, 1185-1204, https://doi.org/10.1002/2014MS000355, 2014.

Cao, G. and Zhang, G. J.: Role of Vertical Structure of Convective Heating in MJO Simulation in NCAR CAM5.3, Journal of Climate, 30, 7423-7439, 10.1175/jcli-d-16-0913.1, 2017.

~~Crawshaw, 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.: On large-scale circulations in convecting atmospheres, Quarterly Journal of the Royal Meteorological Society, 120, 1111-1143, 10.1002/qj.49712051902, 1994.

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, 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, 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 with the Large-Scale Dynamics Using the Cloud-Resolving Convection Parameterization (CRCP), 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 for 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 Cloud Resolving Convection Parameterization for modeling the tropical convecting atmosphere, Physica D: Nonlinear Phenomena, 133, 171-178, https://doi.org/10.1016/S0167-2789(99)00104-9, 1999.

Han, Y., Zhang, G. J., Huang, X., and Wang, Y.: A Moist Physics Parameterization Based on Deep Learning, Journal of Advances in Modeling Earth Systems, 12, e2020MS002076, 10.1029/2020ms002076, 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, 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., 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. 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., 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.

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, 2013, 5, 2013.

Lin, J.-L.: The Double-ITCZ Problem in IPCC AR4 Coupled GCMs: Ocean–Atmosphere Feedback Analysis, Journal of Climate, 20, 4497-4525, 10.1175/jcli4272.1, 2007.

Ling, J., Li, C., Li, T., Jia, X., Khouider, B., Maloney, E., Vitart, F., Xiao, Z., and Zhang, C.: Challenges and Opportunities in MJO Studies, 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 Closure for Eddy-Diffusivity Mass-Flux Schemes of Turbulence and Convection, Journal of Advances in Modeling Earth Systems, 12, e2020MS002161, https://doi.org/10.1029/2020MS002161, 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.: Description of Global-Scale Circulation Cells in the Tropics with a 40–50 Day Period, Journal of Atmospheric Sciences, 29, 1109-1123, 10.1175/1520-0469(1972)029<1109:Dogscc>2.0.Co;2, 1972.

Mooers, G., Pritchard, M., Beucler, T., Ott, J., Yacalis, G., Baldi, P., and Gentine, P.: 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, e2020MS002385, https://doi.org/10.1029/2020MS002385, 2021.

Morrison, H. and Gettelman, A.: A New Two-Moment Bulk Stratiform Cloud Microphysics Scheme in the Community Atmosphere Model, Version 3 (CAM3). Part I: Description and 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., Conley, A. J., Garcia, R., Kinnison, D., and Lamarque, J.-F.: Description of the NCAR community atmosphere model (CAM 5.0), NCAR Technical Note, 1, 1-12, 2012.

Oleson, K. W., Lawrence, D. M., Gordon, B., Flanner, M. G., Kluzek, E., Peter, J., 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, 2010.

Ott, J., Pritchard, M., Best, N., Linstead, E., Curcic, M., and Baldi, P.: A Fortran-Keras Deep Learning Bridge for Scientific Computing, Scientific Programming, 2020, 8888811, 10.1155/2020/8888811, 2020.

Park, S. and Bretherton, C. S.: The University of Washington Shallow Convection and Moist Turbulence Schemes and Their Impact on Climate Simulations with the Community Atmosphere Model, Journal of Climate, 22, 3449-3469, 10.1175/2008jcli2557.1, 2009.

Randall, D., Khairoutdinov, M., Arakawa, A., and Grabowski, W.: Breaking the Cloud 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. Model Dev., 13, 2185-2196, 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.

Song, X. and Zhang, G. J.: The Roles of Convection Parameterization in the Formation of Double ITCZ Syndrome in the NCAR CESM: I. Atmospheric Processes, Journal of Advances in Modeling Earth 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, Monthly Weather 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 Waves: Analysis of 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.

Yu, Tianhe, et al. "Gradient surgery for multi-task learning." Advances in Neural Information Processing Systems, 33 (2020): 5824-5836.

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, 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 centre general circulation model, Atmosphere-Ocean, 33, 407-446, 10.1080/07055900.1995.9649539, 1995.

Zhang, G. J. and Song, X.: Convection Parameterization, Tropical Pacific Double ITCZ, and Upper-Ocean Biases in the NCAR CCSM3. Part II: Coupled Feedback and the Role of Ocean Heat Transport, Journal of Climate, 23, 800-812, 10.1175/2009jcli3109.1, 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, Atmospheric Research, 229, 255-268, https://doi.org/10.1016/j.atmosres.2019.06.023, 2019.

~~Zhang, Y. and Yang, Q.: A Survey on Multi-Task Learning, IEEE Transactions on Knowledge and Data Engineering, 1-1, 10.1109/TKDE.2021.3070203, 2021.~~

Table 1. Input and output variables. For the inputs, $\boldsymbol{q_v(z)}$ is the vertical water vapor profile. $\boldsymbol{T(z)}$ is the temperature profile. $\boldsymbol{dq_{v\,l.s.}(z)}$and $\boldsymbol{dT_{l.s.}}$ are the large scale forcings of the water vapor and temperature, respectively. $\boldsymbol{P_s}$ is the surface pressure; and *Solin* is the TOA solar insolation. For the outputs, $\boldsymbol{dq_v(z)}$ and $\boldsymbol{ds(z)}$ are the tendencies of the water vapor and dry static energy due to moist physics and radiative processes calculated using the NN-Parameterization. The net longwave and shortwave fluxes at the surface and the TOA are the surface net longwave flux (FLNS), surface net shortwave flux (FLNT), TOA net longwave flux (FLNT), and TOA net shortwave fluxes (FSNT). The four downwelling shortwave solar radiation fluxes are the 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) fluxes reaching the surface.

