Development of fast and accurate neural network emulations of mode physics parameterizations

Vladimir Krasnopolsky
NOAA/NCEP & University of Maryland
and
Michael Fox-Rabinovitz,
University of Maryland

Acknowledgments: S. Lord, H.-L. Pan, Y.-T. Hou, S. Saha, A. Belochitski, P. Rasch, C. Bretherton, and M. Khairoutdinov

Outline

- Recent developments
- Background
- Two Modes of NN Applications:
  - NN **Emulations**: Speeding up Calculations of Model Physics
  - NN **Parameterizations**: Improvement of Model Physics
- Hybrid Models: Synergetic Combination of Deterministic and Statistical Learning Model Components
  - Robustness of Hybrid Models
  - Compound Parameterizations: Coherence of NN and Deterministic Components
- NN Ensembles
- Conclusions

Summary of Recent Developments

- **Developed NN Emulation for the NCAR CAM (T42L26) LWR Parameterization**
  - High accuracy of NN emulation vs. the original CAM LWR parameterization
  - NN emulation is **150 times faster** than the original CAM LWR
- **Developed NN Emulation for the NCAR CAM SWR Parameterization**
  - High accuracy of NN emulation vs. the original CAM SWR parameterization
  - NN emulation is **20 times faster** than the original CAM SWR

Summary of Recent Developments

- **Development of NN emulations for RRTM LWR & SWR, i.e., the full radiation block, of the coupled NCEP CFS (T126L64)**
  - High accuracy of LWR and SWR NN emulations vs. the original CFS LWR and SWR parameterizations
  - LWR NN emulation is **20-60 times faster** than the original RRTM LWR
  - SWR NN emulation is being assessed (the work in progress)

Summary of Recent Developments (2)

- **Developed a Hybrid Model Approach and Demonstrated its Feasibility:**
  - Performed **50** year parallel runs of Hybrid NCAR CAM with the full NN radiation block and the original NCAR CAM radiation block: no significant differences between the runs
  - Performed **15** year parallel runs of Hybrid NCEP CFS with the NN LWR and the original NCEP CFS LWR: no significant differences between the runs
  - Developed a **compound parameterization** approach for robust hybrid models

Summary of Recent Developments (3)

- **Developed a NN *Parameterization* Approach**
  - Approach has been conceptually formulated
  - Proof of concept has been produced
- **Developed NN Ensemble Approaches to:**
  - Improve the accuracy of NN emulation
  - Reduce uncertainties of NN Jacobians
- **Transferring the Developed Methodology to NCEP CFS (NN emulations for the radiation block)**

Background

- Any parameterization of model physics is a relationship or **MAPPING** (continuous or almost continuous) between two vectors: a vector of input parameters, *X*, and a vector of output parameters, *Y*,

$$Y = F(X); \quad X \in \Re^n \; and \; Y \in \Re^m$$

- NN is a generic approximation for **any** continuous or almost continuous mapping given by a set of its input/output records:

$$\text{SET} = \{X_i, \; Y_i\}_{i = 1, \ldots, N}$$

Neural Network

Continuous Input to Output Mapping

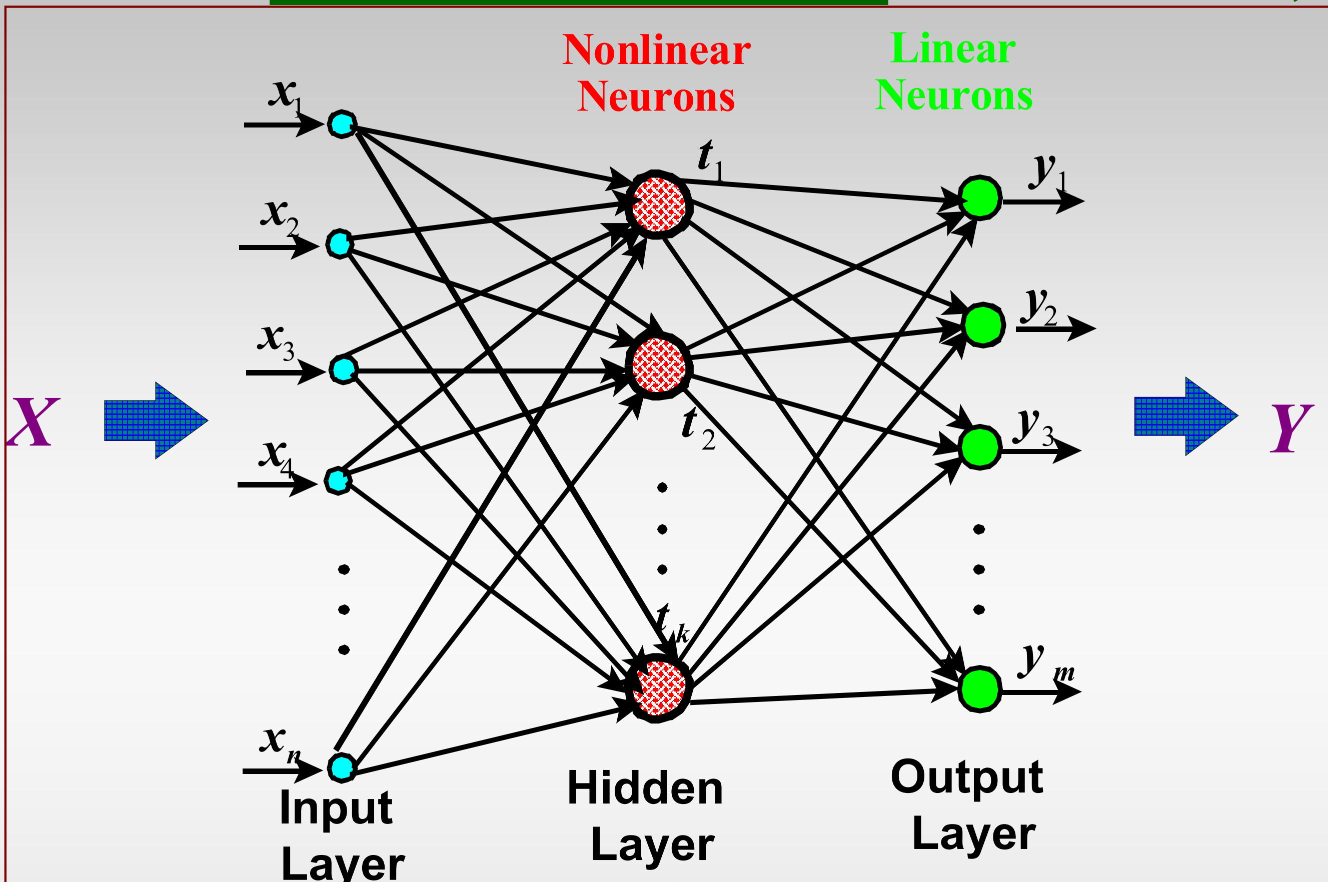
$$Y = F_{NN}(X)$$

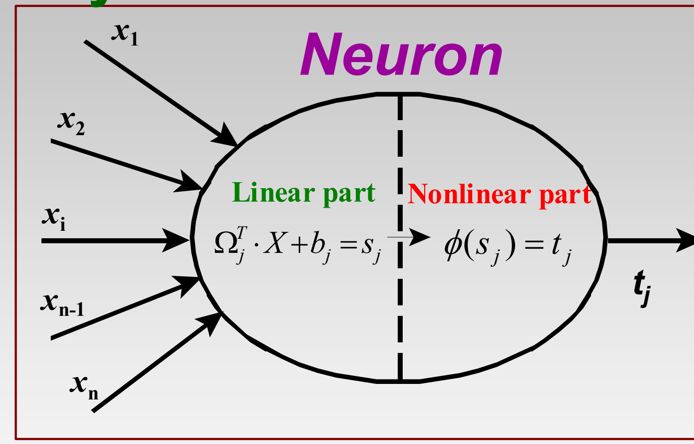
$$y_q = a_{q0} + \sum_{j=1}^{k} a_{qj} \cdot t_j$$

Neuron ➡ $$t_j = \tanh(b_j + \sum_{i=1}^{n} \Omega_{ji} \cdot x_i)$$

NN - Continuous Input to Output Mapping

Multilayer Perceptron: Feed Forward, Fully Connected





$$t_j = \phi(b_j + \sum_{i=1}^{n} \Omega_{ji} \cdot x_i) =$$

$$= \tanh(b_j + \sum_{i=1}^{n} \Omega_{ji} \cdot x_i)$$

Y = F$_{NN}$(X)

$$\begin{cases} y_q = a_{q0} + \sum_{j=1}^{k} a_{qj} \cdot t_j = a_{q0} + \sum_{j=1}^{k} a_{qj} \cdot \phi(b_j + \sum_{i=1}^{n} \Omega_{ji} \cdot x_i) = \\ = a_{q0} + \sum_{j=1}^{k} a_{qj} \cdot \tanh(b_j + \sum_{i=1}^{n} \Omega_{ji} \cdot x_i); \quad q = 1, 2, \ldots, m \end{cases}$$

Jacobian !

Major Advantages of NNs:

- NNs are *generic, very accurate and convenient* mathematical (statistical) models which are able to emulate numerical model components, which are complicated nonlinear input/output relationships (continuous or almost continuous mappings).
- NNs are ***robust*** with respect to random noise and fault-tolerant.
- NNs are *analytically differentiable* (training, error and sensitivity analyses): an almost free Jacobian!
- NNs emulations are **accurate and fast but there is**

  NO FREE LUNCH!

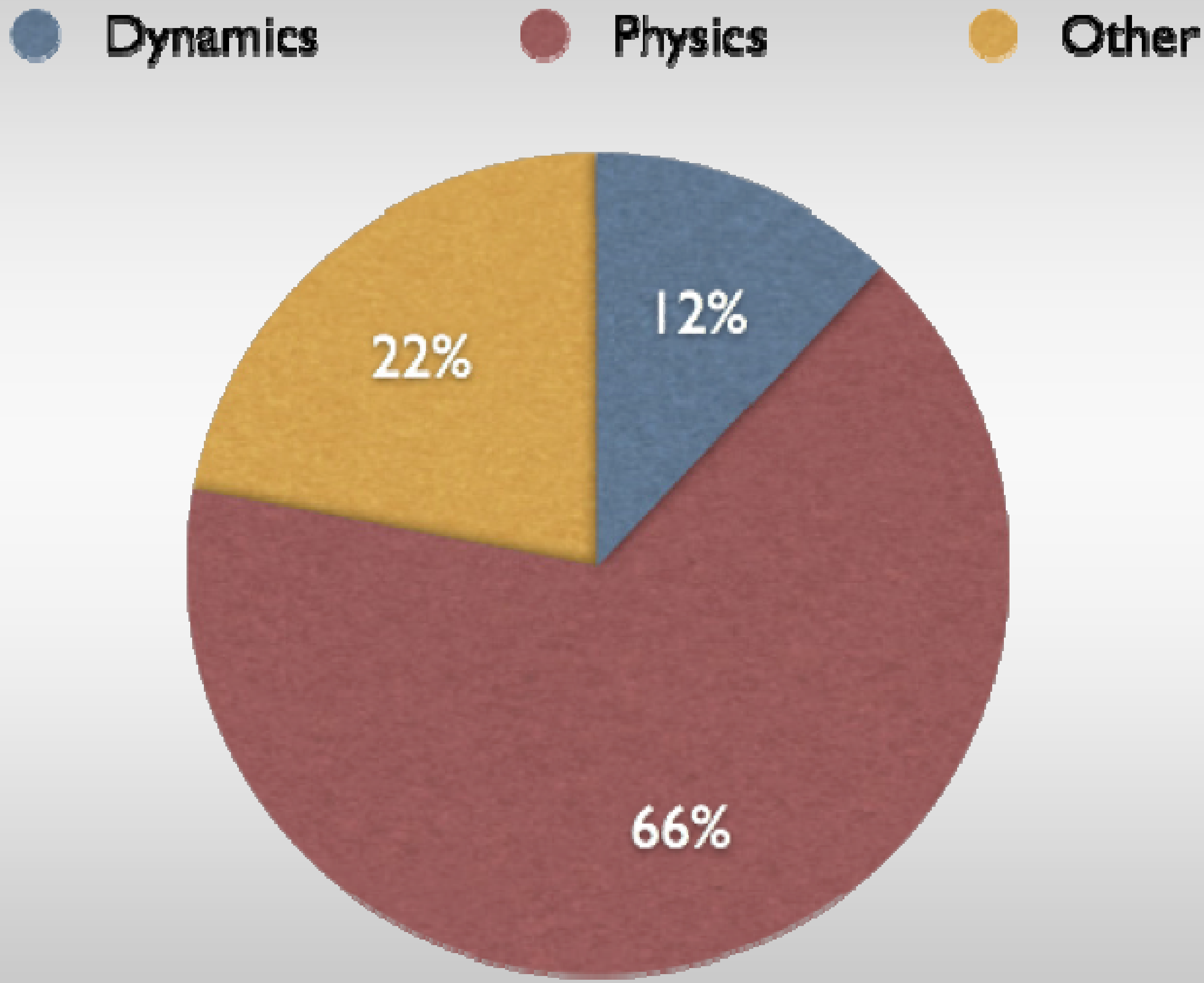
  – Training is complicated and time consuming nonlinear optimization task; *however, training should be done **only once** for a particular application!*
- NNs are well-suited for parallel and vector processing

Three Modes of NN Applications

- **NN emulations** of **existing** model physics parameterizations allow us to:
  - Speed up calculation of model physics 10^2 to 10^5 times
  - Improve model physics by introducing NN emulations of more sophisticated (but time consuming) parameterizations
- Develop **new NN parameterizations** of model physics based on:
  - Observations
  - Data simulated by first principle process models (like cloud resolving models)
- Using NNs as a tool for continuum (seamless) modeling
- All NN applications are ***approximations of the original model physics (parameterizations, CRM, etc.) !*** They are ***NOT an alternative*** to ongoing developments of first principle model physics but rather based on them!

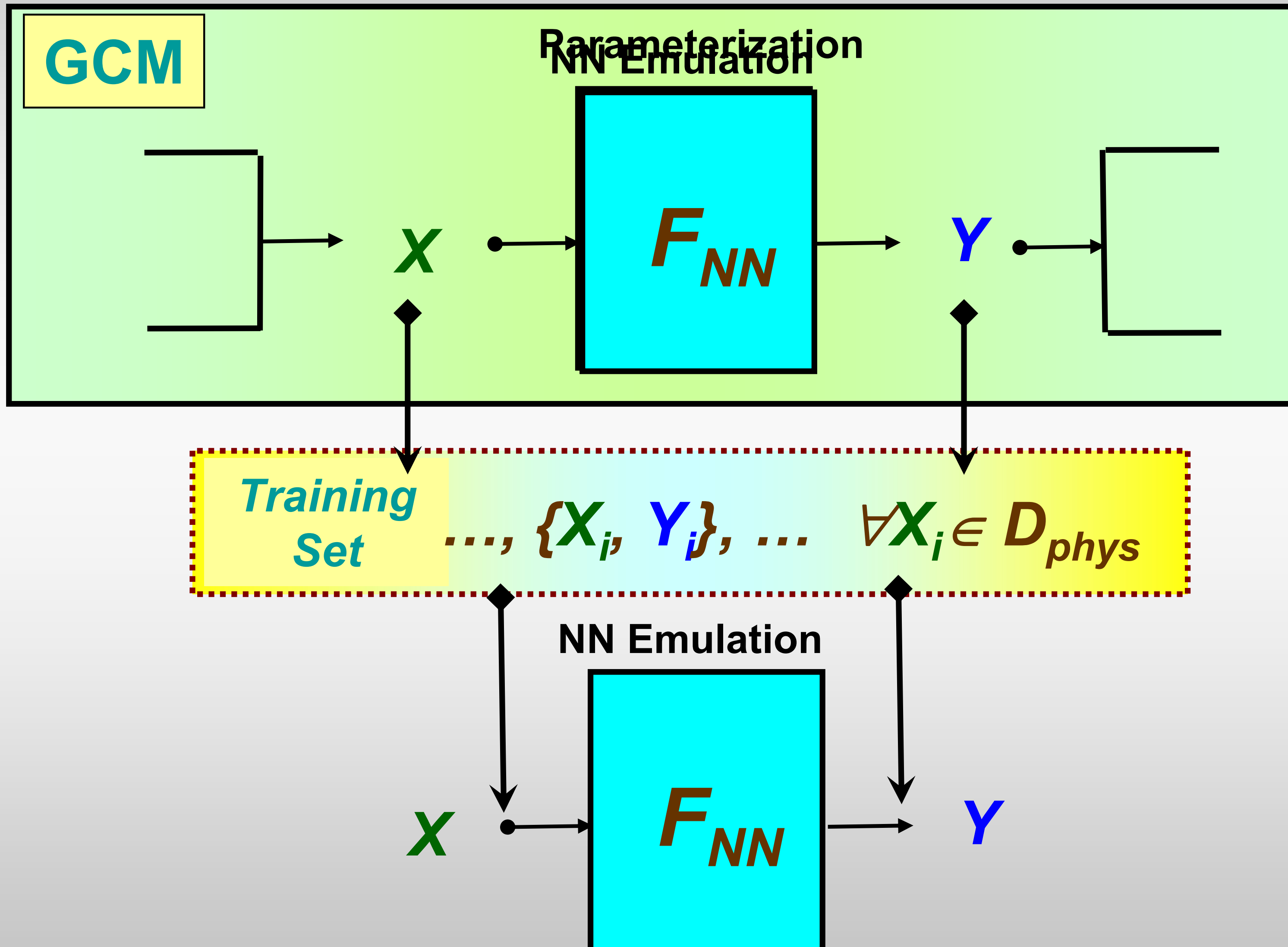
Distribution of NCAR CAM Calculation Time

NCAR CAM 2.0 T42L26

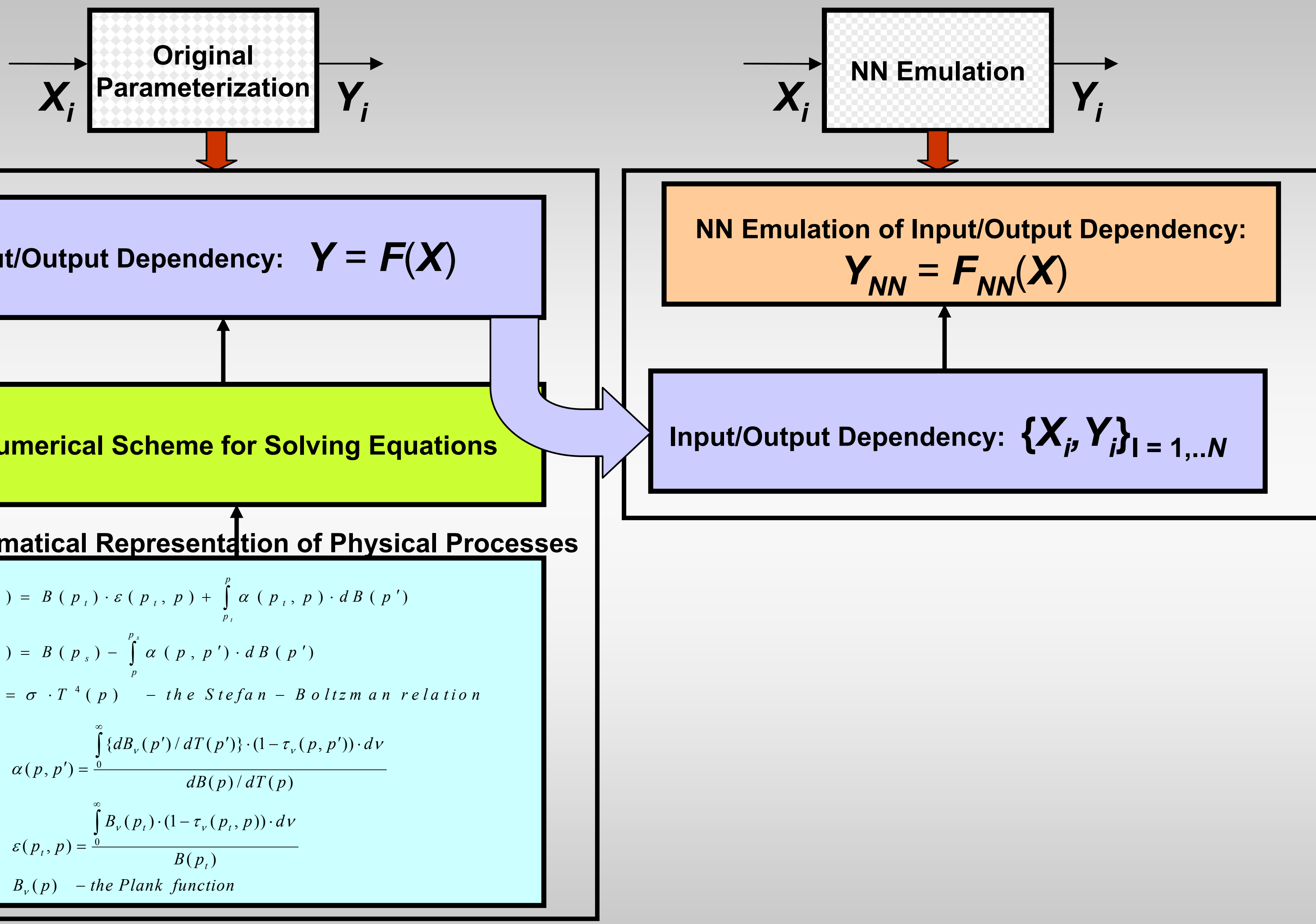


NN Emulations of Model Physics Parameterizations

Learning from Data



The Magic of NN Performance



NN for NCAR CAM Physics

CAM Long Wave Radiation

- Long Wave Radiative Transfer:

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

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

$$B(p) = \sigma \cdot T^4(p) \quad - \mathit{the\ Stefan - Boltzman\ relation}$$

- Absorptivity & Emissivity (optical properties):

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

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

$$B_\nu(p) \quad - \mathit{the\ Plank\ \ function}$$

Neural Network for NCAR LW Radiation

NN characteristics

- 220 (**612 for NCEP**) Inputs:
  - *10 Profiles:* **temperature; humidity; ozone, methane, cfc11, cfc12, & N_2O mixing ratios, pressure, cloudiness, emissivity**
  - *Relevant surface characteristics*: **surface pressure, upward LW flux on a surface - *flwupcgs***
- 33 (**69 for NCEP**) Outputs:
  - **Profile of heating rates (26)**
  - **7 LW radiation fluxes: *flns, flnt, flut, flnsc, flntc, flutc, flwds***
- Hidden Layer: One layer with 50 to 300 neurons
- Training: *nonlinear optimization in the space with dimensionality of 15,000 to 100,000*
  - **Training Data Set: Subset of about 200,000 instantaneous profiles simulated by CAM for the 1-st year**
  - **Training time: about 1 to several days (SGI workstation)**
  - **Training iterations: 1,500 to 8,000**
- Validation on Independent Data:
  - **Validation Data Set (independent data): about 200,000 instantaneous profiles simulated by CAM for the 2-nd year**

