Neural Network Applications to Developing Hybrid Atmospheric and Oceanic Numerical Models

11

Vladimir M. Krasnopolsky

List of Acronyms

4-DVar – Four Dimensional Variational (about DAS)
CAM – Community Atmospheric Model
DAS – Data Assimilation System
DIA – Discrete Interaction Approximation
ECMWF – European Center for Medium-Range Weather Forecast
EOF – Empirical Orthogonal Functions
ENM – Environmental Numerical Model
GCM – General Circulation (or Global Climate) Model
HEM – Hybrid Environmental Model
HGCM – Hybrid GCM
HP – Hybrid Parameterization
LWR – Long Wave Radiation LWR
NCAR – National Center for Atmospheric Research
NNIA – Neural Network Interaction Approximation
NNIAE – Neural Network Interaction Approximation using EOF
NSIPP – Natural Seasonal-to-Interannual Predictability Program
Prmse(i) – RMSE for the *i*th profile, see equation (11.10)
PRMSE – Profile RMSE, see equation (11.11)
RMSE – Root Mean Square Error, see equation (11.8)
SWR – Short Wave Radiation

11.1 Introduction

The past several decades revealed a well pronounced trend in geosciences. This trend marks a transition from investigating simpler linear or weakly nonlinear single-disciplinary systems like simplified atmospheric or oceanic systems that include a limited description of the physical processes, to studying complex nonlinear multidisciplinary systems like coupled atmospheric-oceanic climate systems that take into account atmospheric physics, chemistry, land-surface interactions, etc. The most important property of a complex interdisciplinary system is that it consists of subsystems that, by themselves, are complex systems. Accordingly, the scientific and practical significance of interdisciplinary complex geophysical/environmental numerical models has increased tremendously during the last few decades, due to improvements in their quality via developments in numerical modeling and computing capabilities.

Traditional complex environmental numerical models (ENM) are deterministic models based on "first principle" equations. For example, general circulation models (GCM) a.k.a. global climate models are numerical atmospheric and oceanic models for climate simulation and weather prediction that are based on solving time-dependent 3-D geophysical fluid dynamics equations on a sphere. The governing equations of these models can be written symbolically as,

$$\frac{\partial \psi}{\partial t} + D(\psi, x) = P(\psi, x) \tag{11.1}$$

Vladimir M. Krasnopolsky (✉)
Science Application International Company at Environmental Modeling Center, National Centers for Environmental Prediction, National Oceanic and Atmospheric Administration, Camp Springs, Maryland, USA

Earth System Science Interdisciplinary Center, University of Maryland, 5,200 Auth Rd., Camp Springs, MD 20746, USA
Phone: 301-763-8000 ext. 7262; fax: 301-763-8545;
email: vladimir.krasnopolsky@noaa.gov

S. E. Haupt et al. (eds.), *Artificial Intelligence Methods in the Environmental Sciences*

where ψ are 3-D prognostic or dependent variable or set of variables (e.g., temperature, wind, pressure, moisture); x is a 3-D independent variable (e.g., latitude, longitude, and pressure or height); D is the model dynamics (the set of 3-D partial differential equations of motion, thermodynamics, etc., approximated with a spectral or grid-point numerical scheme); and P is the model physics and chemistry (e.g., the long- and short-wave atmospheric radiation, turbulence, convection and large scale precipitation processes, clouds, interactions with land and ocean processes, etc., and the constituency transport, chemical reactions, etc., respectively). These environmental models are either fully coupled atmosphere-ocean-land/biosphere-chemistry models or partially coupled models (e.g., with the chemistry component calculated off-line, driven by the flow simulated by an atmosphere-ocean-land model).

Another example of a complex ENM is an ocean wind wave model developed for simulation and forecast purposes (Tolman 2002). It is based on a form of the spectral energy or action balance equation

$$\frac{DF}{Dt} = S_{in} + S_{nl} + S_{ds} + S_{sw} \qquad (11.2)$$

where F is the 2-D Fourier spectrum of the ocean surface waves, S_{in} is the input source term, S_{nl} is the nonlinear wave-wave interaction source term, S_{ds} is the dissipation or "whitecapping" source term, and S_{sw} represents additional shallow water source terms.

It is important to emphasize that the subsystems of a complex climate (or weather, or ocean) system, such as physical, chemical, and other processes, are so complicated that it is currently possible to include them into GCMs only as 1-D (in the vertical direction) simplified or parameterized versions (a.k.a. parameterizations). These parameterizations constitute the right hand side forcing for the dynamics equations (11.1, 11.2). Some of these parameterizations are still the most time consuming components of ENMs (see examples in the next subsection).

Thus the parameterizations have a very complicated internal structure, are formulated using relevant first principles and observational data, and are usually based on solving deterministic equations (like radiation equations) and some secondary empirical components based on traditional statistical techniques like regression. Accordingly, for widely used state-of-the-art GCMs all major model components (subsystems) are predominantly deterministic; namely, not only model dynamics but the model physics and chemistry are also based on solving deterministic first principle physical or chemical equations.

In the next section, we discuss the concepts of hybrid parameterization (HP) and hybrid environmental models (HEM). HEMs are based on a synergetic combination of deterministic numerical modeling (first principle equations) with NN emulations of some model physics components. We discuss the conceptual and practical possibilities of developing a hybrid GCM (HGCM) and HEM; namely, the possibility of combining accurate and fast NN emulations of model physics components with the deterministic model dynamics of a GCM or ENM, which are the types of complex environmental models used for modern atmospheric and ocean climate modeling and weather prediction.

11.2 Concepts of a Hybrid Model Component and a Hybrid Model

One of the main problems in the development and implementation of modern high-quality high-resolution environmental models is the complexity of the physical, chemical, and other processes involved. Here we will discuss NN emulations for model physics, keeping in mind that the approach is applicable to other model components (chemical, hydrological and other processes) as well. Parameterizations of model physics are approximate schemes, adjusted to model resolution and computer resources, and based on simplified physical process equations and empirical data and relationships. The parameterizations are still so time-consuming, even for the most powerful modern supercomputers, that some of the parameterizations have to be calculated less frequently than the model dynamics. Also, different physical parameterizations are calculated at different frequencies inversely proportional to their computational complexity. This may negatively affect the accuracy of climate and other environmental simulations and predictions.

For example, in the case of a complex GCM, calculation of a physics package (including the atmospheric and land physics) at typical (a few degrees) resolution as in the National Center for Atmospheric Research (NCAR) Community Atmospheric Model (CAM) takes about 70% of the total model computations. This is despite the fact that while the model dynamics is calculated every 20 min, some computationally expensive parts of the model physics (e.g., short wave radiation) are calculated every hour. The most time consuming calculations of the model atmospheric physics, full long wave radiation including calculation of optical properties, are done only once every 12 h while the heating rates and radiative fluxes are calculated every hour. More frequent model physics calculations, desirable for temporal consistency with model dynamics, and the future introduction of more sophisticated model physics parameterizations will result in a further increase in the computational time spent calculating model physics.

In the wind wave model (11.2), the calculation of the source term, S_{nl} requires roughly 10^3 to 10^4 times more computational effort than all other aspects of the wave model combined. Present operational constraints require that the computational effort for the estimation of S_{nl} should be of the same order of magnitude as for the remainder of the wave model.

This situation is a generic and important motivation in looking for alternative, faster, and most importantly very accurate ways of calculating model physics, chemistry, hydrology and other processes. During the last decade, a new statistical learning approach based on NN approximations or emulations was applied for the accurate and fast calculation of atmospheric radiative processes (e.g., Krasnopolsky (1997); Chevallier et al. (1998)) and for emulations of model physics parameterizations in ocean and atmospheric numerical models (Krasnopolsky et al. 2000, 2002, 2005a, b). *In these works, the calculation of model physics components has been accelerated by 10 to 10^5 times as compared to the time needed for calculating the corresponding original parameterizations of the model physics.*

Approaches formulated by Chevallier et al. (1998, 2000) and Krasnopolsky et al. (2000, 2002, 2005) represent two different ways of introducing a hybridization of first principle and NN components in the physics parameterizations as well as in complex ENMs. These approaches introduce hybridization at two different system levels, at the level of the subsystem (a single parameterization) and at the level of the entire system (ENM). These two approaches lead to the concepts of a hybrid parameterization (HP) (Chevallier et al. 1998, 2000) and a hybrid environmental model (HEM) or hybrid GCM (HGCM) (Krasnopolsky et al. 2000, 2002, 2005; Krasnopolsky and Fox-Rabinovitz 2006a, b). These two concepts are discussed in the following sections.