Inputs	Outputs
$q_v(z)$, $T(z)$, $dq_{vls}(z)$, $\mathrm{d}T_{ls}(z)$, P_s, *Solin*	$dq_v(z)$, $ds(z)$, *FLNS, FSNS, FLNT, FSNT, SOLS*, *SOLL*, *SOLSD*, *SOLLD*

Figures 1 – 15

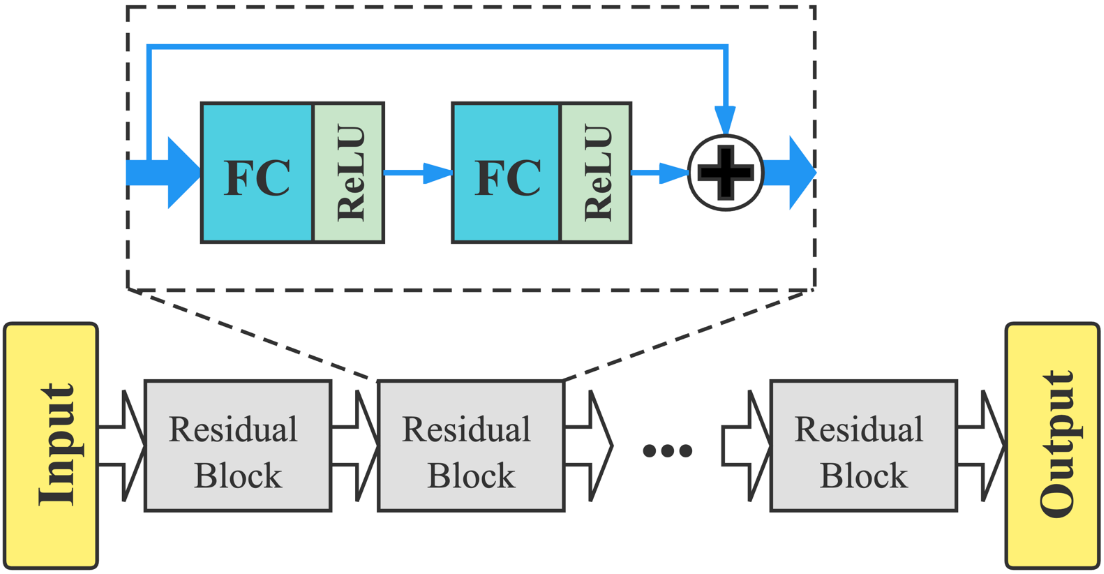

Figure 1. Schematic diagram of the structure of the ResDNN. It consists of seven residual blocks, each of which (dashed box) contains two 512 node-wide dense (fully connected) layers with an ReLU as the activation and a layer jump. The inputs and outputs are discussed in Section 2.2.2.

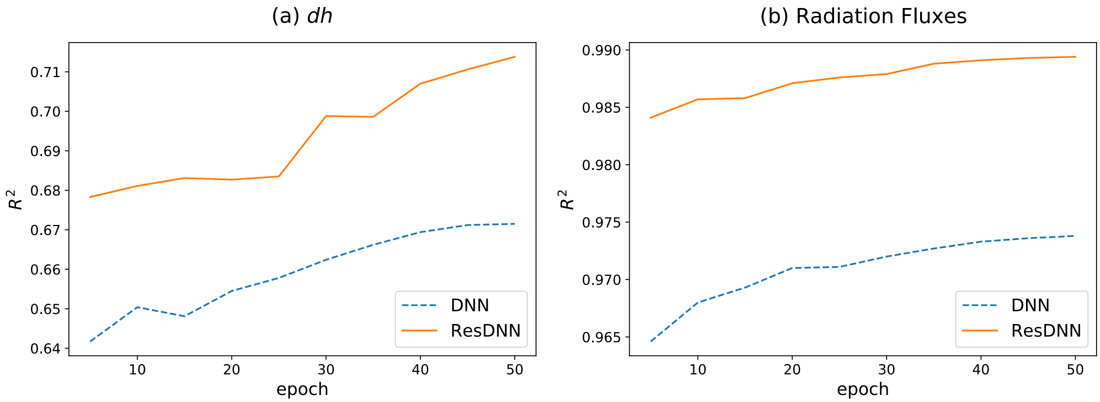

Figure 2. Fitting accuracies (R^2) of both the proposed ResDNN (solid orange lines) and the DNN (dashed blue lines) for different outputs. (a) The R^2 of the moist static energy changing rate (dh) versus the training epochs; and (b) the fitting accuracy of the average R^2 for the eight radiation fluxes. Note: the R^2 values are calculated for both space and time in the validation dataset.

