Research Article

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

Vladimir M. Krasnopolsky,[1,2] Michael S. Fox-Rabinovitz,[2] and Alexei A. Belochitski[3,4]

[1] National Centers for Environmental Prediction, NOAA, College Park, MD 20740, USA

[2] Earth System Sciences Interdisciplinary Center, University of Maryland, College Park, MD 20740, USA

[3] Geophysical Fluid Dynamics Laboratory, NOAA, Princeton, NJ 08540, USA

[4] Brookhaven National Laboratory, Upton, NY 11973, USA

Correspondence should be addressed to Vladimir M. Krasnopolsky; vladimir.krasnopolsky@noaa.gov

A novel approach based on the neural network (NN) ensemble technique is formulated and used for development of a NN stochastic convection parameterization for climate and numerical weather prediction (NWP) models. This fast parameterization is built based on learning from data simulated by a cloud-resolving model (CRM) initialized with and forced by the observed meteorological data available for 4-month boreal winter from November 1992 to February 1993. CRM-simulated data were averaged and processed to implicitly define a stochastic convection parameterization. This parameterization is learned from the data using an ensemble of NNs. The NN ensemble members are trained and tested. The inherent uncertainty of the stochastic convection parameterization derived following this approach is estimated. The newly developed NN convection parameterization has been tested in National Center of Atmospheric Research (NCAR) Community Atmospheric Model (CAM). It produced reasonable and promising decadal climate simulations for a large tropical Pacific region. The extent of the adaptive ability of the developed NN parameterization to the changes in the model environment is briefly discussed. This paper is devoted to a proof of concept and discusses methodology, initial results, and the major challenges of using the NN technique for developing convection parameterizations for climate and NWP models.

1. Introduction

Clouds and convection are among the most important and complex phenomena of the Earth's physical climate system. In spite of intense studies for centuries, clouds still provide an intellectual and computational challenge. Because of the vast range of time and space scales involved, researchers and models that they use typically focus on a particular component of a cloud system, with a narrow range of time and space scales, and prescribe features of the cloud that operate outside of that range. For example, microphysical models describing drop scale motions (e.g., drop coagulation) deal with the fine spatial and temporal scales (of order of millimeters (drop size) and seconds). For more detailed discussion of atmospheric moisture physics, see [1–4]. At the other end of the spectrum of representations of clouds is their representation in large-scale models, for example, in general circulation or global climate models (GCMs), which resolve atmospheric features with spatial scales of the order of 100 km, and temporal scales of the order of 10 minutes.

Numerical atmospheric and coupled atmospheric-oceanic-land models, or GCMs, used for climate and

numerical weather predictions, are based on solving time-dependent 3-D geophysical fluid dynamics equations on the sphere. The governing equations of these complex models, based on conservation lows, can be written symbolically as

$$\frac{\partial \psi}{\partial t} + D\left(\psi, x\right) = P\left(\psi, x\right), \qquad (1)$$

where ψ is a 3-D-prognostic or -dependent variable (e.g., temperature, wind, pressure, and moisture); D is model dynamics (the set of 3-D partial differential equations of motion, thermodynamics, etc., approximated with a spectral or grid-point numerical scheme); x is a 3-D-independent variable (e.g., latitude, longitude, and height); P is model physics (e.g., long and short-wave atmospheric radiation, turbulence, convection and large-scale precipitation processes, clouds, interactions with land and ocean processes, etc.) and chemistry (constituency transport, chemical reactions, etc.). While scientific problems using these models are among the most complex and computationally intensive applications in the history of scientific exploration, the models employ drastic simplifications in their treatment of many physical processes important in climate and weather.

Physical and other processes included in model physics, P, are so complicated that it is practically possible to include them into GCMs only as 1-D (in the vertical direction) simplified or parameterized versions (usually called parameterizations). Thus, the model physics is composed of parameterizations as $P = \sum_k P_k$. These parameterizations constitute the right hand side forcing for the model dynamics equations (1). From the mathematical point of view, each parameterization can be considered as a mapping, which is a relationship between two vectors:

$$\mathbf{Y} = P_k\left(\mathbf{X}\right), \qquad (2)$$

where $\mathbf{X}$ is a vector consisting of profiles of atmospheric parameters describing the state of the atmosphere at a particular time at a particular location (a grid point) and $\mathbf{Y}$ is a vector of parameters providing an effective feedback to the atmosphere from the physical processes described by the parameterization P_k at the same location.

It is noteworthy that, after the very significant simplifications mentioned above and as it is formulated in (2), a parameterization does not depend on time and location explicitly. However, throughout model integration it is put in the environment, which changes in time and space when the parameterization is applied at different times and different horizontal locations (grid points) over the globe. The changes of the environment include temporal changes like diurnal, annual, other atmospheric and solar cycles, global climate changes, and spatial changes like the transition of the underlying surface from ocean to land and from one climate zone to another one (e.g., from the tropics to extra tropics). These internal changes constantly occurring throughout model integration reflect the actual external changes in the climate or weather system described by the model. In this paper, we investigate if our developed NN convection parameterization demonstrates the practically meaningful temporal and spatial generalization capability. The generalization capability allows

the NN convection parameterization to adapt to changing atmospheric states produced throughout climate model simulations.

A GCM does not resolve multiple subgrid processes that occur on temporal and spatial scales much finer than the GCM resolution. However, subgrid processes and scales are the processes and scales of which physics represents and is defined on. Because of that both $\mathbf{X}$ and $\mathbf{Y}$ in (2) have significant uncertainties and are actually stochastic variables. $\mathbf{X}$ and $\mathbf{Y}$ in (2) are defined on a GCM grid, and their uncertainties are due to and represent the subgrid scale variability, which is not resolved by GCM.

Thus, the mapping (2), which establishes the relationship between two stochastic variables $\mathbf{X}$ and $\mathbf{Y}$, is a stochastic mapping, and the parameterization P_k is a stochastic parameterization. Actually, the stochastic mapping is a family of mappings, each of which describes a relationship between two considered stochastic variables $\mathbf{X}$ and $\mathbf{Y}$ inside a range determined by the uncertainties of these variables with a probability determined by a joint probability distribution of these variables. The stochasticity of model physics parameterizations is a natural consequence of the finite model resolution, which leaves unresolved very important sub-grid scale processes [5]. Uncertainties of stochastic parameterization (2) carry important physical information about these subgrid (unresolved on a GCM grid) physical processes and should be properly taken into account in GCMs.

Usually, parameterizations, P_k, are formulated using relevant first principles and observational data and are based on solving deterministic equations (like radiative transfer equations). They also contain some secondary empirical components based on traditional statistical techniques like regression. As the result, for widely used the state-of-the-art GCMs, all major components of model physics and chemistry, are based on solving deterministic first principle physical or chemical equations. Thus, in the process of development of such a physically based parameterization, the family of mappings that represents the stochastic parameterization (2) is collapsed to one member of the family, completely neglecting the stochastic nature of the parameterization. Therefore, a physically based parameterization represents only one arbitrarily selected member of the family of mappings that represents the stochastic parameterization (2).

In this study, we develop a stochastic convection parameterization based on the learning from data approach, using a neural network (NN) technique. If sufficient amount of observations related to atmospheric convection, cloudiness, and precipitation was available, we could have attempted to learn the parameterization directly from observations. Unfortunately, in reality the available observations are sparse in space and time and not sufficient for such developments. To alleviate the problem, we use data simulated by models that explicitly resolve processes at the smaller temporal and spatial scales; that is, these models have resolution of a couple of orders of magnitude higher than that of GCMs. These models are capable of representing and resolving processes that are relevant to many major features of cloud systems with

the spatial scales of the order of several kilometers and temporal scales of seconds to minutes. These are so called cloud resolving models (CRM), which simulate component aspects and evolution of cloud systems much more realistically than large-scale models. We employ the CRM [6–8] initialized and forced by observational data to simulate a limited (but more expanded, in terms of the spatial and temporal resolution and the number of variables, than available observational data) amount of data (they are called "pseudo-observations"). Then we use a NN technique to develop a stochastic convection parameterization (2) by learning from the simulated pseudo-observations.