11.3 Hybrid Parameterizations of Physics

Chevallier et al. (1998, 2000) considered a component of the complex GCM (the ECMWF global atmospheric model) – the long wave radiation (LWR) parameterization. Putting it in terms of the system levels, this single parameterization is considered to be the system and its constituents, with the blocks calculating fluxes, the blocks calculating cloudiness, etc., as the subsystems. The hybridization of first principle components with NN emulations is introduced on the level of these constituents and inside the system, which in this case is the LWR parameterization. A generic LWR parameterization can be represented as a mapping (see Chapter 9, Section 9.1.2),

$$Y = M(X) \tag{11.3}$$

in this particular case the input vector $X = (S, T, V, C)$, where the vector S represents surface variables, T is a vector (profile) of atmospheric temperatures, C is a profile of cloud variables, and the vector V includes all other variables (humidity profile, different gas mixing ratio profiles, etc.). The output of the LWR parameterization, vector Y, is composed of two vectors Q and f, $Y = (Q, f)$. Here Q is a profile of cooling rates $Q = (C_r^1, C_r^2, \ldots, C_r^L)$, where C_r^j is the cooling rate at the j-th vertical model level, and f is a vector of auxiliary fluxes computed by the LWR parameterization. Because of the presence of the cloud variable C, the mapping (11.3) may have finite discontinuities, that is, it is almost continuous.

The ECMWF LWR parameterization considered by Chevallier et al. (1998, 2000) is based on the Washington and Williamson (1977) approach which allows separate cloud variables C. In this parameterization, level fluxes are calculated as,

$$F(S, T, V, C) = \sum_i \alpha_i(C) F_i(S, T, V) \tag{11.4}$$

where each partial or individual flux $F_i(S, T, V)$ is a continuous mapping and all discontinuities related to the cloudiness are included in $\alpha_i(C)$. In their hybrid parameterization "NeuroFlux", Chevallier et al. (1998, 2000) combined calculations of cloudiness functions $\alpha_i(C)$ based on first principle equations with NN approximations for a partial or individual flux $F_i(S, T, V)$. Thus, the flux at each level (11.4) is a linear combination of approximating NNs and cloud physics coefficients $\alpha_i(C)$. As the result, the "NeuroFlux" hybrid LWR parameterization developed by Chevallier et al. (1998, 2000) is a battery of about 40 NNs (two NNs – one for the upward and another one for the downward radiation fluxes – for each of vertical level where clouds are possible). To calculate "NeuroFlux" outputs, namely the cooling rates C_rs, linear combinations of the individual approximating NNs F (equation 11.4) are differentiated at each vertical level,

$$C_r(P) = \frac{\partial F(P)}{\partial P}, \tag{11.5}$$

where P is atmospheric pressure.

The "NeuroFlux" has a very good accuracy; its bias is about 0.05 K/day and RMS error is about 0.1 K/day compared to the LWR parameterization by Washington and Williamson (1977). It is eight times faster than the parameterization by Washington and Williamson (1977). This HP approach has already led to the successful operational implementation of "NeuroFlux" in the ECMWF 4-DVar data assimilation system.

As for limitations of the HP approach, the main one stems from a basic feature of the HP approach; it is based on the analysis of the internal structure of a particular parameterization. The final design of HP is based on and follows this internal structure. Because all parameterizations have different internal structures, the approach and design of a HP developed for one parameterization usually cannot be used, without significant modifications, for another parameterization. For example, the approach used by Chevallier et al. (1998, 2000) and the design of the HP "NeuroFlux" is completely based on the possibility of separating the dependence on the cloudiness (see equation 11.4). Many other LWR parameterizations, like the NCAR CAM LWR parameterization (Collins 2001; Collins et al. 2002) or the LWR parameterization developed by Chou et al. (2001), do not allow for such separation of variables. Thus, for these LWR parameterizations as well as the short wave radiation (SWR) and the moisture model physics block parameterizations, the HP approach developed by Chevallier et al. (1998, 2000) cannot be applied directly; it should be significantly modified or redesigned for each particular new parameterization.

11.4 Hybrid Numerical Models: Accurate and Fast NN Emulations for Parameterizations of Model Physics

A new concept of a complex hybrid environmental model (HEM) has been formulated and developed by Krasnopolsky et al. (2000, 2002, 2005a) and by Krasnopolsky and Fox-Rabinovitz (2006a, b). The hybrid modeling approach considers the whole GCM or ENM as a system. Dynamics and parameterizations of physics, chemistry, etc., are considered to be the components of the system. Hybridization in this case is introduced at the level of components inside the system (ENM). For example, the entire LWR (or SWR) parameterization is emulated by a single NN as a single/elementary object or block. The NN emulation approach is based on the general fact that any parameterization of model physics can be considered as a continuous or almost continuous mapping (11.3) (see Chapter 9, Section 3.1.2).

Here we use the NCAR CAM (see *Journal of Climate* (1998) for the description of the model), a widely recognized state-of-the-art GCM used by a large modeling community for climate predictions, and the state-of-the-art NCEP wind wave model (Tolman 2002) as examples of a complex GCM and ENM. After applying the hybridization approach to the first principle based components of these models by developing NN *emulations* of model physics parameterizations, these models become the examples of an HGCM and HEM, correspondingly.

Krasnopolsky and Fox-Rabinovitz (2006a, b) formulated a developmental framework and test criteria that can be recommended for developing and testing the statistical learning components of HGCM, i.e., NN emulations of model physics components. The developmental process consists of three major steps:

1. Problem analysis or analysis of the model component (i.e., the original parameterization) to be approximated to determine the optimal structure and configuration of the NN emulations – the number of inputs and outputs and the first guess of the functional complexity of the original parameterization that determines an initial number of hidden neurons in one hidden layer of (see Chapter 9, Eqs. 9.2, 9.3).
2. Generation of representative data sets for training, validation, and testing. This is achieved by using data for NN training that are simulated by running an original GCM, i.e., a GCM with the original parameterization. When creating a representative data set, the original GCM must be run long enough to produce all possible atmospheric model simulated states, phenomena, etc. Here, due to the use of simulated data, it is not a problem to generate the sufficiently representative (and even redundant) data sets required to create high quality NN emulations. Using model-simulated data for NN training allows a high accuracy of emulation to be achieved because simulated data are almost free of the problems typical in empirical data (like a high level of observational noise, sparse spatial and temporal coverage, poor representation of extreme events, etc.).
3. Training the NN. Several different versions of NNs with different architectures, initialization, and training algorithms should be trained and validated. As for the NN architecture, the number of hidden neurons k should be kept to the minimum number that provides a sufficient emulation accuracy to create the high quality NN emulations required.

Testing the HGCM that uses the trained NN emulation consists of two major steps. The *first step* is testing the accuracy of the NN approximation against the original parameterization using the independent test data set. In the context of the hybrid approach, the accuracy and improved computational performance of NN emulations, and eventually the HGCM is always measured against the corresponding controls, namely the original parameterization and its original GCM. Both the original parameterization and its NN emulation are complicated multidimensional mappings. Many different statistical metrics of the emulation accuracy should be calculated to assure that a sufficiently complete evaluation of the emulation accuracy is obtained. For example, total, level, and profile statistics have to be evaluated (see Section 11.5.1). The *second* test *step* consists of a comprehensive comparison and analysis of parallel HGCM and GCM runs. For the parallel model simulations all relevant model prognostic (i.e., time-dependent model variables) and diagnostic fields should be analyzed and carefully compared to assure that the integrity of the original GCM and its parameterization, with all its details and characteristic features, is precisely preserved when using a HGCM with NN emulation (see Section 11.5). This test step involving model simulations is crucially important. GCMs are essentially nonlinear complex systems; in such systems, small systematic, and even random, approximation errors can accumulate over time and produce a significant impact on the quality of the model results. Therefore, the development and application framework of the new hybrid approach should be focused on obtaining a high accuracy in both NN emulations and HGCM simulations.

11.5 Atmospheric Applications: NN Emulation Components and HGCM

The NCAR CAM and NASA NSIPP (Natural Seasonal-to-Interannual Predictability Program) GCM are used in this section as examples of GCMs. The NCAR CAM is a spectral model that has 42 spectral components (or approximately $3° \times 3.5°$ horizontal resolution) and 26 vertical levels. The NSIPP model is a grid point GCM that has $2° \times 2.5°$ latitude × longitude horizontal resolution and 40 vertical levels. Note that the model vertical levels are distributed between the surface and upper stratosphere, which is at approximately 60–80 km. NN emulations were developed for the two most time consuming components of model physics, LWR and short wave radiation (SWR). The NCAR and NSIPP models have different LWR and SWR parameterizations. The complete description of the NCAR CAM atmospheric LWR is presented by Collins (2001) and Collins et al. (2002), and the NSIPP LWR by Chou et al. (2001). The full model radiation (or total LWR and SWR) calculations take $\sim$70% of the total model physics calculations. It is noteworthy that the results presented in this section were obtained using the two latest versions of NCAR CAM – the CAM-2 and CAM-3. The version of CAM used in the calculations is specified in the corresponding subsections below.

11.5.1 NCAR CAM Long Wave Radiation

The function of the LWR parameterization in atmospheric GCMs is to calculate the heating fluxes and rates produced by LWR processes. As was already mentioned, the entire LWR parameterization can be represented as an almost continuous mapping (equation 11.3). Here a very general and schematic outline of the internal structure of this parameterization is given in order to illustrate the complexity that makes it a computational "bottleneck" in the NCAR CAM physics. This information about the internal structure of the LWR parameterization was not used when creating the LWR NN emulation.