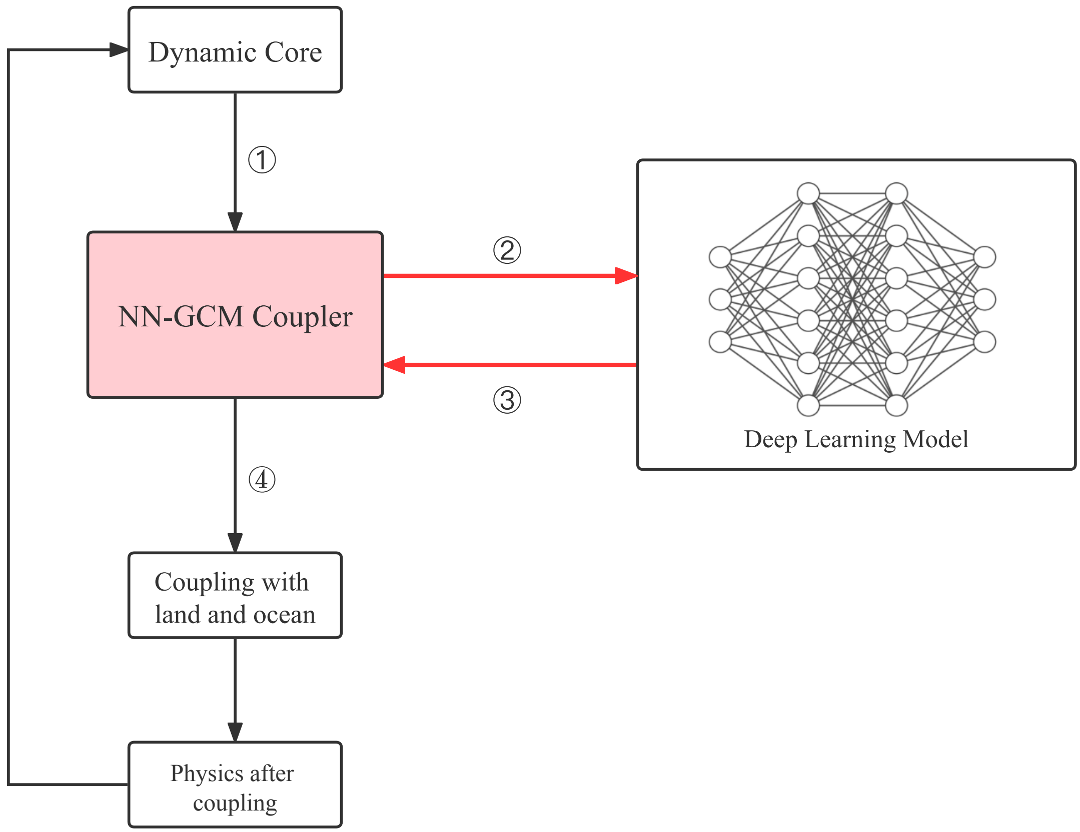

Figure 3. A flow chart of NNCAM, including the NN-GCM coupler. NNCAM runs in the direction of the arrow, and each box represents a module. Among them, the NN-GCM coupler is indicated by the pink box. The NN-Parameterization is shown in the box on the right. ① The dynamic core transmits data to the NN-GCM coupler; ② and ③the data communication between the NN-GCM coupler and the NN-Parameterization; and ④the host GCM accepts the results from the NN-Parameterization.

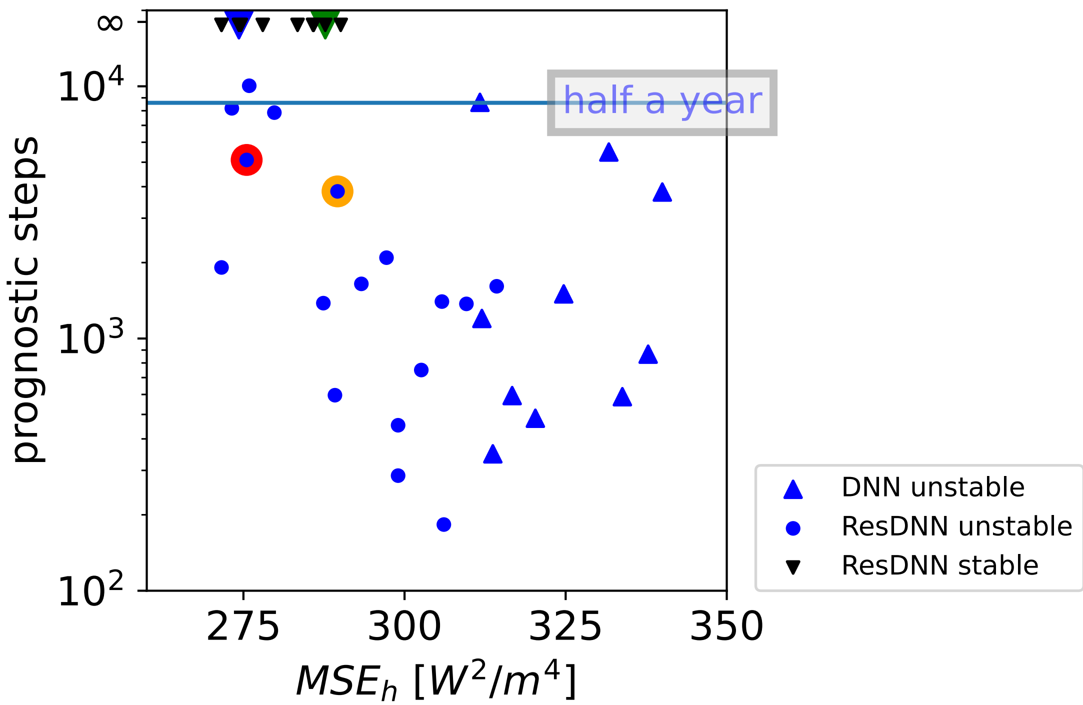

Figure 4. The mean square error of the offline moist static energy vs. the prognostic steps. The black inverted triangles (10 in total but some overlap due to their close MSE_h values) denote stable NN coupled prognostic simulations that last for more than 10 years. The blue dots denote unstable simulations, and the blue triangles denote unstable DNNs. The dots with colored outlines are shown in Figure 5 for the time evolution of the globally averaged energy.

