

## **Development of Neural Network Emulations of Model Radiation for Improving the Computational Performance of the NCEP Climate Simulations and Seasonal Forecasts**

V. M. Krasnopolsky<sup>1,2</sup>, M. S. Fox-Rabinovitz<sup>2</sup>, Y. T. Hou<sup>1</sup>, S. J. Lord<sup>1</sup>, and A. A. Belochitski<sup>2</sup>

<sup>1</sup>*Environmental Modeling Center, NOAA/NWS/NCEP*

<sup>2</sup>*Earth System Science Interdisciplinary Center (ESSIC), University of Maryland*

### **1. Introduction**

Calculation of model physics in a GCM (General Circulation Model) usually takes a very significant part of the total model computations. Evidently, this percentage is model dependent but full model radiation is the most time-consuming component of GCMs (*e.g.*, Morcrette *et al.* 2008, Manners *et al.* 2008). In both climate modeling and NWP, the calculation of radiative transfer is necessarily a trade-off between accuracy and computational efficiency. Very accurate methods exist, such as line-by-line procedures that could be employed ideally to calculate radiative fluxes for every grid-point at every time-step. However, if the radiation transfer were to be computed for every grid point and at all time steps, it would generally require as much CPU time or more than the rest of the model components, *i.e.*, model dynamics and other physical parameterizations (Morcrette *et al.* 2008). Therefore a number of simplifications are usually made to reduce this cost to manageable levels.

For example, in the majority of modern radiative schemes, the correlated-k method (Lacis and Oinas 1991) is typically used to reduce the integration over wavelength by effectively binning wavelengths with similar absorption coefficients (k-terms). This simplification reduces greatly the number of monochromatic radiative transfer calculations required. The number of k-terms can be adjusted, which provides a trade-off between the accuracy and efficiency required for a given application. However, the correlated-k methods cannot be made sufficiently computationally efficient to allow calculations for every grid-point at every time-step.

To reduce the cost further, calculations are usually made at lower temporal and/or spatial resolutions. Quite drastic reductions in temporal resolution are often made (*e.g.*, radiation calculations are made every three hours for the climate and global forecast models at NCEP and UKMO (Manners *et al.* 2008)). Between radiative transfer calculations major changes may occur in the radiative profiles (caused primarily by two factors: changes in clouds and changes in the angle of incident solar radiation) that are not represented. A reduced horizontal resolution approach (the radiative calculations are performed on a coarser grid with a following interpolation of the results to an original finer grid) is used to speed up radiation calculations at ECMWF (Morcrette *et al.* 2007, 2008). A reduced vertical resolution approach (the full radiation is calculated at every other vertical level and interpolated on the intermediate levels) is used in the Canadian operational Global Environmental Multiscale model (*e.g.* Côté *et al.* 1998). Such approaches reduce horizontal or vertical variability of radiation fields. Thus, these approaches may reduce the accuracy of a model's radiation calculation and its spatial or/and temporal consistency with other parts of model physics and with model dynamics, which may, in turn, affect negatively the accuracy of climate simulations and weather predictions.

Such a situation is an important motivation for developing new alternative numerical algorithms that provide faster calculations of model physics while carefully preserving their accuracy. In our previous studies (Krasnopolsky *et al.* 2005, 2008) we demonstrated that the neural network (NN) emulation approach can be successfully used to speed up significantly (by one to two orders of magnitude) the calculations of model radiation while providing a sufficient accuracy of decadal (50 years) climate simulations. We also demonstrated that this approach is a generic one; namely it can be used not only for emulating any formulation of the long wave radiation (LWR) physics but also for emulating any formulation of short wave radiation (SWR) physics.

In this study we applied the NN emulation approach to the higher complexity NCEP CFS (Climate Forecasting System), which required further development of the neural network emulation methodology. We demonstrate that the NN emulation approach for model radiation can be successfully applied to the *significantly higher resolution coupled ocean-atmosphere-land-ice model with time dependent CO2*. The atmospheric part of CFS has spectral T126 horizontal resolution and 64 vertical levels (T126L64); it is coupled with the 40-level interactive MOM4 ocean model, with a state-of-the-art 3D land model, and with an ice model.