The method for calculating LWR in the NCAR CAM is based on LW radiative transfer equations in an absorptivity/emissivity formulation (see Collins 2001 and references there),

$$F^{\downarrow}(p) = B(p_t) \cdot \varepsilon(p_t, p) + \int_{p_t}^{p} \alpha(p, p') \cdot \, dB(p')$$

$$F^{\uparrow}(p) = B(p_s) - \int_{p}^{p_s} \alpha(p, p') \cdot \, dB(p') \tag{11.6}$$

where $F^{\uparrow}(p)$ and $F^{\downarrow}(p)$ are the upward and the downward heat fluxes, $B(p) = \sigma \cdot T^4(p)$ is the Stefan-Boltzmann relation; pressures p_s and p_t refer to the top and surface atmospheric pressures, and α and ε are the atmospheric absorptivity and emissivity. To solve the integral equations (11.6), the absorptivity and emissivity have to be calculated by solving the following integro-differential equations,

$$a(p, p') = \frac{\int_0^\infty \{dB_\nu(p')/dT(p')\} \cdot [1 - \tau_\nu(p, p')] \cdot d\nu}{dB(p)/dT(p)}$$

$$\varepsilon(p_t, p) = \frac{\int_0^\infty B_\nu(p_t) \cdot [1 - \tau_\nu(p_t, p)] \cdot d\nu}{B(p_t)} \tag{11.7}$$

where the integration is over wave number ν, and $B \cdot (p_t)$ is the Planck function. To solve equations (11.7) for the absorptivity and emissivity, additional calculations have to be performed and the atmospheric transmission τ_ν has to be calculated. This calculation involves a time consuming integration over the entire spectral range of gas absorption. Equations (11.6, 11.7) illustrate the complexity of the LWR internal structure and explain the poor computational performance of the original NCAR CAM LWR parameterization, which in this case is determined by the mathematical complexity of the original LWR parameterization.

The input vectors for the NCAR CAM LWR parameterization include ten vertical profiles (atmospheric temperature, humidity, ozone, CO_2, N_2O, CH_4, two CFC mixing ratios (the annual mean atmospheric mole fractions for halocarbons), pressure, and cloudiness) and one relevant surface characteristic (upward LWR flux at the surface). The CAM LWR parameterization output vectors consist of the vertical profile of heating rates (HRs) and several radiation fluxes, including the outgoing LWR flux from the top layer of the model atmosphere (the outgoing LWR or OLR). The NN emulation of the NCAR CAM LWR parameterization has the same number of inputs (220 total) and outputs (33 total) as the original NCAR CAM LWR parameterization.

NCAR CAM was run for 2 years to generate representative data sets. The first year of the model simulation was divided into two independent parts, each containing input/output vector combinations. The first part was used for training and the second for validation (control of overfitting, control of a NN architecture, etc.). The second year of the simulation was used to create a test data set completely independent from both the training and validation sets. This data set was used for testing only. All approximation statistics presented in this section were calculated using this independent test data set.

The NN emulations developed were tested against the original NCAR CAM LWR parameterization. Both the original LWR parameterization and its NN emulation are complex multidimensional mappings. Because of their complexity, many different statistics and statistical cross-sections were calculated to obtain a complete enough comparison between these two objects and to evaluate the accuracies of the NN emulations. The mean difference B (bias or systematic error of approximation) and the root mean square difference *RMSE* (a root mean square error of approximation) between the original parameterization and its NN emulation are calculated as follows:

$$B = \frac{1}{N \times L} \sum_{i=1}^{N} \sum_{j=1}^{L} [Y(i,j) - Y_{NN}(i,j)]$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \sum_{j=1}^{L} [Y(i,j) - Y_{NN}(i,j)]^2}{N \times L}} \tag{11.8}$$

where $Y(i, j)$ and $Y_{NN}(i, j)$ are outputs from the original parameterization and its NN emulation, respectively, where i = *(latitude, longitude)*, $i = 1, \ldots, N$ is the horizontal location of a vertical profile; N is the number of horizontal grid points; and $j = 1, \ldots, L$ is the vertical index where L is the number of the vertical levels.

These two error characteristics (equations (11.8)) describe the accuracy of the NN emulation integrated over the entire 4-D (latitude, longitude, height, and time) data set. Using a minor modification of equations (11.8), the bias and RMSE for the mth vertical level of the model can be calculated:

$$B_m = \frac{1}{N} \sum_{i=1}^{N} [Y(i,m) - Y_{NN}(i,m)]$$

$$RMSE_m = \sqrt{\frac{\sum_{i=1}^{N} [Y(i,m) - Y_{NN}(i,m)]^2}{N}} \tag{11.9}$$

The root mean square error can also be calculated for each ith profile:

$$prmse(i) = \sqrt{\frac{1}{L} \sum_{j=1}^{L} [Y(i,j) - Y_{NN}(i,j)]^2} \tag{11.10}$$

This error is a function of the horizontal location of the profile. It can be used to calculate a mean profile root mean square error *PRMSE* and its standard deviation σ_{PRMSE} which are location independent:

$$PRMSE = \frac{1}{N} \sum_{i=1}^{N} prmse(i)$$

$$\sigma_{PRMSE} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} [prmse(i) - PRMSE]^2} \tag{11.11}$$

The statistics (11.11) and (11.8) both describe the accuracy of the NN emulation integrated over the entire 4-D data set. However, because of a different order of integration it reveals different and complementary information about the accuracy of the NN emulations. The root mean square error profile can be calculated:

$$prmse(j) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} [Y(i,j) - Y_{NN}(i,j)]^2} \tag{11.12}$$

Several NNs have been developed that all have one hidden layer with 20 to 300 neurons. Varying the number of hidden neurons allows one to demonstrate the dependence of the accuracy of NN emulation on this parameter, which is actually the complexity of the NN emulation, as well as selecting an optimal NN emulation (Krasnopolsky et al. 2005) with the minimal complexity that still provides an emulation accuracy sufficient for a successful multi-decadal climate model integration.

All NN emulations (Krasnopolsky et al. 2005; Krasnopolsky and Fox-Rabinovitz 2006a, b) developed for the NCAR CAM LWR have almost zero or negligible systematic errors (biases). Figure 11.1 illustrates convergences of root mean square errors (11.8, 11.9, and 11.11) that are random errors in the case of negligible biases. The figure shows that an error convergence has been reached when the number of hidden neurons $k \approx 100$. However, the convergence becomes slow and non-monotonic at $k \approx 50$. The final decision about the optimal NN emulation (in terms of sufficient accuracy and minimal complexity) to be implemented into the model is based on decadal (40 year) integrations using the NN emulations within HGCM (Krasnopolsky et al. 2005; Krasnopolsky and Fox-Rabinovitz 2006a, b). For assessing the impact of using an NN emulation of the LWR parameterization in the HGCM, parallel climate simulation runs were performed with the original GCM (NCAR CAM including the original LWR parameterization) as the control run and with the HGCM (NCAR CAM including the NN emulations of LWR described above). The climate simulations were run for 50 years. As is usually done in climate simulations the simulated fields for the first 10 years, that potentially include the climate model spin-up effects, are not used for the

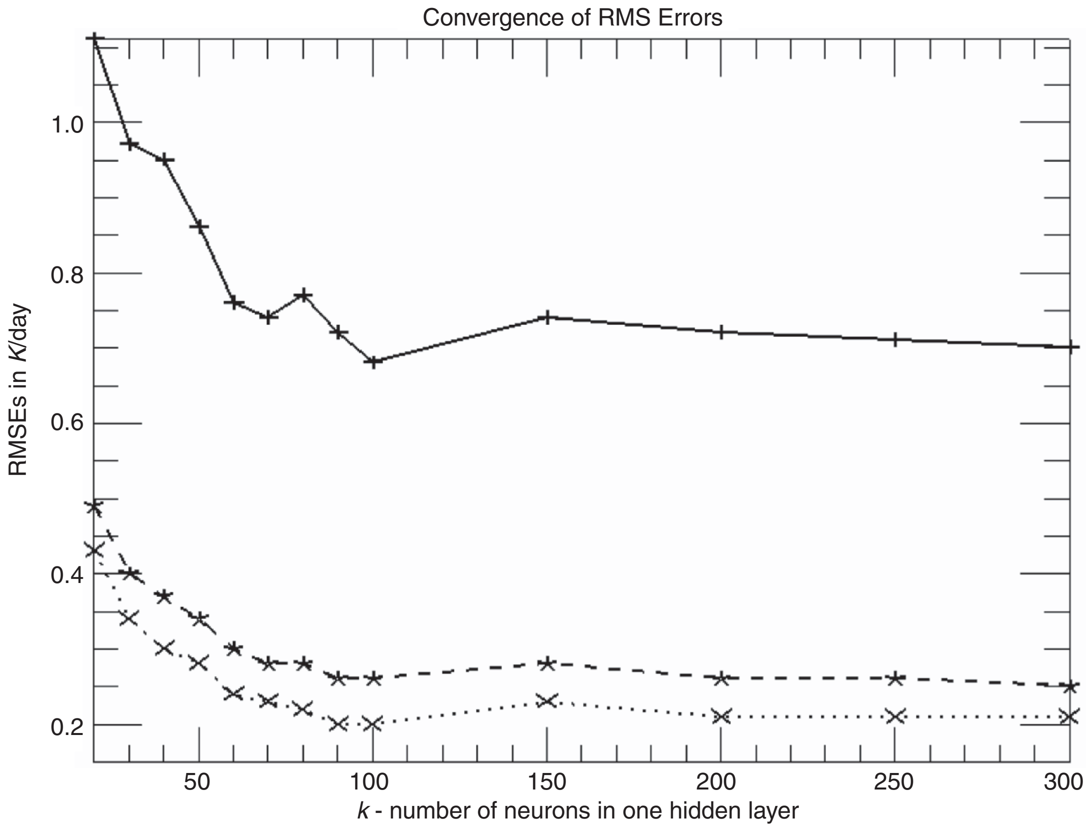

Fig. 11.1 The convergence of root mean square errors (11.8, 11.9, and 11.11). Solid line – RMSE_{26} (11.9) dashed line – RMSE (11.8), and dotted line – PRMSE (11.11)

analysis of the simulation results, leaving the remaining 40 year period to be used for that purpose.