Neural Network for NCAR SW Radiation

NN characteristics

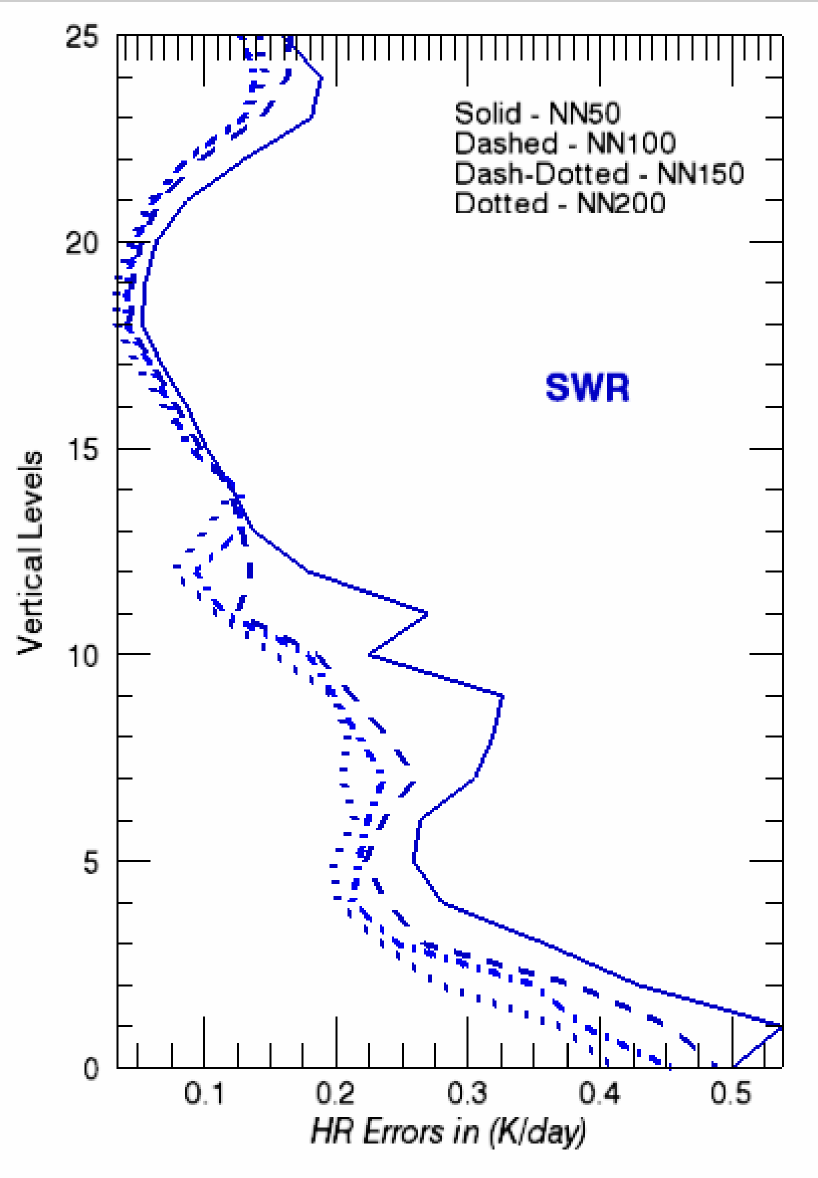
- 451 Inputs:
  - *21 Profiles:* **specific humidity, ozone concentration, pressure, cloudiness, aerosol mass mixing ratios, etc**
  - *7 Relevant surface characteristics*
- 33 Outputs:
  - **Profile of heating rates (26)**
  - **7 LW radiation fluxes: *fsns, fsnt, fsdc, sols, soll, solsd, solld***
- Hidden Layer: One layer with 50 to 200 neurons
- Training: *nonlinear optimization in the space with dimensionality of 25,000 to 130,000*
  - **Training Data Set: Subset of about 100,000 instantaneous profiles simulated by CAM for the 1-st year**
  - **Training time: about 1 to several days (SGI workstation)**
  - **Training iterations: 1,500 to 8,000**
- Validation on Independent Data:
  - **Validation Data Set (independent data): about 100,000 instantaneous profiles simulated by CAM for the 2-nd year**

NN Approximation Accuracy and Performance vs. Original Parameterization (*on independent data set*)

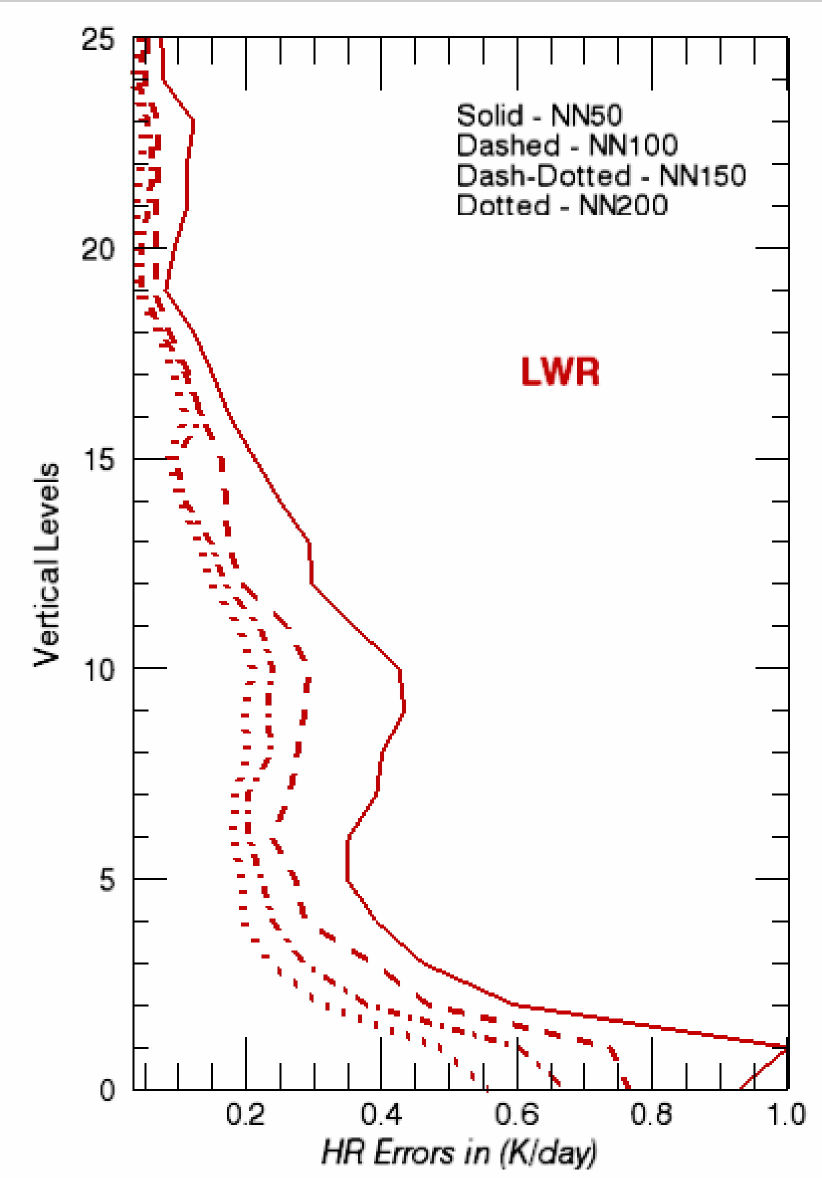
Parameter	Model	Bias	RMSE	Mean	σ	Performance
LWR (°*K/day*) NN50	NASA **M-D. Chou**	$1.\ 10^{-4}$	0.32	-1.52	1.46	
	NCEP **AER *RRTM***	$7.\ 10^{-5}$	0.40	-1.88	2.28	
	NCAR **W. Collins**	$3.\ 10^{-5}$	0.28	-1.40	1.98	~ 150 times faster
SWR (°*K/day*) NN55	NCAR **W. Collins**	$-4.\ 10^{-4}$	0.15	1.47	1.89	~ 20 times faster

Error Vertical Variability Profiles

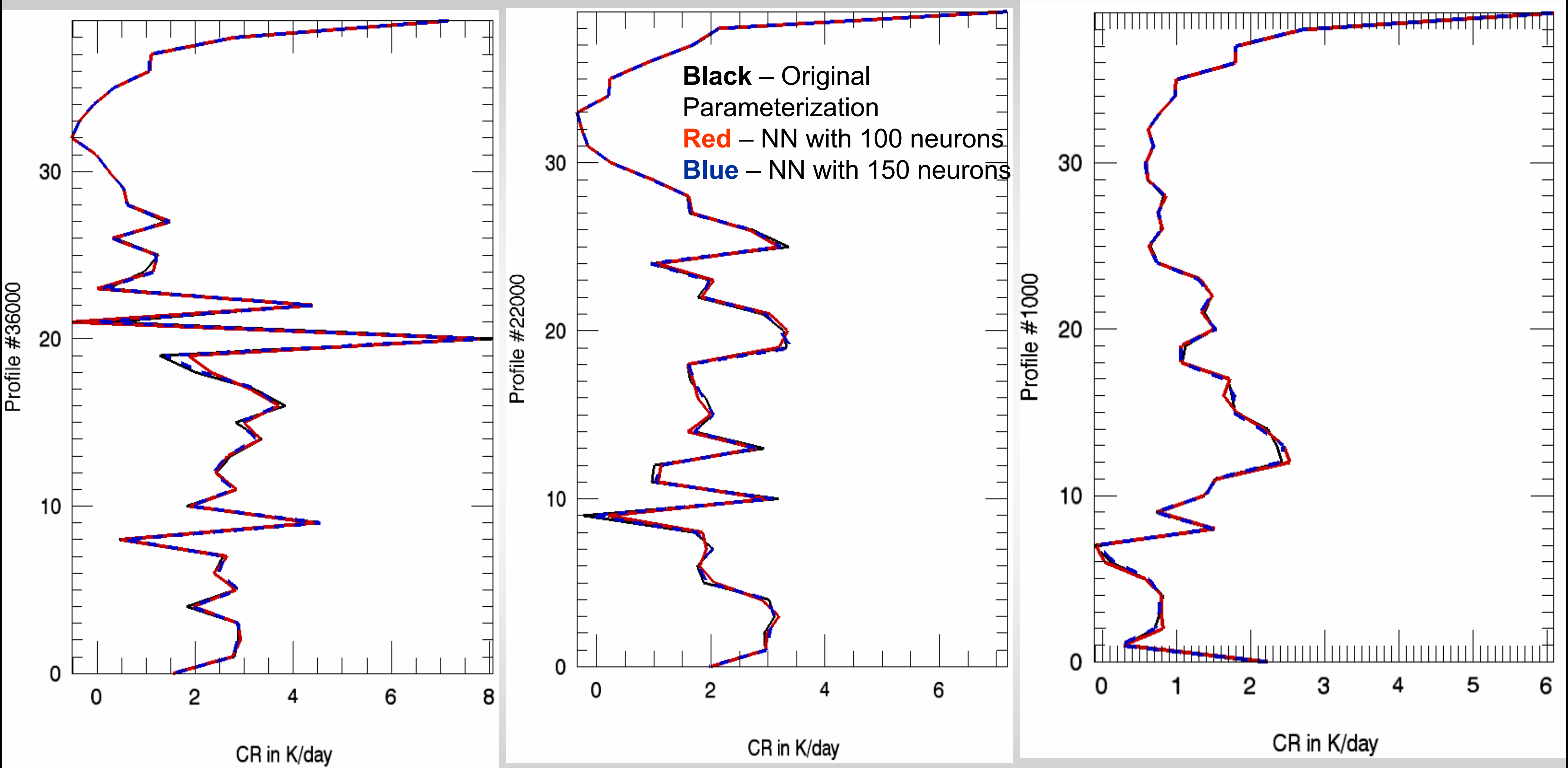
RMSE profiles in K/day



RMSE Profiles in K/day

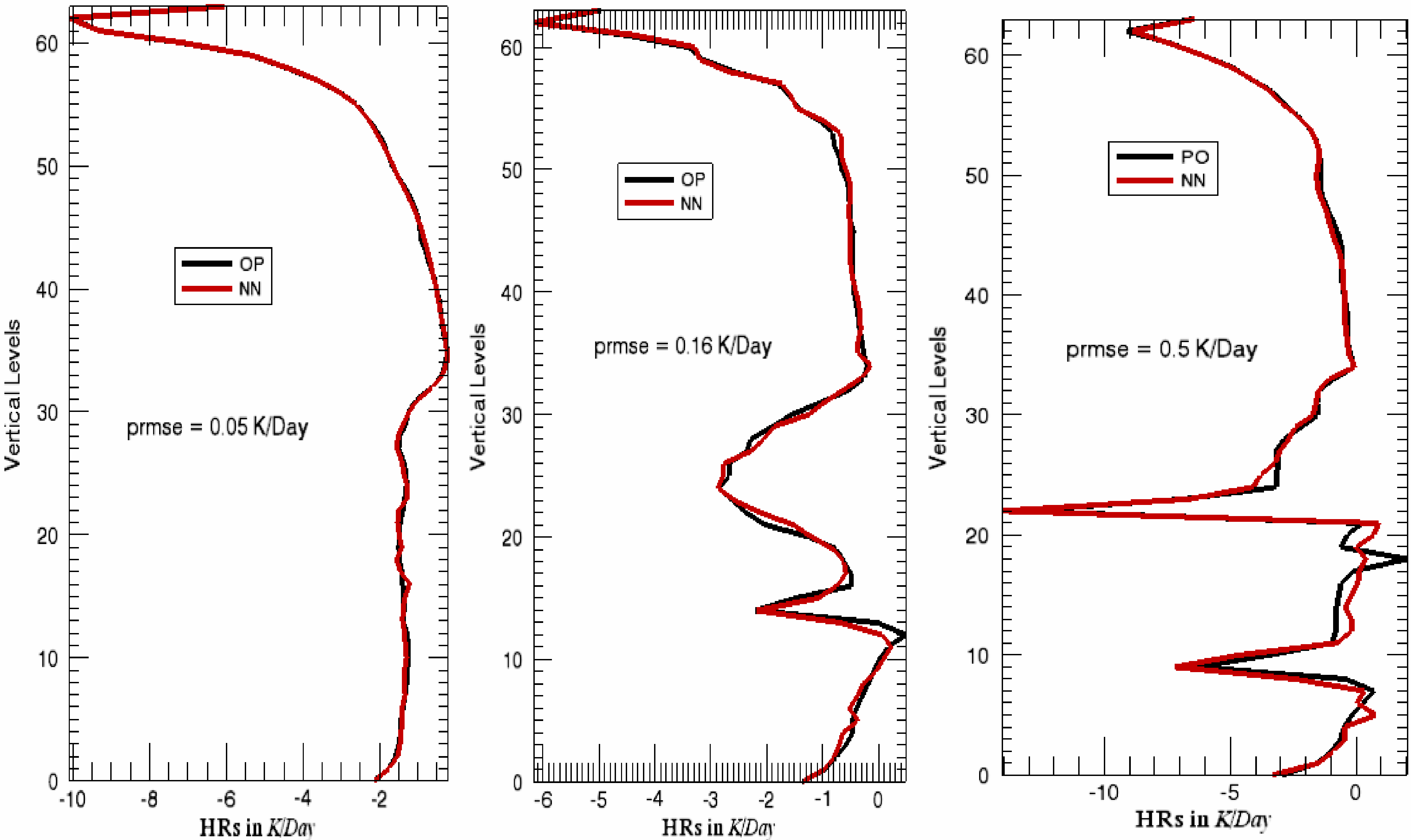


Individual Profiles



PRMSE = 0.18 & 0.10 K/day **PRMSE = 0.11 & 0.06 K/day** **PRMSE = 0.05 & 0.04 K/day**

Individual Profiles (NCEP)



Hybrid Numerical Models (HNM)

- Hybrid Numerical Models (HNM) – combine deterministic components with statistical components
- Deterministic components are based on first principles or physically based
- Statistical components – are either NN emulations of physically based components or NN parameterizations.

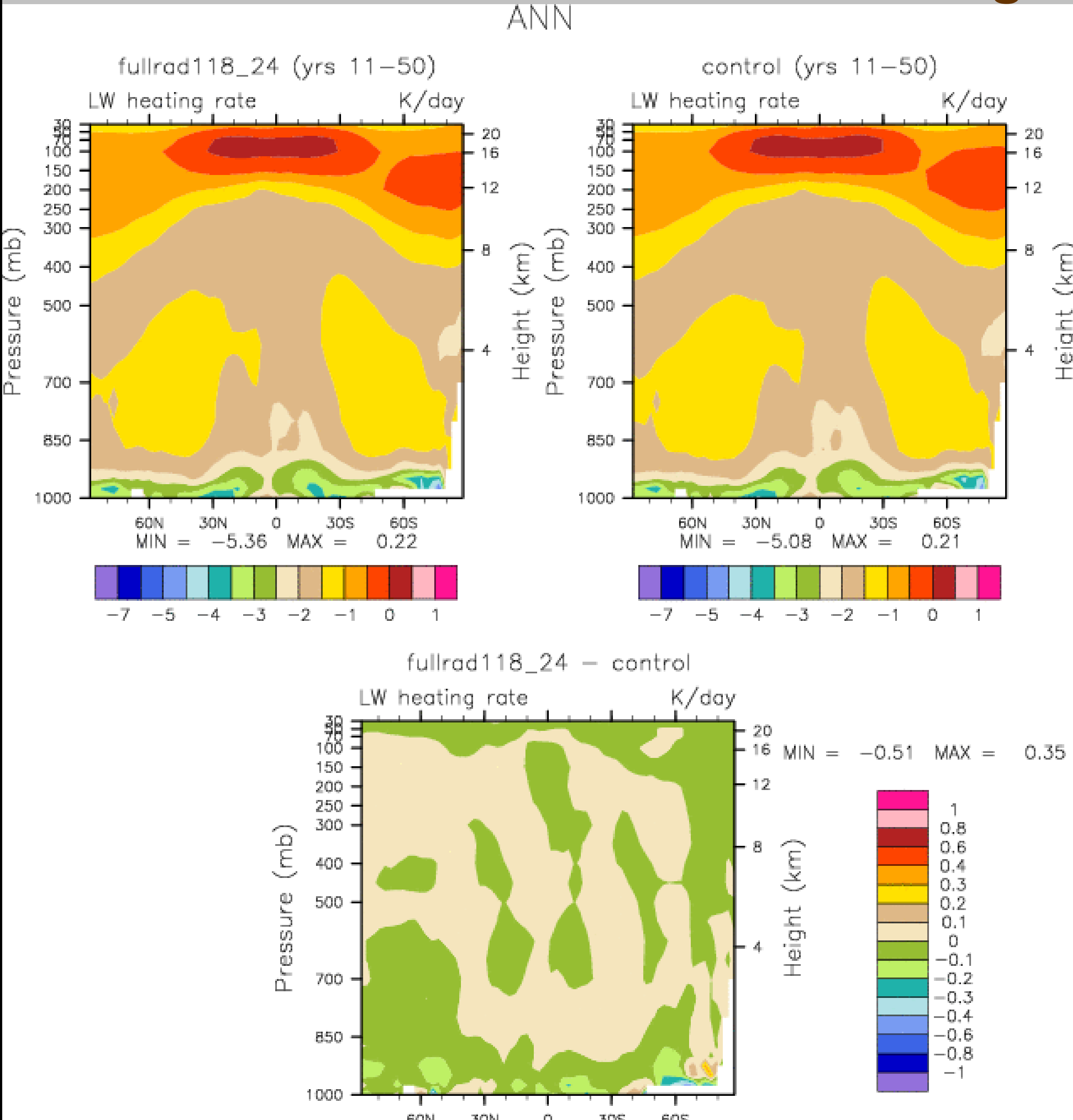
Validation of Hybrid NCAR CAM Model

- The control NCAR CAM with the original LWR and SWR parameterizations is run for 40 years (in the 50 year run first the 10 years are skipped to account for spin-up effects).
- The Hybrid NCAR CAM with LWR and SWR NN emulations is run for 40 years.
- Validation of the Hybrid NCAR CAM with NN emulations of the full radiation block (LWR & SWR) is done against the control run. The following is the comparison of the two parallel runs (11-50 year products).

Time and Global Means for Model Diagnostics From the 40 year NCAR CAM Climate Simulations With the Original LWR and SWR and With NN Emulations

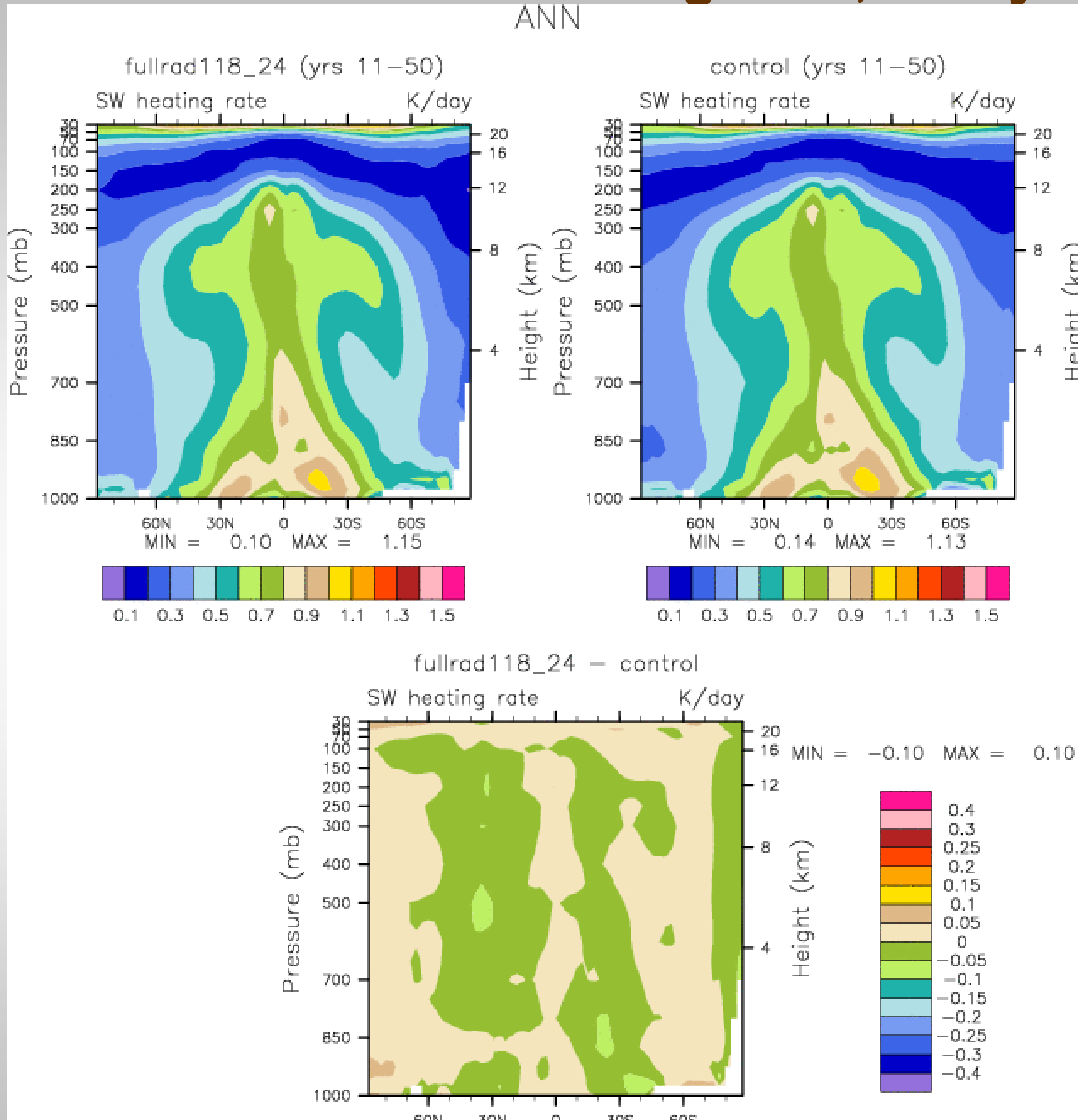
Field	NCAR CAM with the original LWR and SWR Parameterizations	NCAR CAM with LWR and SWR NN Emulations	Difference in %
Mean Sea Level Pressure (hPa)	1011.48	1011.50	0.002
Surface Temperature (°K)	289.02	288.92	0.03
Total Precipitation (mm/day)	2.86	2.89	1.04
Total Cloudiness (fraction, %)	60.71	61.12	0.6
Outgoing LWR (W/m^2)	234.56	233.86	0.3

Zonal Mean LW Heating Rate, K/day.