In Section 2, the improved NN emulation approach and developed NN emulations for the NCEP CFS long-wave radiation (LWR) and short-wave radiation (SWR) are briefly described in terms of their design, accuracy, and computational performance. In Section 3, the results of the parallel decadal model simulations, one using both LWR and SWR NN emulations for calculation of full model radiation and the other using the original model radiation (the control run) are compared in terms of similarity of their spatial and temporal variability characteristics. Section 4 contains conclusions.

## 2. NN emulations for the NCEP CFS radiation

### 2.1 Background information on the NN emulation approach

NN emulations of model physics are based on the two following facts. First, any parameterization of model physics is a continuous or almost continuous mapping (input vector vs. output vector dependence) and can be symbolically written as:

$$Y = M(X); \quad X \in \mathfrak{R}^n, Y \in \mathfrak{R}^m \quad (1)$$

where  $M$  denotes the mapping,  $n$  is the dimensionality of the input space, and  $m$  is the dimensionality of the output space. And second, NNs (multilayer perceptrons) are generic tools for approximation of such mappings (Funahashi 1989).

NN is an analytical approximation that uses a family of functions like:

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

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  $\Phi$  is usually a hyperbolic tangent,

$n$  and  $m$  are the numbers of inputs and outputs (the same  $n$  and  $m$  as in Eq. (1)), respectively, and  $k$  is the number of neurons in the hidden layer. Definitions of NN terminology can be found in many places, for example in the recent book by Bishop (2006) and in the review paper by Krasnopolsky (2007); however, eq. (2) is sufficient to understand the subject of this paper. The numerical complexity of NN (2) can be well approximated by a number of NN weights (Krasnopolsky 2007):

$$N_C = k \cdot (n + m + 1) + m \quad (3)$$

The NN numerical complexity  $N_C$  determines the time,  $T_{NN}$ , required for the estimating NN (2),

$$T_{NN} = c \cdot N_C$$

$T_{NN}$ , is directly proportional to  $N_C$  with the coefficient of proportionality  $c$  depending mainly on a hardware and software environment of the computer used.

### 2.2 NN emulations for full model radiation

The LWR and SWR parameterizations together comprise the full model radiation. The LWR and SWR parameterizations or the full model radiation for the NCEP CFS have been emulated using two NNs, one for LWR and another for SWR.

The input and output vectors for NNs, emulating the LWR or SWR parameterizations, include the same parameters as those of the input and output vectors for the original LWR or SWR parameterizations, respectively. For the LWR NN emulation, these input parameters are the following nine profiles: atmospheric pressure, temperature, specific humidity, ozone mixing ratio, total cloud fraction, cloud liquid

water path, mean effective radius for liquid cloud, cloud ice water path, and mean effective radius for ice cloud. The LWR parameterization (and LWR NN emulation) output vectors consist of the profile of heating rates (HRs) and five radiation fluxes: the total sky outgoing LW radiation flux from the top layer of the model atmosphere (the outgoing LWR or OLR), the clear sky upward flux at the top of the model atmosphere, the total sky upward flux at the surface, the total sky downward flux at the surface, and the clear sky downward flux at the surface.

The NN emulation of the LWR parameterization includes all non-constant inputs of the original LWR (total 556 inputs;  $n = 556$  in Eq. (1)). It has the same outputs (total 69 outputs;  $m = 69$  in Eq. (1)) as the original LWR parameterization. We have developed several NNs, all of which have the same aforementioned inputs and outputs, with the number  $k$  changing from 50 to 200 in Eq. (2). Varying  $k$ , the number of terms (or neurons) in Eq. (2), allows us to demonstrate the dependence of the accuracy of approximation on this parameter as well as its convergence, and as a result, to provide a sufficient accuracy of approximation for the model (*e.g.* Krasnopolsky *et al.* 2005).