The NN emulation with $k = 50$ (NN50) is the simplest NN emulation that could be integrated into the model for decadal (40 years or longer) climate simulations without any visible (significant) accumulations of errors in climate simulations, compared to the control run with the original LWR parameterization. This is the main indicator (in the framework of this NN application) that the accuracy of this NN emulation is sufficient for this application. Figure 11.2 shows the vertical error profile (11.12) *prmse(j)* for the "optimal" NN emulation with 50 hidden neurons (NN50). It shows that the errors are very small; at the top 10 levels the error does not exceed 0.2 K/day, at the top 20 levels it does not exceed 0.3 K/day and reaches just about 0.6–0.8 K/day at the lowest level, which does not lead to significant errors in the 40 year climate simulations with HGCM. In addition to having sufficient emulation accuracy, the NN50 NN emulation performs about 150 times faster than the original NCAR CAM LWR parameterization in a code by code comparison.

Comparisons between the control and NN emulation runs are presented in Table 11.1. They are done by analyzing the time (40-year) and global mean differences between the results of the parallel runs, as is routinely done in climate modeling. In the climate simulations performed with the original GCM and with HGCM, the time and global mean mass or mean surface pressure are precisely preserved, which is the most important preservation property for climate simulations. For the NN50 run, there is no difference in mean sea surface pressure between the NN and control runs (see Table 11.1). Other time global means, some of which are also presented in Table 11.1, show a profound similarity between the parallel simulations for these terms. These very small differences indicate the very close results from the parallel climate simulations. Other

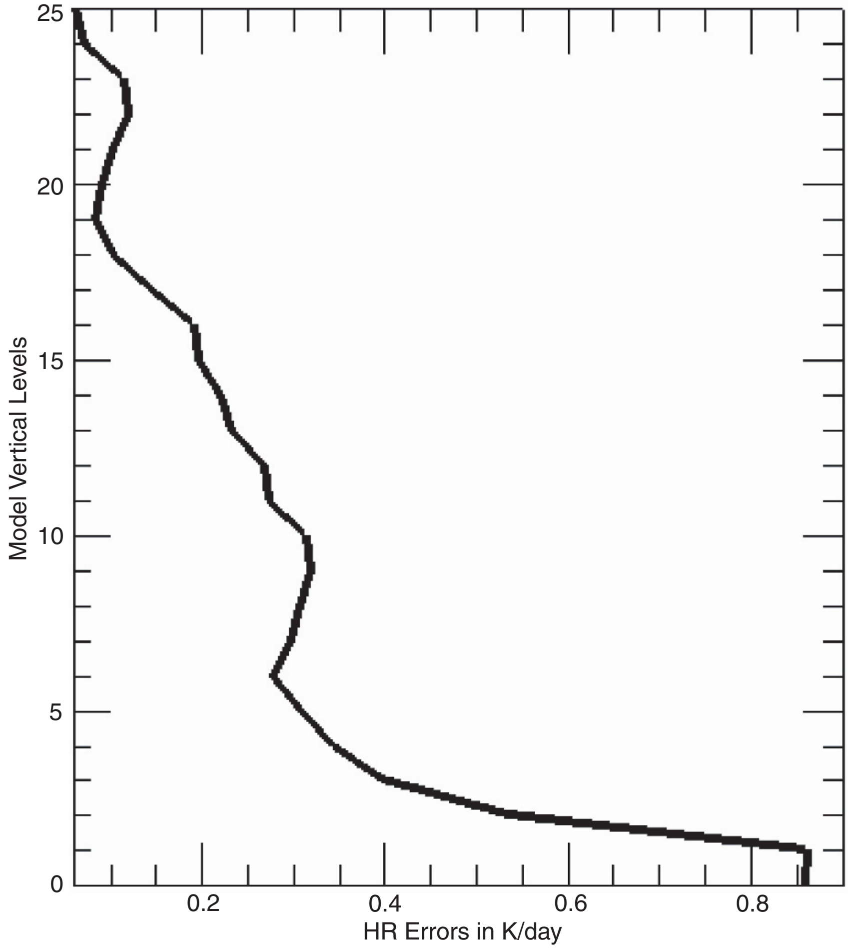

Fig. 11.2 The vertical error profile (11.10), *prmse(j),* for the "optimal" LWR NN emulation with 50 hidden neurons (NN50)

simulations (with NN90, NN150, NN200, etc.) also show that the HGCM results are profoundly similar to those of the original GCM (Krasnopolsky et al. 2005; Krasnopolsky and Fox-Rabinovitz 2006a, b). It is noteworthy that the differences between these parallel runs (HGCM and GCM) do not exceed the differences seen in two identical GCM runs performed on different supercomputers.

11.5.2 NASA NSIPP Long Wave Radiation

The robustness of the NN emulation approach was investigated using another GCM. The NASA NSIPP GCM (with a different LWR parameterization and other different model components compared to the NCAR CAM and its LWR parameterization) was used for this purpose. The input vector for the NSIPP LWR

Table 11.1 Time (40-years) and global means for mass (mean sea level pressure) and other model diagnostics for the NCAR CAM-2 climate simulations with the original LWR parameterization (in GCM), and its NN emulation (in HGCM) using NN50 and their differences (in %)

Field	GCM with the original LWR parameterization	HGCM with NN emulation	Difference (in %)
Mean sea level pressure (hPa)	1,011.48	1,011.48	$<10^{-3}$
Surface temperature (K)	289.02	288.97	0.02
Total precipitation (mm/day)	2.86	2.89	1.04
Total cloudiness (fractions, %)	60.71	61.26	0.9
Wind at 12 km (m/s)	16.21	16.16	0.3

includes surface temperature and five vertical profiles of cloud fraction, pressure, temperature, specific humidity and ozone mixing rate, for a total of 202 inputs. The NSIPP LWR output vector consists of a profile of heating rates and one surface parameter, for a total of 41 outputs.

The NN emulation accuracy and complexity results in this case (Krasnopolsky et al. 2005; Krasnopolsky and Fox-Rabinovitz 2006a, b) are very similar to the ones presented above for NCAR CAM. This illustrates the robustness of the NN emulation approach.

11.5.3 NCAR CAM Short Wave Radiation

The second component of atmospheric radiation is short wave radiation (SWR). LWR and SWR together comprise the total atmospheric radiation. The function of the SWR parameterization in atmospheric GCMs is to calculate the heating fluxes and rates produced by SWR processes. A description of the NCAR CAM atmospheric SWR parameterization is presented in a special issue of *Journal of Climate* (1998). The input vectors for the NCAR CAM SWR parameterization include twenty-one vertical profiles (specific humidity, ozone concentration, pressure, cloudiness, aerosol mass mixing ratios, etc.) and several relevant surface characteristics. NN emulations for the CAM-2 and CAM-3 versions of NCAR CAM SWR parameterizations have been developed (Krasnopolsky and Fox-Rabinovitz 2006a, b). The major difference between the CAM-2 and CAM-3 SWR versions is that CAM-3 uses significantly more information about aerosols. This extended aerosol information is responsible for a substantial increase in the number of inputs into the CAM-3 SWR parameterization as compared with CAM-2. The CAM SWR parameterization output vectors consist of a vertical profile of heating rates (HRs) and several radiation fluxes.

The data sets for training, validating, and testing SWR emulating NNs were generated in the same way as those for the LWR NN emulations described above. SWR NN emulations were tested against the original NCAR CAM SWR parameterizations using the independent test set.