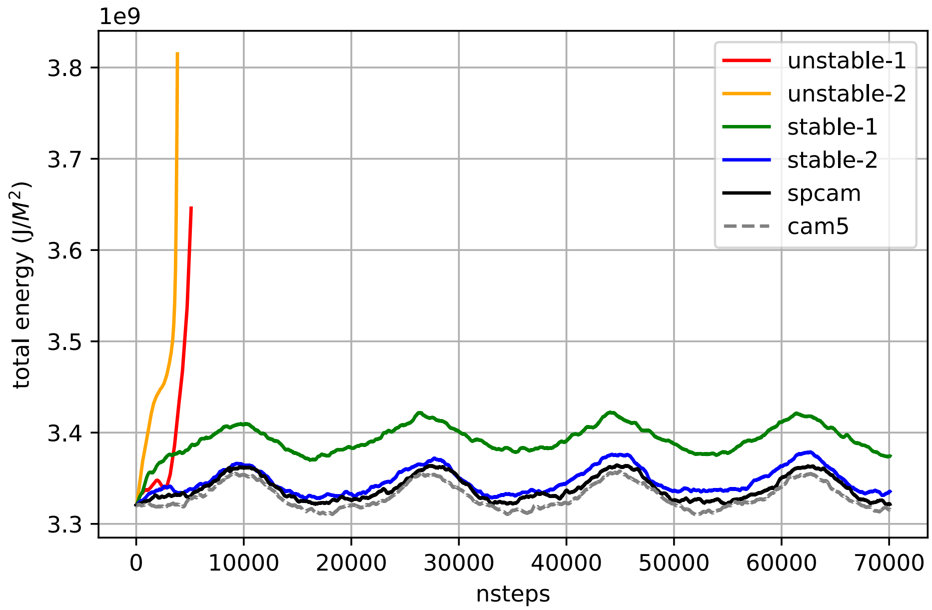

Figure 5. Time evolution of the globally averaged column of the integral total energy of NNCAM with different ResDNN-parameterizations (marked with the same colors as in Figure 4), SPCAM target (black line), and CAM5 control run (grey dashed line). The blue line indicates the stable and accurate ResDNN, the green line indicates the stable but deviating ResDNN, and the orange and red lines indicate unstable ResDNNs.

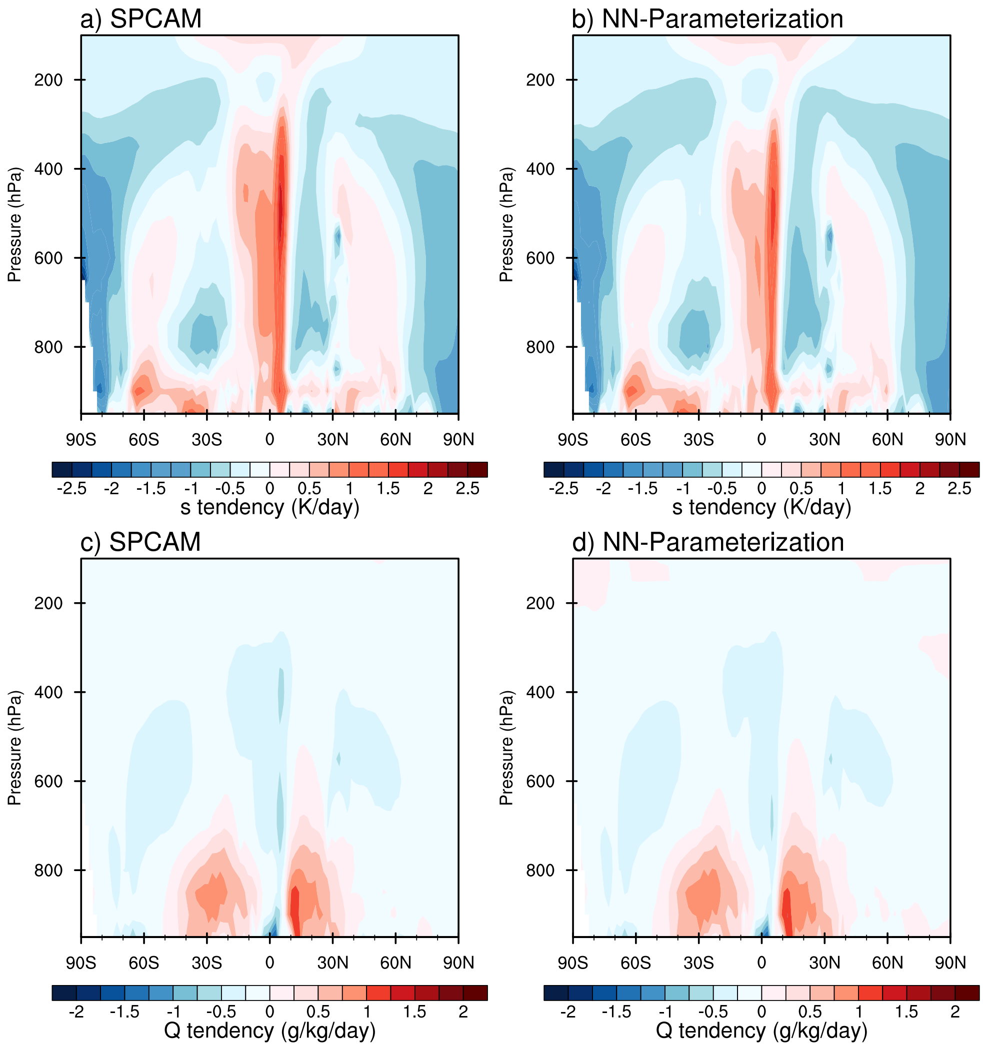

Figure 6. Latitude-pressure cross sections of the annual and zonal mean heating (top) and moistening (bottom) due to moist
815 physics during the year 2000 for (a, c) SPCAM simulations, and (b, d) the offline test using the NN-Parameterizations.