The input vectors for the SWR parameterization include 55 vertical profiles: atmospheric pressure, temperature, specific humidity, ozone, CO<sub>2</sub>, N<sub>2</sub>O, O<sub>2</sub>, and CH<sub>4</sub> volume mixing ratios, total cloud fraction, cloud liquid water path, mean effective radius for liquid cloud, cloud ice water path, mean effective radius for ice cloud, and three profiles (optical depth, single scattering albedo, and asymmetry parameter) for each of 14 different species of aerosols. The input vectors include also the solar zenith angle, the solar constant and the surface albedo for four different bands. The SWR parameterization output vectors consist of a vertical profile of heating rates (HRs) and nine radiation fluxes: three fluxes at the top layer of the model atmosphere (the total sky outgoing SW radiation flux, the total sky downward flux, the clear sky upward flux), four radiation fluxes at the surface (the total sky upward and downward fluxes and the clear sky upward and downward fluxes), and the downward (the total and clear sky) fluxes in the UV-B spectral band.

The NN emulations of the SWR parameterization have 562 inputs and 73 outputs. We have developed several NNs, with the number  $k$  changing from 50 to 200 in Eq. (2). It is noteworthy that, as in the case of the NN emulation of LWR, the number of NN inputs is less than the number of input profiles multiplied by the number of vertical layers plus the number of relevant single level characteristics. Many input variables (*e.g.*, almost all gases) have zero or constant values for the upper vertical layers, and for some gases the entire volume mixing ratio profile is a constant (obtained from climatological data).

### 2.3 Generating data sets for NN training and validation

The NCEP CFS (T126L64) has been run for seventeen years to generate representative data sets. The representative data set samples adequately the atmospheric state variability, *i.e.*, it represents all possible states produced by the model as fully as possible (including the states introduced due to time dependent CO<sub>2</sub> concentration). All inputs and outputs of original LWR and SWR parameterizations have been saved for two days per month, *i.e.*, for one day at the beginning and one day in the middle of the month, every three hours (eight times per day) to cover the annual and diurnal cycles. From each three hour global data set three hundred events (the set of input and output profiles) have been selected. The data set was divided into three independent parts, each containing input/output vector combinations. Each part consists of about 200,000 input/output records. The first part has been used for training and the second one for tests (control of overfitting, control of NN architecture, *etc.*). The third part of the data set was used to create a validation data set independent of both the training and test data sets. The third part or the validation set was used for validation only. All approximation statistics presented in this section are calculated using this independent validation data set. The accuracy of the NN emulation, *i.e.*, biases and rmse, are calculated against the control (the original parameterization).

It is noteworthy that along with the aforementioned requirement of representing all possible states produced by the model, the size of the training data set is limited mainly by the training time, which, in turn, is determined by the processor type and the amount of memory available. The training time is approximately proportional to the size of the training data set. In our case, the selection of about 200,000 input/output records for training is a result of an optimal choice providing a sufficient representativeness of the training set and a reasonable training time. We selected the size of the test set equal to the size of the training set

because the training and test sets are supposed to have close statistical properties. There are no serious limitations to the size of the validation set; we selected it equal to the size of the first two sets.

#### 2.4 Bulk approximation error statistics

To ensure a high quality of representation of the LWR and SWR processes, the accuracy of their NN emulations has been carefully investigated. The NN emulations have been validated against the original NCEP CFS LWR and SWR parameterizations. To calculate the error statistics presented in Table 1; the original parameterizations and their NN emulations have been applied to the validation data set. Two sets of the corresponding HR profiles have been generated for both LWR and SWR. Total and level bias (or a mean error), total and level RMSE, profile RMSE or PRMSE, and  $\sigma_{\text{PRMSE}}$  have been calculated (see Krasnopolsky 2007).