In our previous works, we successfully applied a NN technique to develop accurate and fast emulations of complex physically based parameterizations. Because of the complexity of the physical processes involved and the complexity of their mathematical and numerical representations, some of these parameterizations are the most time-consuming components of GCMs. We have developed NN emulations for the most time-consuming part of model physics: model radiation [9–13]. Because, as it was mentioned above, a physically based parameterization is represented by a single mapping, we successfully used a single NN to emulate a physically based parameterization. However, a single NN is not an adequate tool for emulating a stochastic mapping (2), which is actually a family of mappings. An adequate tool in this case is an ensemble of NNs, which can effectively emulate the stochastic mapping (2) [4].

In this study, we use the CRM developed and provided by Khairoutdinov and Randall [7]. We also use the GCM developed by the National Center for Atmospheric Research (NCAR) that is called the Community Atmospheric Model (CAM). Thus, in this study we (a) apply a NN ensemble technique to learn a NN stochastic convection parameterization from the pseudo-observations simulated by the CRM, (b) introduce this NN stochastic convection parameterization into the NCAR CAM, and (c) run climate simulations to test the validity of the new NN stochastic convection parameterization. In Section 2, we formulate our approach and describe details of the training set creation and NN training. In Section 3, the results of validation of the developed NN ensemble convection parameterization are described. Section 4 presents discussion of results and Section 5 contains conclusions.

2. Formulation of the Approach: Development of a NN Ensemble Convection Parameterization from CRM Data

In this study, we develop an ensemble of NNs which emulates the behavior of fine-scale CRM simulations at larger GCM scales in a variety of regimes and initial conditions. The resulting ensemble NN parameterization can be used as a novel, and computationally viable convection parameterization in GCMs. This approach has a realistic potential of producing a parameterization of a similar or better quality to the existing physically based parameterizations that are used in GCMs, effectively taking into account subgrid scale (in terms of

a GCM) effects at a fraction of the computational cost of the existing approaches.

As we have shown in our previous works (e.g., [12]) any parameterization of model physics (2) can be emulated employing multilayer perceptron NNs using learning-from-data approach. This NN is an analytical approximation that uses a family of functions like

$$y_q = a_{q0} + \sum_{j=1}^{k} a_{qj} \cdot \phi \left(b_{j0} + \sum_{i=1}^{n} b_{ji} \cdot x_i \right); \quad q = 1, 2, \ldots, m,$$

(3)

where x_i and y_q are components of the input and output vectors X and Y, respectively, a and b are fitting parameters, and $\phi(b_{j0} + \sum_{i=1}^{n} b_{ji} \cdot x_i)$ is a "neuron." The activation function ϕ is usually a hyperbolic tangent, n and m are the numbers of inputs and outputs, respectively, and k is the number of neurons in (3). In the case of a stochastic parameterization, an ensemble of NNs (3) provides an adequate tool for representing a parameterization (2) [4].

2.1. Design and Development of NN Convection Parameterizations and Training Sets. Figure 1 summarizes the process of development of the NN parameterization. The CRM simulations use the TOGA-COARE (the international observational experiment in the tropics conducted for the 4-month period from November 1992 to February 1993) observational data for initialization and forcing and have the horizontal resolution ρ of 1 km, 64 or 96 vertical layers extending from the surface to ~30 km, and the time integration step of 5 s. We integrate the CRM over the domain of 256×256 km.

The development of the NN parameterization is a multistep process. These steps are as follows.

(1) CRM simulated data: the CRM has been run for the 4-month period (from November 1992 to February 1993 or for 120 days of the TOGA-COARE observational experiment), and the high 1 km resolution output of the model has been obtained. The CRM-simulated temperature, wind, humidity, and other data are in a good agreement with those of the TOGA-COARE observational data. Also, the CRM produces additional prognostic and diagnostic fields not observed in the TOGA-COARE experiment.

(2) Reducing the horizontal resolution of the CRM simulated data: the CRM-simulated data are averaged in space and time. The data are averaged to a reduced horizontal resolution to r where $\rho < r \leq R$, and ρ and R are the CRM and GCM resolutions correspondingly. Also the data are interpolated/averaged onto the number of vertical layers $l = L$, where L is the number of vertical layers in the GCM.

(3) Projecting a CRM space of atmospheric states onto a GCM space of atmospheric states: the CRM has many variables that describe fine-scale processes not resolved by the GCM, for example, the condensed water in each CRM column. Such variables have no analogs in the GCM. From the point of view of
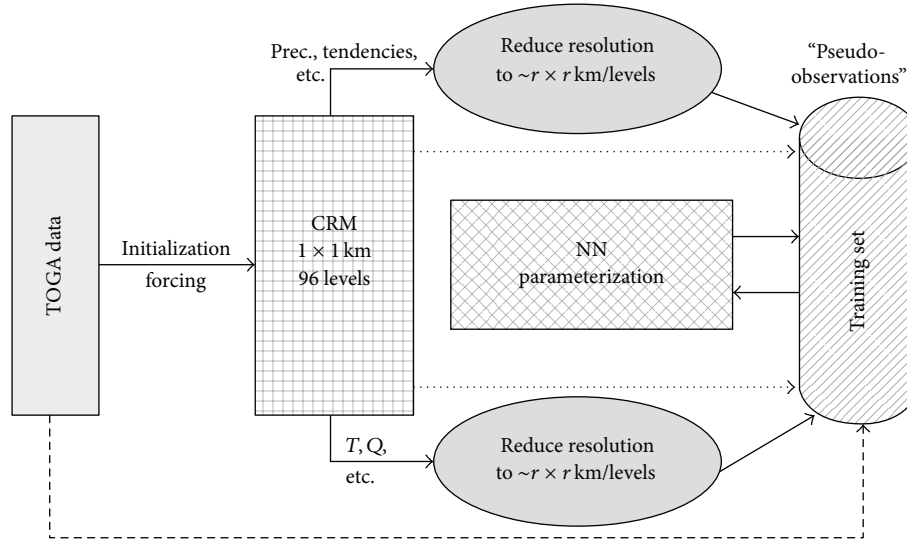
Figure 1: Development design of a NN convection parameterization. *T, Q,* and so forth refer to selected CRM-simulated fields used as inputs for NN parameterization; Prec., Tendencies, and so forth refer to selected CRM simulated variables used as NN outputs. The dotted and dashed lines indicate that observational and high-resolution simulated data can be added to the training data set if necessary.

a GCM "model reality" these variables are "hidden" variables responsible for sub-grid scale variability. The acknowledgement of this challenge requires development of the concept of uncertainty and "stochasticity"; it leads to recognition of a significant level of uncertainty in the pseudo-observations and of the stochasticity of the convection parameterization (2) learned from these data. The obtained set of "pseudo-observations" implicitly represents a stochastic convection parameterization with an uncertainty, which is an inherent feature of such a parameterization. Thus, only the variables that can be identified with the corresponding GCM variables or can be calculated from or converted to prognostic or diagnostic variables available in the GCM are selected to be included in the development set (called "pseudo-observations" in Figure 1; actually they are obtained from the averaged CRM simulated data). Only these variables are used as inputs and outputs of a NN convection parameterization. Thus, a subset of variables is selected from the reduced resolution CRM-simulated data created at the previous step (2), and this subset constitutes the NN development set. The dotted and dashed lines in Figure 1 show that, in principle, if it is found to be desirable, the high-resolution CRM-simulated data and/or even observed data can be added to the development set to enrich subgrid variability in the development data.

(4) The developed "pseudo-observations" are separated into two sets, one set being used for training and another independent set for testing/validation. Then the NN parameterization is trained using the training set. Due to the inherent uncertainty of pseudo-observations, the parameterization represented by these data is a stochastic parameterization; it should

be considered as a stochastic mapping. Thus, the NN parameterization is implemented as an ensemble of NNs.

All the aforementioned issues are discussed in detail in [4].

The validation procedure for the NN parameterization consists of two steps. First, the trained NN is applied to the test set and error statistics are calculated. Second, the tested NN parameterization is included into the GCM. This last step is the most important step of validation and of our approach.