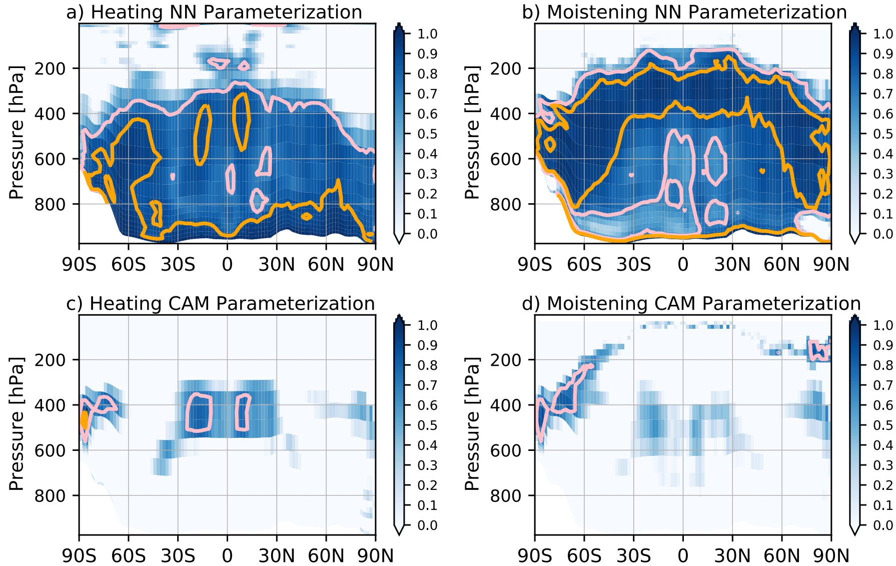

Figure 7. Latitude-pressure cross sections of the coefficient of determination (R^2) for the zonally averaged heating (left panels) and moistening (right panels) predicted using (a & b) the NN-Parameterization in the offline one-year SPCAM run, and (c & d) the offline CAM5 parameterizations. Both were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

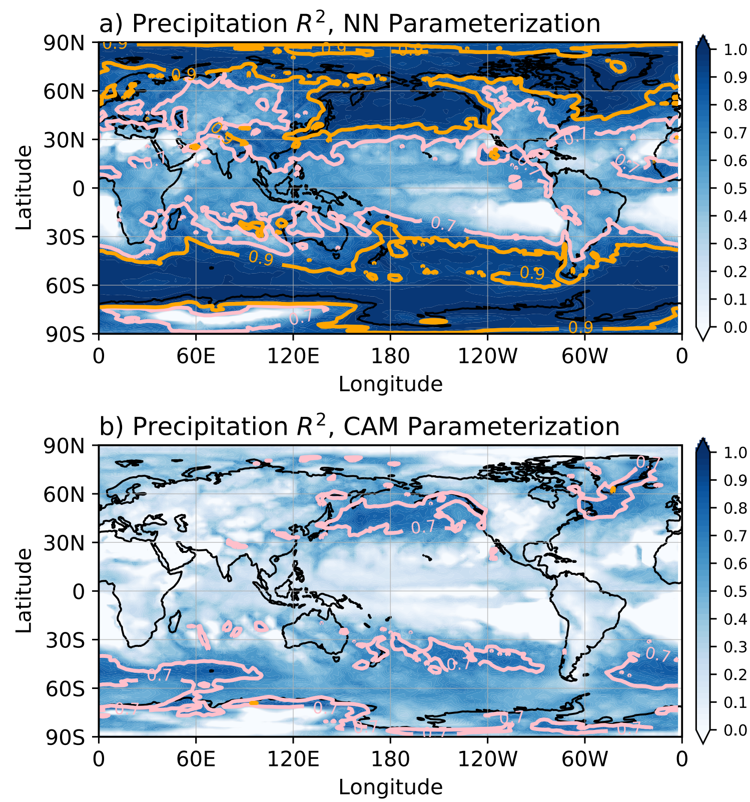

Figure 8. Latitude-pressure cross sections of the coefficient of determination (R^2) for the time sequence at each location for (a) the derived precipitation predicted using the NN-Parameterization and (b) the total precipitation from the CAM5 parameterization compared to the offline one-year SPCAM run. The predictions and SPCAM targets are for a 30 min time step interval. Note: the areas where R^2 is greater than 0.7 are contoured in pink, and the areas where R^2 is greater than 0.9 are contoured in orange.