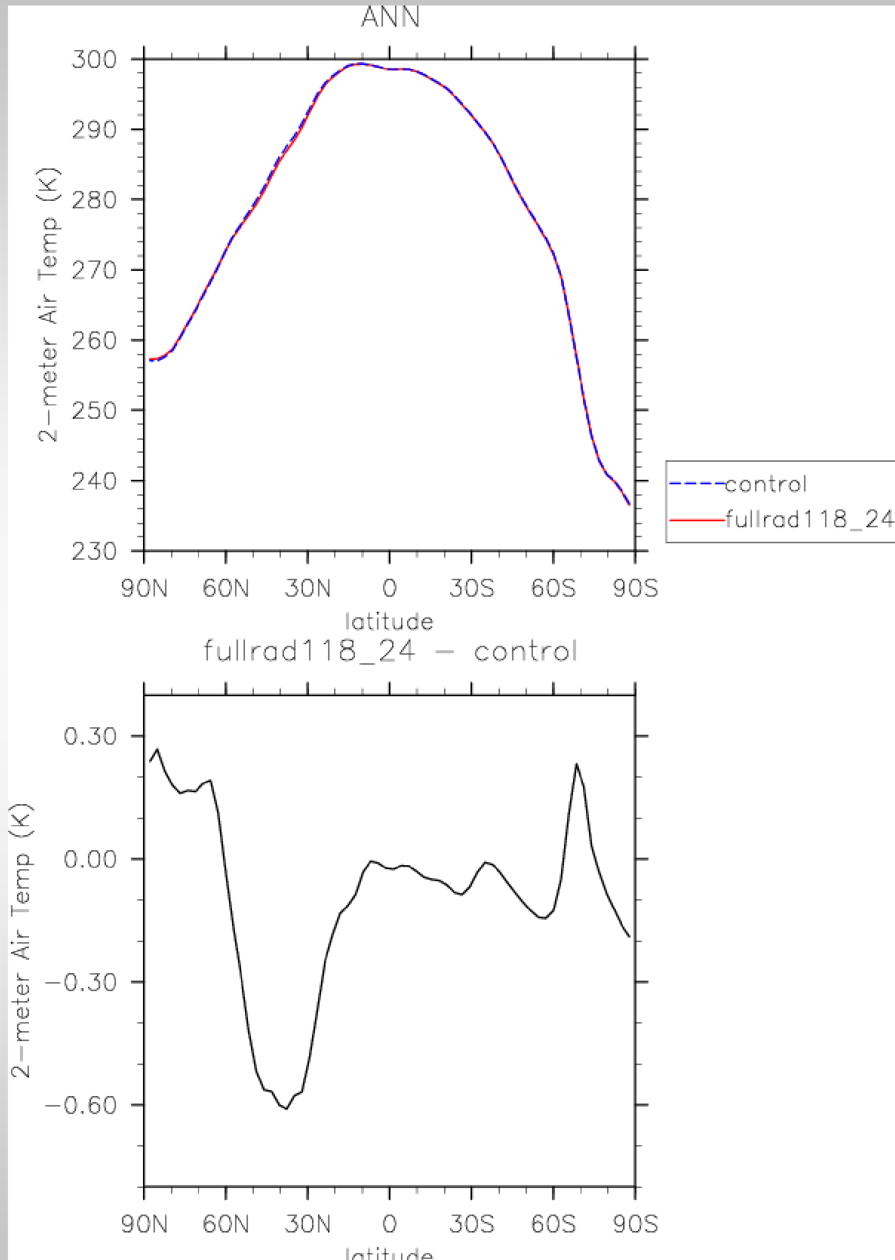


Note: light green and light beige colors correspond to the near 0 values!

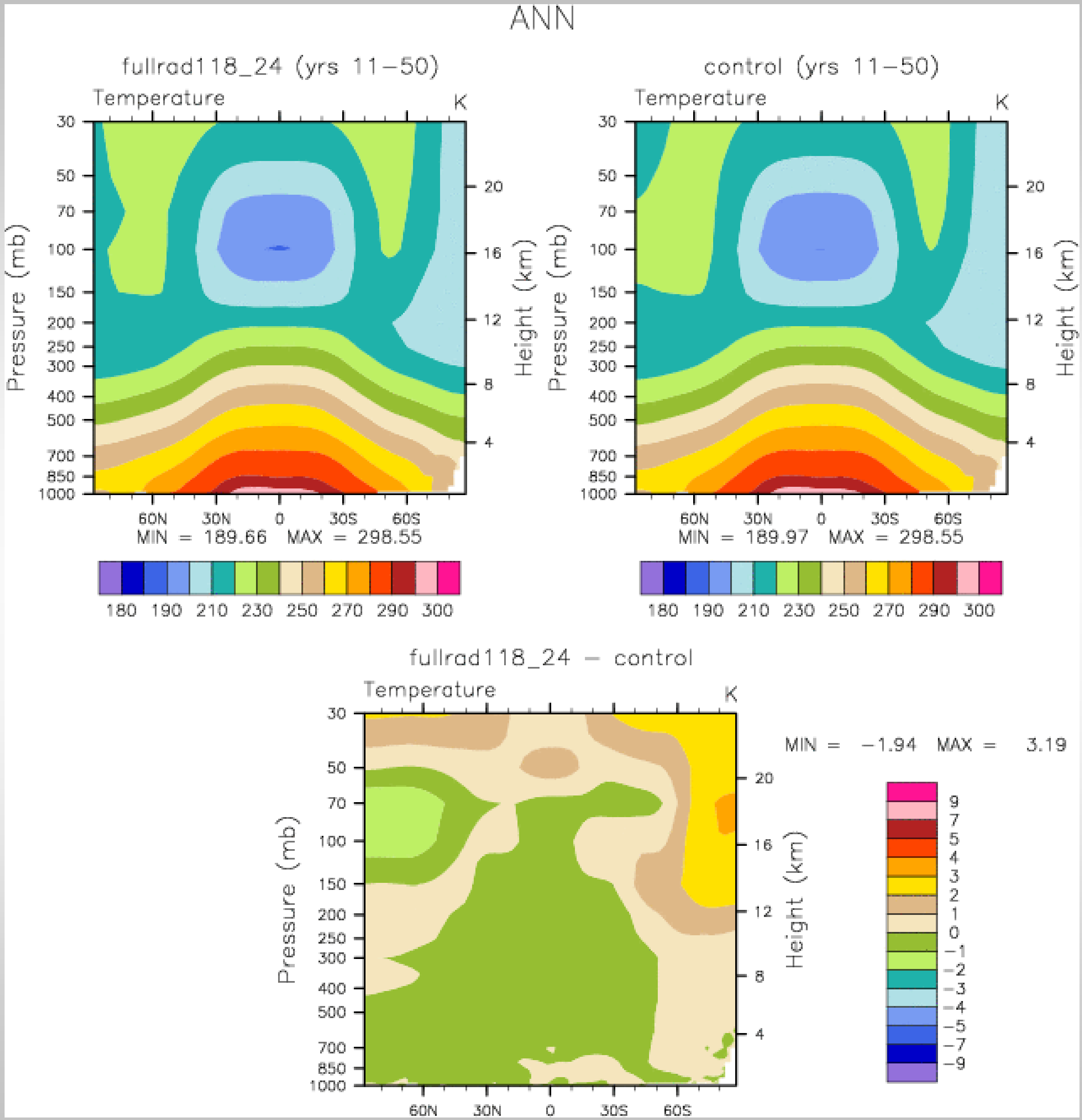
Zonal Mean SW Heating Rate, K/day.



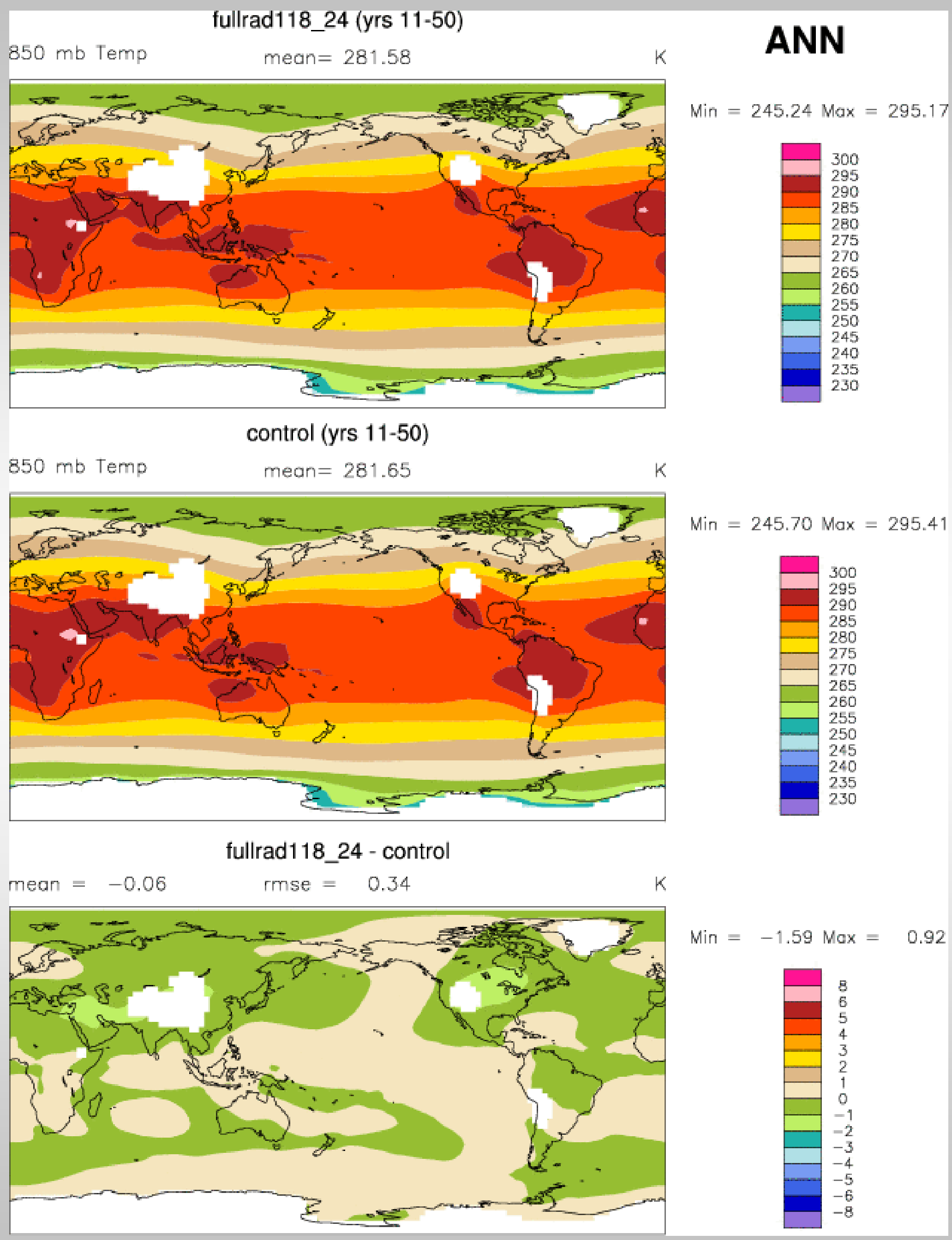
Zonal Mean 2-meter T, K.



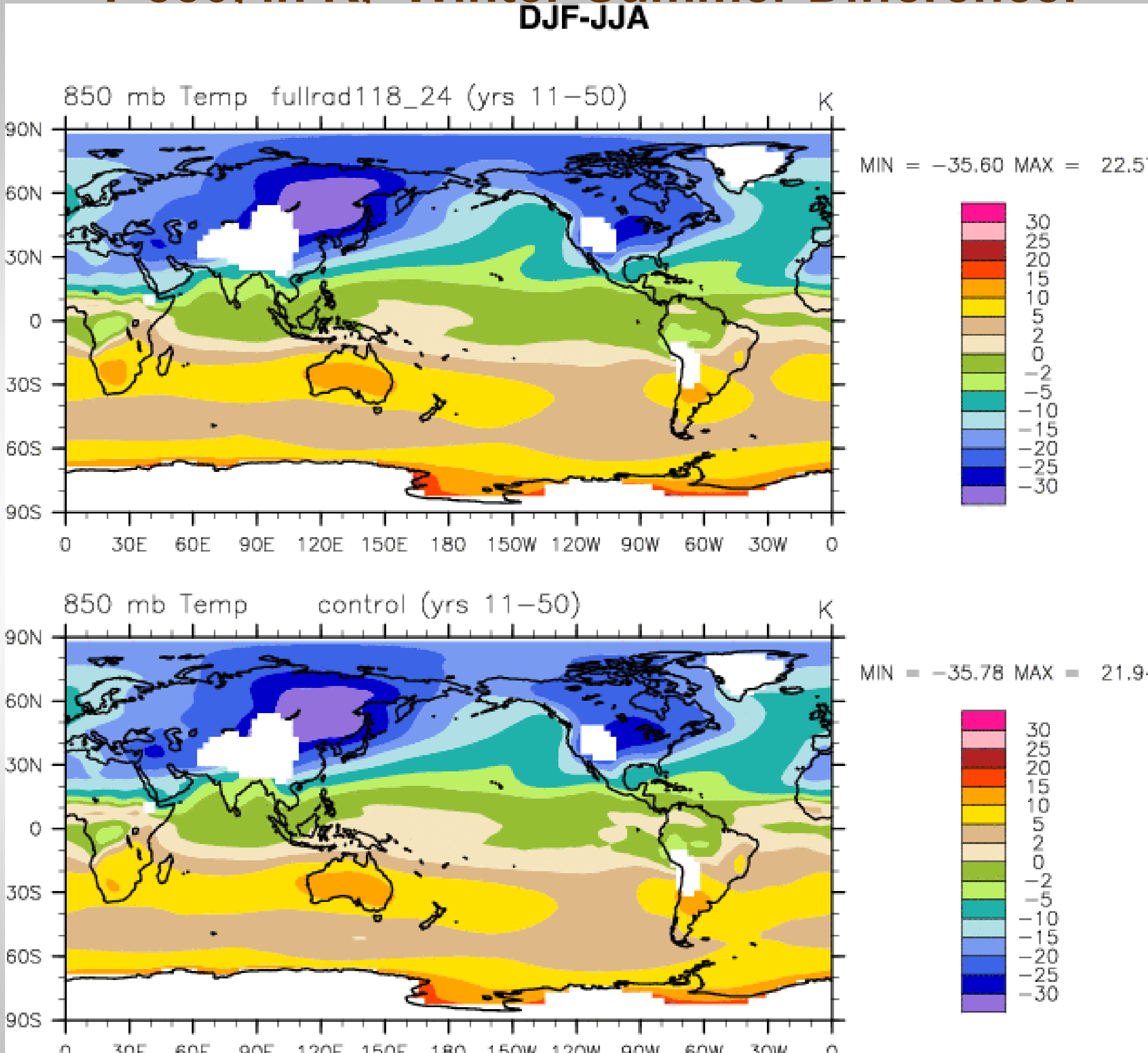
Zonal Mean T, K.



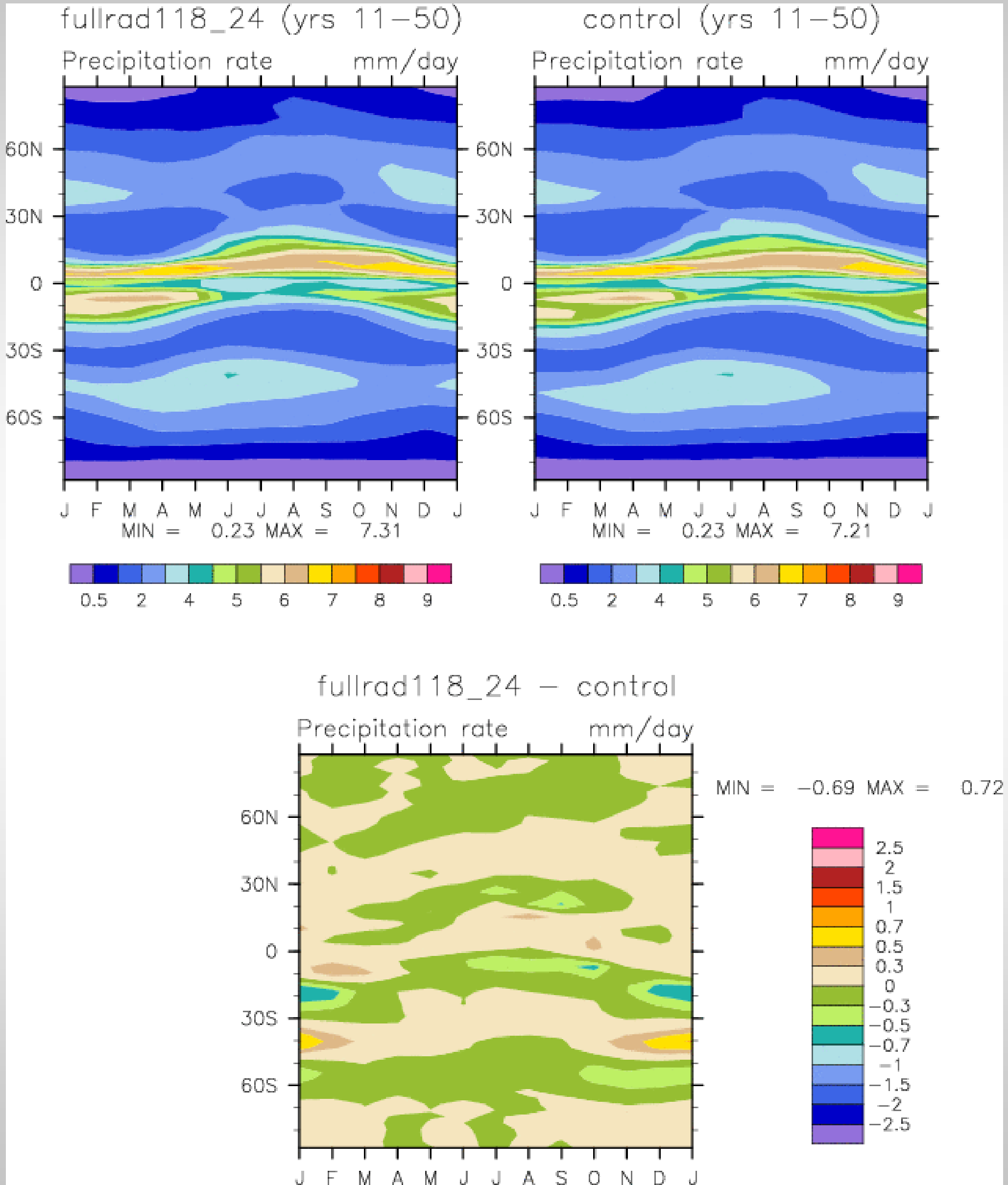
T-850, K.


fullrad118_24 (yrs 11-50)
850 mb Temp
mean= 281.58
K
ANN
Min = 245.24 Max = 295.17
300
295
290
285
280
275
270
265
260
255
250
245
240
235
230
control (yrs 11-50)
850 mb Temp
mean= 281.65
K
Min = 245.70 Max = 295.41
300
295
290
285
280
275
270
265
260
255
250
245
240
235
230
fullrad118_24 - control
mean = −0.06
rmse = 0.34
K
Min = −1.59 Max = 0.92
8
6
5
4
3
2
1
0
−1
−2
−3
−4
−5
−6
−8

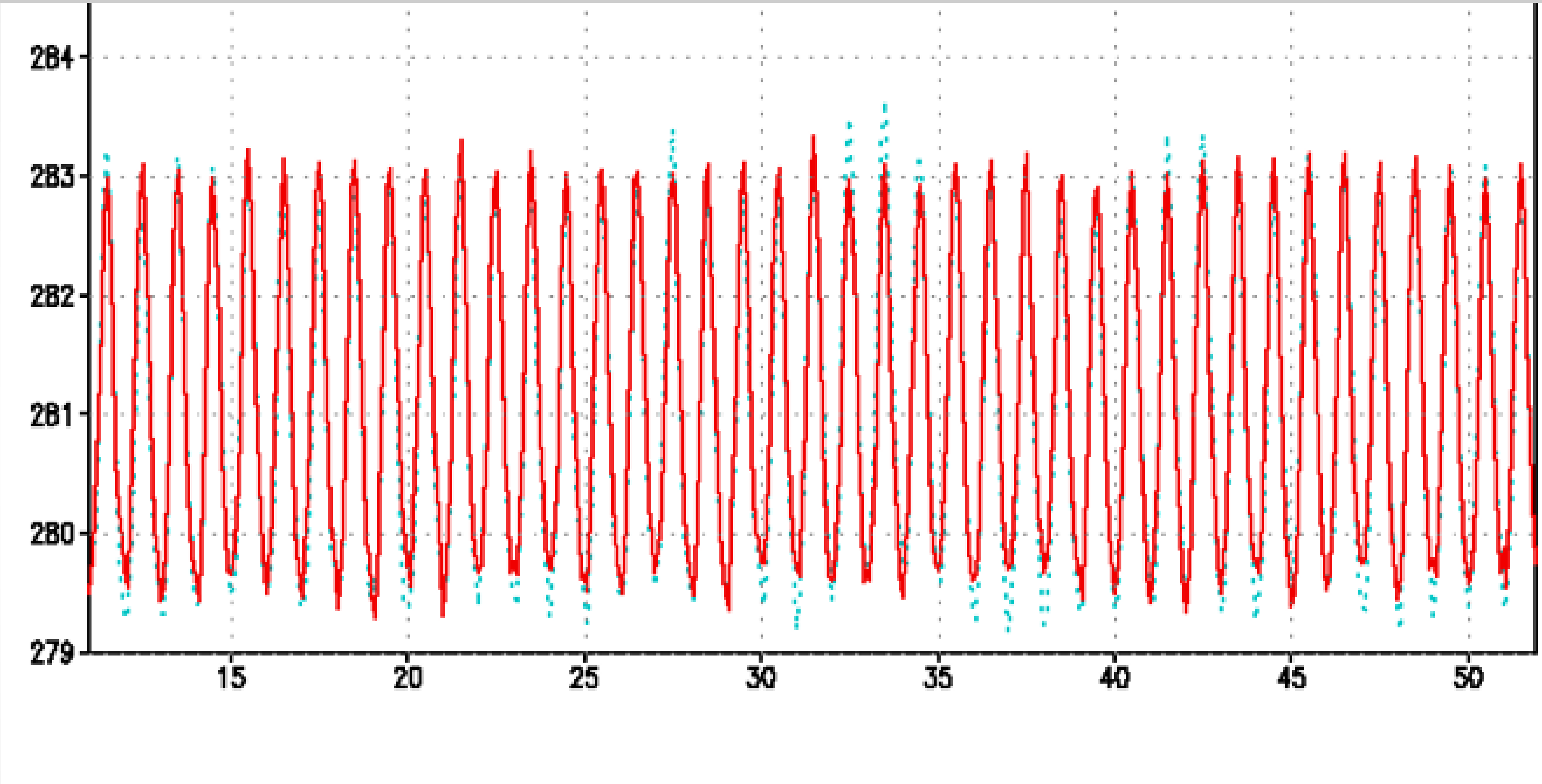
T-850, in K, Winter-Summer Difference.