The NN emulations of NCAR CAM-2 and CAM-3 SWR parameterizations have 173 and 451 inputs, respectively, and 33 outputs, which are the same numbers as the inputs and outputs for the original NCAR CAM-2 and CAM-3 SWR parameterizations. As in the case of the LWR parameterizations, several NNs were developed that all have one hidden layer with 20 to 300 neurons. In the case of the SWR parameterizations, the convergence of root mean square errors (11.8, 11.9, and 11.11) is very similar to that for the LWR parameterization shown in Fig. 11.1. The convergence is reached when the number of hidden neurons $k \approx 100$. However, it becomes slow and nonmonotonic at $k \approx 50$. The NN emulation with $k = 55$ (NN55) is the simplest NN emulation that satisfies the sufficient accuracy criterion; it could be integrated in the HGCM for multi-decadal simulations without visible (significant) accumulations of errors in climate simulations as compared to the control run with the original SWR parameterization. Figure 11.3 shows the vertical error profile (11.12) *prmse(j)* for the "optimal" NN emulation NN55. It shows that the errors are very small; at the top 20 levels the error does not exceed 0.2 K/day and reaches just about 0.45 K/day at the lowest level, which does not lead to significant errors in the HGCM climate simulations. In addition to sufficient emulation accuracy, the NN55 SWR NN emulation performs about 20 times faster than the original NCAR CAM SWR parameterization in a code by code comparison.

Comparisons between the control and NN emulation runs are also presented in Table 11.2 (see the Table 11.1 explanation in the text above). For the NN55 run there is a negligible difference between the NN and control runs for sea surface pressure (see Table 11.2). Other time global means, some of which are also presented in Table 11.2, show a profound similarity between the parallel simulations for these terms, with differences usually within about 0.3%. These very small differences indicate the very close results from the parallel climate simulations. Other simulations (with NN100, NN150, etc.) also show that the HGCM results are profoundly similar to those of the original GCM (Krasnopolsky and Fox-Rabinovitz 2006a, b).

11.5.4 NCAR CAM Full Radiation

It was shown in the previous subsections that both components of radiation, LWR and SWR, can be successfully emulated using the NN approach. This means

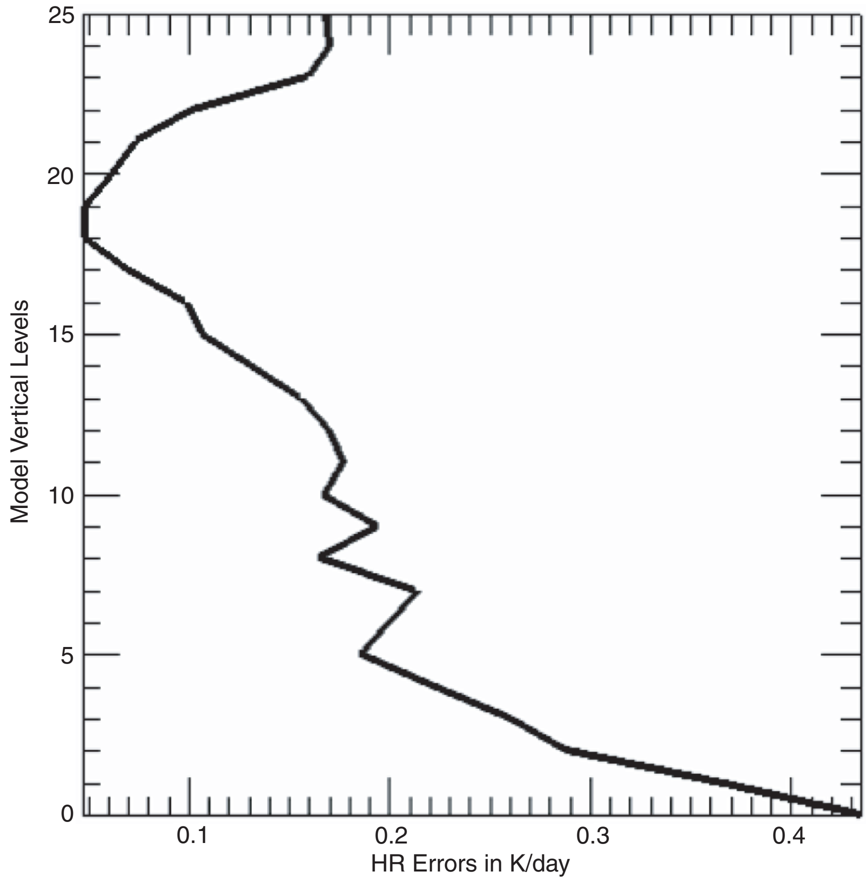

Fig. 11.3 The vertical error profile (11.10), *prmse(j),* for the "optimal" SWR NN emulation with 55 hidden neurons (NN55)

that these most time consuming components of model physics can be significantly sped up without any negative impact on the accuracy of the climate simulations. The next logical step is to combine these two NN emulations (LWR and SWR) to emulate the total model radiation. The NN50 LWR emulation and NN55 SWR emulation described in the previous subsections were combined together in one HGCM. This HGCM with the NN emulations of the total model radiation was integrated for 40 years and the results of the climate simulation were compared with those of the NCAR CAM-2 GCM simulation control run with the original NCAR CAM LWR and SWR parameterizations. In addition to having a sufficient emulation accuracy, the total radiation NN emulations perform about 12 times faster in the model than the original NCAR CAM parameterizations in terms of model time spent to calculate total radiation.

Comparisons between the control and NN emulation runs are presented in Table 11.3 (see the Table 11.1 explanation in the text above). For the total radiation run there is a negligible difference of

Table 11.2 Time (40-years) and global means for model diagnostics from NCAR CAM-2 climate simulations with the original SWR (in GCM), its NN emulation (in HGCM) using NN55, and their differences (in %)

Field	GCM with the original SWR parameterization	HGCM with SWR NN emulation	Difference (in %)
Mean sea level pressure (hPa)	1,011.48	1,011.49	0.001
Surface temperature (K)	289.01	288.97	0.01
Total precipitation (mm/day)	2.86	2.86	<0.1
Total cloudiness (fractions, %)	60.73	60.89	0.3
Wind at 12 km (m/s)	16.21	16.20	0.06

Table 11.3 Time (40-years) and global means for model diagnostics from NCAR CAM-2 climate simulations with the original LWR and SWR (in GCM), their NN emulations (in HGCM) using NN50 (LWR) and NN55 (SWR), and their differences (in %)

Field	GCM with the original LWR and SWR parameterizations	HGCM with LWR and SWR NN emulations	Difference (in %)
Mean sea level pressure (hPa)	1,011.48	1,011.50	0.002
Surface temperature (K)	289.02	288.92	0.03
Total precipitation (mm/day)	2.86	2.89	1.04
Total cloudiness (fractions, %)	60.71	61.12	0.6
Wind at 12 km (m/s)	16.21	16.29	0.5

0.002% between the NNs and control runs for sea surface pressure (see Table 11.3). Other time global means, some of which are also presented in Table 11.3, show a profound similarity between the parallel simulations for these terms. These very small differences indicate the very close results from the parallel climate simulations.

11.6 Ocean Application of the Hybrid Model Approach: Neural Network Emulation of Nonlinear Interactions in Wind Wave Models

The ocean wind wave model used for simulation and forecast purposes is another example of an ENM. It is based on a form of the spectral energy or action balance equation (11.2) and has the nonlinear wave-wave interaction source term S_{nl} as a part of the model physics in the right hand side of the equation. In its full form (e.g., Hasselmann and Hasselmann 1985) the calculation of the S_{nl} interactions requires the integration of a six-dimensional Boltzmann integral:

$$\begin{aligned} S_{nl}(\vec{k}_4) &= T \otimes F(\vec{k}) \\ &= \omega_4 \int G(\vec{k}_1, \vec{k}_2, \vec{k}_3, \vec{k}_4) \cdot \delta(\vec{k}_1 + \vec{k}_2 - \vec{k}_3 - \vec{k}_4) \\ &\quad \times \delta(\omega_1 + \omega_2 - \omega_3 - \omega_4)\left[n_1 \cdot n_3 \cdot (n_4 - n_2)\right. \\ &\quad \left. + n_2 \cdot n_4 \cdot (n_3 - n_1)\right] d\vec{k}_1 d\vec{k}_2 d\vec{k}_1 \qquad (11.13) \\ n(\vec{k}) &= \frac{F(\vec{k})}{\omega}; \quad \omega^2 = g \cdot k \cdot \tanh(kh) \end{aligned}$$

where the complicated coupling coefficient G contains moving singularities; T is a symbolic representation for the mapping. This integration requires roughly 10^3 to 10^4 times more computational effort than all other aspects of the wave model combined. Present operational constraints require that the computational effort for the estimation of S_{nl} should be of the same order of magnitude as the remainder of the wave model. This requirement was met with the development of the Discrete Interaction Approximation (DIA, Hasselmann et al. 1985). Two decades of experience with the DIA in wave models has identified significant shortcomings in the DIA (Tolman et al. 2005). It is not sufficiently accurate and, in many physically important cases, significantly deviates from the equation (11.13) deteriorating the prediction capabilities of the wind wave model (11.2).

When considering the above, it is crucially important for the development of third generation wave models to develop *an economical yet accurate approximation* for S_{nl}. A Neural Network Interaction Approximation (NNIA) was explored to achieve this goal (Krasnopolsky et al. 2002; Tolman et al. 2005). NNs can be applied here because the nonlinear interaction (11.13) is essentially a nonlinear mapping, symbolically represented by T, which relates two vectors F and S_{nl} (2-D fields in this case). Discretization of S and F (as is necessary in any numerical approach) reduces (11.13) to a continuous mapping of two vectors of finite dimensions. Modern high resolution wind wave models use descretization on a two dimensional grid which leads to S and F vector dimensions on the order of $N \sim 1{,}000$. It seems unreasonable to develop a NN emulation of such a high dimensionality (about 1,000 inputs and outputs). Moreover, such a NN will be grid dependent.