Statistics Types	Statistics	LWR			SWR	
		NCAR CAM	NCEP CFS		NCAR CAM	NCEP CFS RRTMG
			RRTMG	RRTMF		
Total Error Statistics	Bias	$3. \cdot 10^{-4}$	$2. \cdot 10^{-3}$	$7. \cdot 10^{-4}$	$-4. \cdot 10^{-3}$	$5. \cdot 10^{-3}$
	RMSE	0.34	0.49	0.42	0.19	0.20
	PRMSE	0.28	0.39	0.30	0.15	0.16
	$\sigma_{\text{PRMSE}}$	0.2	0.31	0.30	0.12	0.12
Bottom Layer Error Statistics	Bias	$-2. \cdot 10^{-3}$	$-1. \cdot 10^{-2}$	$6. \cdot 10^{-3}$	$-5. \cdot 10^{-3}$	$9. \cdot 10^{-3}$
	RMSE	0.86	0.64	0.67	0.43	0.22
Top Layer Error Statistics	Bias	$-1. \cdot 10^{-3}$	$-9. \cdot 10^{-3}$	$2. \cdot 10^{-3}$	$2. \cdot 10^{-3}$	$1. \cdot 10^{-2}$
	RMSE	0.06	0.18	0.09	0.17	0.21
NN Complexity	$N_C$ See eq. (3)	12,733	33,294	93,969	11,418	45,173
Speedup, $\eta$	Times	150	12	21	20	45

**Table 1** Statistics estimating the accuracy of HRs (in K/day) calculations and the computational performance for NCEP CFS (T126L64) LWR and SWR using NN emulation vs. the original parameterization. For comparison, NCAR CAM (T42L26) LWR and SWR statistics are also shown. Total statistics show the bias, RMSE, PRMSE, and  $\sigma_{\text{PRMSE}}$  for the entire 3-D HR fields. Layer (for the top and bottom layers) statistics show the bias and RMSE for one horizontal layer (the top or bottom layer). Also, the NN complexity  $N_C$  (3) and speedup  $\eta$  (how many times NN emulation is faster than the original parameterization) are shown. RRTMG and RRTMF are different versions of the radiation code developed by AER Inc.

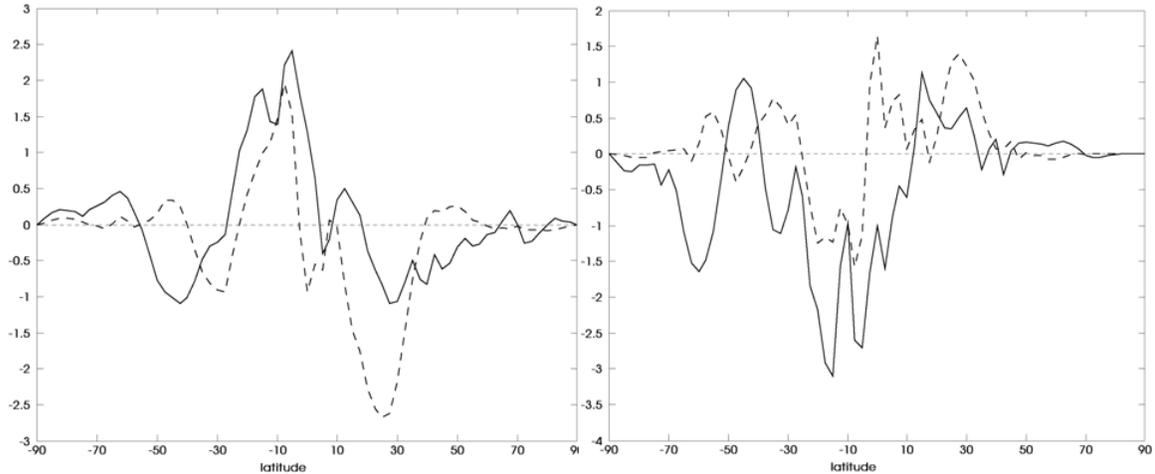
Using NN emulations simultaneously for LWR and SWR or for the full model radiation results in an overall significant, about 20 – 25% speedup of NCEP CFS climate simulations when both LWR and SWR are calculated every hour. The speedup  $\eta$  provided by NN emulations (see Table 1) can be also used for more frequent calculations of model radiations. For example, for calculations with higher (T382) model horizontal resolution, if full NN radiation is calculated 10 times more frequently, *i.e.*, every six minutes, at every model dynamics time step (instead of every hour), the time required for the climate simulation using full NN radiation will be still less than the time needed for the climate simulation using the original radiation with the one hour frequency.