Annual Cycle Precipitation, mm/day

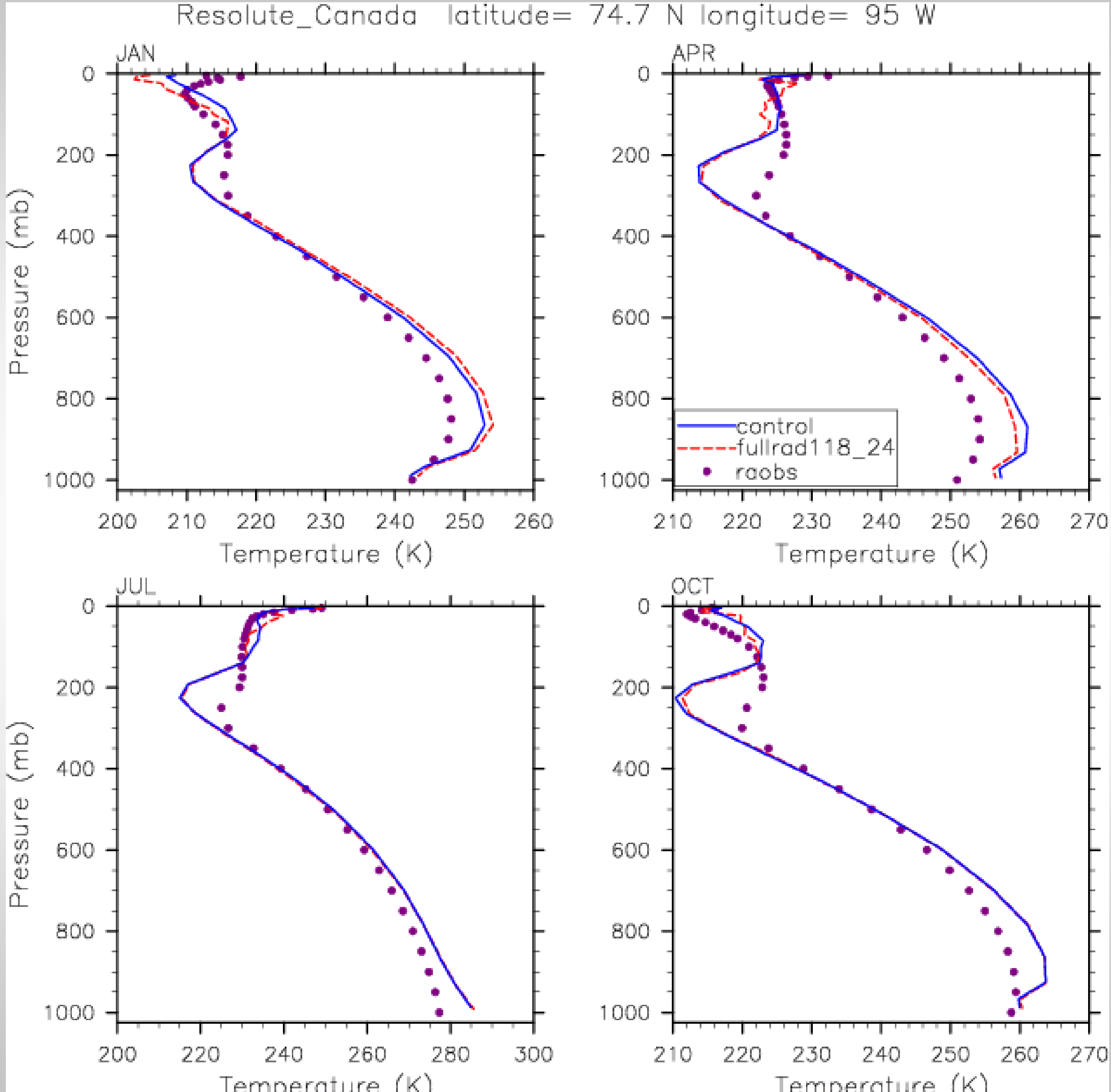


T-850, in K, Global Time Series.



Red – NCAR CAM, Blue – Hybrid NCAR CAM (with NN radiation)

Vertical Profile of T (Resolute, Canada), K



NCEP CFS LWR NN175 – Training

- NN175 emulates LWR **RRTM2** that is the most time consuming physics component in the coupled NCEP CFS/GFS model
- NN architecture: ***High dimensional*** training set consists of LWR inputs and outputs:
  - the training set was selected from 12-year (1995 – 2006) T126L64 run
  - half of the data is used for training and another half for validation or NN accuracy estimation vs. the original LWR
  - every 2 weeks, one day, i.e. eight 3 hourly global files have been recorded; totally – **2,080** files
  - **100** events/profiles have been selected randomly from each 3 hourly global file
  - Totally – **208,000** events/profiles is used for NN training

CFS Model: LWR NN emulation

NN dimensionality and other parameters:

- 591 inputs: 12 variables (pressure, T, moisture, cloudiness parameters, surface emissivity, gases (ozone, CO2)
- 69 outputs: 6 variables (heating rates, fluxes)
- Number of neurons for NN versions: 50 to 200
- **NN dimensionality** for the complex system: 50,000 to 250,000

NN Approximation Accuracy (*on independent data set*) vs. Original Parameterization (all in K/day).

Parameterization	NN	Bias	RMSE	Mean HR	σ_{HR}
LWR (°*K/day*)	**NN75**	$1. 10^{-3}$	0.40	-1.88	2.28
	NN95	$4. 10^{-3}$	0.38		
	NN175	$-5. 10^{-4}$	0.32		

NN Computational Performance: LWR NN emulations are *20-60 times faster* than the original LWR

Overall CFS model computational performance: *~ 20-25% faster* when using LWR NN emulations vs. the original LWR

Note: The practically zero bias obtained is crucially important for non-accumulating errors during long model integrations. The obtained mean PRMSE is quite small/limited.

Validation of the parallel 10-Year Climate Run and Seasonal Predictions Using NN emulations vs. the Control Run

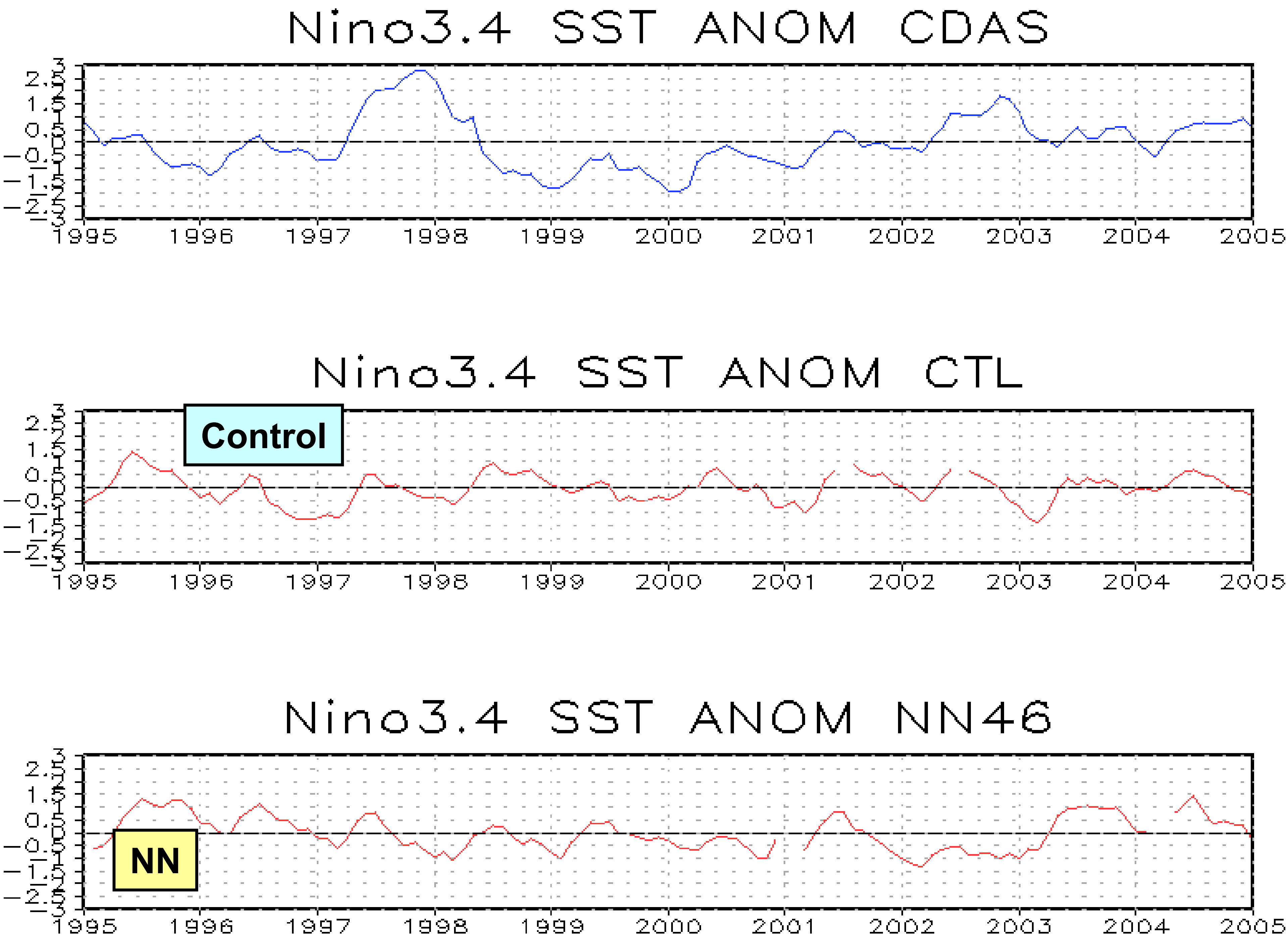
Coupled NCEP CFS
With NN RRTM2 LWR

Acronyms:
CTL – control run
NN175 – RRTM2 LWR NN with 175 hidden neurons

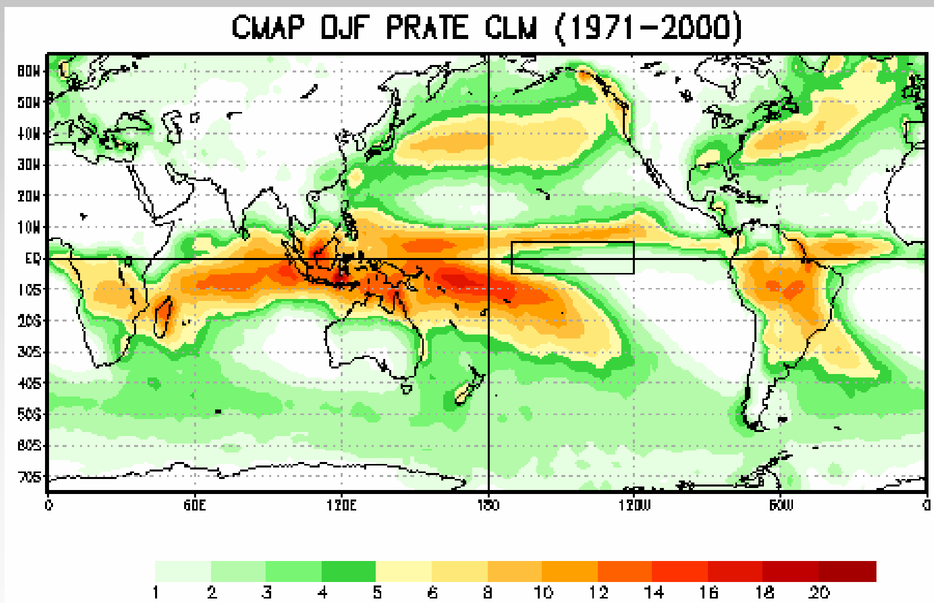
Validation: 10-year Parallel Runs

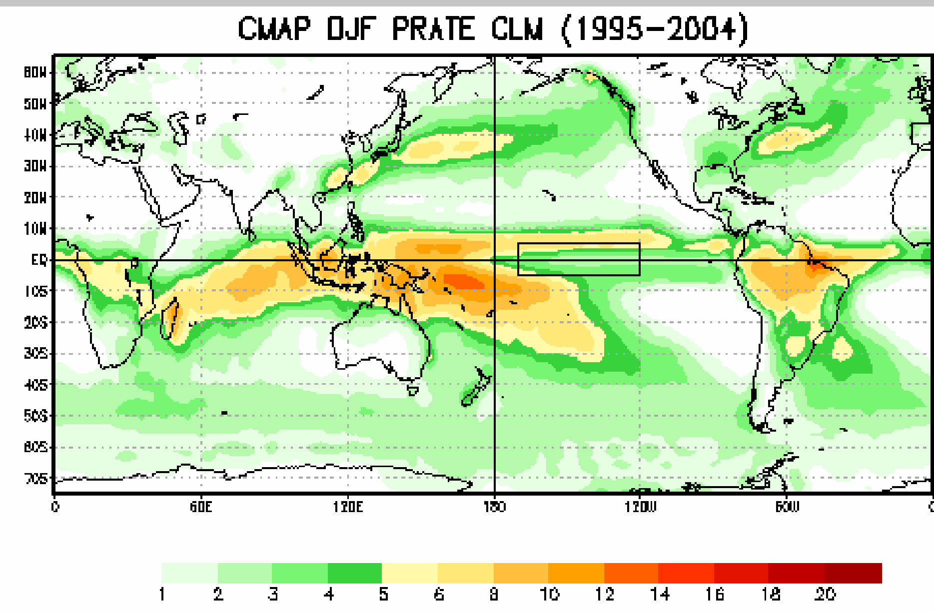
- Control run with RRTM2 LWR for 1995 – 2005
- NN run with NN175 (175 hidden neurons); 1995 – 2005
- Initialization: Control initialization run started in 1990 form initial conditions then:
  - Control run (1995-2005) started from the Jan 1995 restart of the initialization run
  - NN run (1995-2005) started from the same Jan 1995 restart
- NN LWR vs. RRTM2 speed up is about 100 times
- Total model speed up is about 20-25%
- **All figures are prepared using the validation software provided by Dr. S. Saha**

Nino 3.4 SST Anomaly



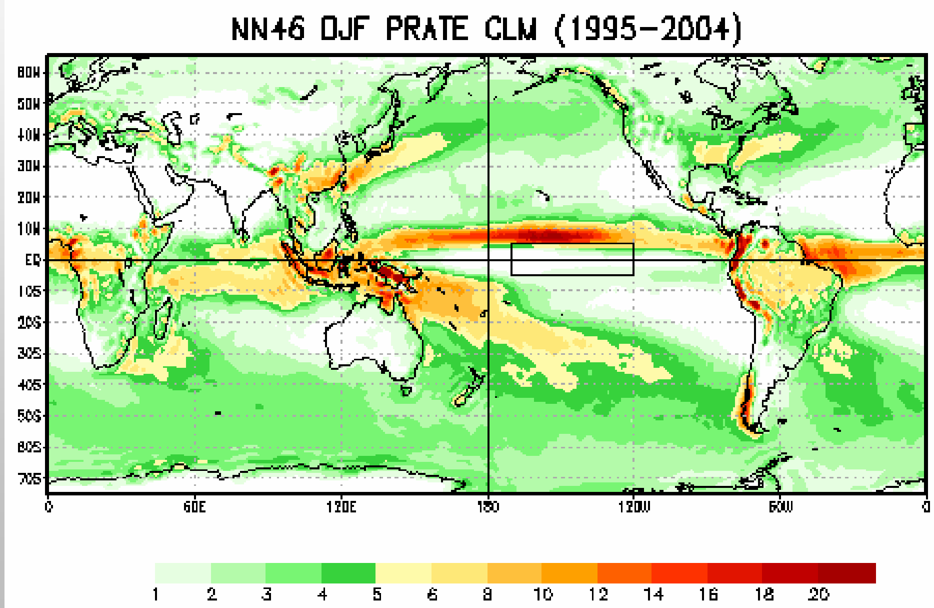
10 year PRATE (djf)

