Acceleration of Calculations and Improvement of Model Physics Using Neural Networks

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

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

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 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 (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
  - 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
- **Started Transfer of Developed Methodology to NCEP CFS (NN emulations for the radiation block)**

Recent Papers Published:

- **V.M. Krasnopolsky, M.S. Fox-Rabinovitz, and A. Beloshitski, 2007, "Compound Parameterization for a Quality Control of Outliers and Larger Errors in NN Emulations of Model Physics", *Neural Networks*, submitted**

- **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*, in press**

- **V.M. Krasnopolsky, 2007: "Reducing Uncertainties in Neural Network Jacobians and Improving Accuracy of Neural Network Emulations with NN Ensemble Approaches", *Neural Networks*, in press**

- **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**

- **H. Tolman, V. M. Krasnopolsky, and D. Chalikov, 2005: "Neural Network Approximations for Nonlinear Interactions in Wind Wave Spectra: Direct Mapping for Wind Seas in Deep Water", *Ocean Modelling*, v. 8, issue 3, 253-278**

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*,

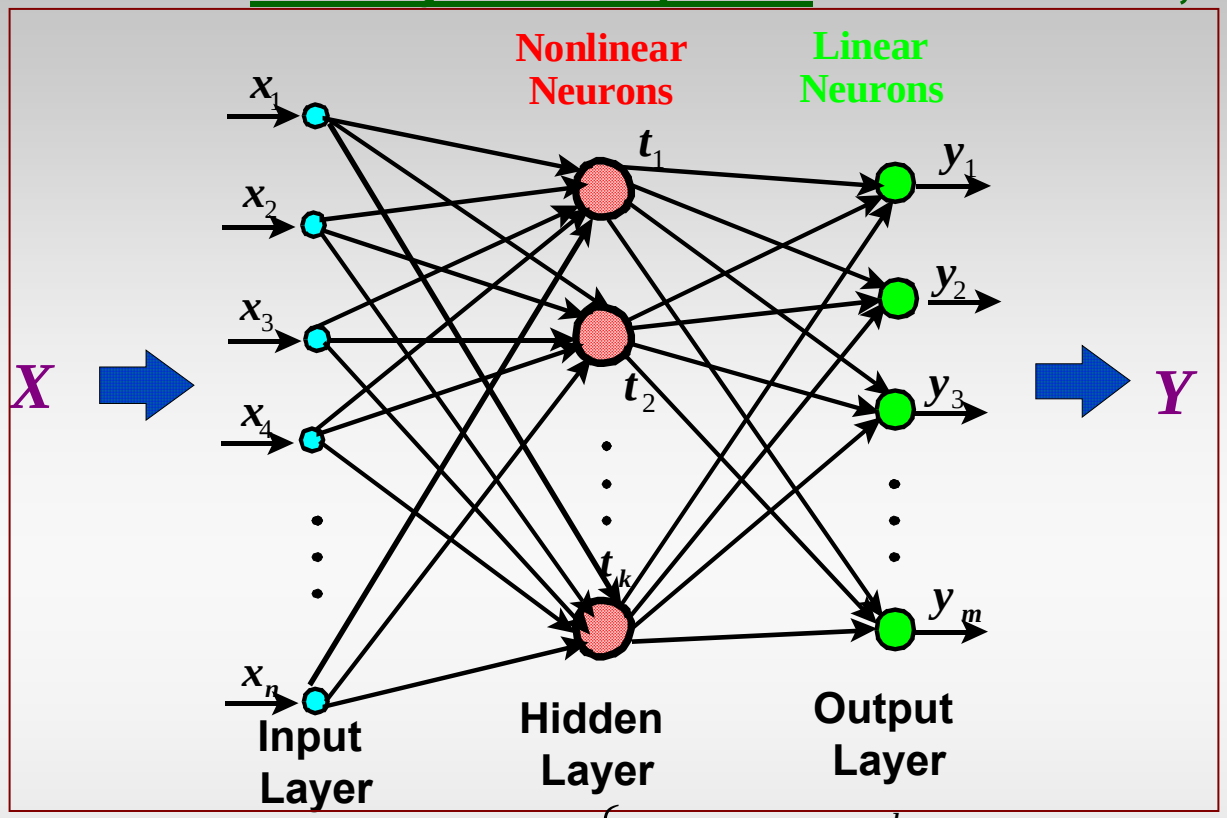
$$Y = F(X); \quad X \in \Re^n \; and \; Y \in \Re^m$$