### 3. Validation of parallel decadal model simulations and seasonal predictions

In this section we present comparisons between two parallel 17-year NCEP CFS model runs: one using the original LWR and SWR (the control run) and another one using their NN emulations. Both spatial and temporal characteristics of prognostic and diagnostic fields are compared for the parallel runs. To better estimate the changes introduced by NN emulations, we compare them with “background changes” between two control runs performed with the original NCEP CFS model configuration, *i.e.*, without NN emulations.

The first run was performed before and the second run after the routine changes (introduced quasi-regularly by system administrators) of the version of the FORTRAN compiler and libraries.

The results of 17-year climate simulations performed with NN emulations for both LWR and SWR, *i.e.*, for the full model radiation, have been validated against the parallel control NCEP CFS simulation using the original LWR and SWR. We analyze the differences between the parallel runs in terms of time and spatial (global) means as well as temporal characteristics.



**Fig. 1** Zonal and time mean Top of Atmosphere Upward Long (left panel) and Short (right panel) Wave Fluxes (in  $\text{W/m}^2$ ) for the winter. The solid line – the difference (the full radiation NN run – the control (CTL)), the dash line – the background differences (the differences between two control runs) presented for comparison. The fluxes' differences are multiplied by  $\cos(\text{lat})$  to equalize the areas.

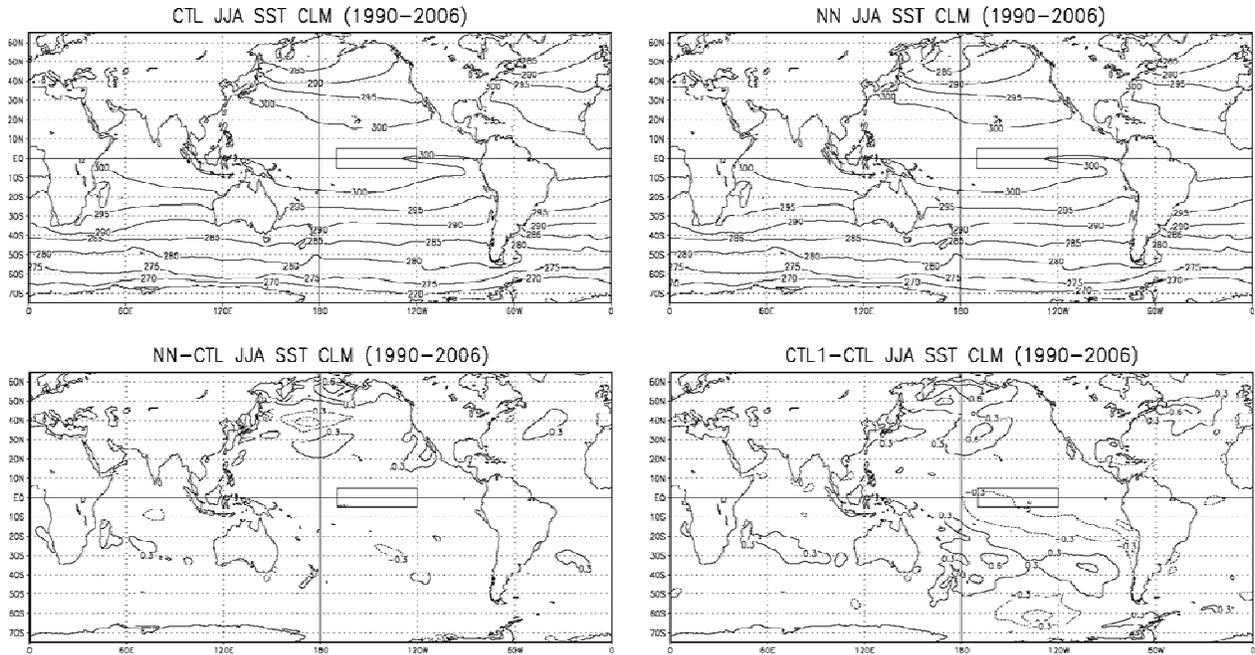
Let us discuss first the differences between the parallel simulations in terms of spatial and temporal radiation characteristics. The differences between the NN radiation and control runs and the differences between two control runs for zonal and time mean LWR and SWR fluxes are presented in Fig. 1. For fluxes presented in Fig. 1 both the differences between the NN radiation and control runs and the differences between two control runs are small and similar by magnitude. They do not exceed  $2\text{-}3 \text{ W/m}^2$  that is overall they are within observational errors and uncertainties of reanalysis (*e.g.* Kalnay *et al.* 1996).