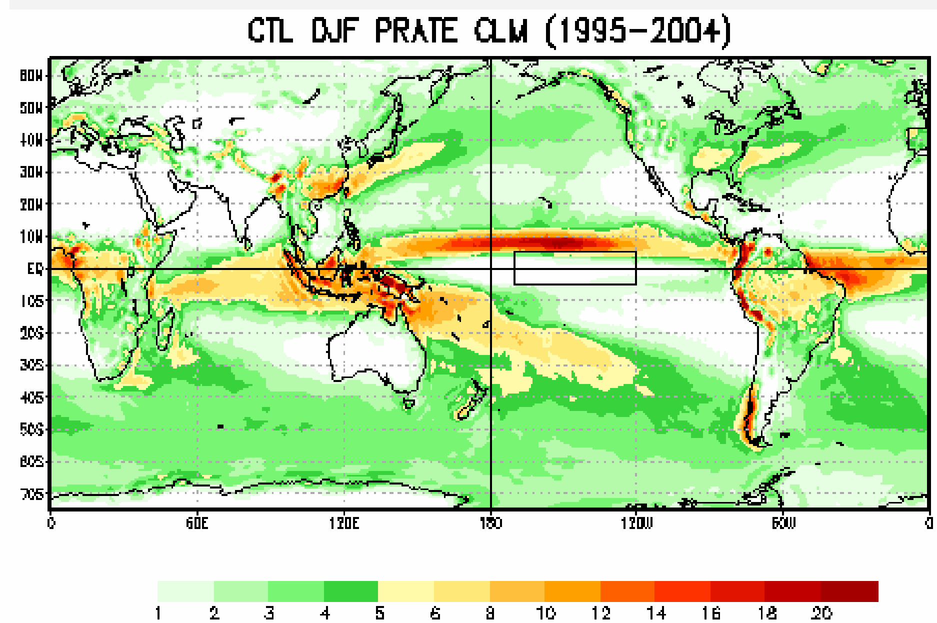




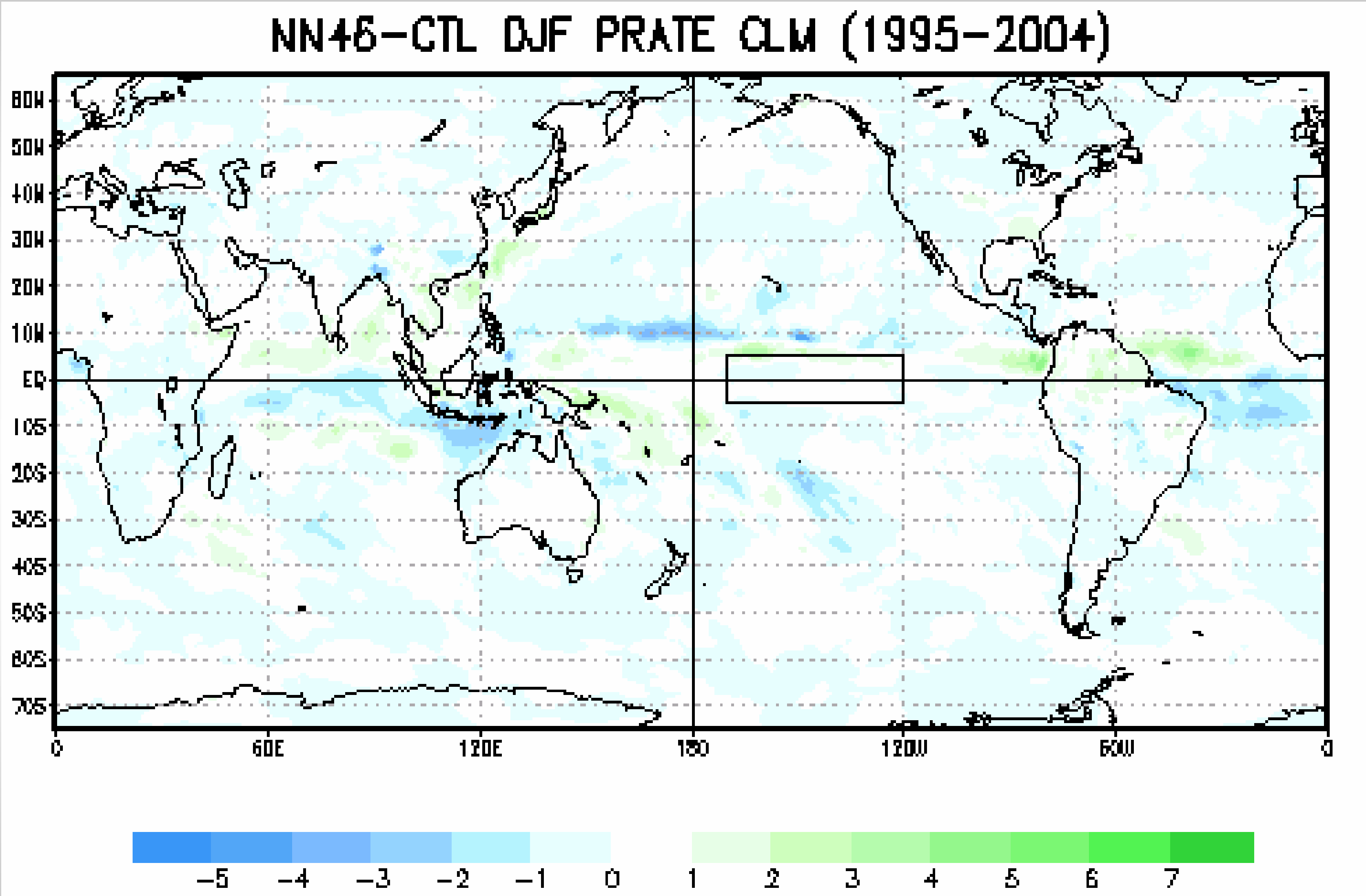
NN

Control



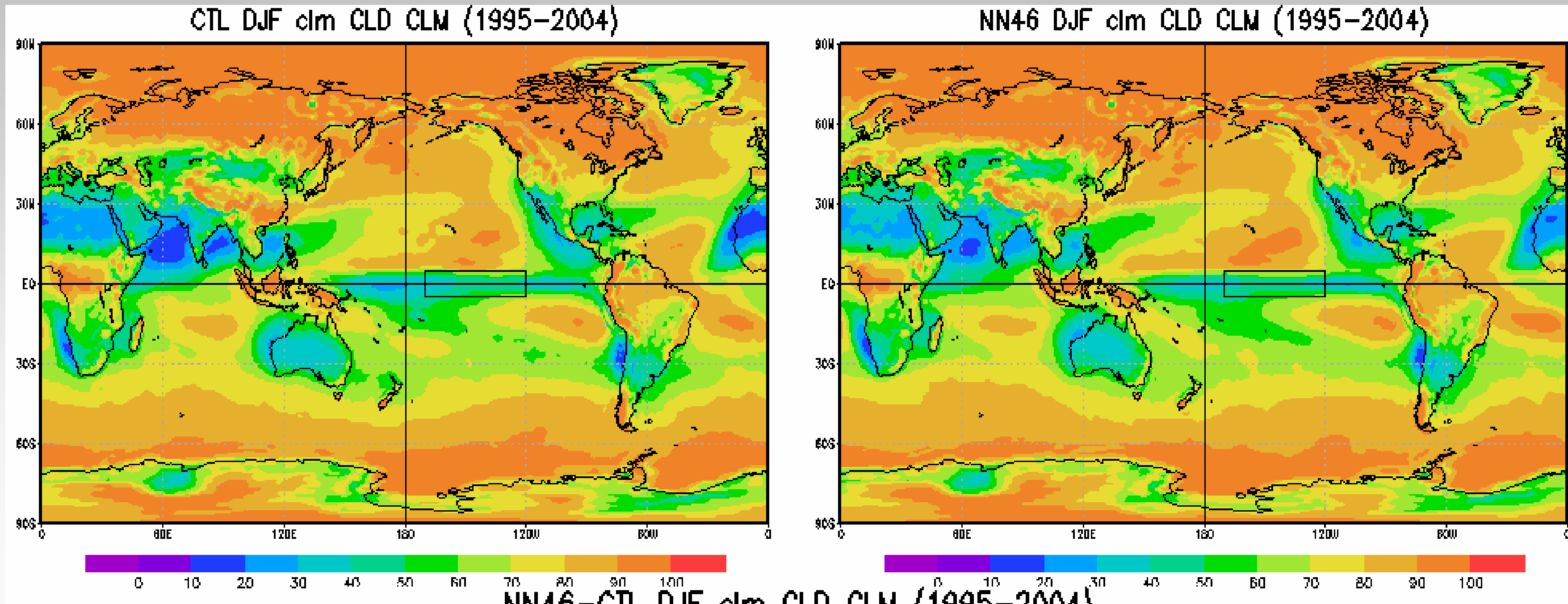


10 year PRATE (djf)



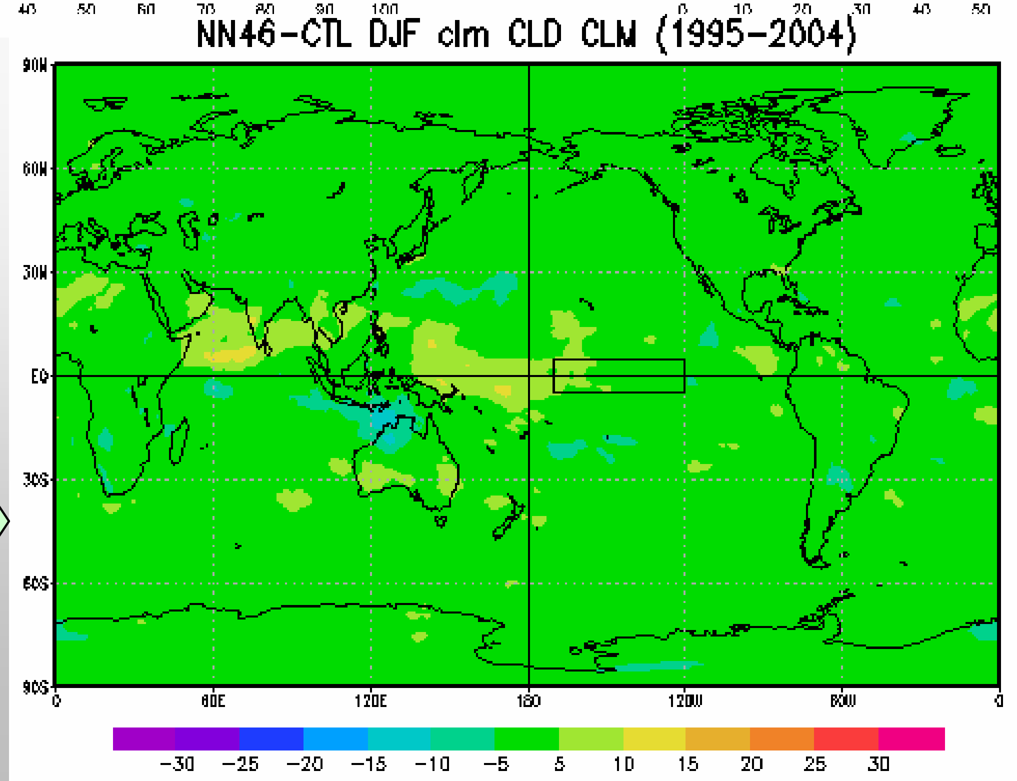
NN - Control

10 year Total Clouds cldclm (djf)



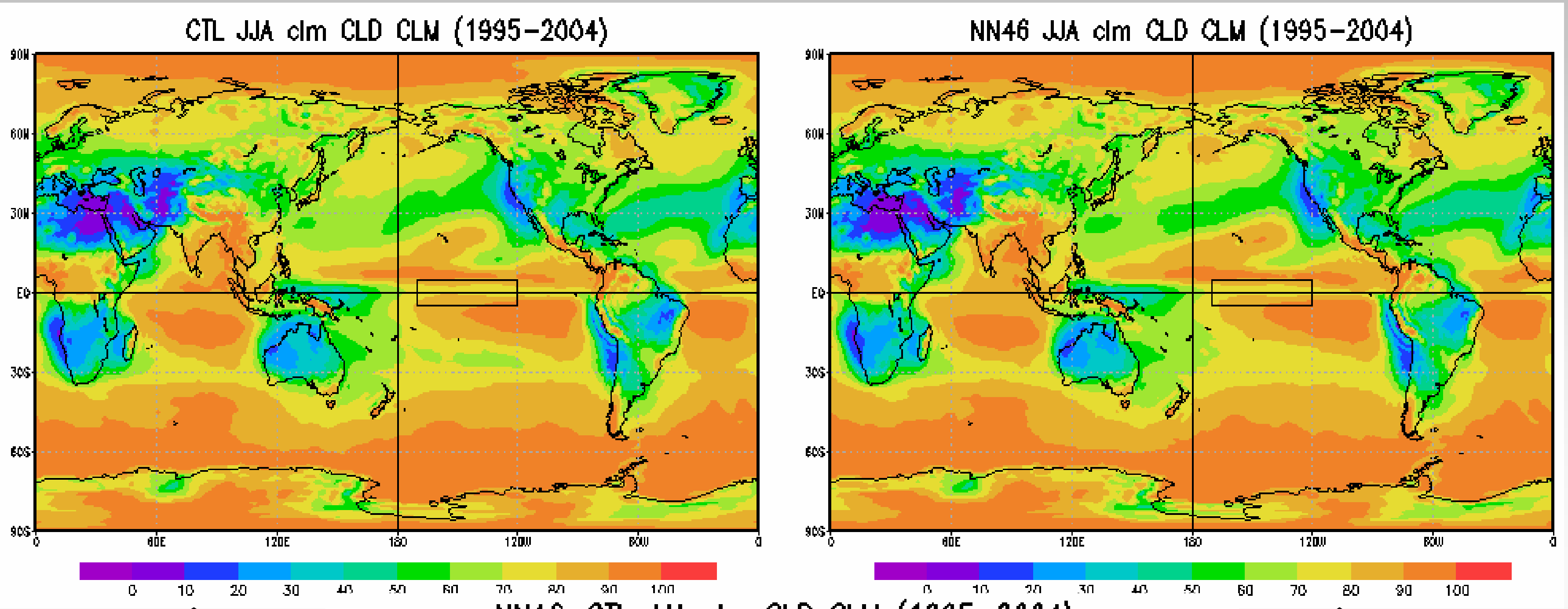
Control

NN



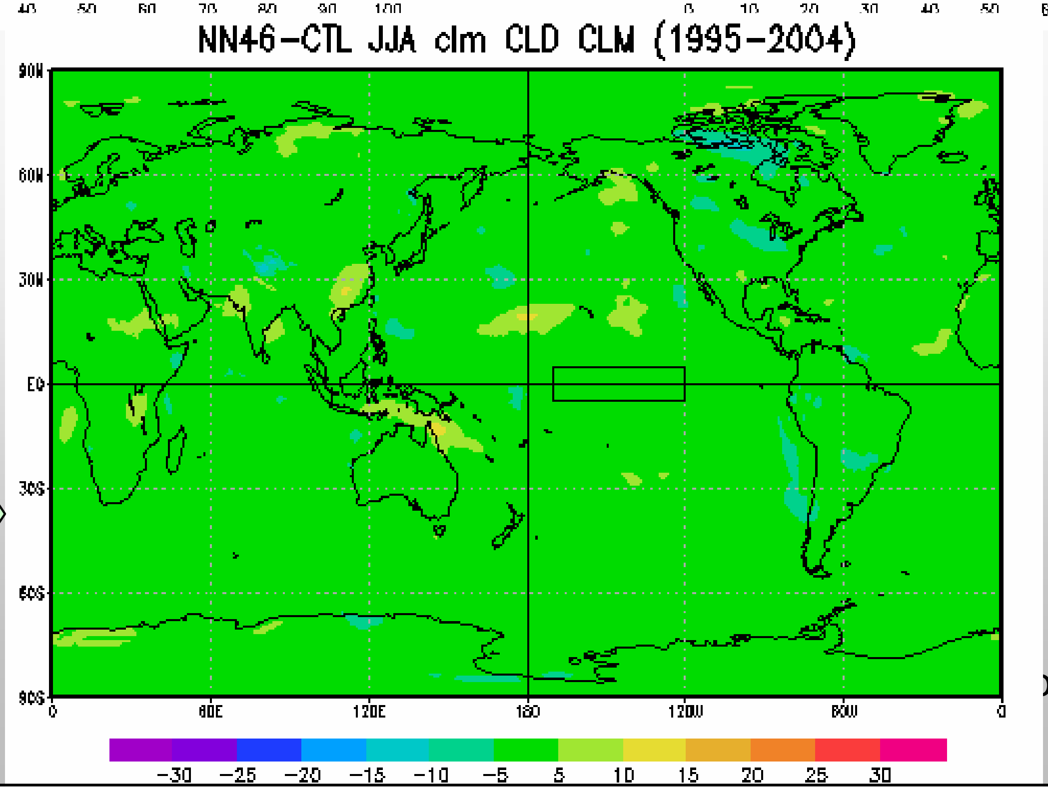
NN - Control

10 year Total Clouds cldclm (jja)



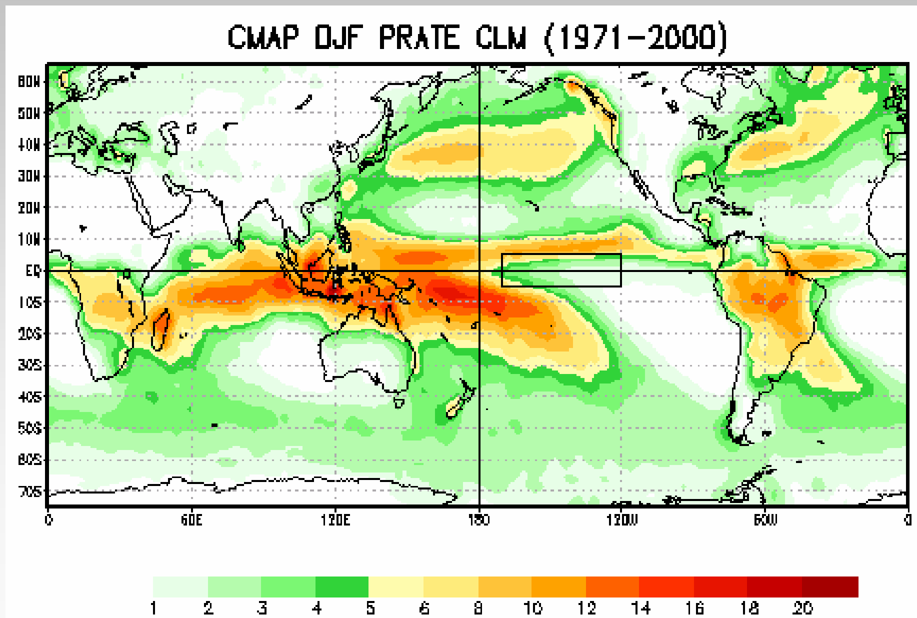
Control

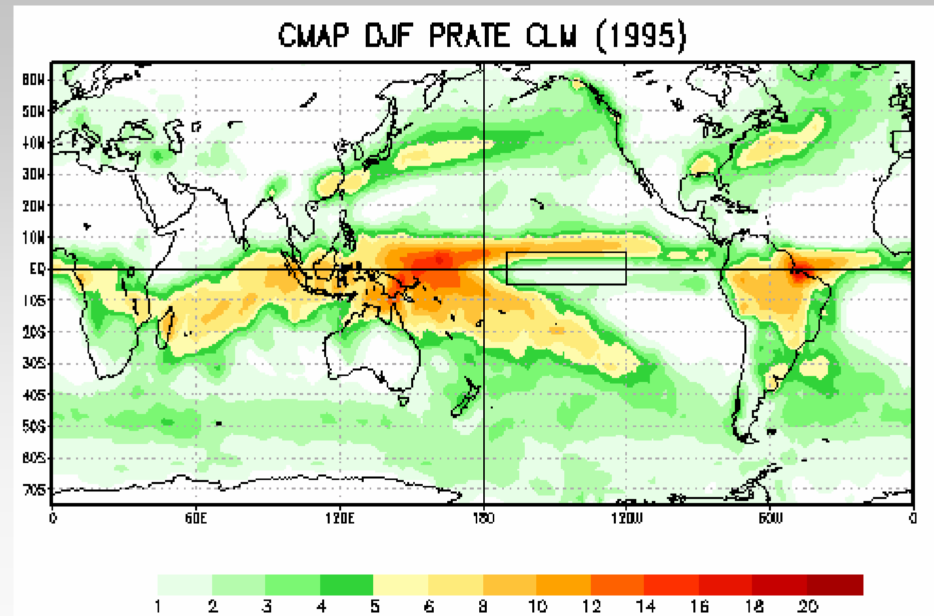
NN



NN - Control

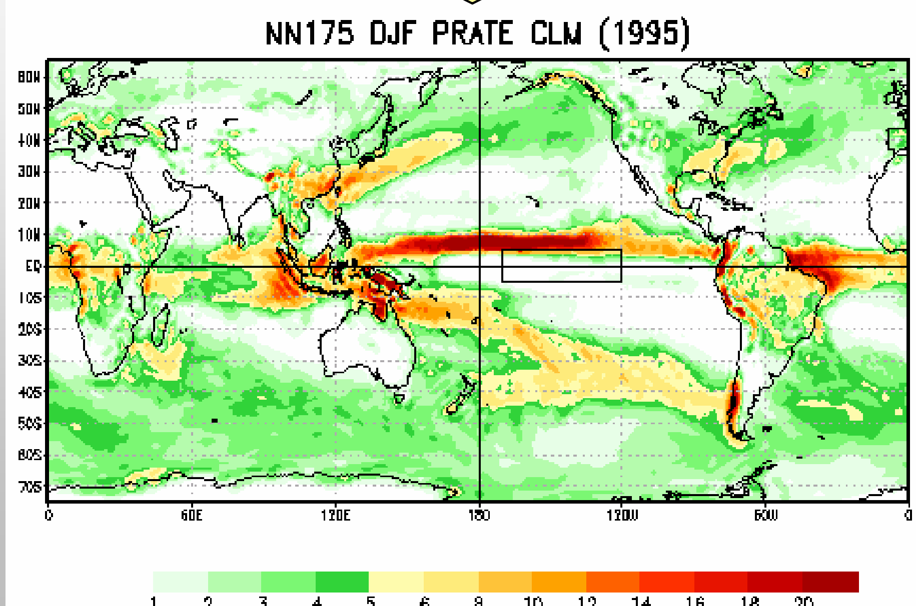
Seasonal PRATE (djf)

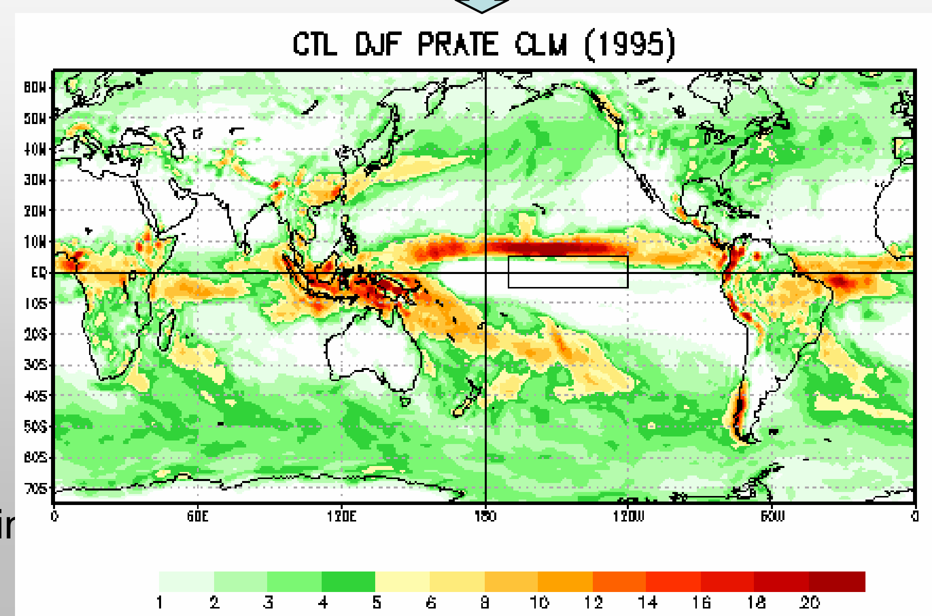




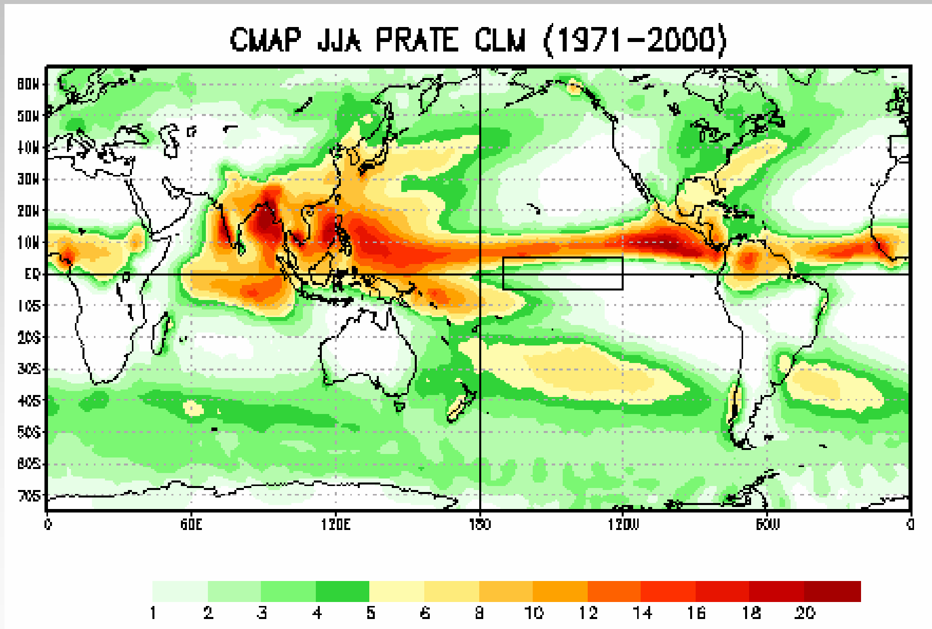
NN

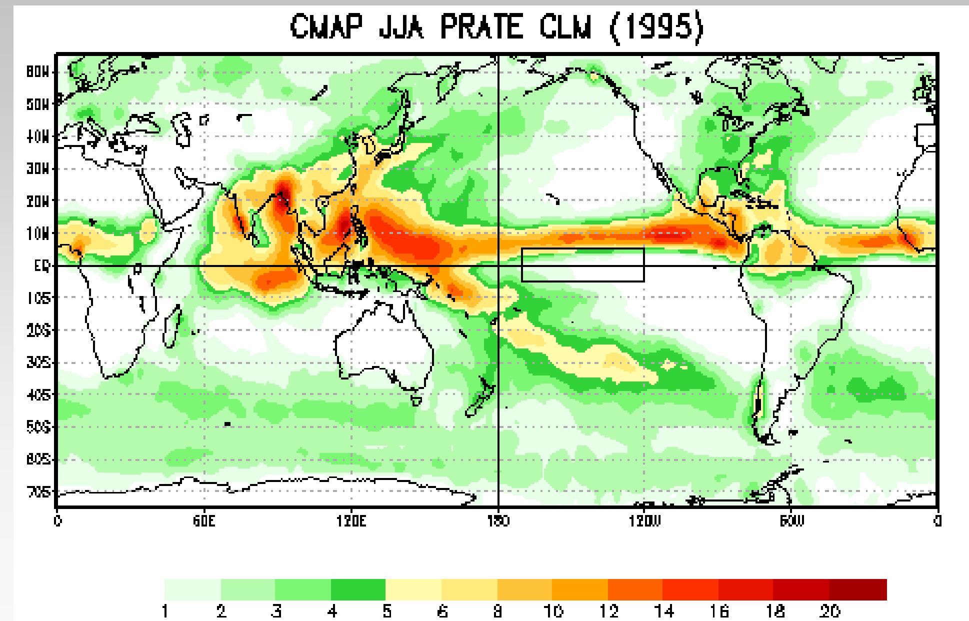
Control





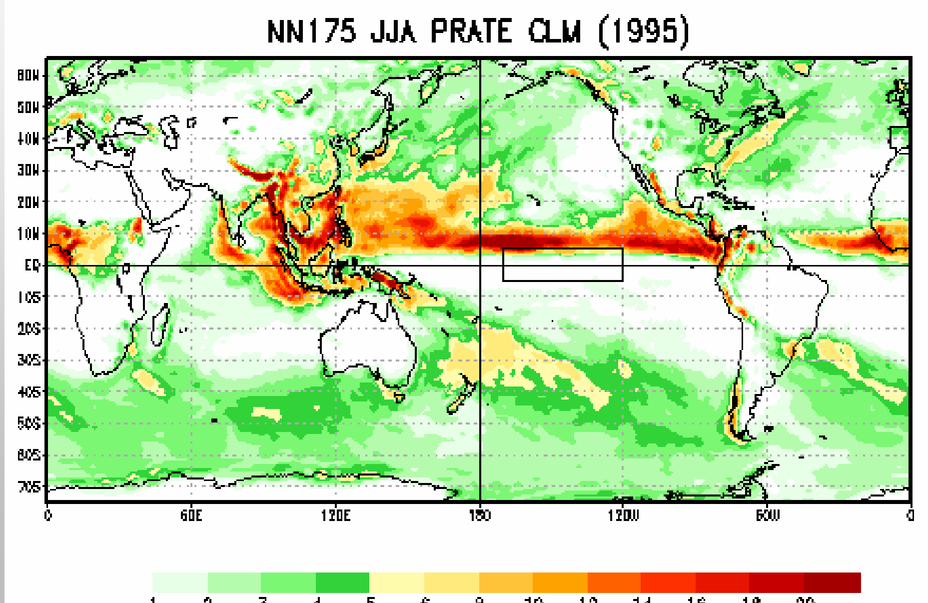
Seasonal PRATE (jja)