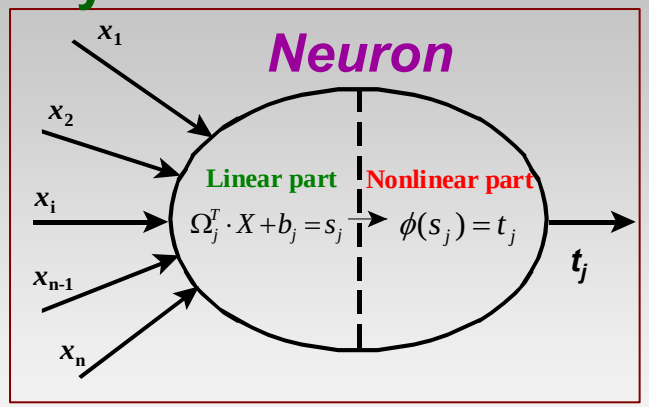
- NN is a generic approximation for **any** continuous or almost continuous mapping given by a set of its input/output records:

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

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:

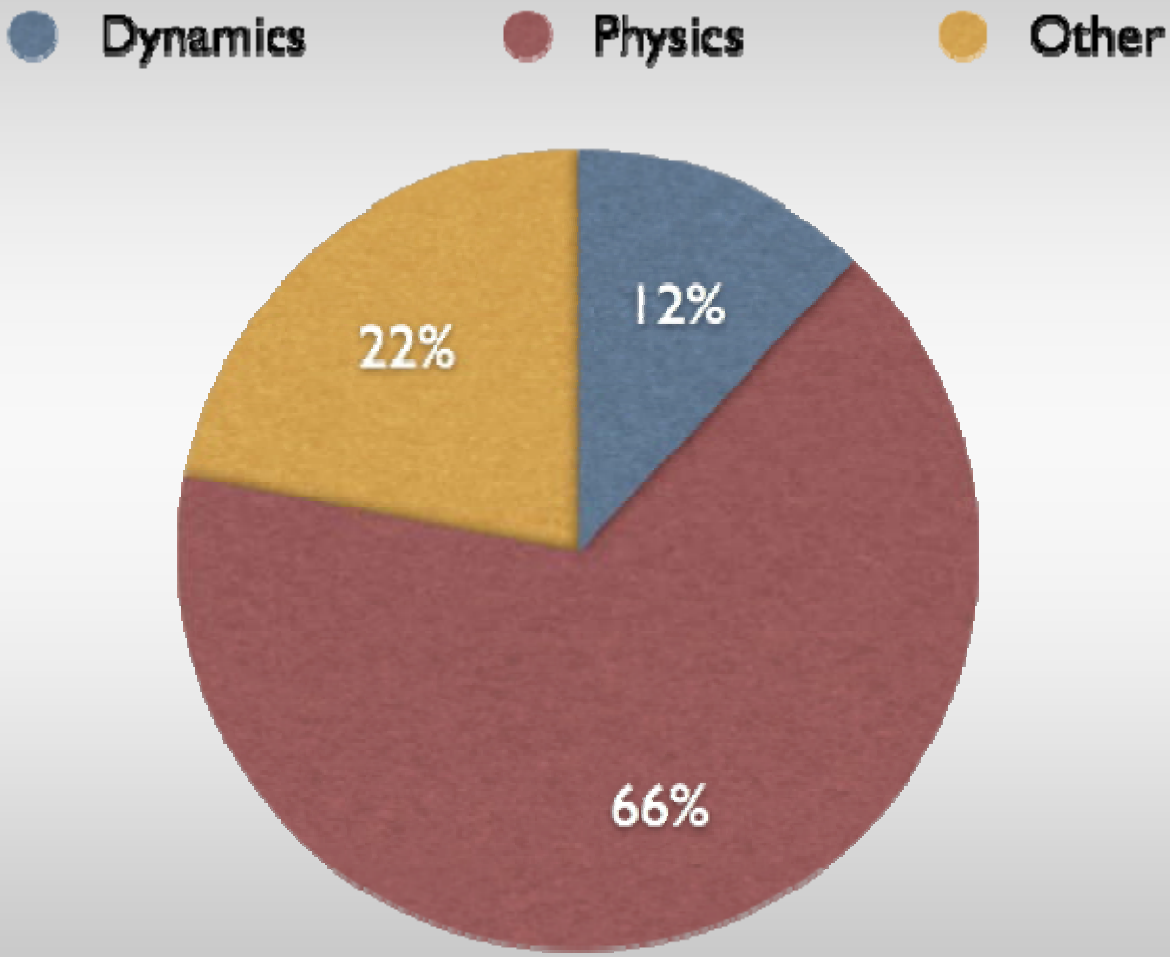
- NNs are *generic,* very *accurate and convenient* mathematical (statistical) models which are able to emulate numerical model components, which are complicated nonlinear input/output relationships (continuous or almost continuous mappings).
- NNs are ***robust*** with respect to random noise and fault-tolerant.
- NNs are *analytically differentiable* (training, error and sensitivity analyses): almost free Jacobian!
- NNs emulations are **accurate and fast but <u>NO FREE LUNCH</u>!**
  - Training is complicated and time consuming nonlinear optimization task; *<u>however, training should be done only once for a particular application!</u>*
- 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!