Let us discuss now prognostic and diagnostic characteristics such as SST, precipitation, different types of clouds, and time series that are sensitive to changes in the model resulted from using NN emulations. Close similarities have also been obtained for these results of parallel runs in terms of time mean spatial fields, which are presented in Figs. 2 to 4. Figs. 2 to 4 have the same design: the upper left panel shows fields produced in the control run (CTL) and the upper right – in the full radiation NN run. The bottom left panel shows the difference (bias) between the full radiation NN and CTL runs, and the bottom right panel shows for comparison the background differences (between two control runs) described above.

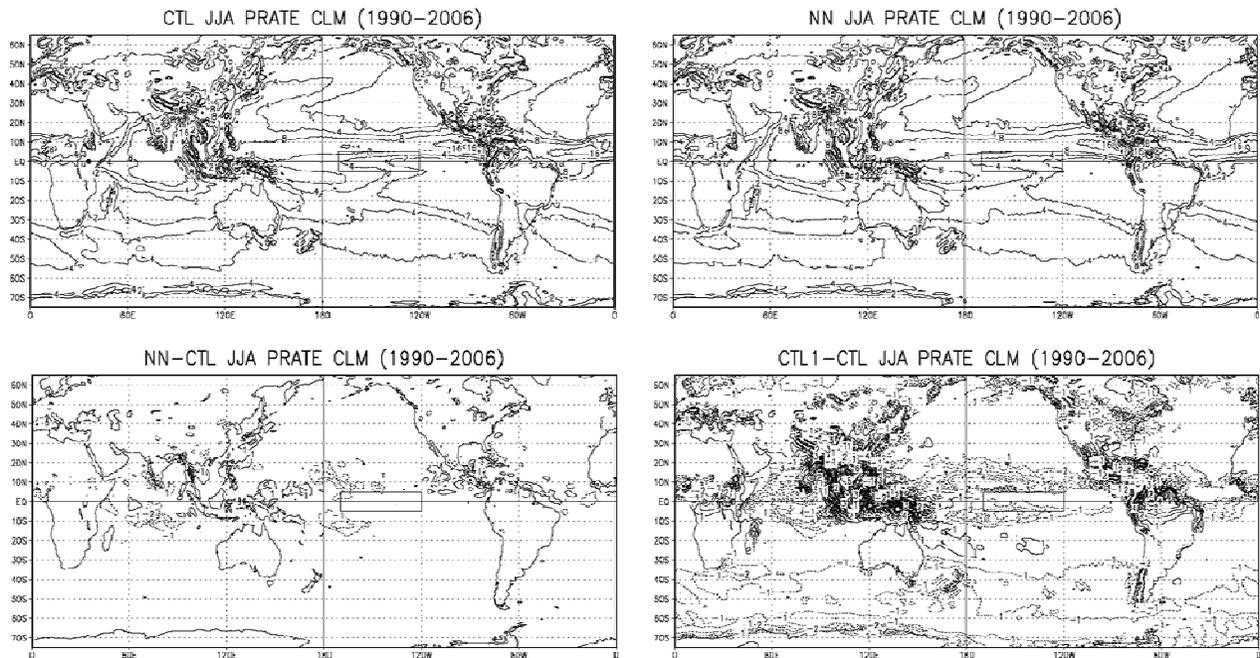
The 17-year (1990-2006) time-mean SST distributions and bias/differences for the full radiation NN run vs. the control run and the background differences between two control runs are presented for summer in Fig. 2. The SST bias is very small; it is not larger than the background differences. The results for other seasons are similar.

Figure 3 shows the 17-year (1990-2006) time-mean distributions and bias/differences for total precipitation (PRATE) for the parallel full radiation NN and control runs for summer. The PRATE bias is quite limited and occurs mostly in the tropics; it is also very close by magnitude and pattern wise to the background differences. The results for other seasons are similar.