In order to reduce the dimensionality of the NN and convert the mapping (11.13) to a continuous mapping of two finite vectors that are less dependent on the actual spectral discretization, the spectrum F and source function S_{nl} are expanded using systems of

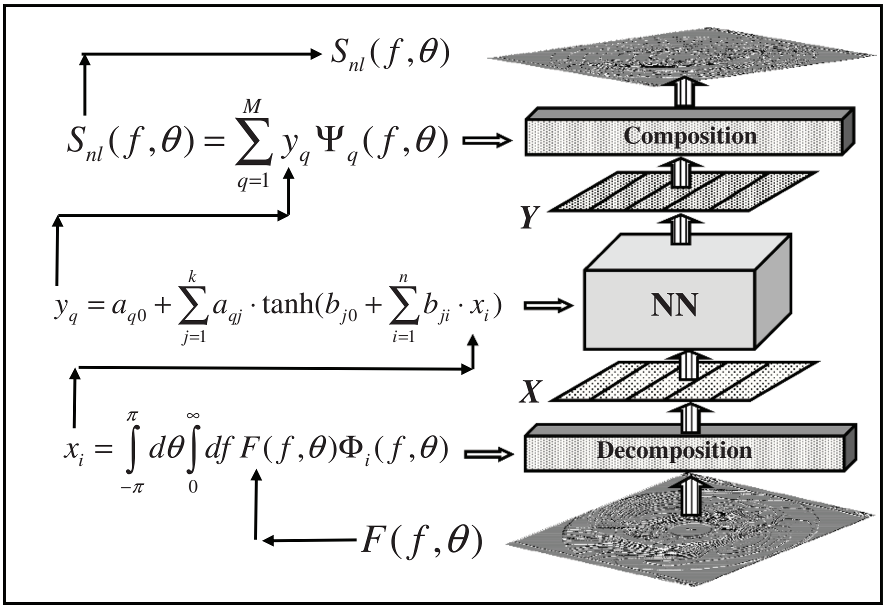

Fig. 11.4 Graphical representation of the NNIA and NNIAE algorithms

two-dimensional functions, each of which (Φ_i and Ψ_q) creates a complete and orthogonal two-dimensional basis

$$F \approx \sum_{i=1}^{n} x_i \Phi_i, \qquad S_{nl} \approx \sum_{q=1}^{m} y_q \Psi_q, \tag{11.14}$$

where for the coefficients of decomposition/composition x_i and y_q,

$$x_i = \iint F\Phi_i, \qquad y_q = \iint S_{nl}\Psi_q, \tag{11.15}$$

where the double integral identifies integration over the spectral space. Now, the developed NN emulation relates vectors of coefficients $\boldsymbol{X}$ and $\boldsymbol{Y}$: $\boldsymbol{Y} = \boldsymbol{T_{NN}}(\boldsymbol{X})$.

To train the NN emulation T_{NN}, a training set has to be created that consists of pairs of the vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$. To create this training set, a representative set of spectra F_p has to be generated with corresponding (exact) interactions $S_{nl,p}$ using equation (11.13). For each pair $(F, S_{nl})_p$, the corresponding vectors $(\boldsymbol{X},\boldsymbol{Y})_p$ are determined using equation (11.15). These pairs of vectors are then used to train the NN to obtain T_{NN}. After T_{NN} has been trained, the resulting NN Interaction Approximation (NNIA) algorithm consists of three steps: (i) decompose the input spectrum F by applying equation (11.15) to calculate $\boldsymbol{X}$; (ii) estimate $\boldsymbol{Y}$ from $\boldsymbol{X}$ using NN; and (iii) compose the output source function S_{nl} from $\boldsymbol{Y}$ using equation (11.14). A graphical representation of the NNIA algorithm is shown in Fig. 11.4.

Two approaches have been used for the basis functions. The first is the mathematical basis used in Krasnopolsky et al. (2002). As is usually done in the parametric spectral description of wind waves, separable basis functions are chosen where the frequency and angular dependence are separate. The advantage of this choice of basis functions is the simplicity of the basis generation. The disadvantage is the slow convergence of the decompositions. As an alternative, a second approach to the basis functions choice has been investigated. In this approach Empirical Orthogonal Functions (EOFs) or principal components (Lorenz 1956; Jolliffe 2002) are used (Tolman et al. 2005).

EOFs compose a statistically optimal basis. In the case considered, the basis functions Φ_i and Ψ_q are functions of two variables f and θ. The set of spectra F and source terms S_{nl}, which are used for the training of the NN, are also used to generate the EOFs for decomposing F and S_{nl}. When using EOFs the basis generation procedure is computationally expensive, with the cost increasing as the resolution of the model increases. However, as in NN training the basis generation needs to be performed only once. Stored results can be used without the need to recalculate in the final NNIA algorithm. The main advantage of EOFs is the fast convergence of the decomposition.

Table 11.4 Approximation RMSes (in nondimensional units) and performance (DIA calculation time is selected as a unit) for DIA, NNIA, NNIAE, and exact S_{nl} calculation (original)

Algorithm	RMSE	Performance
DIA	0.312	1
NNIA	0.088	4
NNIAE	0.035	7
Original parameterization	0.	$\sim 8. \times 10^5$

To distinguish between NN algorithms using different basis functions for decomposition, we use the abbreviation NNIAE for our NN algorithm that used the EOF basis. Table 11.4 demonstrates comparisons of the accuracy and performance of DIA with the two NN emulations NNIA and NNIAE, all versus the exact calculation of S_{nl} original parameterization. Approximation errors (RMSEs) are calculated in nondimensional units and performance is measured in DIA calculation times (taken as a unit). The NNIAE is nearly ten times more accurate than DIA. It is about 10^5 times faster than the original parameterization. As in the case of the atmospheric long wave radiation, a careful investigation of the parallel runs with the original ENM (the wave model with the original wave-wave interaction) and the HEM run with the NN emulation should be performed for the final test of the NN emulation (Tolman et al. 2005).

11.7 Discussion

11.7.1 Summary and Advantages of the Hybrid Modeling Approach

In this chapter, we reviewed a new hybrid paradigm in environmental numerical modeling. Within the framework of this paradigm a new type of ENM – a hybrid environmental model (HEM) based on a synergetic combination of deterministic modeling and statistical learning within an HEM (using a NN technique) is introduced. This approach uses NNs to develop highly accurate and fast emulations of model physics components. The presented results show:

(i) The conceptual and practical possibility of developing HEMs with accurate NN emulations of model components, which preserve the integrity and all the detailed features of the original ENM.

(ii) NN emulations of model physics parameterizations developed by Krasnopolsky et al. (2000, 2002, 2005) are practically identical to the original physical parameterizations, due to the capability of NN techniques to very accurately emulate complex systems like the model physics. This fact allows the integrity and level of complexity of the state-of-the-art parameterizations of model physics to be preserved. As a result, for example, a HGCM using these NN emulations produces climate simulations that are practically identical to those of the original GCM. It is noteworthy that the NN emulation developed has the same inputs and outputs as the original parameterization and is used precisely as its functional substitute within the model.

(iii) That accurate NN emulations are robust and very fast (10 to 10^5 times faster than the original parameterization) so the significant speed-up of HEM calculations can be achieved without compromising accuracy.

(iv) That statistical (NN) components can be successfully combined with deterministic model components within the HEM so their synergy can be efficiently used for environmental and climate modeling without any negative impacts on simulation quality.

(v) That this productive synergy or new combination of state-of-the-art deterministic and NN emulation approaches leads to new opportunities in using HEMs for environmental and climate simulations and prediction. For example new more sophisticated parameterizations, or even "superparameterizations" such as a cloud resolving model, that are extremely time consuming or even computationally prohibitive if used in their original form will become computationally "affordable" in ENMs when using their accurate and computationally much more efficient NN emulations in HEMs.

11.7.2 Limitations of the Current Hybrid Modeling Framework

The development of NN emulations, the core of the hybrid modeling approach, depends significantly on our ability to generate a representative training set to avoid using NNs for extrapolation far beyond the domain covered by the training set. Because of high dimensionality of the input domain that is on the order of several hundreds or more, it is rather difficult to cover the entire domain, especially the "far corners" associated with rare events, even when we use simulated data for the NN training. Another related problem is that NN emulations are supposed to be developed for an environmental or climate system that changes in time. This means that the domain configuration for a climate simulation may evolve over time, for example, when using a future climate change scenario. In both situations described the emulating NN may be forced to extrapolate beyond its generalization ability and may lead to errors in NN outputs and result in simulation errors in the corresponding HEM. The next subsection is devoted to addressing these issues.