2.2. NN Emulation of the Convection Parameterization and Estimation of Its Uncertainties

2.2.1. Data. The data set simulated for the NN development is limited by the length of the observational data set needed for forcing the CRM simulations (see Figure 1). The CRM was run for 120 days, using the TOGA-COARE forcing for the 256×256 km domain with 1 km resolution and 96 vertical layers (0–28 km). Then the simulated data were averaged at every hour of model integration to produce the simulation data set with an effective horizontal resolution of 256 km and 26 vertical levels. Finally, only variables that are available in the GCM (NCAR CAM) or can be calculated there have been selected. The final data set consists of 2,800 records of hourly mean data.

The simulation dataset was partitioned into two parts: a training set consisting of 2,240 records or 80% of data and a test set consisting of 560 records or 20% of data. Namely, first the 2,240 records are included in the training set and the last 560 records in the independent test set.

These two data sets have been used for the NN training and test/validation. As was noticed in the previous section, these data implicitly represent a stochastic parameterization

and inherently contain an uncertainty, η, which is not a useless noise. However, in the process of learning the NN convection parameterization from pseudo-observations, from the point of view of a single NN trained using the data (both components X and Y of the data are stochastic variables), the situation is similar to the case when the data contain a significant level of noise.

Symbolically, the NN emulation of the stochastic parameterization (2) can be written as

$$Y = P_{\mathrm{NN}}(X) + \eta + \varepsilon, \tag{4}$$

where P_{NN} is a NN emulation of the mapping P_k (2) and ε is a NN approximation error. Thus, in the case of the stochastic parameterization, the NN emulation task is different from the task of emulating a *deterministic* radiation parameterization in the GCM [9–11]. For example, in the GCM the radiation parameterization, which is a closed analytical expression or a computer code (mapping) is usually considered as an "exact" source of radiation information (it is not considered as a stochastic parameterization with an uncertainty); thus, for the NN emulation approach the goal is to emulate it as accurate as possible. This can be done because, in this case, the simulated data can be produced using the given parameterization (mapping), and considered as accurate data (with no noise greater than the round off errors).

In the current work, the situation is different. We do not have an expression (or computer code) for the mapping (2) that we want to emulate with NN (3). We can only assume that it exists and, in this case, it is a stochastic mapping, which is represented by pseudo-observations. Because we derive pseudo-observation, using a rather complex data processing described in the previous section, from the data simulated by the CRM, the pseudo-observations include a significant level of uncertainty. The uncertainty and stochasticity are the essential conceptual features of the NN parameterization that we are going to learn from the pseudo-observations. In a sense, emulating stochastic parameterization is closer to the task of learning from noisy empirical data [14]. This important difference should be taken into account when the NN approximation is trained, the approximation error statistics are analyzed and interpreted, and the NN architecture is selected. For example, in the case of training, the usually used criterion of minimum of the root mean square error should be substituted by the requirement that the root mean square error should not exceed the uncertainty η or

$$\frac{1}{N}\sum_{i=1}^{N}[Y_i - P_{\mathrm{NN}}(X_i)]^2 < \eta^2, \tag{5}$$

where N is the number of records in the training set.

All NNs that satisfy the condition (5) *are valid emulations of the stochastic parameterization* (2). Actually, each of these NNs can be considered as an emulation of a member of the family of mappings that together represent the stochastic parameterization (2). Therefore, *all NNs satisfying* (5) *together—the entire ensemble of NNs—represent the stochastic parameterization* (2). It is clear now that any estimate of the magnitude of the uncertainty η is important for our

Table 1: NN architecture (inputs and outputs) investigated in the paper.

NN architecture	NN inputs		NN outputs			
In : out	T	QV	Q1C	Q2	PREC	CLD
36 : 55	18	18	18	18	1	18

T is temperature, QV is atmospheric moisture—vapor mixing ratio, Q1C: the "apparent heat source," Q2: the "apparent moist sink," PREC: precipitation rates, and CLD: cloudiness. Numbers in the table show the dimensionality of the corresponding input and output parameters. In : Out stand for NN inputs and outputs and show their corresponding numbers.

approach. We will attempt to derive such an estimate in the next sections.

2.2.2. NN Architectures, NN Training, and Validation. Selecting an emulating NN architecture includes two different aspects and types of decisions: (i) the selection of inputs and outputs and their numbers (n and m in (3)), which, as we have already mentioned, are determined by the availability of the variables in the GCM, and (ii) the selection of the number of hidden neurons (k in (3)) in the emulating NN, which is determined by many factors (the length of the training set, the level of uncertainty in the data, the characteristics of conversions of the training and test errors, etc.).

Table 1 shows (in terms of inputs and outputs) the architecture we have experimented with here. The major inputs are the vertical profiles of the following model prognostic and diagnostic fields: T—a profile of temperature—and QV—a profile of water vapor concentration. We experimented also with additional inputs: time, latitude, and longitude to take into account the changes in the data environment where the NN is applied; however, because in this particular case we work with data from a relatively small area and over a relatively short period of time (120 days), explicit introduction of the time and location dependencies does not make any difference and does not complement the indirect dependence on the time and location introduced by inputs T and QV. These inputs themselves depend on time and location. The major outputs for NN architecture shown in Table 1 are the vertical profiles (or vectors) of the following model prognostic and diagnostic fields: Q1C (a profile of the "apparent heat source"), Q2 (a profile of the "apparent moist sink"), PREC (precipitation rates, a scalar), and CLD (a profile of cloudiness).

The numbers in Table 1 show how many vertical levels of the corresponding profile (26 levels maximum as in the NCAR CAM) have been included as inputs in the NN. Many profiles have zeros, or constants, or values that are almost constant (their standard deviations are very small) for the entire data set at some levels (usually for the upper model levels in the stratosphere). Zeros and constants should not be included in inputs or outputs because (1) they carry no information about input/output functional dependence and (2) if not removed they introduce additional noise in training. As for small values that are almost constant, these small signals may be in some cases not a noise but very important signals; however, taking into account the level of uncertainty
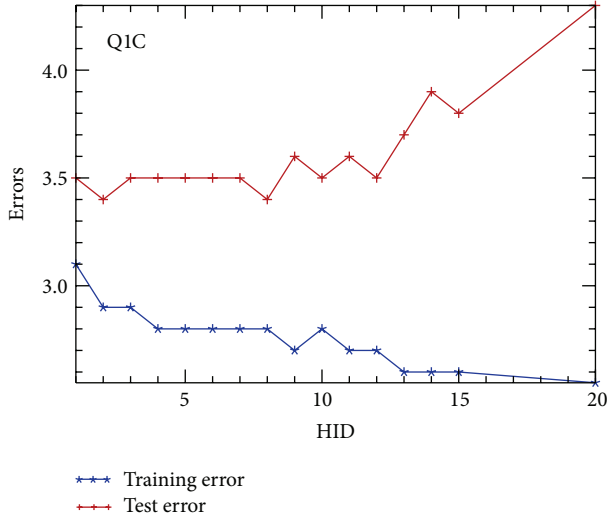
FIGURE 2: NN approximation errors on training (blue) and test (red) sets for Q1C. HID is the number of hidden neurons.

FIGURE 3: Same as in Figure 2 but for precipitation.

in the problem, information that these small signals may provide is well below the level of uncertainty and is practically useless. Moreover, some of these variables were included in training and no improvement was observed. In general, if they are important, they should be normalized differently or weighted.

Next, the number of hidden neurons (HID) has to be selected. We varied HID, trained a corresponding NN, and tested it. Figures 2 and 3 show the results of these experiments for two output parameters, Q1C and PREC.

It is noteworthy that the NN training (a least square minimization) attempts to minimize the total $(\eta + \varepsilon)^2$, that is, the sum of the approximation error and the uncertainty. Because of very different statistical properties of these components, they can be considered as independent random variables and approximately separated as

$$(\eta + \varepsilon)^2 \approx \eta^2 + \varepsilon^2. \tag{6}$$

Thus, η can be roughly estimated using detailed information about the training and test statistics. This issue is discussed in more detail below.