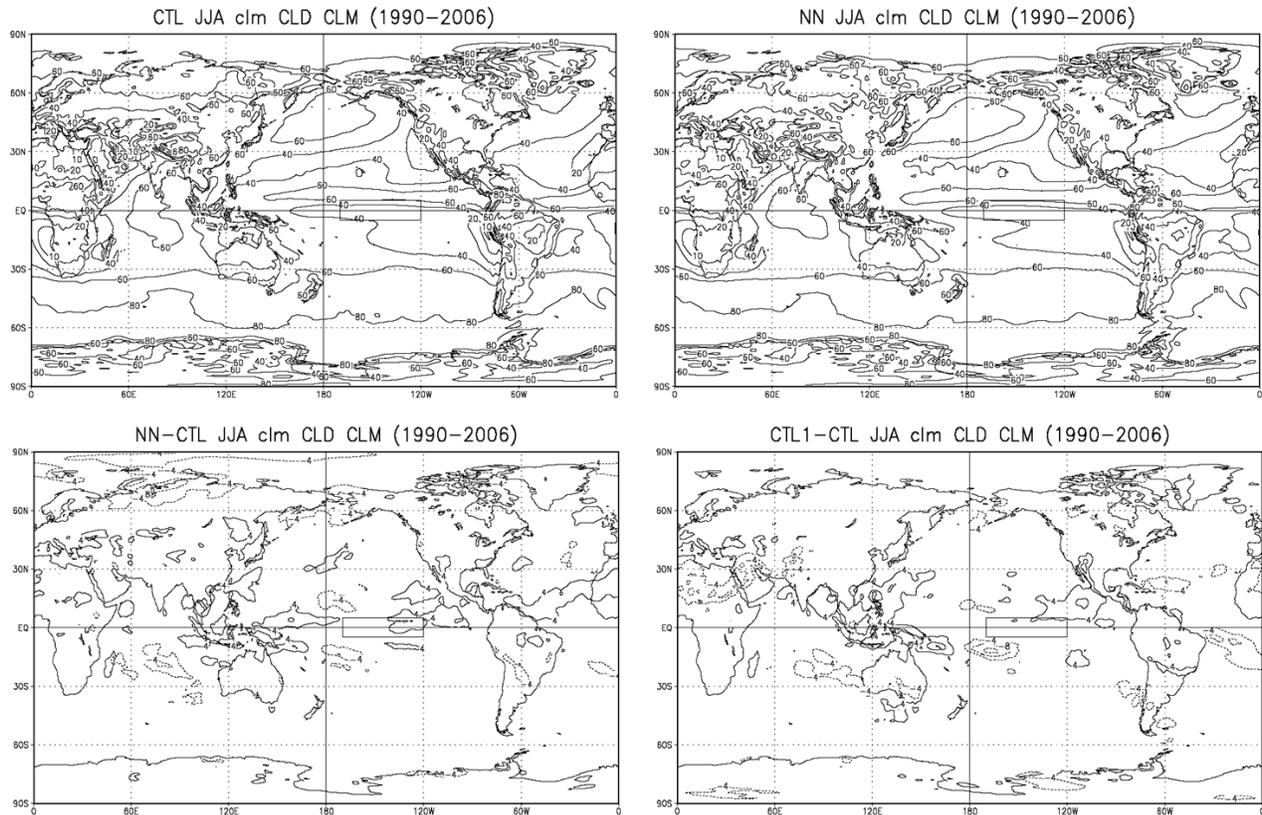
Figure 4 shows comparisons for the parallel full radiation NN and control runs for total clouds, which are very close for the above runs. The results for other seasons and for other types of clouds are similar.



**Fig. 2** The 17-year (1990-2006) time-mean SST distributions and bias/differences for summer (JJA: June-July-August) for the full radiation NN run vs. the control run. The upper row panels: left – the control (CTL) run, right – full radiation NN run. The bottom row panels: left – bias or the difference (full radiation NN run – CTL), right – the background differences between two control runs shown for comparison. The contour intervals for the SST fields are 5° K and for the SST bias and difference are 0.3° K.