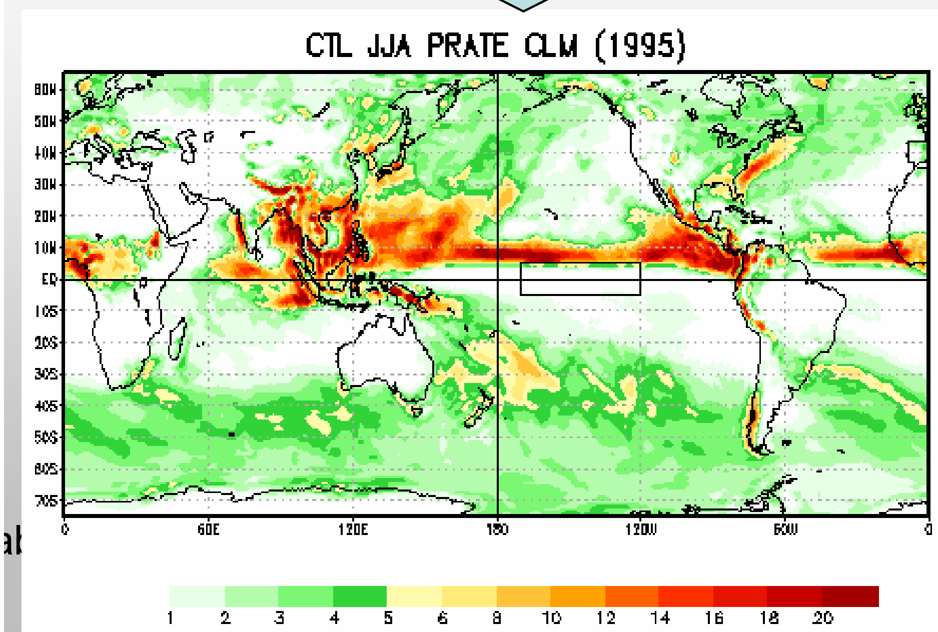




NN

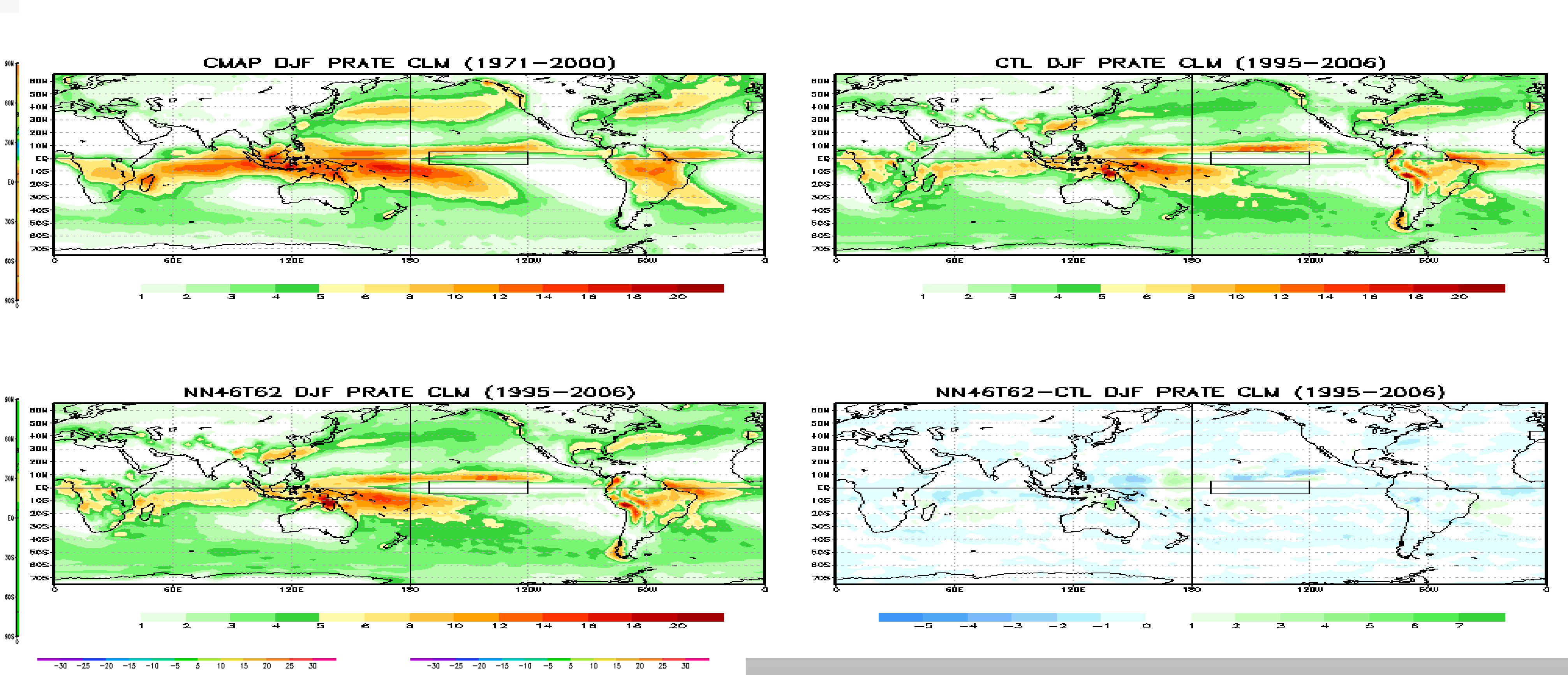
Control



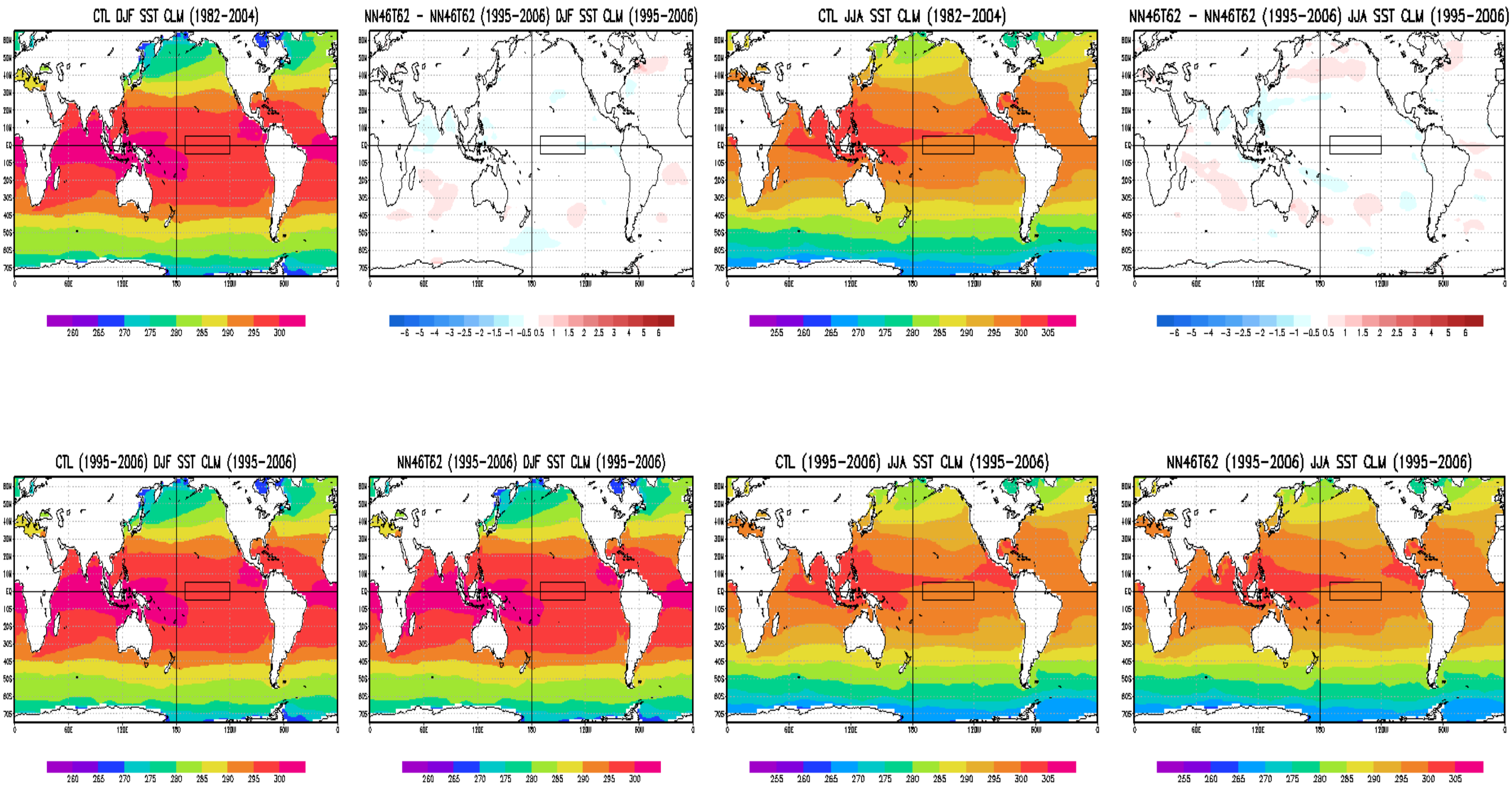


Robustness of NNs to model horizontal resolution

- NN emulation developed for T126L64 has been used for the 10-year T62L64 and 4-month T382L64 parallel runs
- Differences for the parallel 10-year T62L64 runs are small and similar to those of the parallel 10-year T126 runs; the analysis of the 4-month parallel T382L64 runs are in progress



Robustness of NNs to model horizontal resolution

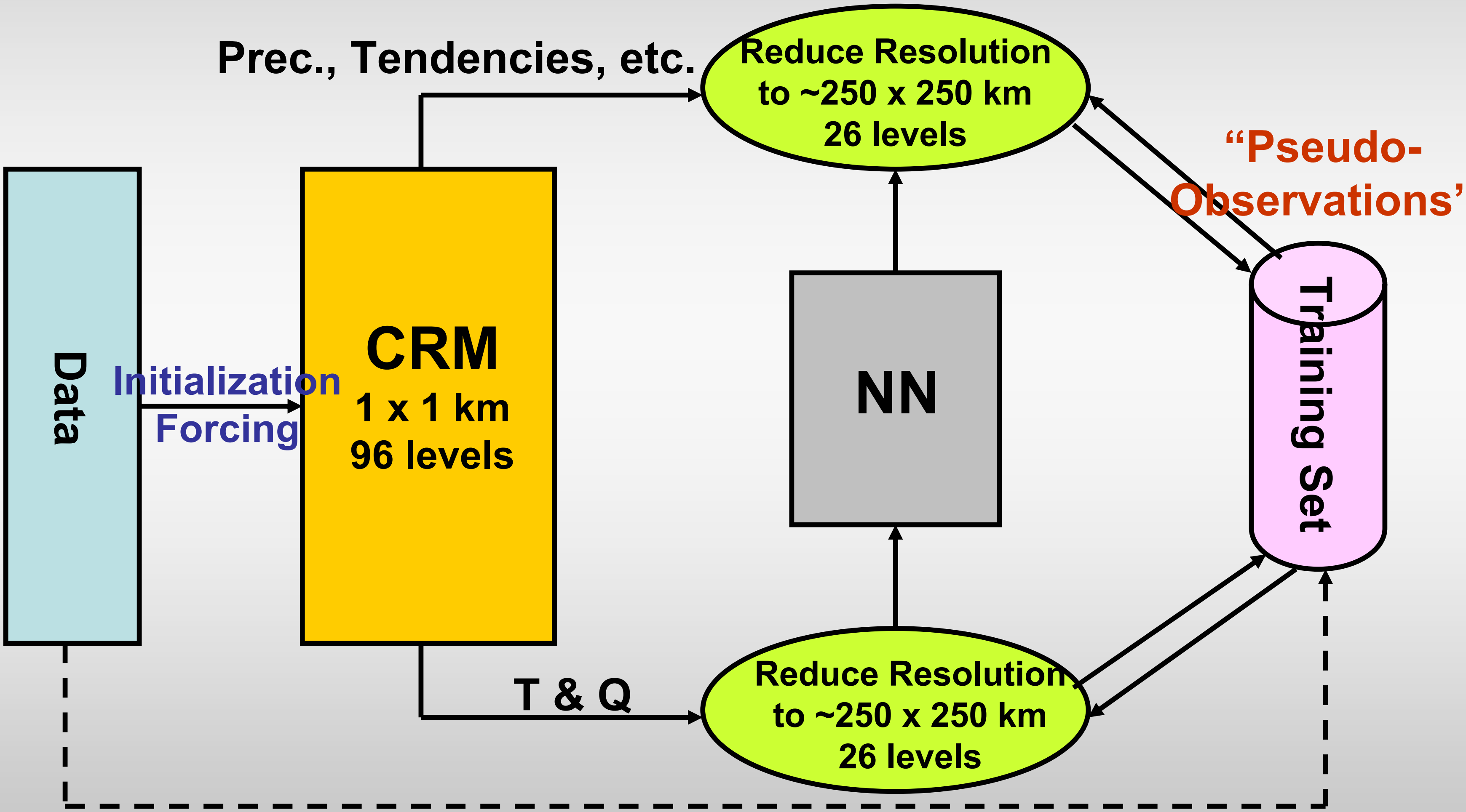


NN Parameterizations

- **New NN parameterizations** of model physics can be developed based on:
  - Observations
  - Data simulated by first principle process models (like cloud resolving models).
- Here NN serves as an interface transferring information about sub-grid scale processes from fine scale data or models (CRM) into GCM (upscaling)

NN convection parameterizations for climate models based on learning from data.

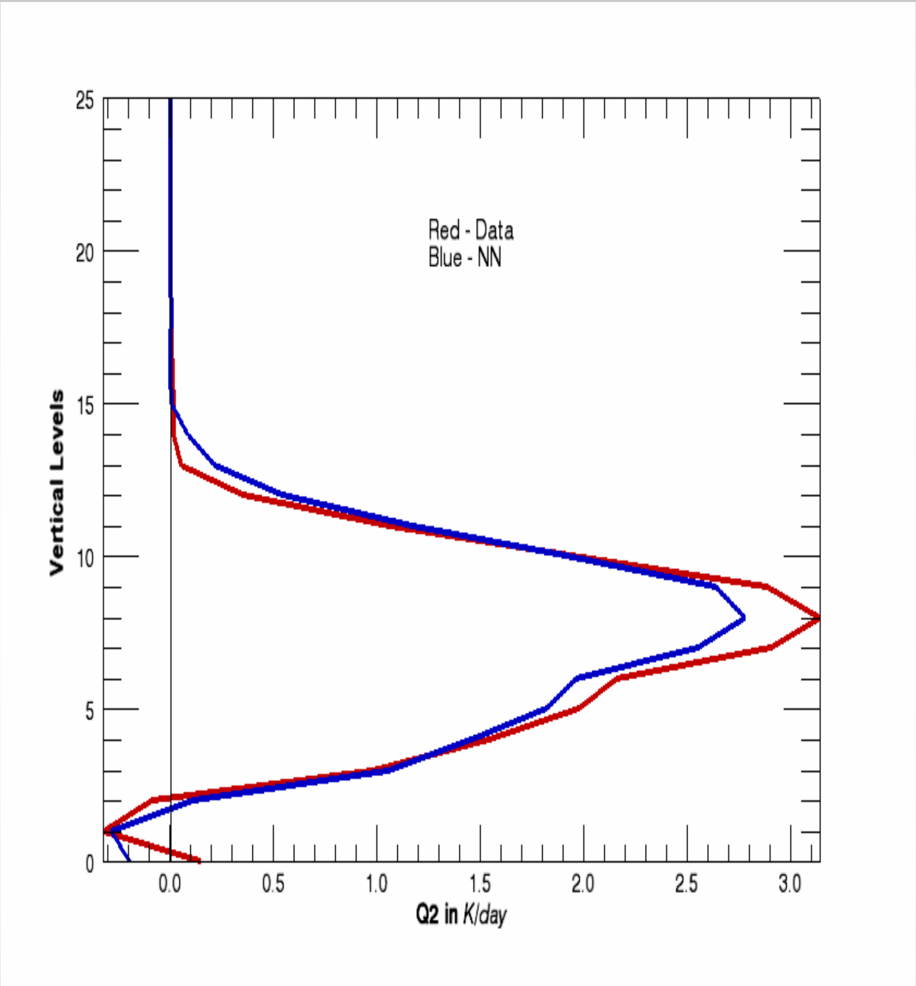
Proof of Concept (POC) -1.