Figures 2 and 3 demonstrate a situation that is usually observed when NN is trained using data with a significant level of noise. The training error, after a sharp initial drop, stays almost constant and then decreases slowly. The test error, after an initial drop, stabilizes and then increases. The interpretation of this behavior is well known. After the initial improvement of the approximation of the data due to an increasing flexibility of an approximating NN, a short interval of stability is reached (at HID ~3 to 7) when NN fits the signal inside the corridor of errors. Then with the increase of the flexibility of the approximating NN, it starts fitting the noise; that is, the overfitting occurs. The training error is slowly decreasing; however, the test error quickly increases. Table 2 shows the number of fitting parameters (NN weights) in NNs with different HID, which were used for
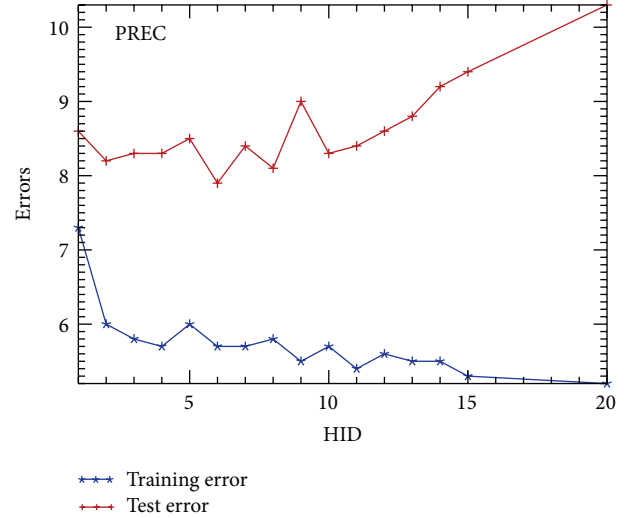
Figures 2 and 3. Taking into account that the training set contains a limited number of records, 2240 records, it is not surprising that clearly pronounced overfitting is observed at HID > 10 when the number of NN weights, N_C, becomes comparable with the number of data records.

Thus, we can conclude that, for a particular simulated data set used, HID = 5 would be an acceptable choice for the number of hidden neurons in emulating NN. This value is inside of the intervals of stability of the training and test errors. Because for different NN outputs the interval of stability is slightly different, the choice of the optimal number of hidden neurons is hardly possible. The best solution of the problem, in our opinion, would be included in the NN ensemble members with various architectures (different numbers of hidden neurons and even with different inputs [4]).

The ensemble of ten NNs has been trained, and error statistics for seven of them that are significantly different are presented in Table 3. All NNs presented in Table 3 have the same number of inputs (36), outputs (55), and hidden neurons (5). They all have been initialized using the same initialization procedure [15] with different small random numbers. The ten different members of NN ensemble correspond to ten different local minima of the error function. Table 3 shows the comparison of NN ensemble member error statistics on the training set (Tr) and on the independent test set (Ts); both sets are described above in Section 2.2.1. For each NN output variable, three statistics were calculated (bias, RMSE, and correlation coefficient) by comparison of NN-generated output variables with the corresponding ones in the training or test set.

The *training errors* (Tr) for all output parameters are significantly closer to each other for different NN ensemble members and less sensitive to the selection of HID (not shown in Table 3) inside the interval of stability (see Figures 2 and 3) than the test errors (Ts). Thus, the training errors can be considered as a rough estimate of the noise in the

TABLE 2: The number of fitting parameters (NN weights), N_C, at different values of HID = k (see (3)).

	HID					
	1	2	5	10	15	20
N_C	166	273	594	1129	1667	2199

data, which is the inherent uncertainty η of the stochastic parameterization (2).

Following this assumption, we can approximately estimate the uncertainty η. For example, for Q1C the average training RMS error (calculated using Table 3) is about 2.4 K/day, which can be attributed to the uncertainty η. This estimate for the uncertainty for one of the model variable, which is introduced by taking into account sub-grid scale effects, is an important result per se. The quantitative information about the uncertainty is instrumental in evaluating the accuracy of the model forecast. Now, using (6) for the test error we can estimate the NN approximation error. For this example, the test error is 2.9 K/day and, following (6), only about 1.6 K/day of this error should be attributed to the NN approximation error. If we perform such a correction for all NN ensemble members presented in Table 3, we find out that, as in the aforementioned example, after the separation of the uncertainty (the training error), the NN approximation errors on the test set do not exceed (often they are smaller than) the uncertainty.

Figures 4, 5, and 6 illustrate performance of different members of the NN ensemble on the independent test set. Figure 4 demonstrates predictions of precipitation time series produced by different NN ensemble members in comparison with "pseudo-observations" (or "Data" in the figure legend). The NN ensemble members produce an envelope (with a rather measurable spread) which on average gives a very good prediction of precipitation on the test set. The spread of the envelope shows that there is still a measurable difference between NN ensemble members, and some of the members of the envelope (e.g., {9}) give results that are closer to the "pseudo-observations." The magnitude of the spread may serve as another measure of the uncertainty of the stochastic parameterization (2). It is in agreement with the measure we introduced above, with the magnitude of the training error for PREC shown in Table 3.

Figure 5 depicts mean profiles for one of the outputs of the NN parameterization, Q1C. As in the case of precipitation, different members of the NN ensemble create an envelope with a significant spread for the mean profiles, and the magnitude of the spread is close to the training error for Q1C shown in Table 3. The differences between members inside the envelope are small as compared with the uncertainty; however, these differences are significant. They give estimates of the differences between members of the family of mappings representing the stochastic parameterization (2) and implicitly available in pseudo-observations.

Figure 6 shows the Hovmöller diagrams (the time evolution of vertical profiles) for the time series of cloudiness (CLD) profiles for the NN ensemble mean as compared with the verification data. The upper panel shows the time
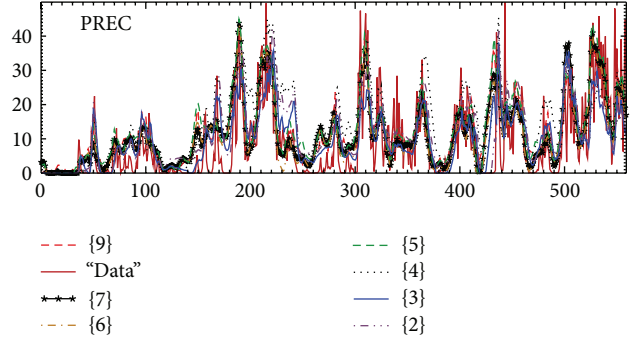

FIGURE 4: NN simulations of precipitation on the test set. Different curves presented in the figure represent seven significantly different members of the NN ensemble (see the numbers in parentheses) and verification data. The NN ensemble member {9} is shown by the thick-dashed red line.
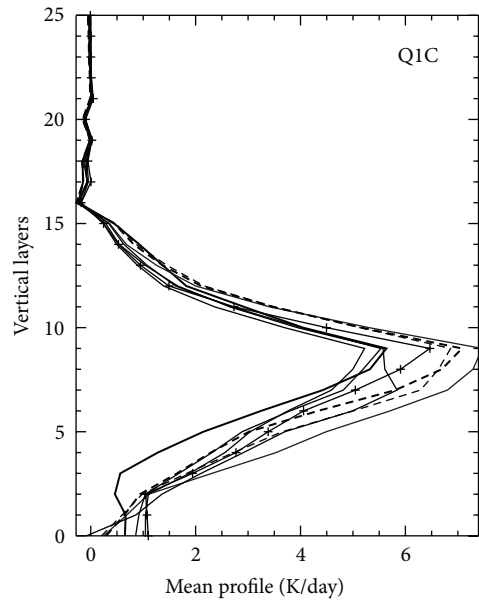

FIGURE 5: Q1C (the apparent heat source from convection) mean profiles on the test set produced by different NN ensemble members. The different curves presented in the figure correspond to different ensemble members; the thick solid line shows the verification data in the test set.

series of the pseudo-observation profiles, and the lower panel shows the time series of the profiles generated by NN ensemble. Each profile in the lower panel is the average of ten profiles generated by ten NN ensemble members. The patterns generated by the NN ensemble are a bit smoothed and diffused; they are not as sharp as the observed ones but are well recognizable. The NN ensemble mean represents the sequence of patterns well and without significant shifts.