11.7.3 Current and Future Developments of the Hybrid Modeling Approach

Two new techniques are being developed to take care of the kind of problems outlined in the previous section and to make the NN emulation approach suitable for long-term climate change simulations and other applications – a *compound parameterization* (CP) and a NN *dynamical adjustment* (DA) (Krasnopolsky and Fox-Rabinovitz 2006a, b). Here they are only briefly outlined.

CP consists of the following three elements: the original parameterization, its NN emulation, and a quality control (QC) block. During a routine HEM simulation with CP, QC block determines (at each time step of integration at each grid point) based on some criteria whether the NN emulation or the original parameterization has to be used to generate physical parameters (parameterization outputs). When the original parameterization is used instead of the NN emulation, its inputs and outputs are saved to further adjust the NN emulation. After accumulating a sufficient number of these records, a DA of the NN emulation is produced by a short retraining using the accumulated input/output records. Thus, the adapted NN emulation becomes dynamically adjusted to the changes and/or new events/states produced by the complex environmental or climate system.

There were different possible designs considered for QC (Tolman and Krasnopolsky 2004, Krasnopolsky and Fox-Rabinovitz 2006a, b). The first and simplest QC design is based on a set of regular physical and statistical tests that are used to check the

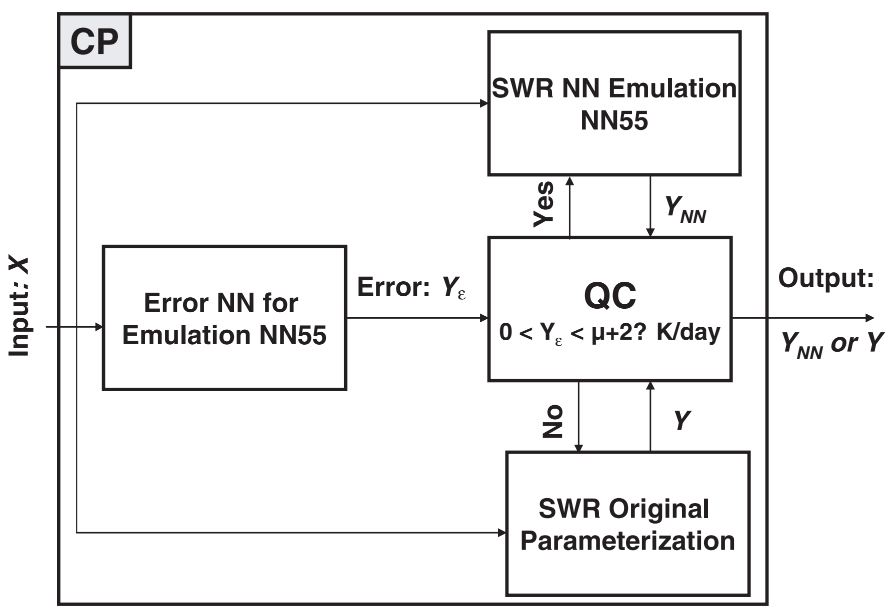

Fig. 11.5 Compound parameterization design for the NCAR CAM SWR. For each NN emulation (NN55 in this case), additional NNs (Error NN) is trained specifically for predicting, for a particular input, X, the errors, Y_ε, in the NN emulation output Y_{NN}. If these errors do not exceed a predefined threshold (mean value plus two standard deviations in this case), the SWR NN emulation (NN55) is used; otherwise, the SWR original parameterization is used instead of the NN emulation

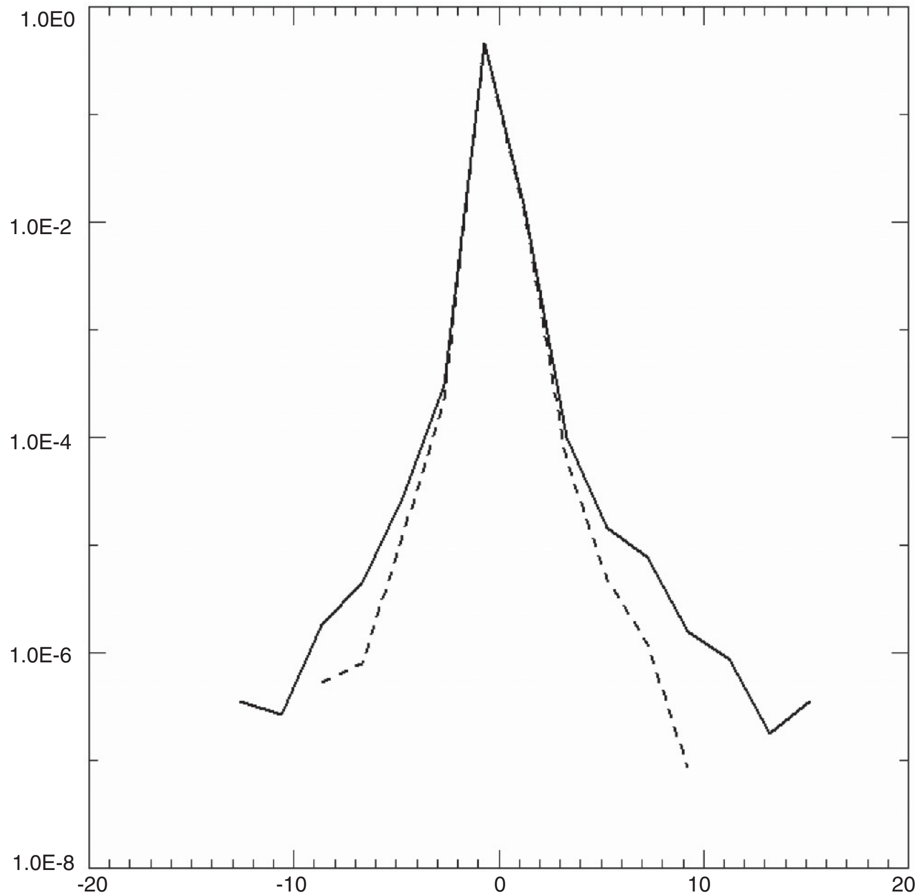

Fig. 11.6 Probability density distributions of emulation errors for the SWR NN emulation NN55 (solid line) and for the compound SWR parameterization (dashed line) are shown in Fig. 11.5. Both errors are calculated vs. the original SWR parameterization. Compound parameterization reduces the probability of medium and large errors an order of magnitude. Vertical axis is logarithmic

consistency of the NN outputs. This is the simplest, mostly generic but not sufficiently focused approach.

The second more sophisticated and effective QC design is based on training, for each NN emulation, an additional NN to specifically predict the errors in the NN emulation outputs from a particular input. If these errors do not exceed a predefined threshold the NN emulation is used; otherwise, the original parameterization is used instead. A CP of this design was successfully tested for the NCAR CAM SWR. For the SWR NN55 (see Section 11.5.3) an error NN was trained which estimated a NN55 output error *prmse(i)* (11.10) for each particular input vector X_i. The design of the CP in this case is shown in Fig. 11.5. Figure 11.6 shows the comparison of two error probability density functions. One curve (solid line) corresponds to the emulation errors of NN55, another (dashed line) corresponds to the emulation errors of the CP shown in Fig. 11.5 (both errors are calculated vs. the original parameterization on the independent test set; vertical axes are logarithmic). Figure 11.6 demonstrates the effectiveness of CP; the application of CP reduces medium and large errors by about an order of magnitude. Figure 11.7 demonstrates the effectiveness of CP in removing outliers, and Table 11.5 shows improvements in other statistical measures. It is noteworthy that for this CP less than 2% of the SWR NN emulation outputs are rejected by QC and calculated using the original SWR parameterization. Further refinement of the criteria used in the QC may result in a reduction in the already small percentage of outliers.

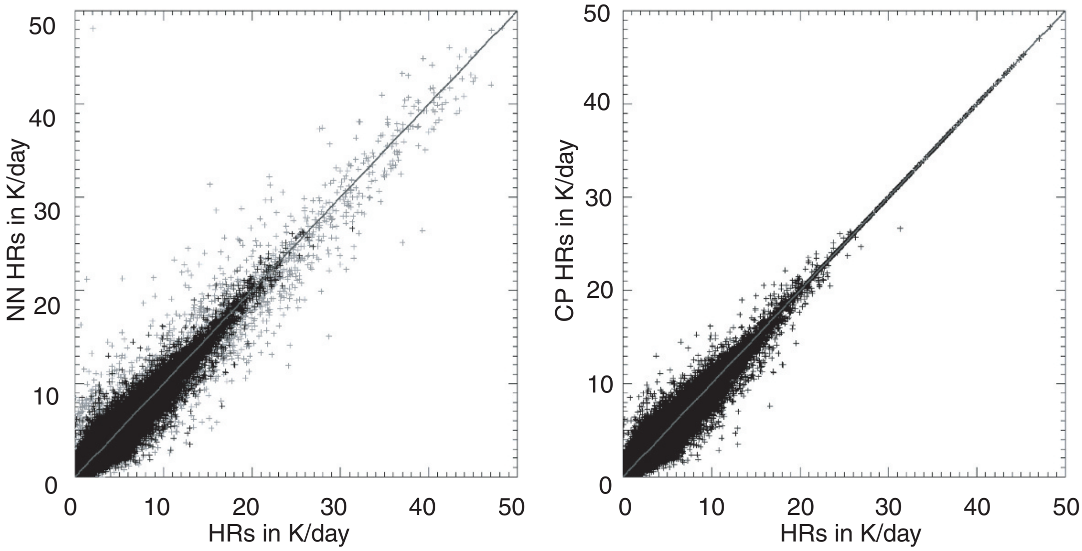

Fig. 11.7 Scatter plot for HRs calculated using the SWR NN emulation NN55 (left figure) vs. the original SWR parameterization (left and right horizontal axes) and for HRs calculated using the SWR compound parameterization (right figure) vs. the original SWR parameterization. Gray crosses (left figure) show outliers that will be eliminated by the compound parameterization (right figure)

The third QC design is based on the domain check technique proposed in the context of NN applications to satellite remote sensing (see Chapter 9, Section 9.5). In this case, QC is based on a combination of forward and inverse NNs. This design has already been successfully applied, as a preliminary study, to the ocean wave model (Section 11.6) (Tolman and Krasnopolsky 2004). Figure 11.8 illustrates the CP design in the case of the NNIA described in Section 11.6.