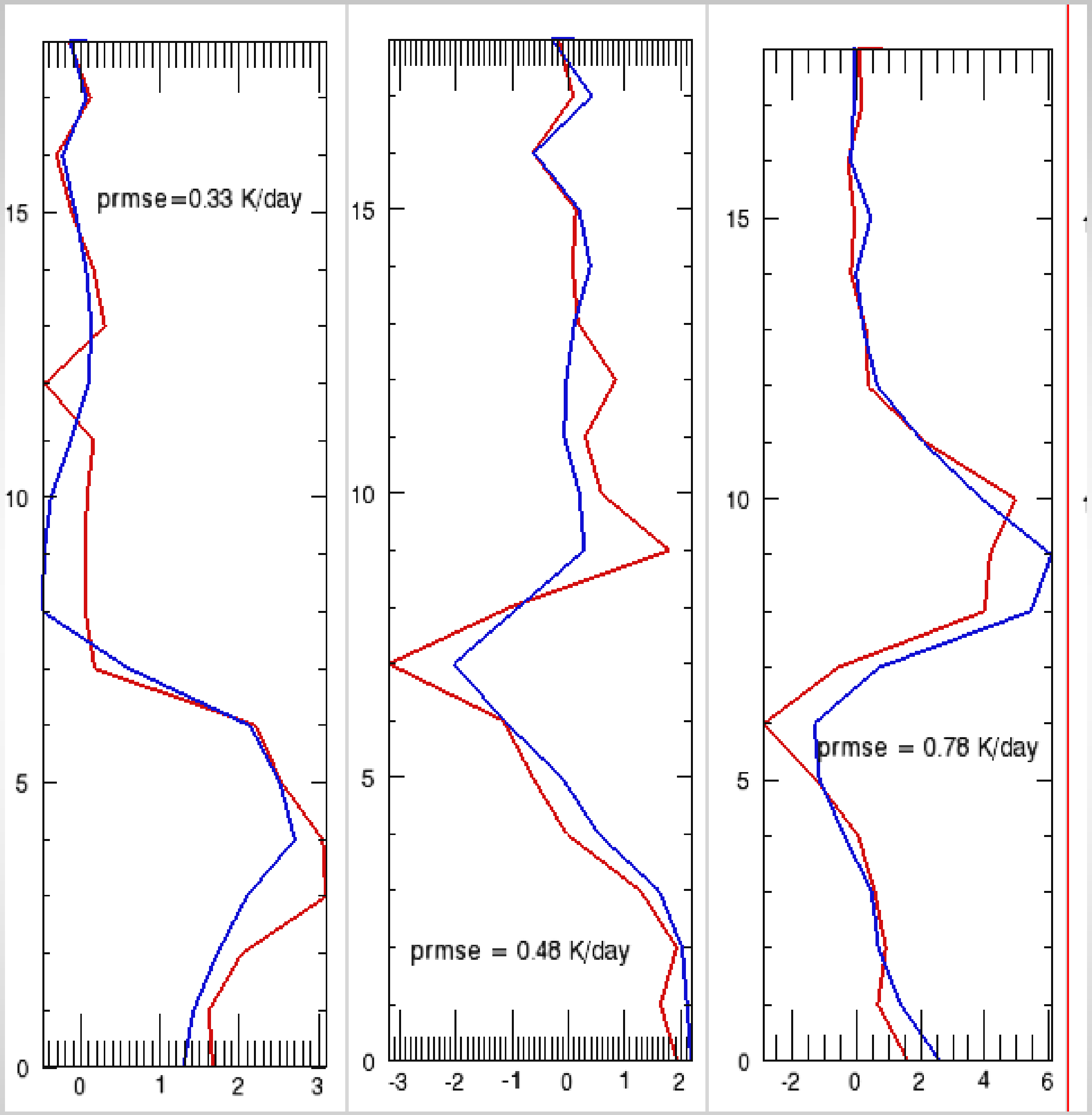
Proof of Concept - 2

- **Data (forcing and initialization):** TOGA COARE meteorological conditions
- **CRM:** the SAM CRM (Khairoutdinov and Randall, 2003).
  - Data from the archive provided by C. Bretherton and P. Rasch (*Blossey et al, 2006*).
  - Hourly data over 90 days
  - Resolution 1 km over the domain of 256 x 256 km
  - 96 vertical layers (0 – 28 km)
- **Resolution of "pseudo-observations" (averaged CRM data)**:
  - Horizontal 256 x 256 km
  - 26 vertical layers
- **NN inputs: only** temperature and water vapor fields; a limited training data set used for POC
- **NN outputs:** precipitation & the tendencies T and q, i.e. "apparent heat source" (Q1), "apparent moist sink" (Q2), and cloud fractions (CLD)

Proof of Concept - 4

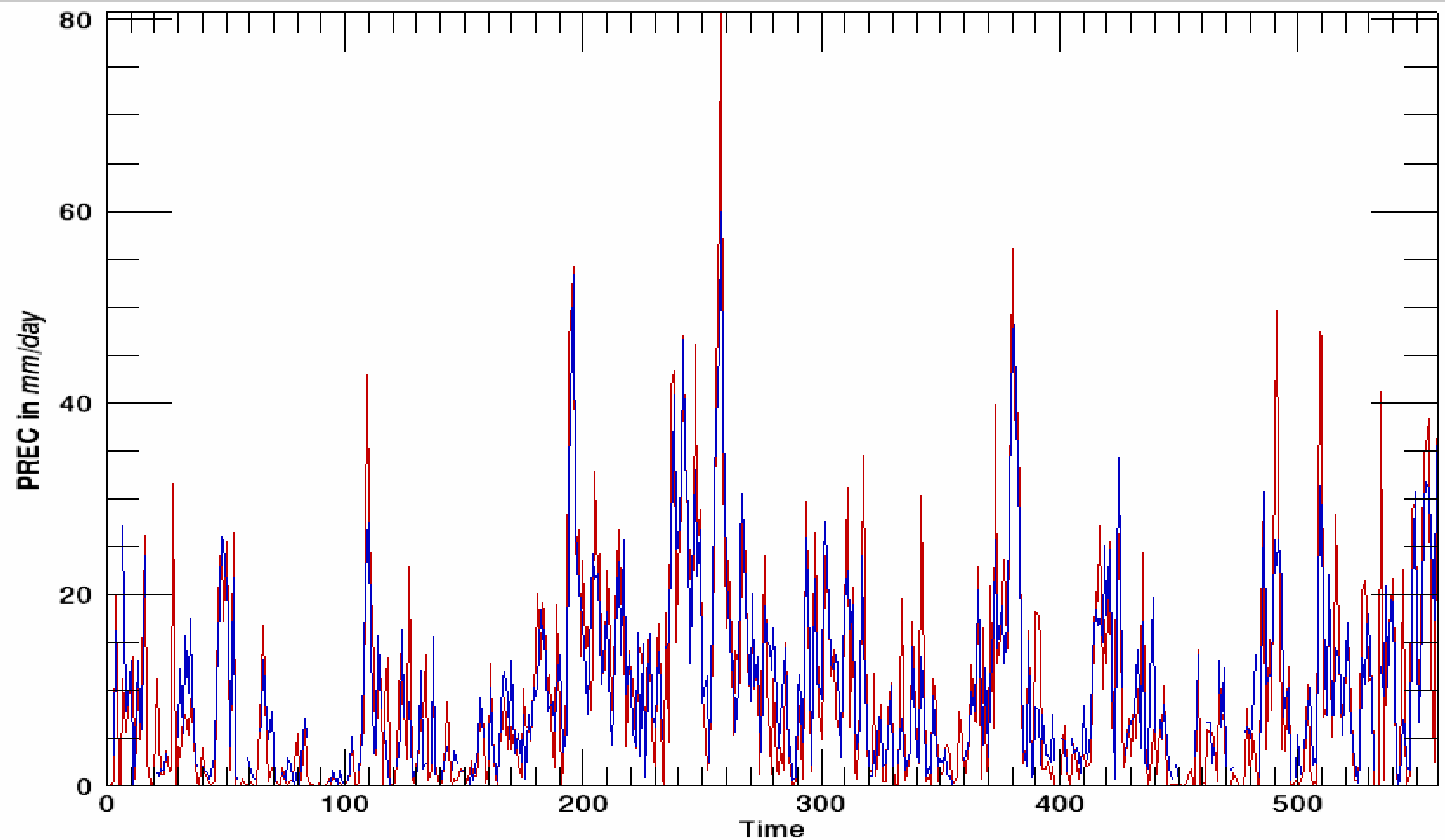


Time averaged water vapor tendency (expressed as the equivalent heating) for the validation dataset.



Q2 profiles (red) with the corresponding NN generated profiles (blue). The profile rmse increases from the left to the right.

Proof of Concept - 3

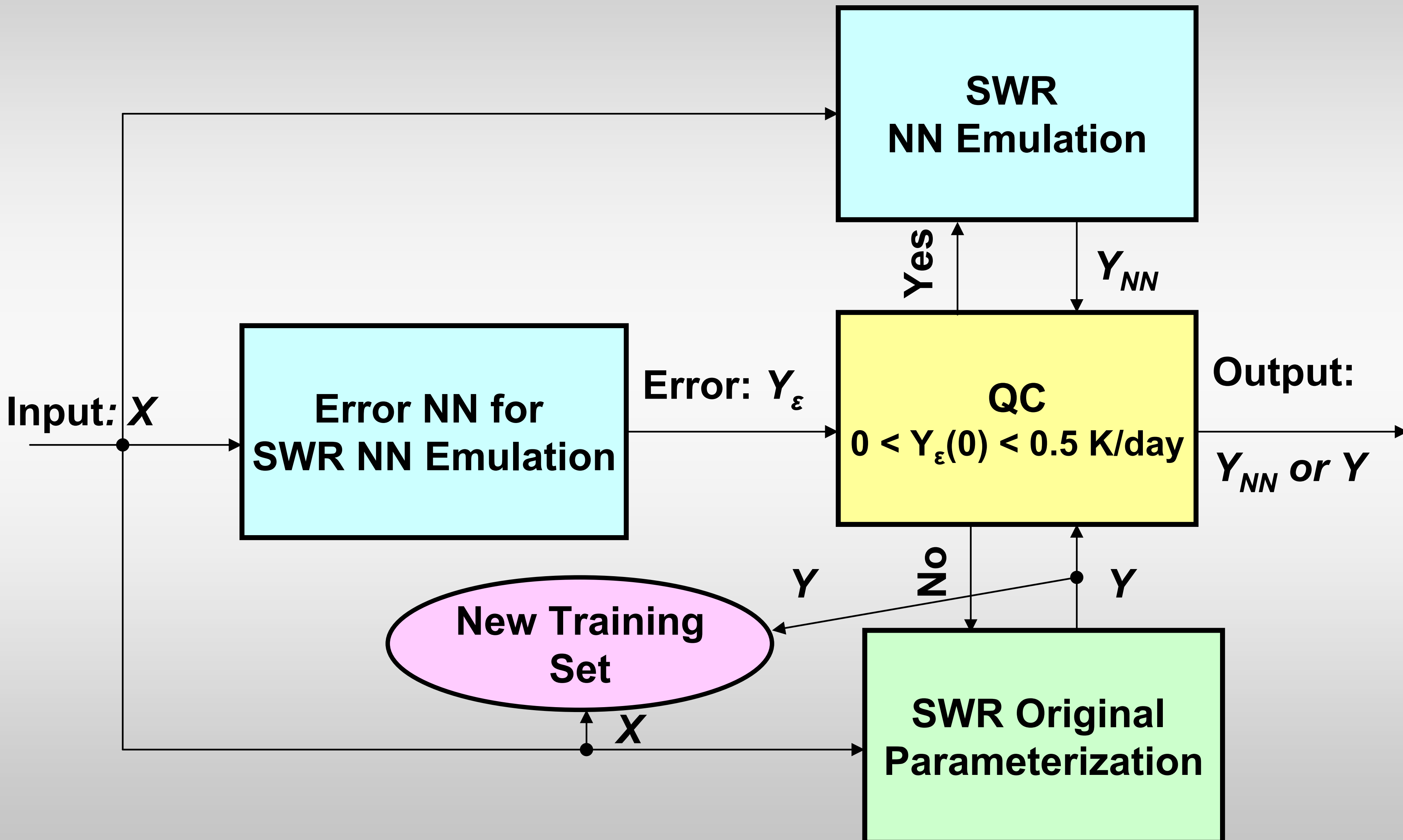


Precipitation rates for the validation dataset. Red – data, blue - NN

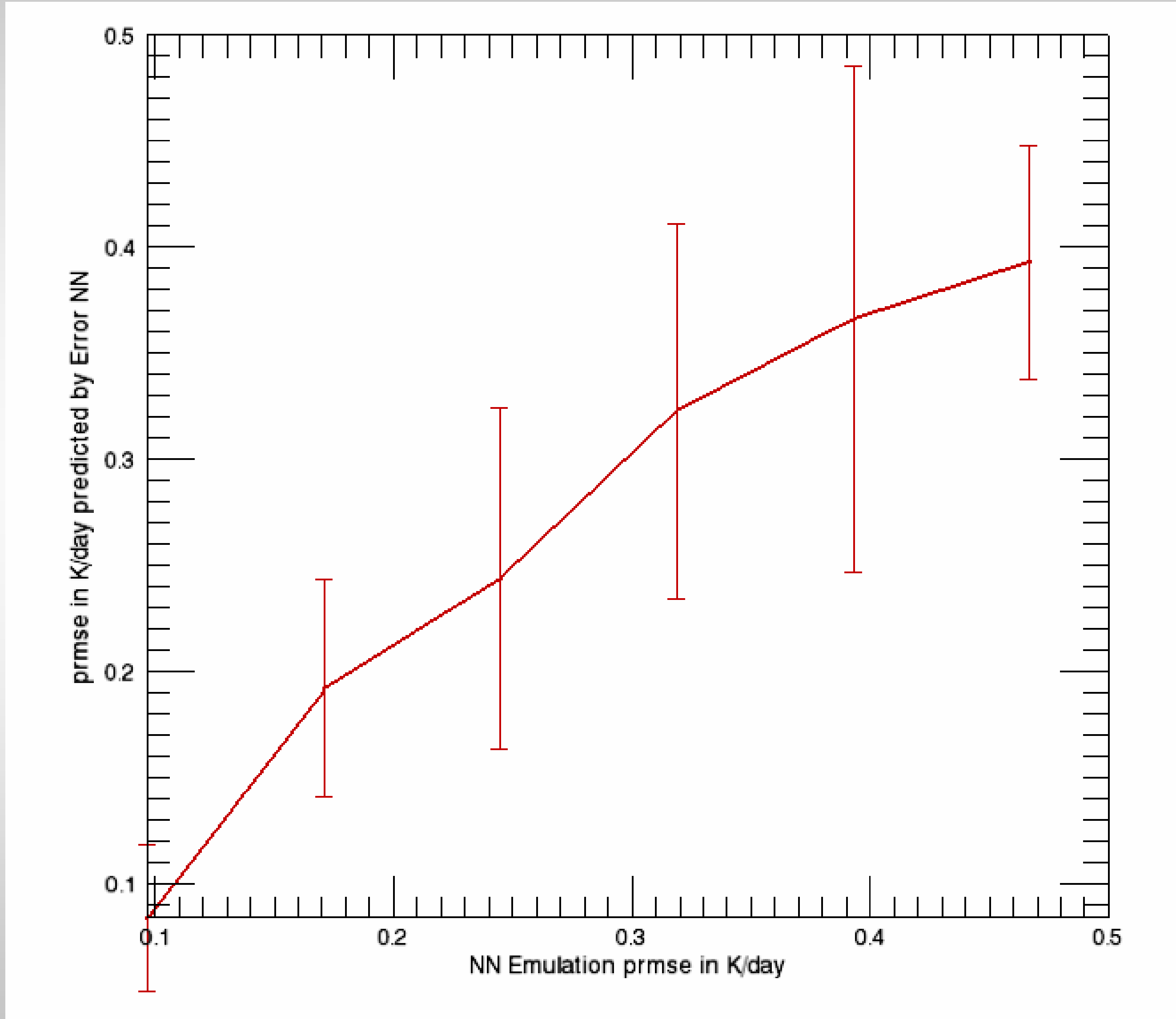
Compound Parameterization (CP): Quality Control (QC) for Outliers and Adaptivity to Climate Change

- NN emulations are very accurate; however, larger errors (outliers) may occur because of several reasons and a quality control (QC) for outliers is needed to account for:
  – Not representative training set
  – High level of noise in training data
  – Redundancy in NN architecture
- Due to climate changes NN can face situations for which it was not trained (e.g. an increased level of CO_2)
- CP addresses both above challenges

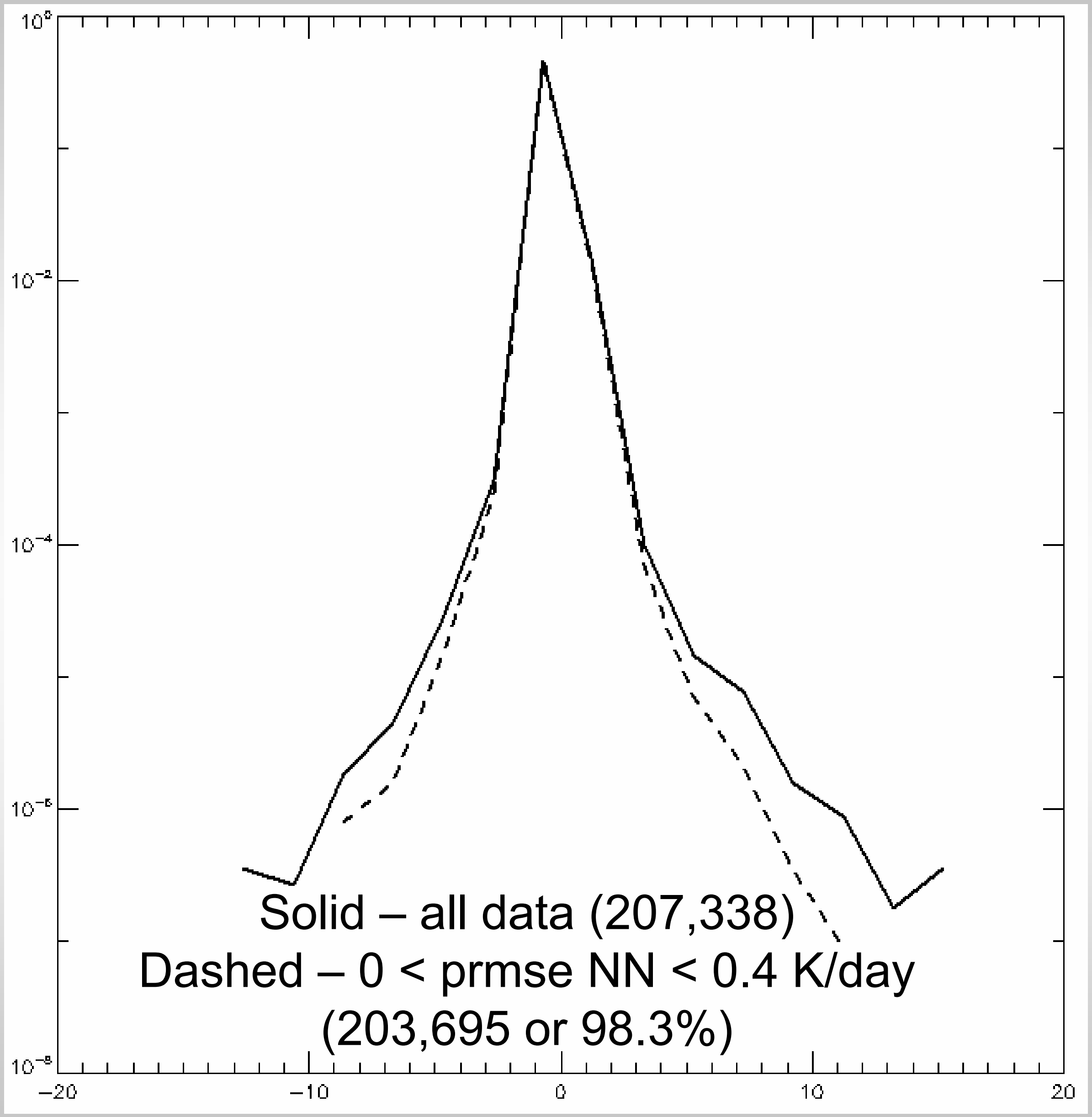
Compound Parameterization Based on Error NN