As mentioned above, the NN errors on the raining set as well as the spread of the envelope created by different NN ensemble members represent the level of uncertainty in pseudo-observation data or the uncertainty of the stochastic parameterization (2). It means that, in the context of the current application (development of NN emulation for

TABLE 3: NN ensemble member error statistics on training (Tr) and independent test (Ts) sets. CC is the correlation coefficient. HID = 5.

Data set	Ens. mem.	Q1C (K/day)			Q2 (K/day)			Prec (mm/day)			CLD (fractions)		
		Bias	RMSE	CC	Bias	RMSE	CC	Bias	RMSE	CC	Bias	RMSE	CC
	2	$1 \cdot 10^{-3}$	2.8	0.75	$2 \cdot 10^{-2}$	4.0	0.63	$1 \cdot 10^{-2}$	6.0	0.85	$1 \cdot 10^{-4}$	0.07	0.91
	3	$2 \cdot 10^{-3}$	2.4	0.78	$2 \cdot 10^{-2}$	3.7	0.66	$1 \cdot 10^{-2}$	5.7	0.86	$2 \cdot 10^{-6}$	0.07	0.92
	4	$1 \cdot 10^{-3}$	2.3	0.81	$2 \cdot 10^{-3}$	3.7	0.68	$4 \cdot 10^{-3}$	5.2	0.89	$1 \cdot 10^{-4}$	0.07	0.92
Tr	5	$2 \cdot 10^{-3}$	2.3	0.80	$1 \cdot 10^{-3}$	3.8	0.66	$2 \cdot 10^{-2}$	5.3	0.88	$3 \cdot 10^{-5}$	0.07	0.91
	6	$4 \cdot 10^{-4}$	2.3	0.80	$1 \cdot 10^{-3}$	3.8	0.64	$1 \cdot 10^{-3}$	5.3	0.88	$6 \cdot 10^{-5}$	0.08	0.89
	7	$2 \cdot 10^{-4}$	2.3	0.81	$3 \cdot 10^{-4}$	3.7	0.67	$7 \cdot 10^{-3}$	5.2	0.89	$5 \cdot 10^{-5}$	0.06	0.93
	9	$1 \cdot 10^{-3}$	3.1	0.73	$4 \cdot 10^{-3}$	4.0	0.64	$2 \cdot 10^{-2}$	5.8	0.86	$1 \cdot 10^{-4}$	0.07	0.90
	2	−0.1	3.5	0.62	0.02	4.7	0.49	−1.1	8.5	0.68	0.03	0.11	0.81
	3	−0.6	3.5	0.62	−0.8	5.0	0.44	−5.1	10.6	0.66	0.01	0.11	0.81
	4	−0.5	3.0	0.70	−0.6	4.5	0.53	−4.0	8.8	0.73	0.00	0.09	0.86
Ts	5	−0.1	2.9	0.71	−0.1	3.9	0.52	−1.8	7.8	0.74	0.01	0.08	0.87
	6	−0.3	2.9	0.70	−0.1	3.9	0.51	−2.6	8.0	0.74	0.01	0.08	0.88
	7	−0.4	2.9	0.73	−0.5	4.3	0.58	−3.3	7.9	0.77	0.00	0.07	0.92
	9	−0.7	3.8	0.65	−0.8	4.7	0.51	−4.1	8.6	0.76	0.01	0.10	0.84
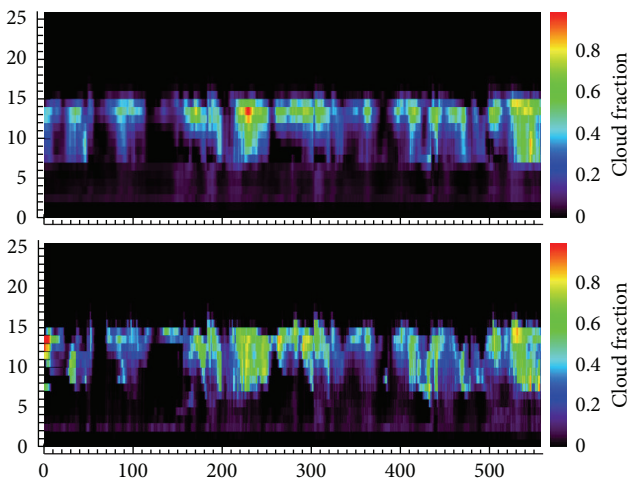
FIGURE 6: Hovmöller diagrams (the time evolution of vertical profiles) for the CLD profile time series: pseudo-observations, the upper panel and the NN ensemble mean, the lower panel. The x-axis shows time in hours for the test/validation set.

a stochastic convection parameterization (2)), selecting the best single emulating NN followed by the use of this "optimal" NN parameterization in the GCM is not the best approach.

All NNs presented here (as well as other NNs, for example, with different architectures, different number of neurons in the hidden layer evaluated in [4]) can be considered as valid emulations of the parameterization (2). These NNs should be considered as members of a NN ensemble realization of the stochastic parameterization (2) represented by a particular data set. The spread in the NN ensemble roughly reflects the skill (error) of the prediction that could be obtained using this NN ensemble.

3. Validation of the Stochastic NN Convection Parameterization in NCAR CAM

We consider the results presented in this section mostly as a proof of concept for our NN approach to developing NN convection parameterizations. Keeping this in mind, we present the results from the standpoint of their general quality with a clear understanding that more precise quantitative climatological results could be expected in our future efforts.

The NN stochastic convection parameterization described in the previous sections has been implemented as the ensemble of NNs, which are trained on the averaged CRM simulated data "pseudo-observations". In this section, we discuss the results of introduction of the NN stochastic parameterization into the NCAR CAM. Here our goal is to verify whether the NN ensemble, emulating the stochastic convection parameterization (2), provides meaningful/realistic outputs when using the CAM inputs. We performed the validation of our NN parameterization in the following two experiments.

(1) Over the TOGA-COARE location, the grid point (−2° S, 155° E) for the time period for which the TOGA-COARE data are available (the TOGA-COARE 4-month period from November 1992 to February 1993) we produced the grid-point time-mean profiles and time series.

(2) Over the large tropical Pacific region (with the area size of 120° × 30° and the following coordinates: 150°E < lon < 90° W; 15° S < lat < 15° N), we performed the parallel runs with the standard CAM and with the diagnostic CAM-NN run (see below) for the decadal (1990–2001) boreal winters, from November to February or NDJF, climate simulations. Throughout the CAM-NN run we applied at each grid point and at every time step the aforementioned ten

NN ensemble members and calculated the ensemble mean for each NN output (Q1C, Q2, and CLD profiles, and for PREC).

Note that the parallel decadal climate simulations have been actually performed for 11 years, but the decadal means (actually 11 boreal winters or NDJF) used below for validation have not included the TOGA-COARE period (see above). The TOGA-COARE data for the 4-month period from November 1992 to February 1993 were used for initializing and forcing CRM simulations, that is, for creating simulated data, which was converted into pseudo-observations used for the NN ensemble training. The validation of the parallel runs has been done for the independent decade.

For simple/initial testing and validation of the NN stochastic convection parameterization (2) in the CAM, we introduced a diagnostic mode of integration. For the diagnostic mode of integration, at every time step, the NN convection parameterization is applied, all ten NN ensemble members are evaluated, and averages of their outputs are calculated and used as NN ensemble convection parameterization outputs. Hereafter this diagnostic run is called CAM-NN. These outputs have been accumulated and the averaged fields have been calculated and compared with those produced by the original CAM convection parameterization and NCEP reanalysis data [16, 17], which provides verification data for climate simulations. Note that reanalysis data (an integrated set of global observations) is produced (every 10 years or so) using a data assimilation system (DAS) which employs a GCM and observational data for past several decades, for example, from 1948 to 2010. A DAS is a complicated procedure, which involves nonlinear optimization in the space of very high dimensionality, combining or blending observational data with the GCM simulations to produce the best possible estimates of atmospheric states for a reanalysis period.