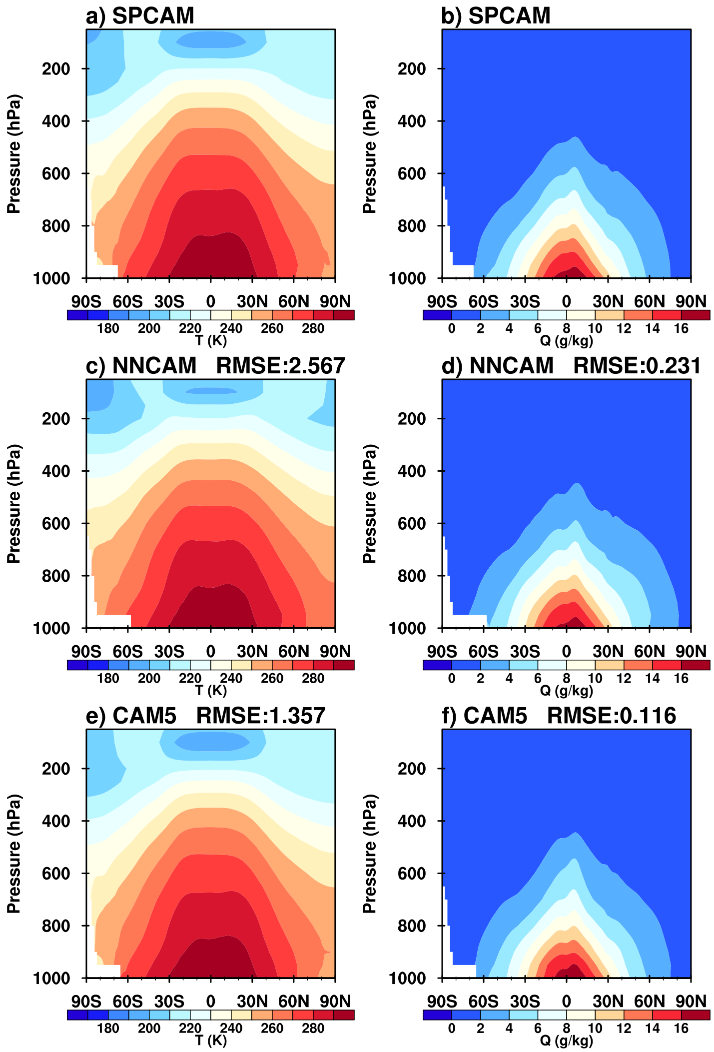

Figure 9. Latitude-pressure cross sections of the zonal mean temperature (left panels) and specific humidity (right panels) averaged from 1999 to 2003 predicted using (a, b) SPCAM, (c, d) NNCAM, and (e, f) CAM5.

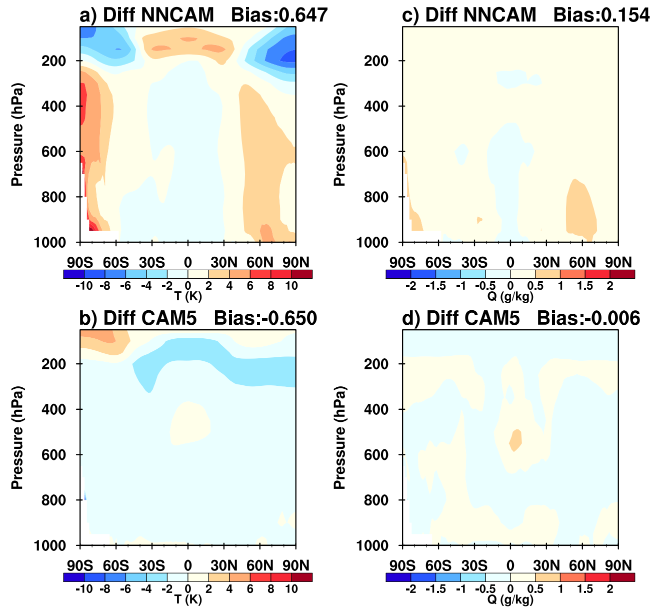

Figure 10. Latitude-pressure cross-section of the zonal and annual mean differences in the temperature (left panels) and specific humidity (right panels) between (a & c) NNCAM and SPCAM and (b & d) CAM5 and SPCAM. The simulation period for all of the models was from 1999 to 2003.