Predicted vs. Real Errors
Mean CC = 0.87

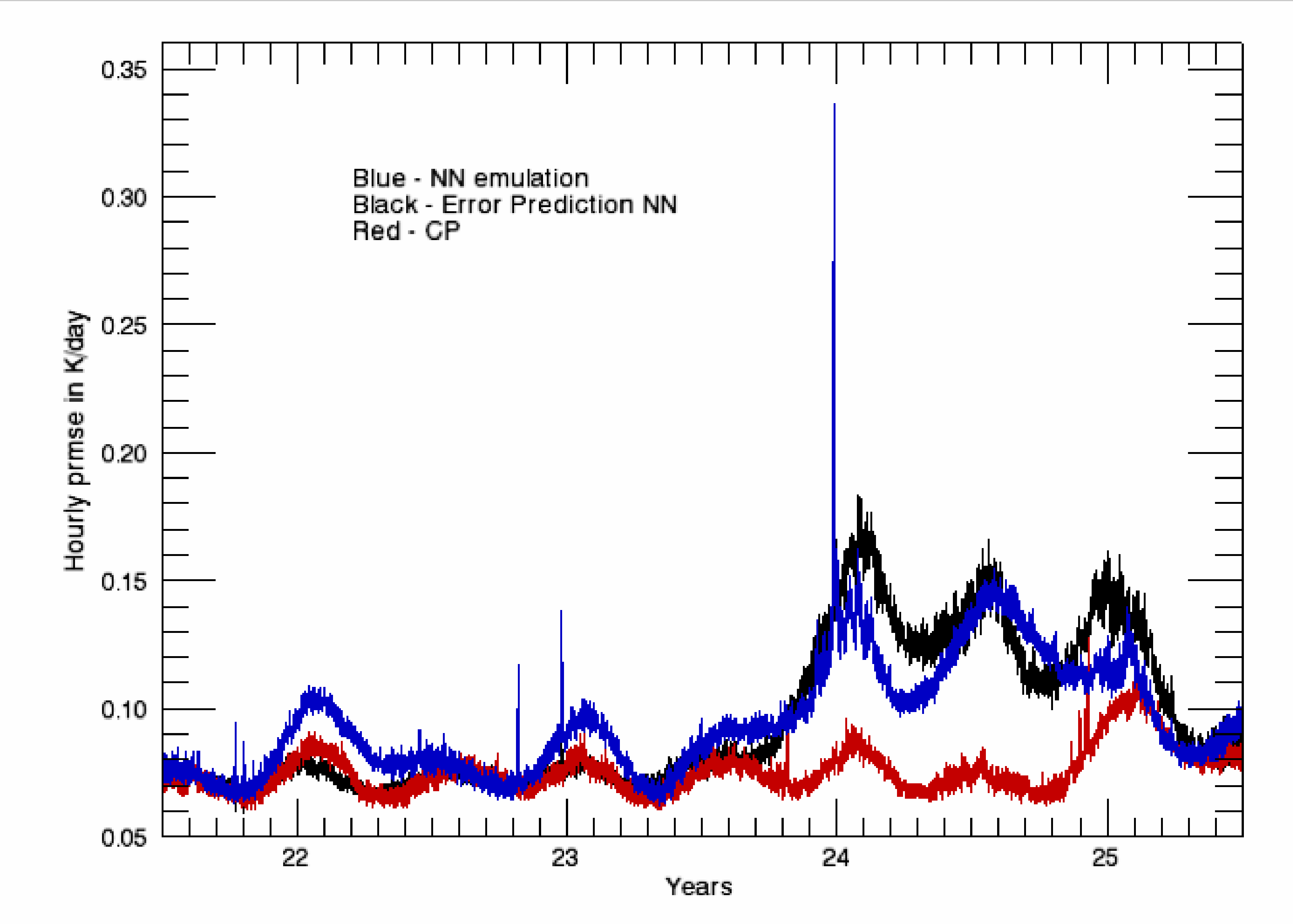


Error PDF



HR Errors in K/day

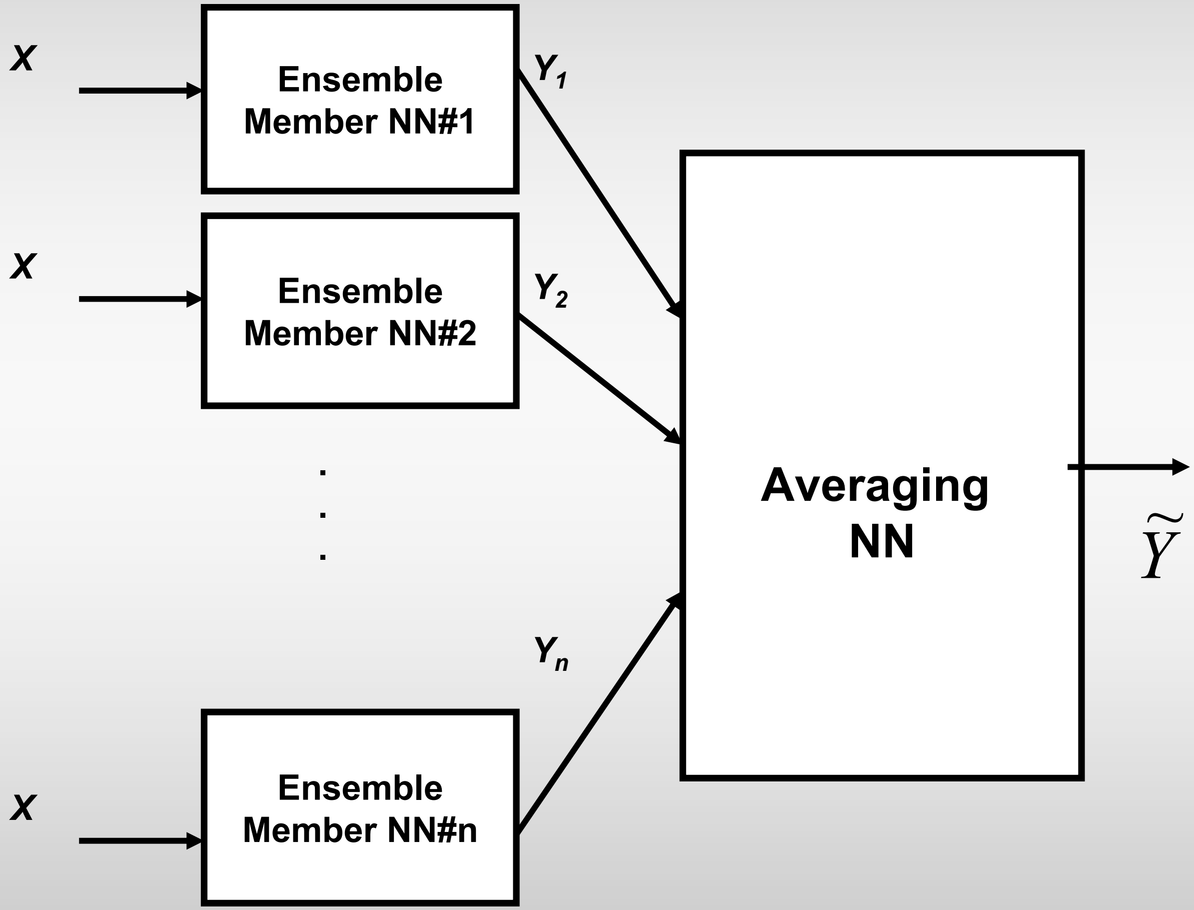
Hourly prmse



NN Ensembles

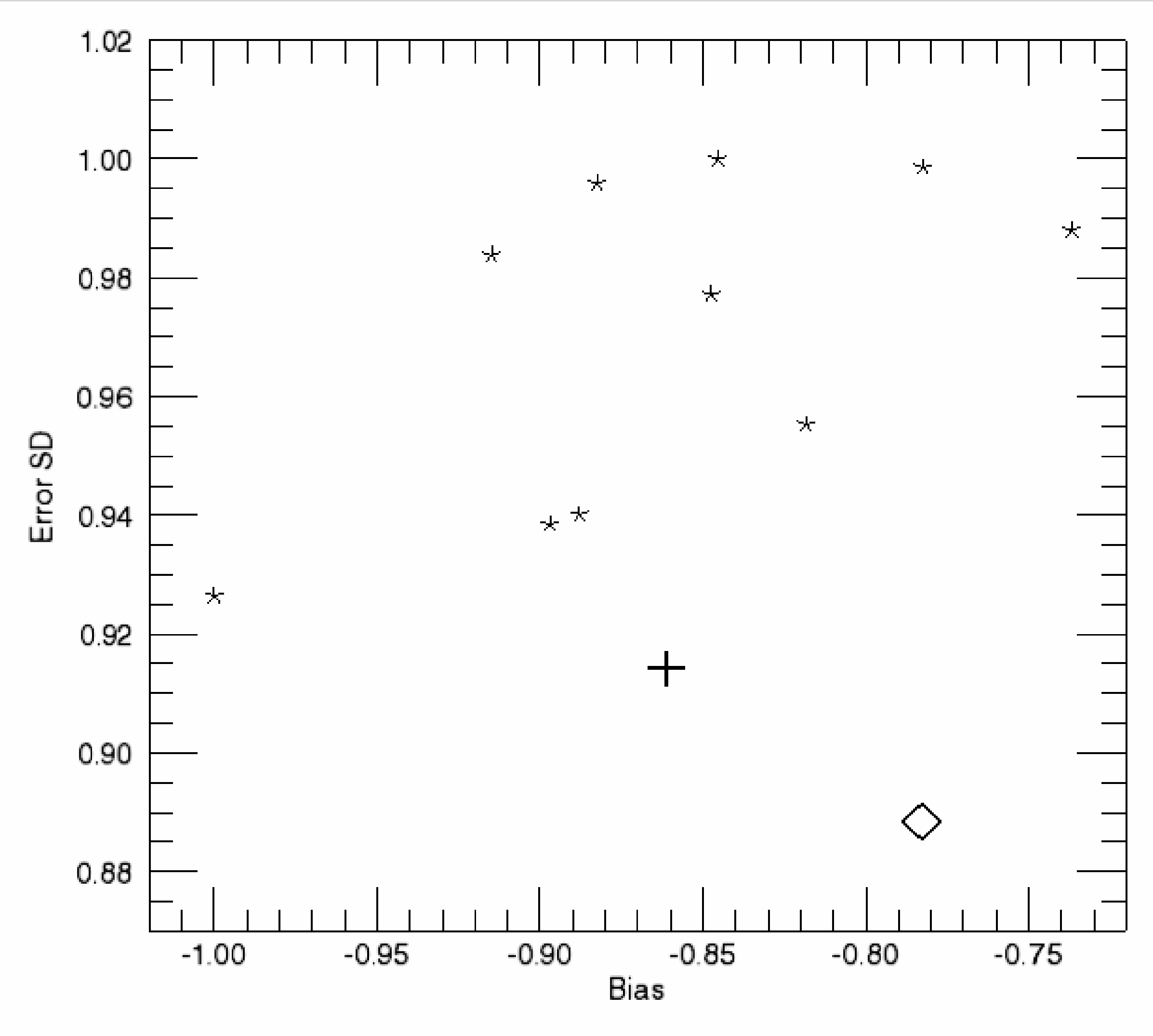
- **Ensemble approach is easily and naturally compatible with NN techniques**
  - Many ways of generating NN ensembles
    - Different initial conditions for NN training
    - Different NN architectures
    - Different partitioning of training set and/or domain
  - NNs can be applied for a nonlinear ensemble averaging
- **NN ensembles can be applied to:**
  - Improve accuracy of NN emulations
  - Improve accuracy of NN parameterizations
  - Reduce uncertainty of climate simulations and projections
  - Reduce uncertainty of NN Jacobian
  - Produce ensembles of perturbed physics
- **NN ensembles are possible due to computational efficiency of NNs**

Ensemble Approaches to Improve the Accuracy of NN Emulations:
Nonlinear Ensemble.



Schematic representation of a nonlinear ensemble that uses an averaging NN. A tilde over the averaging NN output $\tilde{Y}$ emphasizes that a nonlinear ensemble average is produced. *X* is an input of the emulation.

Ensemble Approaches to Improve the Accuracy of NN Emulations:
Nonlinear Ensemble.



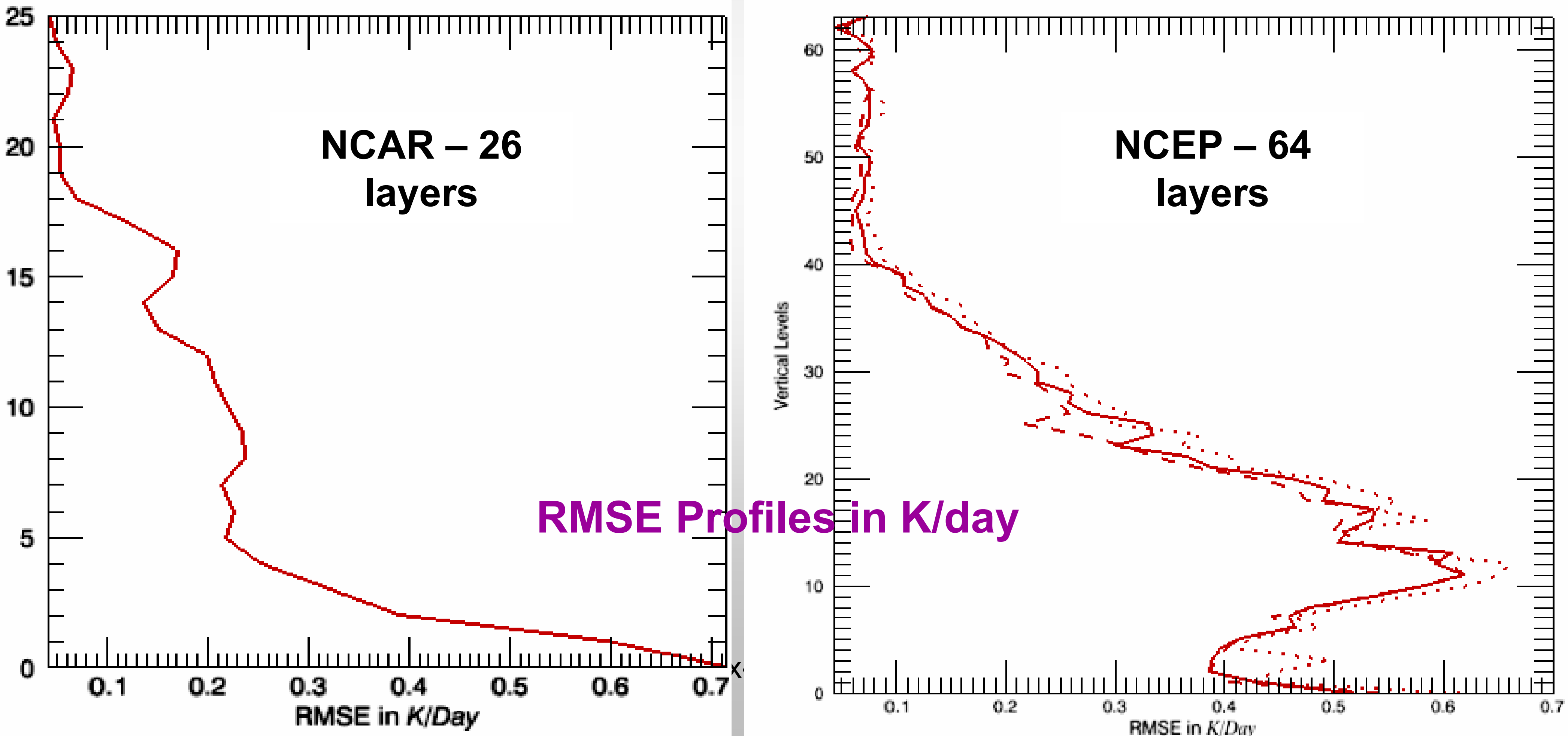
The random part of the emulation error (the standard deviation, SD, of the error) normalized to the maximum member error (the vertical axis) and the systematic error (bias) also normalized to the maximum member error (the horizontal axis). Each ensemble member is represented by a star, the conservative ensemble average – by the cross, and the nonlinear ensemble using the averaging NN by the diamond.

From: ”A Reduced Radiation Grid for the ECMWF Integrated Forecast System” by Morcrette et al., MWR, 2008

A **neural network approach** to the long-wave radiation transfer was also tested in the ECMWF model (Chevallier et al., 2000) and was found to **give adequate results** (sufficient accuracy together with a six-fold decrease in the computer time for the long-wave radiative calculations) at low to medium vertical resolution (up to 50 layers). **At 60 layers and above (the vertical resolution since December 2001), both accuracy and rapidity could not be kept at once** given the increased non-linearity in the lowest and uppermost atmospheric layers. Consequently the neural network approach is used only for the 4D-Var linearised physics (Janiskova et al., 2002) when the accuracy requirements are weaker.

LWR NN Emulation – Approximation Errors in K/Day

	Bias	RMSE	$RMSE_{highest\ lev}$	$RMSE_{lowest\ lev}$
NCAR – 26 vertical levels	$1.\ 10^{-4}$	0.29	0.05	0.85
NCEP – 64 vertical levels	$-5.\ 10^{-4}$	0.32	0.07	0.52



Recent Journal and Conference Papers

Journal Papers:

- Krasnopolsky, V. M., M.S. Fox-Rabinovitz, H.L. Tolman, and A. A. Belochitski, 2008: "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*, 21, 535–543; *doi:10.1016/j.neunet.2007.12.019*
- V. M. Krasnopolsky, M.S. Fox-Rabinovitz, and A. A. Belochitski, 2008: "Decadal Climate Simulations Using Accurate and Fast Neural Network Emulation of Full, Long- and Short Wave, Radiation.", *Monthly Weather Review*, Vol. 136, No. 10.
- Krasnopolsky, V. M., M. S. Fox-Rabinovitz, Y. T. Hou, S. J. Lord, and A. A. Belochitski, 2008: Accurate and Fast Neural Network Emulations of Full, Long- and Short- Wave Radiation for the NCEP Coupled Climate Forecast System Model: Decadal Climate Simulation and Seasonal Forecasting" in preparation.
- V.M. Krasnopolsky, 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
- V.M. Krasnopolsky, 2007: "Reducing Uncertainties in Neural Network Jacobians and Improving Accuracy of Neural Network Emulations with NN Ensemble Approaches", *Neural Networks*, 20, pp. 454-46
- V.M. Krasnopolsky and M.S. Fox-Rabinovitz, 2006: "Complex Hybrid Models Combining Deterministic and Machine Learning Components for Numerical Climate Modeling and Weather Prediction", *Neural Networks*, 19, 122-134
- V.M. Krasnopolsky and M.S. Fox-Rabinovitz, 2006: "A New Synergetic Paradigm in Environmental Numerical Modeling: Hybrid Models Combining Deterministic and Machine Learning Components", *Ecological Modelling*, v. 191, 5-18
- V.M. Krasnopolsky, M.S. Fox-Rabinovitz, and D.V. Chalikov, 2005: "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*, v. 133, No. 5, 1370-1383

Conference Papers;

- V.M. Krasnopolsky, M. S. Fox-Rabinovitz, Y.-T. Hou, S. J. Lord, and A. A. Belochitski, 2007, "Development of Fast and Accurate Neural Network Emulations of Long Wave Radiation for the NCEP Climate Forecast System Model", NOAA 32nd Annual Climate Diagnostics and Prediction Workshop
- V. M. Krasnopolsky, M. S. Fox-Rabinovitz, Y. T. Hou, S. J. Lord, and A. A. Belochitski, 2008: "Accurate and Fast Neural Network Emulations of Long Wave Radiation for the NCEP Climate Forecast System Model", *Proc.of 20th Conference on Climate Variability and Change,* 88th AMS Annual Meeting, New Orleans, LA, 20-24 January 2008, CD-ROM, P3.10
- V. M. Krasnopolsky, M.S. Fox-Rabinovitz, and A. A. Belochitski, 2008: "Ensembles of Numerical Climate and Weather Prediction Models Using Neural Network Emulations of Model Physics", Proc. of the 2008 IEEE World Congress on Computational Intelligence, Hong Kong, June 1-6, 2008, CD-ROM, paper NN0498, pp. 1524-1531
- *Krasnopolsky, V. M., M. S. Fox-Rabinovitz, Y. T. Hou, S. J. Lord, and A. A. Belochitski, 2009: Fast Neural Network Emulations of Long Wave Radiation for the NCEP Climate Forecast System Model: Seasonal Prediction and Climate Simulation, 89th AMS Annual Meeting, Phoenix, AZ, 11-15 January, 2009, accepted.*
- M. Fox-Rabinovitz, V. Krasnopolsky, and A. Belochitski, 2006: "Ensemble of Neural Network Emulations for Climate Model Physics: The Impact on Climate Simulations", *Proc., 2006 International Joint Conference on Neural Networks,* Vancouver, BC, Canada, July 16-21, 2006, pp. 9321-9326, CD-ROM

Conclusions - 1

- **NN is a powerful tool for speeding up calculations of model physics or for developing NN emulations**
  - Accurate and fast NNs emulations have been successfully developed for:
    - NCAR CAM LWR & SWR parameterizations
    - NASA LWR parameterization
    - NCEP LWR & SWR parameterizations (work in progress)
  - NN emulations of the NCAR CAM LWR & SWR are 150 to 20 times faster and very close to the original parameterizations.
  - The simulated diagnostic and prognostic fields are very close for the parallel NCAR CAM climate runs with NN emulations and the original parameterizations

Conclusions-1 (cont.): NCEP CFS

- Developed NN emulations for the NCEP CFS model show ***high accuracy and computational efficiency***
- Validation of NN emulations through 10-year and seasonal NCEP CFS model runs using the NN emulations vs. the control model run show a ***close similarity*** of the runs, namely the differences are mostly within observational errors or the ***uncertainty of observational data or reanalyses***
- NN emulations are robust: can be used for different horizontal resolutions
- Increase of tropical inputs/outputs to be used for NN training has to be considered
- Potential applications to NCEP GFS for NWP & DAS

Conclusions - 2

- **NN is a promising tool for improving model physics or developing new NN parameterizations**
  - With an extremely simple characterization of the atmospheric state, useful information about the CRM model behavior can be produced (i.e. we can emulate CRM convective and cloud physics) at the GCM scales
  - This is the first step in the construction of a NN convective parameterization.

Conclusions - 3

- Potential challenges can be successfully resolved exploiting the tremendous flexibility of NN techniques:
  - A higher accuracy of approximation can always be achieved by the expense of a speed up because the later will remain significant anyway
  - NNs can be used for developing Compound Parameterizations with sophisticated QC procedures
  - NN can be dynamically adjusted to climate changes
  - NN ensembles can be efficiently used
- Alternative Statistical (or Machine) Learning Techniques (like Support Vector Machines or other Kernel Methods) are being explored

Continuum (seamless) modeling using NNs

- The need for continuous multi-scale interaction for global modeling: from short term forecasting to climate simulation and projection
- Tropical convection and its impact on extratropical flow or tropics-extratropics exchange
- Micro – meso – synoptic – large scale interaction: upscaling and downscaling

The Magic of NN Performance

$Y = F(X)$

$Y_{NN} = F_{NN}(X)$

- **OP Numerical Performance is Determined by:**
  - **Numerical Complexity (NC) of OP**
    - **Complexity of OP Mathematics**
      - **Complexity of Physical Processes**

- **NN Emulation Numerical Performance is Determined by:**
  - **NC of NN emulation**
    - **Functional Complexity (FC) of OP, i.e. Complexity of I/O Relationship: *Y = F(X)***

- Explanation of the Magic of NN Performance:
  - **Usually, FC of OP << NC of OP**

    AS A RESULT
  - **NC of NN Emulation ~ FC of OP**

    and

    NC of NN Emulation << NC of OP

Our NN vs. NeuroFlux

- We emulate the mapping (1) by <u>a single NN</u>, *N*, with *n* inputs and *m* outputs:

  $\mathbf{Y} = N(\mathbf{X}),\ \mathbf{X} \in \Re^n,\ \mathbf{Y} \in \Re^m$.

Following the notation C98, our NN emulation is represented as [in this particular case, the input vector **X =** (**S***,* **T***,* **V***,* **C**)],

$\mathbf{Y} = N(\mathbf{S}, \mathbf{T}, \mathbf{V}, \mathbf{C})$

where the vector **S** is representing surface variables, **T** is a vector (profile) of atmospheric temperatures, **C** is a profile of cloud variables, and vector **V** includes all other variables (humidity profile, different gas mixing ratio profiles, etc.). The output of our NN emulation, the vector **Y**, is identical to the output of the original parameterization. It is composed of two vectors **Q** and **f**, **Y =** (**Q***,* **f**). Here **Q** is a profile of cooling rates, **Q =** (*Cr=1*, *Cr-2, . . . , Cr-L)*, where *Cr-j* is the cooling rate at the *j*th vertical level, and *f* is a vector of auxiliary fluxes computed by the LWR parameterization.

The NeuroFlux LWR parameterization developed by C98 is based on the Washington and Williamson (1977) approach, which allows us to <u>separate cloud Variables</u> **C**. In the NeuroFlux parameterization level fluxes are calculated as

$F(\mathbf{S}, \mathbf{T}, \mathbf{V}, \mathbf{C}) = \Sigma\, \mathbf{A}(\mathbf{C})\, F(\mathbf{S}, \mathbf{T}, \mathbf{V})$

and NN approximations are built for each partial or individual flux $F_i(\mathbf{S}, \mathbf{T}, \mathbf{V})$. As a result, <u>the flux at each level is a linear combination of approximating NNs</u>. To calculate NeuroFlux outputs, namely, the cooling rates *Cr*'s, the linear combinations of individual approximating NNs, *F*, are differentiated at each vertical level

$Cr(P) = dF(P)/dP$

where *P* is the atmospheric pressure.

Our NN vs. NeuroFlux …

- Summarizing the differences between our approaches, we can point out at least three major differences:
- 1) *The* C98 *approach and the* KFC05 *approach are conceptually different.*

The C98 approach closely follows the internal structure of the corresponding LWR parameterization and uses its internal properties.

Our approach treats an entire parameterization as an elementary/single object and emulates its functionality (input–output relationship) as a whole. This conceptual difference between the approaches leads to the following important practical differences.

- 2) The C98 approach is based on an internal property of the ECMWF LWR parameterization, namely separation of variables that limits the applicability of the C98 approach only to the LWR parameterizations, which are formulated using such a separation of variables. Our approach does not rely on or require any separation properties.
- 3) The NeuroFlux numerical design follows the internal structure of the ECMWF LWR

parameterization. It is based on the development of a set or a "battery" of NNs (C98) to approximate partial fluxes *Fi* with a following linear combination of these fluxes (for the clear-sky case the linear combination is reduced to a single NN) and differentiation of the combination to calculate *Cr*.

Our NN vs. NeuroFlux …

- In our approach, the LWR parameterization outputs, including *Cr*, are the direct outputs of a single NN, which emulates the entire LWR parameterization. In the C98 approach multiple NNs are trained to approximate continuous partial fluxes *Fi*. In our approach a single NN is trained to emulate the entire LWR parameterization, which is an almost continuous mapping (may contain stepfunction-type discontinuities). As a result, in our approach a single NN is presented with a more complicated task than each individual NN in the C98 approach and it has higher number of hidden neurons than each individual NN (but significantly smaller number of hidden neurons than all of them totally) in the C98 approach. The training of such a large single NN may present greater challenges than training of each individual NN in the C98 approach.
- Let us discuss some consequences of these major differences between our approaches. Because our approach does not follow the internal structure of the original parameterization and does not use any additional assumptions about the internal properties of the original parameterization, it is generic and portable (can be applied to other parameterizations and models).

Our NN vs. NeuroFlux …

- It is also important that *the difference between the approaches* mentioned above, namely approximating the entire parameterization with a single NN in our approach versus using multiple NNs for approximating multiple internal components of the LWR parameterization in C98, results in different numerical procedures for calculating the parameterization outputs and parameterization Jacobian. Generally, the parameterization Jacobian in the case of the LWR parameterization is composed of first derivatives of the cooling rates, *Cr*. In the C98 approach, calculation of the first derivatives of the cooling rates, which themselves are proportional to the first derivatives of the individual NNs, leads to calculating linear combinations (for the clear-sky case the linear combination is reduced to a single NN) of *second* derivatives of approximating NNs. In our approach, the parameterization output is the output of a single NN emulator and the parameterization Jacobian is composed of *first* derivatives of a single emulating NN.