3.1. Validation of the NN Convection Parameterization Using the NCAR CAM for the TOGA-COARE Location and 4-Month Period from November 1992 to February 1993. At the first step of our validation, the outputs generated by the ensemble of ten NNs have been compared with CAM-simulated data for one grid point at the TOGA-COARE location ($-2°$ S, $155°$ E) during the TOGA-COARE period, from November 1992 to February 1993. Thus, the CAM-simulated data were collocated in space and time with the averaged CRM simulated data.

We used the CAM-simulated T and QV as inputs for the NN ensemble trained on the averaged CRM-simulated data (pseudo-observations). The major NN outputs (CLD and PREC) obtained in this experiment have been compared with CAM CLD and PREC and with pseudo-observations. Figure 7 shows mean CLD profiles for the aforementioned experiment. The CAM-NN profile deviates from the pseudo-observation profile because the NN ensemble has been trained for pseudo-observation inputs, not for the CAM ones. It is also different from the CAM profile, which suggests that our stochastic NN convection parameterization effectively introduces in the CAM-NN run the convection and cloud
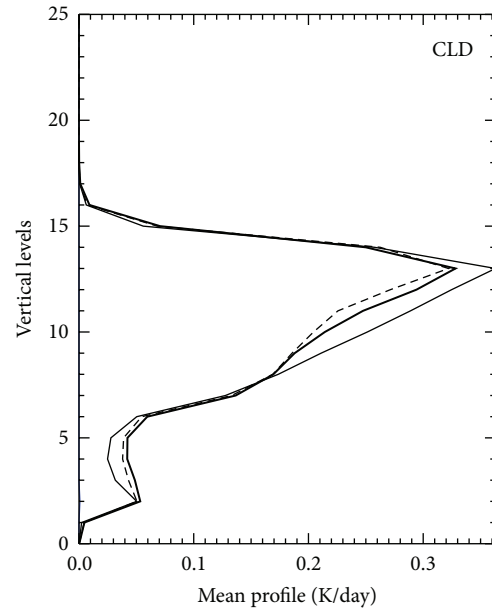

FIGURE 7: Three different mean cloud (CLD) profiles for the TOGA-COARE period: CAM-NN (thick solid), pseudo-observations (dashed), and CAM (solid).
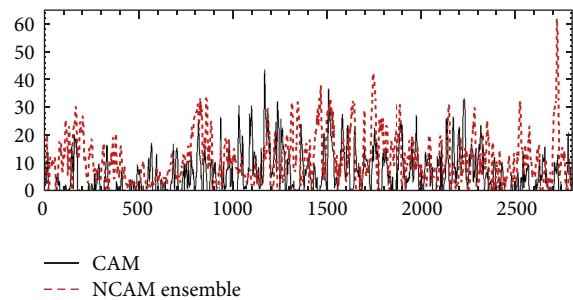

FIGURE 8: Precipitation (PREC, in mm/day) time series: CAM (black solid) and CAM-NN (or NCAM) ensemble mean (red dashed).

physics, which is different from that of introduced in the parallel CAM run employing an existing CAM convection parameterization.

Figure 8 shows the precipitation (PREC) time series produced by the original CAM run and the CAM-NN run using the NN ensemble mean. The scope, mean, and frequencies of the time series are quite similar for both models and look reasonable. Table 4 shows the bulk statistics for CLD and PREC variables for the CAM, CAM-NN runs and for pseudo-observations. Like in Figure 7 the statistics for the CAM-NN run are in between those of the CAM run and pseudo-observations; they are very reasonable and physically meaningful.

Let us stress that we should not expect full similarity here between CAM-NN and CAM statistics, profiles, and time series. The CAM-NN results are generated by our NN convection parameterization learned from CRM cloud physics, which is different from the cloud physics implemented currently in the CAM. Full similarity of the CAM and CAM-NN

TABLE 4: Bulk statistics for CLD and PREC outputs for CAM, CAM-NN, and pseudo-observations (PO).

	Mean	Standard deviation	Min	Max
PREC in mm/day				
PO	9.22	11.24	0.	80.8
CAM-NN	8.50	8.14	0.	63.6
CAM	6.41	7.23	0.	43.5
CLD in fraction				
PO	0.072	0.154	0.	1.00
CAM-NN	0.104	0.240	0.	1.00
CAM	0.159	0.256	0.	1.00

results would mean that our NN convection parameterization is not different form the convection parameterization used in the CAM and has no value in terms of introducing new physics in the CAM. Even in this case, it may still be valuable in terms of higher computational performance providing also a possibility of using NN ensembles.

3.2. Evaluating the NN Convection Parameterization Generalization Ability in Parallel Decadal Climate Simulations for a Large Tropical Pacific Region. Encouraged by the aforementioned success of our NN parameterization in the CAM-NN, we extended our diagnostic tests beyond the TOGA-COARE location and beyond the time interval covered by the CRM simulated data to test the NN parameterization generalization ability and its ability to adapt to the changing data environment. Namely, we performed the parallel decadal CAM and CAM-NN simulations and analyzed their results over a large tropical Pacific region.

We would like to emphasize that the NN convection parameterization has been developed for the TOGA-COARE location, which is represented by just one grid point in the CAM and which is actually a small area in the Equatorial Pacific (marked by a star in the middle panel of Figure 11). Also the TOGA-COARE data have been produced only over a short 4-month period (from November 1992 to February 1993 or NDJF). To evaluate the NN convection parameterization generalization and adaptive ability, we applied the NN ensemble convection parameterization in the CAM-NN run for the entire large tropical Pacific region (with the area size of $120° \times 30°$ and the following coordinates: $150°$ E < lon < $90°$ W; $15°$ S < lat < $15°$ N) during the decadal (1990–2001, with the TOGA-COARE 4-month period from November 1992 to February 1993 excluded) run. This is a very hard test for the generalization and adaptive ability of the NNs trained over a single location and a short 4-month period.

As described above, the developed NN convection parameterization has been introduced into the CAM-NN and run in the aforementioned diagnostic mode, for which CAM inputs have been used for calculating NN convection outputs. Below we compare the parallel decadal CAM-NN and CAM simulations and validate them against the NCEP reanalysis. Because NN convection was trained using simulated data
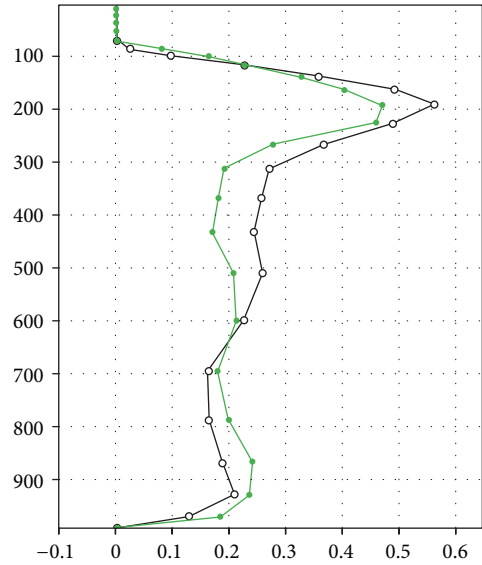

FIGURE 9: Vertical profiles of decadal boreal winter mean CLD for the TOGA-COARE location, in fractions, for the CAM-NN (open circles) and CAM (full circles) runs. Atmospheric pressure in hPa is the vertical coordinate.
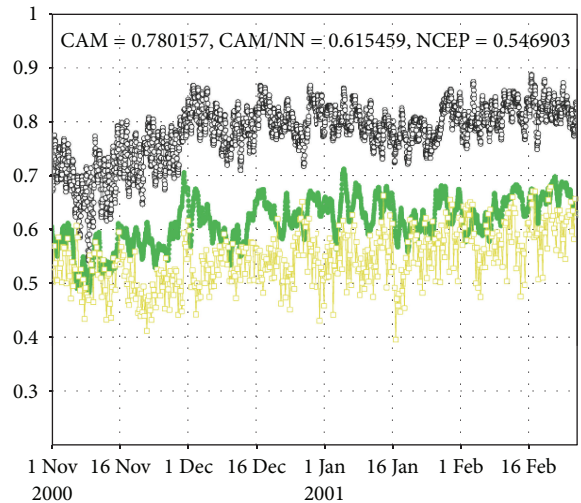

FIGURE 10: Time series of decadal boreal winter mean total cloudiness (CLD, in fractions) for the TOGA-COARE location for the CAM (black) and CAM-NN (green) runs, and for the NCEP reanalysis (yellow). Numbers in the figure show the mean values for the time series.

for the TOGA-COARE 4-month period (November 1992–February 1993), below we will analyze the decadal simulations for 4-month boreal winter seasons only.