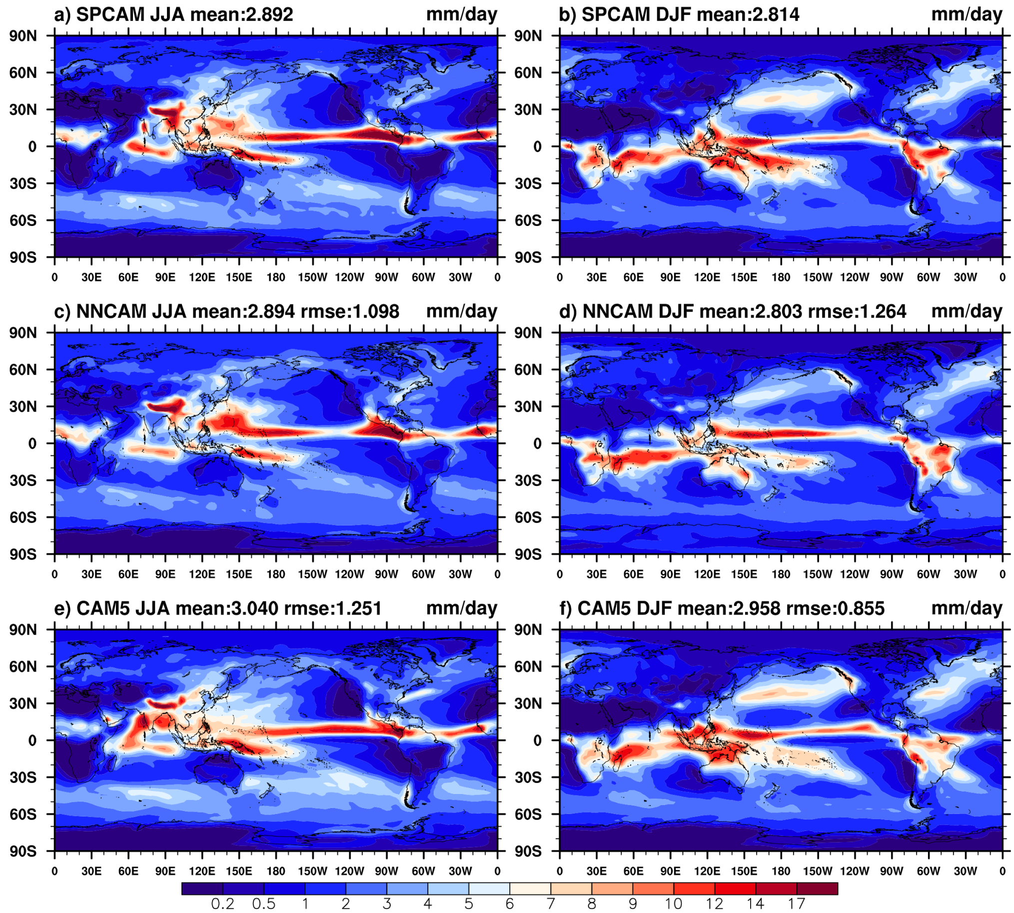

Figure 11. The mean precipitation rate (mm day^{-1}) averaged from 1999 to 2003 for June-July-August (left panels) and December-January-February (right panels) predicted using (a, b) SPCAM, (c, d) NNCAM, and (e, f) CAM5.

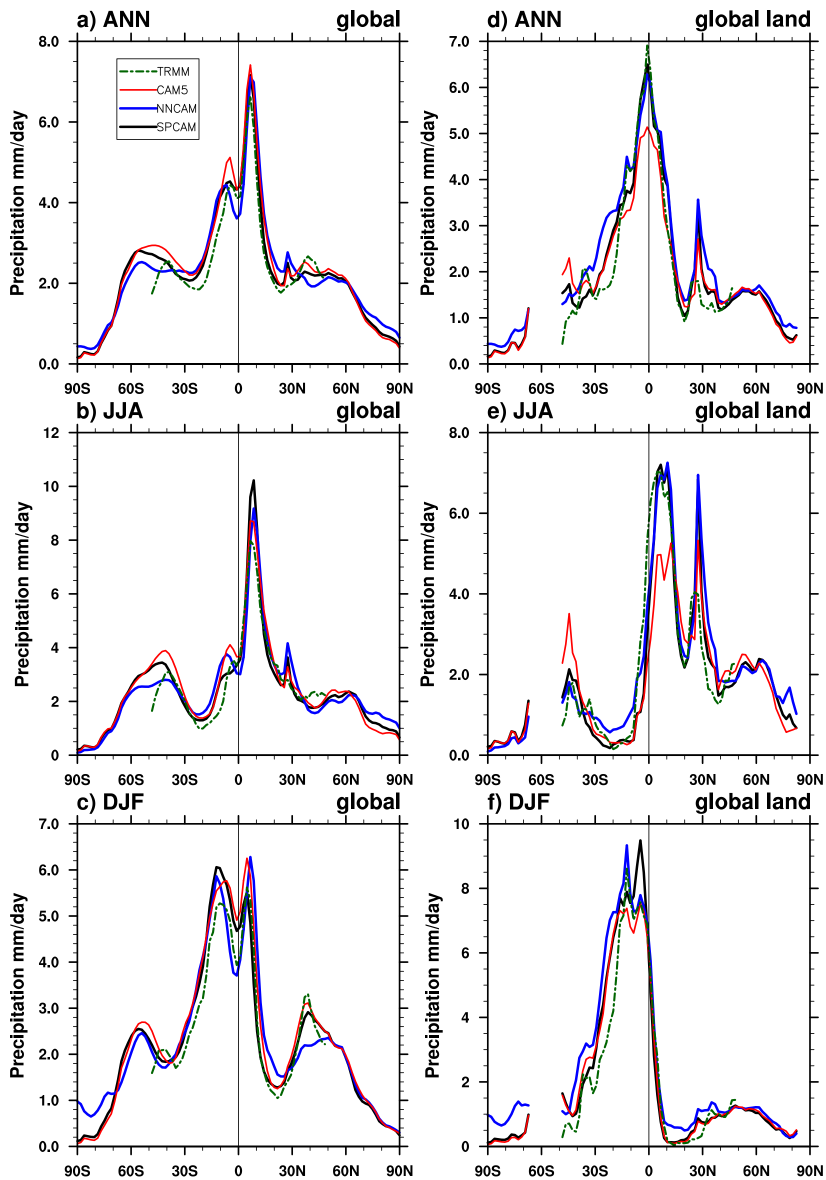

Figure 12. The zonal mean precipitation rate (mm/day) averaged from 1999 to 2003 for (a, d) the annual mean, (b, e) June-July-August, and (c, f) December-January-February. The black, blue, and red solid lines denote SPCAM, NNCAM, and CAM5, respectively. The dark green dashed line denotes the averaged results of the TRMM 3B42 daily rainfall product.