**Fig. 3** The same as in Fig. 2 but for total precipitation (PRATE). The contour levels for the PRATE fields are 2, 4, 8, 16 and 32 mm/day. The contour intervals for the PRATE differences (the bottom panels) are 1 mm/day with 0 mm/day contour skipped for clarity.



**Fig. 4** The same as in Fig. 2 but for total clouds. The contour intervals for the cloud fields are 20% and for the differences 4% with 0 % contour skipped for clarity.

#### 4. Conclusions

In this study, the NN emulation approach (Krasnopolsky *et al.* 2005, 2008) is implemented in the state-of-the-art, high resolution, coupled NCEP CFS. The developed highly accurate neural network emulations of long-wave (RRTMG-LW and RRTMF-LW) and short-wave (RRTMG-SW) radiation parameterizations are 12 to 45 times faster than the original/control long-wave and short-wave radiation parameterizations, respectively. The use of the full NN model radiation results in: (1) an overall speedup of about 20 – 25% for climate simulations and seasonal predictions, and (2) an opportunity to increase significantly the frequency of radiation calculations (for example, to calculate model radiation at every model dynamic time step) without increasing the total model calculation time.

*Acknowledgments.* The authors would like to thank Drs. H.-L. Pan, S. Saha, S. Moorthi, and M. Iredell for a valuable help with practical use of NCEP CFS and for useful discussions and consultations. We also thank Drs. S. Moorthi and G. White for reading and commenting on the manuscript. This study is based upon the work supported by the NOAA/CDEP/CTB grant NA06OAR4310047.

#### References

- Bishop, Ch. M., 2006: Pattern Recognition and Machine Learning. *Springer*, 738 pp.
- Côté, J., S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998: The operational CMC-MRB global environmental multiscale (GEM) model. Part I: Design considerations and formulation, *Mon. Wea. Rev.*, **126**, 1373-1395.
- Funahashi, K., 1989: On the Approximate Realization of Continuous Mappings by Neural Networks. *Neural Networks*, **2**, 183-192.

- Kalnay, E., *et al.*, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteorol. Soc.*, **77**: 437-472.
- Krasnopolsky, V.M., 2007: 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*, **45**, RG3009, doi:10.1029/2006RG000200.
- Krasnopolsky, V.M., M.S. Fox-Rabinovitz, and D.V. Chalikov, 2005: Fast and Accurate Neural Network Approximation of Long Wave Radiation in a Climate Model, *Mon. Wea. Rev.*, **133**, pp. 1370-1383.
- Krasnopolsky, V.M., M.S. Fox-Rabinovitz, A. A. Belochitski, 2008: "Decadal Climate Simulations Using Accurate and Fast Neural Network Emulation of Full, Long- and Short Wave, Radiation.", *Mon. Wea. Rev.*, **136**, 3683-3695, doi: 10.1175/2008MWR2385.1.
- Krasnopolsky, V.M., M. S. Fox-Rabinovitz, Y. T. Hou, S. J. Lord, and A. A. Belochitski, 2009: Accurate and Fast Neural Network Emulations of Model Radiation for the NCEP Coupled Climate Forecast System: Climate Simulations and Seasonal Predictions, submitted.
- Lacis, A. A., V. Oinas, 1991: A description of the correlated k-distribution method for modeling non-gray gaseous absorption, thermal emission and multiple scattering in vertically inhomogeneous atmospheres. *J. Geophys. Res.* **96**: 9027-9063.
- Manners, J., J.-C. Thelen, J. Petch, P. Hill & J.M. Edwards, 2008: Two fast radiative transfer methods to improve the temporal sampling of clouds in NWP and climate models, *Q. J. R. Meteorol. Soc.* **00**: 1-11.
- Morcrette, J.-J., G. Mozdzyński and M. Leutbecher, 2008: A reduced radiation grid for the ECMWF Integrated Forecasting System, *Mon. Wea. Rev.*, **136**, 4760-4772, doi: 10.1175/2008MWR2590.1