The results of these decadal parallel climate simulations for the tropical Pacific region for decadal means of boreal winter (NDJF) distributions for total cloudiness (CLD) are shown in Figures 9, 10, and 11.

The decadal mean CLD profiles for the TOGA-COARE location for the CAM-NN and CAM runs shown in Figure 9
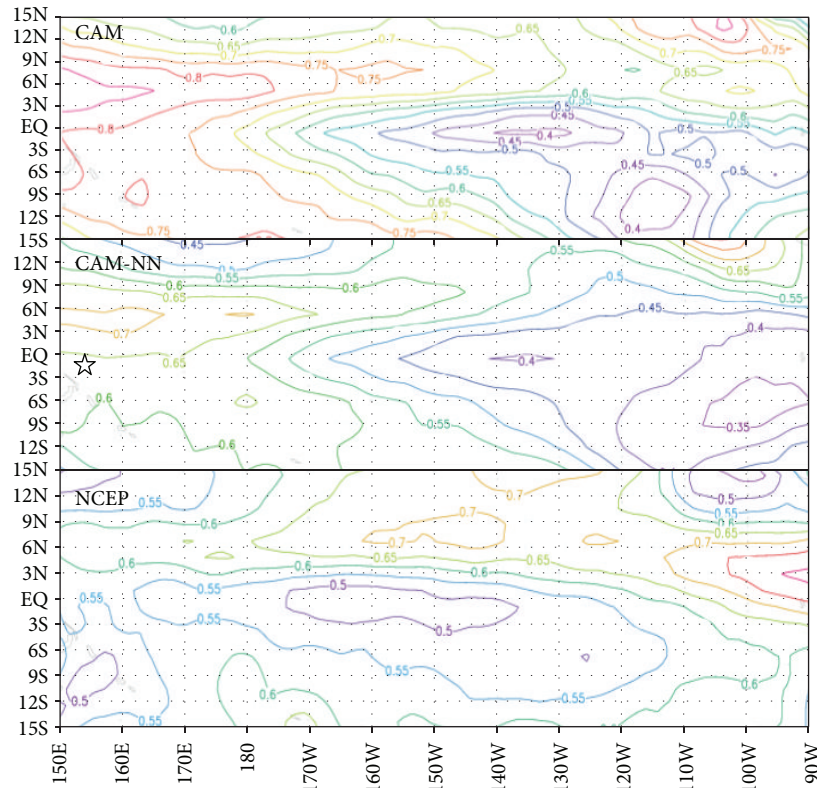
FIGURE 11: Decadal boreal winter mean cloudiness (CLD, in fractions) distribution for the CAM (upper panel) and CAM-NN (middle panel) runs over tropical Pacific region (with the area size of 120° × 30° and the following coordinates: 150° E < lon < 90° W shown with the 10° interval; 15° S < lat < 15° N shown with the 3° interval). The lower panel shows the corresponding NCEP reanalysis decadal mean distribution. The TOGA-COARE location, for which the pseudo-observations were generated and the NN ensemble was trained, is shown by a star in the middle panel. The contour interval is 0.05°C.

are close to each other. Note that the decadal mean profiles are consistent with those shown for the CAM-NN and CAM runs in Figure 7 for the TOGA-COARE period. When comparing these two figures, the difference in the vertical coordinates should be taken into account. In Figure 7 the model vertical level number is used as the vertical coordinate, whereas in Figure 9 the atmospheric pressure in hPa is used. The conversion from one coordinate to another one is essentially nonlinear. The pressure of about 200 hPa corresponds to the vertical level number 13.

The frequencies and magnitudes of the decadal mean CLD time series for the CAM and CAM-NN runs presented in Figure 10 are similar and consistent. The time series for the CAM run show measurably higher magnitudes, with the mean of 0.78, compared to those of the time series for the CAM-NN run, with the mean of 0.61. The time series of the NCEP reanalysis show lower magnitudes, with the mean of 0.54, which are significantly closer to those of the time series for CAM-NN. Note that the CLD results presented in Figure 6 have shown a close agreement of the CRM-simulated (and grid-box averaged) data and the NN ensemble mean, at the TOGA-COARE location. In our view, the improvement of the CLD time series for the decadal CAM-NN run for a large tropical region can be attributed to both a good quality of the CRM-simulated data, which implicitly represent a better

CRM cloud physics, and the positive impact from using the NN ensemble.

The horizontal distribution of total cloudiness for the large tropical Pacific Ocean region for the CAM-NN run versus the CAM control run and the NCEP reanalysis (Figure 11) have been produced and analyzed. For the region, the precipitation and cloudiness patterns for the parallel decadal CAM-NN and CAM simulations have been qualitatively and quantitatively compared.

The major result is that the regional CLD distributions for the decadal parallel runs presented in Figure 11 show a consistency and similarity, in terms of both the pattern and especially the magnitude, between the CAM-NN and CAM runs and, to some extent, with the NCEP reanalysis [17] used for validation. However, both the CAM and CAM-NN run patterns show some noticeable deviations from the NCEP reanalysis pattern. This is definitely the subject for the future improvements (see the future work outlined in Sections 4 and 5).

The CLD magnitudes for the CAM-NN run (Figure 11) are mostly closer to those of the NCEP reanalysis than the CLD magnitudes of the CAM run. However, such a positive feature should be mentioned cautiously because this is just an initial result. It is noteworthy that the CLD decadal time series for the TOGA-COARE location (Figure 10) and

the CLD distribution for the tropical Pacific Ocean (Figure 11) are consistent in the sense that for both characteristics the CAM run shows measurably higher magnitudes compared to those of the CAM-NN run, the latter being closer to those of the NCEP reanalysis.

Similar results have been obtained for the decadal boreal winter precipitation distribution over the tropical Pacific Ocean region (see [4]).

4. Discussion

At this initial stage of our development of the stochastic NN convection parameterizations, which is mostly the proof of concept, it seems reasonable to compare the CAM and CAM-NN runs mostly in terms of their general consistency between themselves and with the NCEP reanalysis. A detailed climatological analysis of regional and global simulations for all seasons will be done at the next stage of our development. It will be based on using extended more representative CRM simulations with broader spatial and temporal coverage for developing stochastic NN convection parameterizations for the CAM, which could be applied globally and for all seasons.

The CAM-NN results are generated by our NN convection parameterization learned from CRM cloud physics, which is different from the cloud physics currently implemented in the CAM convection parameterization. The CAM and CAM-NN results are consistent and quite similar, with some differences discussed above.

In our view, the results presented above in Sections 2 and 3 demonstrate a realistic potential of the presented NN ensemble approach for developing stochastic NN convection parameterizations. Our first attempt in this direction led to meaningful results despite the fact that for our development we used for the NN training a limited amount of data available over a small area in the tropical Pacific Ocean (the TOGA-COARE site) and during only four month (the TOGA-COARE period from November 1992 to February 1993). We obtained physically meaningful results not only over this particular location and time interval, but our decadal climate simulation for cloudiness and precipitations over this location and over the extended large tropical Pacific Ocean region look meaningful even without introduction in NNs an explicit time and location dependencies. These results demonstrate: (1) a very good generalization ability of the NN ensemble in this application and (2) a good ability of the NN ensemble to adapt to a changing data environment using implicit dependencies of NN inputs on time and location without introducing these dependencies explicitly.