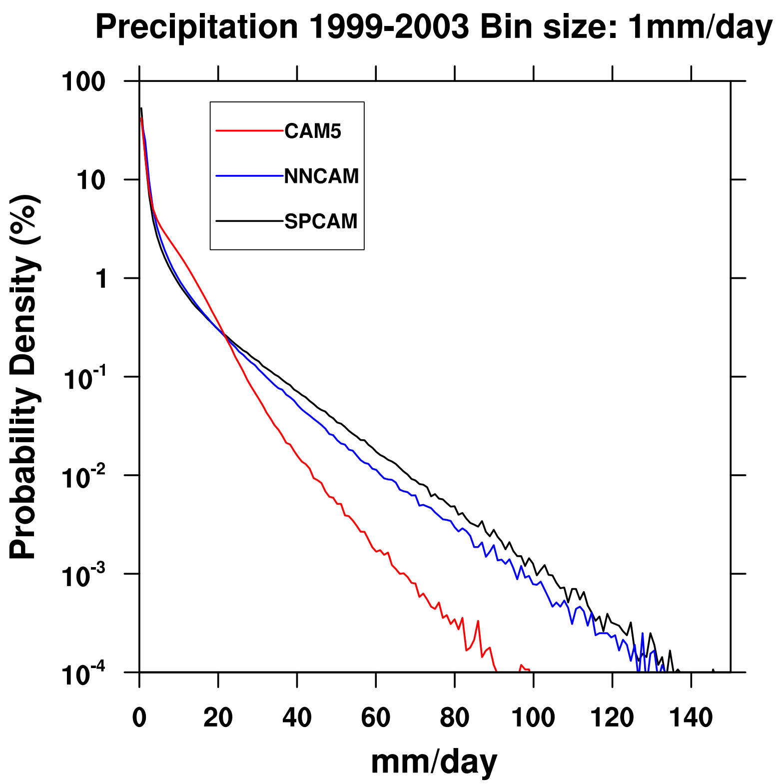

Figure 13. Probability densities of the daily mean precipitation in the tropics (30°S−30°N) obtained from the three model simulations. The black, blue, and red solid lines denote SPCAM, NNCAM, and CAM5, respectively.

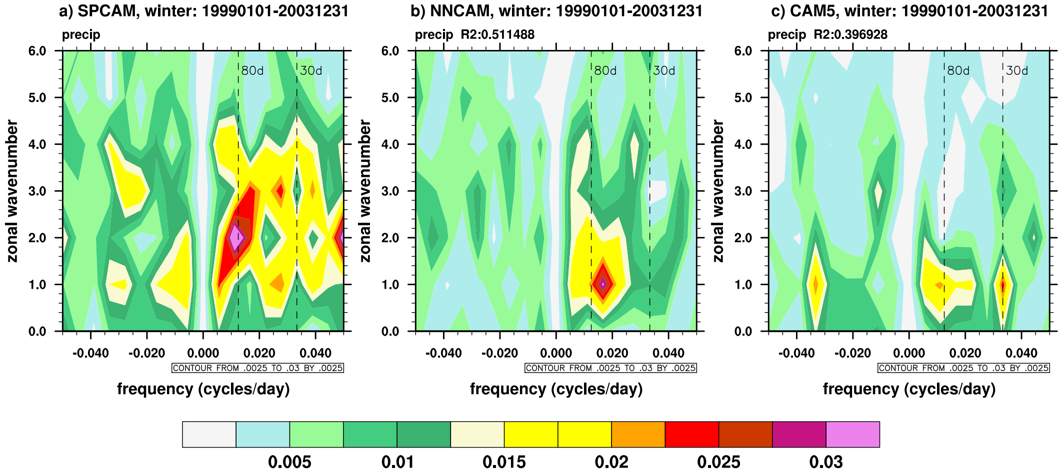

Figure 14. The wavenumber-frequency spectra for the daily precipitation anomalies at 10°S−10°N for (a, b) SPCAM, (c, d) NNCAM, and (e, f) CAM5 simulations in boreal winter.

860

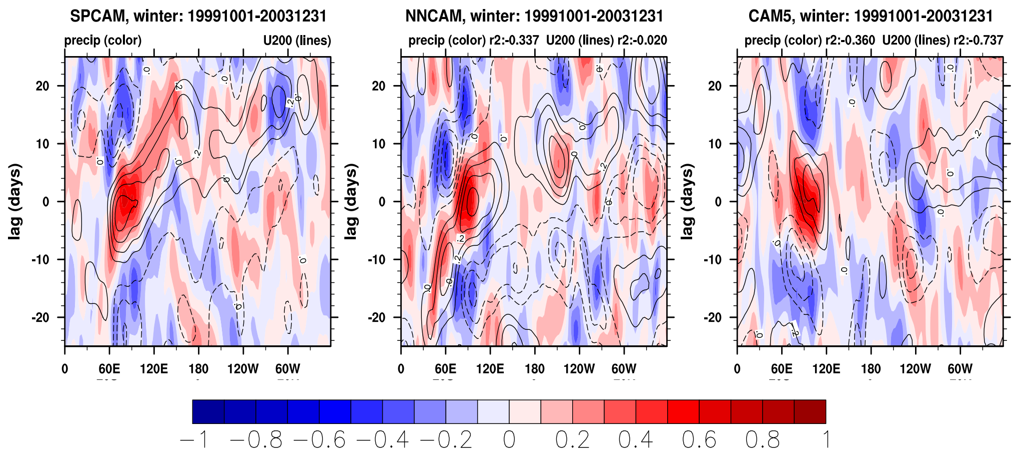

Figure 15. Longitude-time evolution of the lagged correlation coefficient for the 20–100-day bandpass-filtered precipitation anomalies (averaged over 10°S−10°N) against the regionally averaged precipitation (shading) and zonal wind at 200 hPa (contours) over the equatorial eastern Indian Ocean (80−100°E, 10°S−10°N). The dashed lines in each panel denote the 5 m s^{-1} eastward propagation speed.