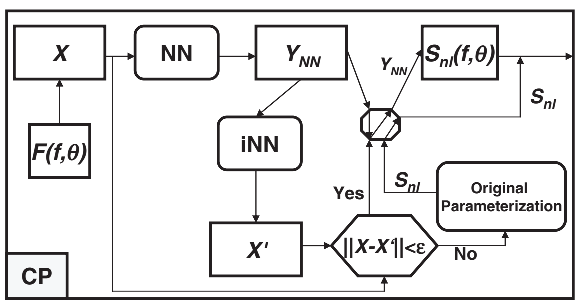

Fig. 11.8 Compound parameterization design for the NNIA and NNIAE algorithms described in Section 11.6 and shown in Fig. 11.4. Due to the use of the EOF decomposition and composition procedures the inverse NN (iNN) and QC block is implemented on the level of composition coefficients X and X'

The parameterization Jacobian, a matrix of the first derivatives of parameterization outputs over inputs, may be useful in many cases. For example, in data assimilation applications (an optimal blending of observational and simulated data to produce the best possible fields) a Jacobian is used to create an adjoint (a tangent-linear approximation). A Jacobian is also instrumental for a statistical analysis of the original parameterization and its NN emulation. An inexpensive computation of the Jacobian when using a NN emulation is one of the advantages of the NN approach. Using this Jacobian in combination with the tangent-linear approximation can additionally accelerate the calculations (Krasnopolsky et al. 2002). However since the Jacobian is not trained, it is simply calculated through the direct differentiation of

Table 11.5 Error statistics for SWR NN emulation NN55 and SWR compound parameterization: Bias and RMSE (25), $RMSE_{26}$ (26), and Extreme Outliers (Min Error & Max Error)

	Bias	RMSE	$RMSE_{26}$	Min error	Max error
SWR NN55	4.10^{-3}	0.193	0.434	−46.1	13.6
SWR CP	4.10^{-3}	0.171	0.302	−9.2	9.5

an emulating NN. In this case the statistical inference of a Jacobian is an ill-posed problem and it is not guaranteed that the derivatives will be sufficiently accurate.

It is noteworthy that for the type of NN applications considered in this section, the NN emulation approach that treats a parameterization of model physics as a single object offers a simple and straightforward solution that alleviates the need for calculating the NN Jacobian explicitly. The adjoint tangent-linear approximation of a parameterization (e.g., of a radiation parameterization) may be considered as an independent/new parameterization, the NN emulation approach can be applied to such a parameterization, and a separate NN emulation can be trained to emulate the adjoint. For other applications that require an explicit calculation of the NN Jacobian, several solutions have been offered and investigated (Aires et al. 1999, 2004; Krasnopolsky 2006).

References

Aires, F., Schmitt, M., Chedin, A., & Scott, N. (1999). The "weight smoothing" regularization of MLP for Jacobian stabilization. *IEEE Transactions on Neural Networks*, *10*, 1502–1510.

Aires, F., Prigent, C., & Rossow, W. B. (2004). Neural network uncertainty assessment using Bayesian statistics: A remote sensing application. *Neural Computation*, *16*, 2415–2458.

Chevallier, F., Chéruy, F., Scott, N. A., & Chedin, A. (1998). A neural network approach for a fast and accurate computation of longwave radiative budget. *Journal of Applied Meteorology*, *37*, 1385–1397.

Chevallier, F., Morcrette, J.-J., Chéruy, F., & Scott, N. A. (2000). Use of a neural-network-based longwave radiative transfer scheme in the EMCWF atmospheric model. *Quarterly Journal of Royal Meteorological Society*, *126*, 761–776.

Chou, M.-D., Suarez, M. J., Liang, X.-Z., & Yan, M. M.-H. (2001). A thermal infrared radiation parameterization for atmospheric studies. *Technical Report Series on Global Modeling and Data Assimilation*, Editor Max J. Suarez (NASA/TM-2001-104606), Vol. 19.

Collins, W. D. (2001). Parameterization of generalized cloud overlap for radiative calculations in general circulation models. *Journal of the Atmospheric Sciences*, *58*, 3224–3242.

Collins, W. D., Hackney, J. K., & Edwards, D. P. (2002). A new parameterization for infrared emission and absorption by water vapor in the national center for atmospheric research community atmosphere model. *Journal of Geophysical Research*, *107*(D22), 1–20.

Hasselmann, S., & Hasselmann, K. (1985). Computations and parameterizations of the nonlinear energy transfer in a gravity wave spectrum. Part I: A new method for efficient computations of the exact nonlinear transfer integral. *Journal of Physical Oceanography*, *15*, 1369–1377.

Hasselmann, S. et al. (1985). Computations and parameterizations of the nonlinear energy transfer in a gravity wave spectrum. Part II: Parameterization of the nonlinear transfer for application in wave models. *Journal of Physical Oceanography*, *15*, 1378–1391.

Jolliffe, I. T. (2002). *Principal component analysis* (502 pp.). New-York: Springer.

Journal of Climate (1998), *11*(6) (the special issue).

Krasnopolsky, V. (1997). A neural network-based forward model for direct assimilation of SSM/I brightness temperatures. *Technical note* (OMB contribution No. 140). NCEP/NOAA, Camp Springs, MD 20746.

Krasnopolsky, V. M. (2006). Reducing uncertainties in neural network Jacobians and improving accuracy of neural network emulations with NN ensemble approaches. *Proceedings of the IJCNN2006*, Vancouver, BC, Canada, July 16–21, 2006, pp. 9337–9344, CD-ROM; (2007), *Neural Networks*, *20*, 454–461.

Krasnopolsky, V. M., & Fox-Rabinovitz, M. S. (2006a). A new synergetic paradigm in environmental numerical modeling: Hybrid models combining deterministic and machine learning components. *Ecological Modelling*, *191*, 5–18.

Krasnopolsky, V. M., & Fox-Rabinovitz, M. S. (2006b). Complex hybrid models combining deterministic and machine learning components for numerical climate modeling and weather prediction. *Neural Networks*, *19*, 122–134.

Krasnopolsky, V. M. et al. (2000). Application of neural networks for efficient calculation of sea water density or salinity from the UNESCO equation of state. *Proceedings of the Second Conference on Artificial Intelligence, AMS*, Long Beach, CA, January 9–14, 2000, pp. 27–30.

Krasnopolsky, V. M, Chalikov, D. V., & Tolman, H. L. (2002). A neural network technique to improve computational efficiency of numerical oceanic models. *Ocean Modelling*, *4*, 363–383.

Krasnopolsky, V. M., Fox-Rabinovitz, M. S., & Chalikov, D. V. (2005a). New approach to calculation of atmospheric model physics: Accurate and fast neural network emulation of long wave radiation in a climate model. *Monthly Weather Review*, *133*, 1370–1383.

Krasnopolsky, V. M, Fox-Rabinovitz, M. S., & Chou, M.-D. (2005b). Robustness of the NN approach to emulating atmospheric long wave radiation in complex climate models. *Proceedings of the International Joint Conference on Neural Networks*, Montréal, Québec, Canada, July 31–August 4, 2005, pp. 2661–2665.

Lorenz, E. N. (1956). *Empirical* orthogonal functions and statistical weather prediction. *Statistical forecasting project* (Sci. Rep. No. 1). Cambridge, MA: MIT Press, 48 pp.

Tolman, H. L. (2002). User manual and system documentation of WAVEWATCH III version 2.22. *Technical note 222*. NOAA/NWS/NCEP/MMAB, p. 133, Camp Springs, MD 20746.

Tolman, H. L., & Krasnopolsky, V. M. (2004). Nonlinear interactions in practical wind wave models. *Proceedings of 8th International Workshop on Wave Hindcasting and Forecasting*, Turtle Bay, Hawaii, 2004, CD-ROM, E.1.

Tolman, H. L., Krasnopolsky, V. M., & Chalikov, D. V. (2005). Neural network approximations for nonlinear interactions in wind wave spectra: Direct mapping for wind seas in deep water. *Ocean Modelling*, *8*, 253–278.