These two issues are extremely important for future development of this approach. Our final goal is to develop a global NN convection parameterization, which can be used in the CAM and other global GCMs. To achieve this goal, a representative global data set of pseudo-observations is required. However, data for initialization and forcing CRM are available only over a few sites (TOGA-COARE, ARM, etc.); thus, CRM simulations initialized and driven by a limited amount of observations are not representative in terms of different global geographical locations and different weather conditions. Hopefully, the aforementioned data could be augmented by data simulated by the CRM, which is initialized and driven not only by observations but also by GCM-simulated data. In principle, the CRM could be run in such a way at each GCM grid point for a long period of time and supply a representative global set of pseudo-observations. However, since the CRM runs are very time consuming, this scenario is not practically feasible. In this context, good generalization and adaptation abilities of the stochastic NN convection parameterization demonstrated in this study become crucial; they will hopefully allow us to reduce the number of locations for generating pseudo-observations to a manageable and computationally affordable number of grid points.

It is noteworthy that the NN ensemble convection parameterization is very fast, contrary to any alternative approaches that have been developed to introduce new cloud and convective physics in GCMs (see [4], for details). These alternative approaches are very time consuming and barely affordable for climate simulations and weather prediction.

5. Conclusions

In this paper we introduce a novel approach to development of NN convection parameterizations based on applying the NN ensemble technique. This approach has been conceptually formulated and developed. Several very important notions are introduced which constitute the conceptual skeleton of the approach:

(1) pseudo-observations which are the result of averaging and projecting of high-dimensional and high-resolution CRM-simulated data. The pseudo-observations contain the uncertainty which is a result of averaging and projection of the original CRM simulated data,

(2) stochastic mapping/parameterization that is implicitly defined by pseudo-observations with uncertainties,

(3) NN ensemble emulation that is an adequate tool for emulating stochastic mappings/parameterizations,

(4) adaptation to temporal and spatial change in the environment, in which the NN ensemble parameterization performs through implicit time and location dependencies of NN inputs.

Our future plans include the following:

(1) running CRM simulations initialized and forced by GCM-simulated data and by reanalysis data to generate a more representative data set that will include a broader range of convection regimes, longer time periods, more locations, and more diverse weather conditions,

(2) using the representative global data set produced in this way to train a global NN convection parameterization,

(3) testing the NN convection parameterization trained using these new data in the CAM in diagnostic and prognostic modes,
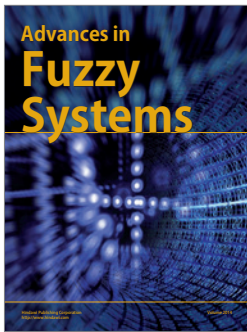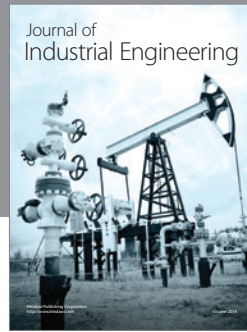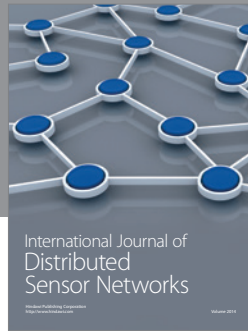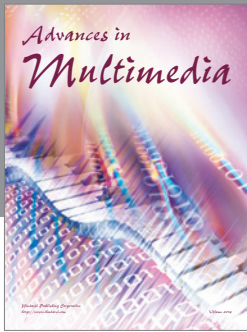
(4) introducing tools allowing the NN parameterization to adapt to changes in the environment by: (1) using time and location as additional inputs in the NN parameterization and (2) using dynamically adjustable NN parameterization based on approaches developed in [18]. The approaches use various procedures to recognize new atmospheric states emerged due to the changes in the environment. These states are used for an online adjustment of the NN parameters.

Acknowledgments

References

[1] L. J. Donner and P. J. Rasch, "Cumulus initialization in a global model for numerical weather prediction," *Monthly Weather Review*, vol. 117, pp. 2654–2671, 1989.

[2] P. J. Rasch, J. Feichter, K. Law et al., "A comparison of scavenging and deposition processes in global models: results from the WCRP Cambridge workshop of 1995," *Tellus B*, vol. 52, no. 4, pp. 1025–1056, 2000.

[3] P. J. Rasch, M. J. Stevens, L. Ricciardulli et al., "A characterization of tropical transient activity in the CAM3 atmospheric hydrologic cycle," *Journal of Climate*, vol. 19, no. 11, pp. 2222–2242, 2006.

[4] V. Krasnopolsky, M. Fox-Rabinovitz, A. Belochitski, P. Rasch, P. Blossey, and Y. Kogan, "Development of neural network convection parameterizations for climate and NWP models using Cloud Resolving Model simulations," NCEP Office Note 469, 2011, http://www.emc.ncep.noaa.gov/officenotes/newernotes/on469.pdf.

[5] T. N. Palmer and P. D. Williams, "Introduction. Stochastic physics and climate modeling," *Philosophical Transactions of the Royal Society A*, vol. 366, pp. 2421–2427, 2008.

[6] M. F. Khairoutdinov and D. A. Randall, "A cloud resolving model as a cloud parameterization in the NCAR community climate system model: preliminary results," *Geophysical Research Letters*, vol. 28, no. 18, pp. 3617–3620, 2001.

[7] M. F. Khairoutdinov and D. A. Randall, "Cloud resolving modeling of the ARM summer 1997 IOP: model formulation, results, uncertainties, and sensitivities," *Journal of the Atmospheric Sciences*, vol. 60, no. 4, pp. 607–625, 2003.

[8] F. Guichard, J. C. Petch, J. L. Redelsperger et al., "Modelling the diurnal cycle of deep precipitating convection over land with cloud-resolving models and single-column models," *Quarterly Journal of the Royal Meteorological Society*, vol. 130, no. 604, pp. 3139–3172, 2004.

[9] V. M. Krasnopolsky, M. S. Fox-Rabinovitz, and D. V. Chalikov, "New approach to calculation of atmospheric model physics: accurate and fast neural network emulation of longwave radiation in a climate model," *Monthly Weather Review*, vol. 133, no. 5, pp. 1370–1383, 2005.

[10] V. M. Krasnopolsky, M. S. Fox-Rabinovitz, and A. A. Belochitski, "Decadal climate simulations using accurate and fast neural network emulation of full, long- and short wave, radiation," *Monthly Weather Review*, vol. 136, pp. 3683–3695, 2008.

[11] V. M. Krasnopolsky, M. S. Fox-Rabinovitz, Y. T. Hou, S. J. Lord, and A. A. Belochitski, "Accurate and fast neural network emulations of model radiation for the NCEP coupled climate forecast system: climate simulations and seasonal predictions," *Monthly Weather Review*, vol. 138, no. 5, pp. 1822–1842, 2010.

[12] V. M. Krasnopolsky, "Neural network emulations for complex multidimensional geophysical mappings: applications of neural network techniques to atmospheric and oceanic satellite retrievals and numerical modeling," *Reviews of Geophysics*, vol. 45, Article ID RG3009, 2007.

[13] V. Krasnopolsky, "Neural network applications to developing hybrid atmospheric and oceanic numerical models," in *Artificial Intelligence Methods in the Environmental Sciences*, S. E. Haupt, A. Pasini, and C. Marzban, Eds., pp. 217–234, Springer, 2009.

[14] V. Krasnopolsky, "Neural network applications to solve forward and inverse problems in atmospheric and oceanic satellite remote sensing," in *Artificial Intelligence Methods in the Environmental Sciences*, S. E. Haupt, A. Pasini, and C. Marzban, Eds., pp. 191–205, Springer, 2009.

[15] D. Nguyen and B. Widrow, "Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '90)*, vol. 3, pp. 21–26, San Diego, Calif, USA, June 1990.

[16] E. Kalnay, M. Kanamitsu, R. Kistler et al., "The NCEP/NCAR 40-year reanalysis project," *Bulletin of the American Meteorological Society*, vol. 77, pp. 437–471, 1996.

[17] S. Saha, S. Moorthi, H.-L. Pan et al., "The NCEP climate forecast system reanalysis," *Bulletin of the American Meteorological Society*, vol. 91, pp. 1015–1057, 2010.

[18] V. M. Krasnopolsky, M. S. Fox-Rabinovitz, H. L. Tolman, and A. A. Belochitski, "Neural network approach for robust and fast calculation of physical processes in numerical environmental models: compound parameterization with a quality control of larger errors," *Neural Networks*, vol. 21, no. 2-3, pp. 535–543, 2008.