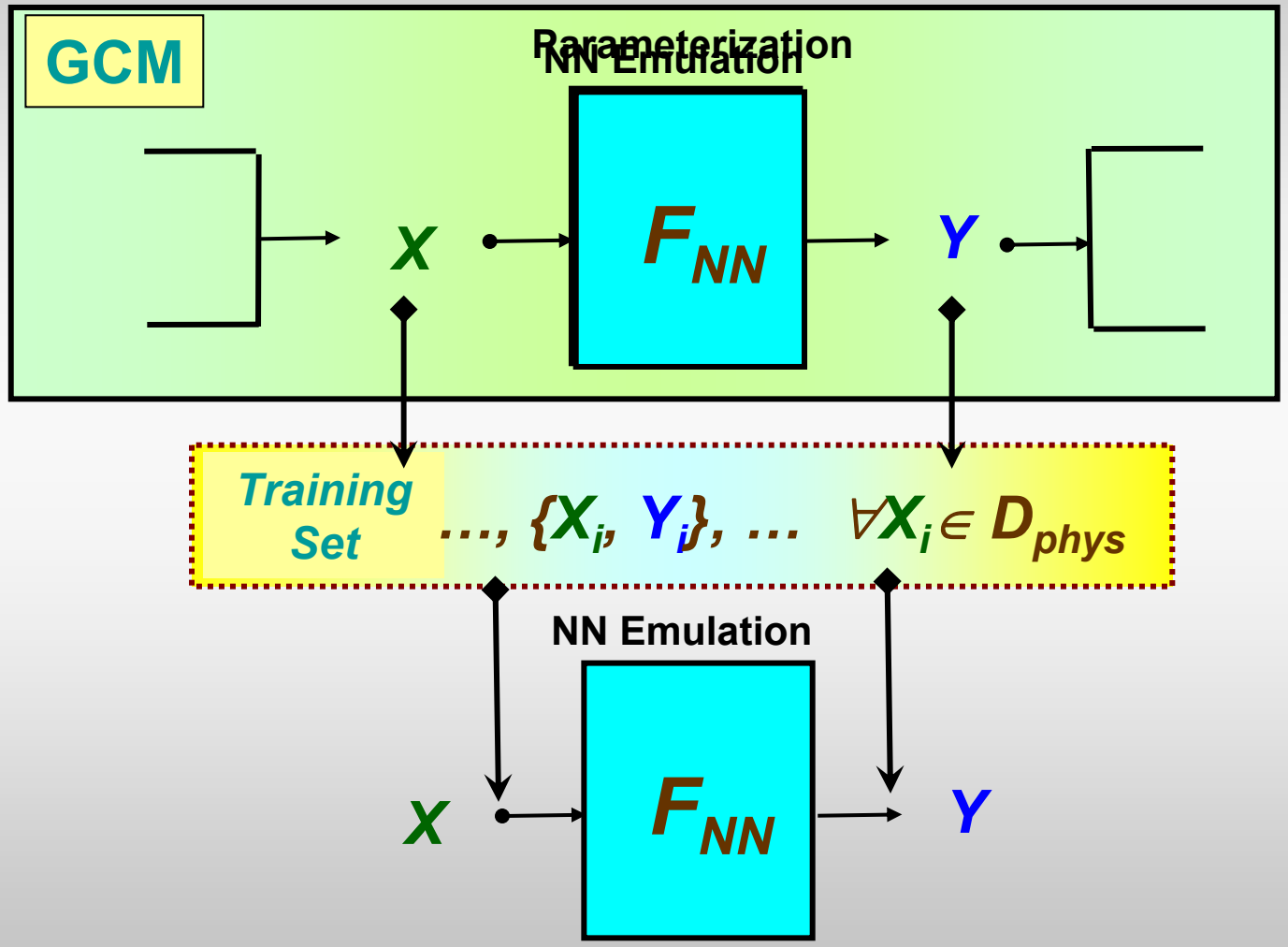
Distribution of NCAR CAM Calculation Time

NCAR CAM 2.0 T42L26

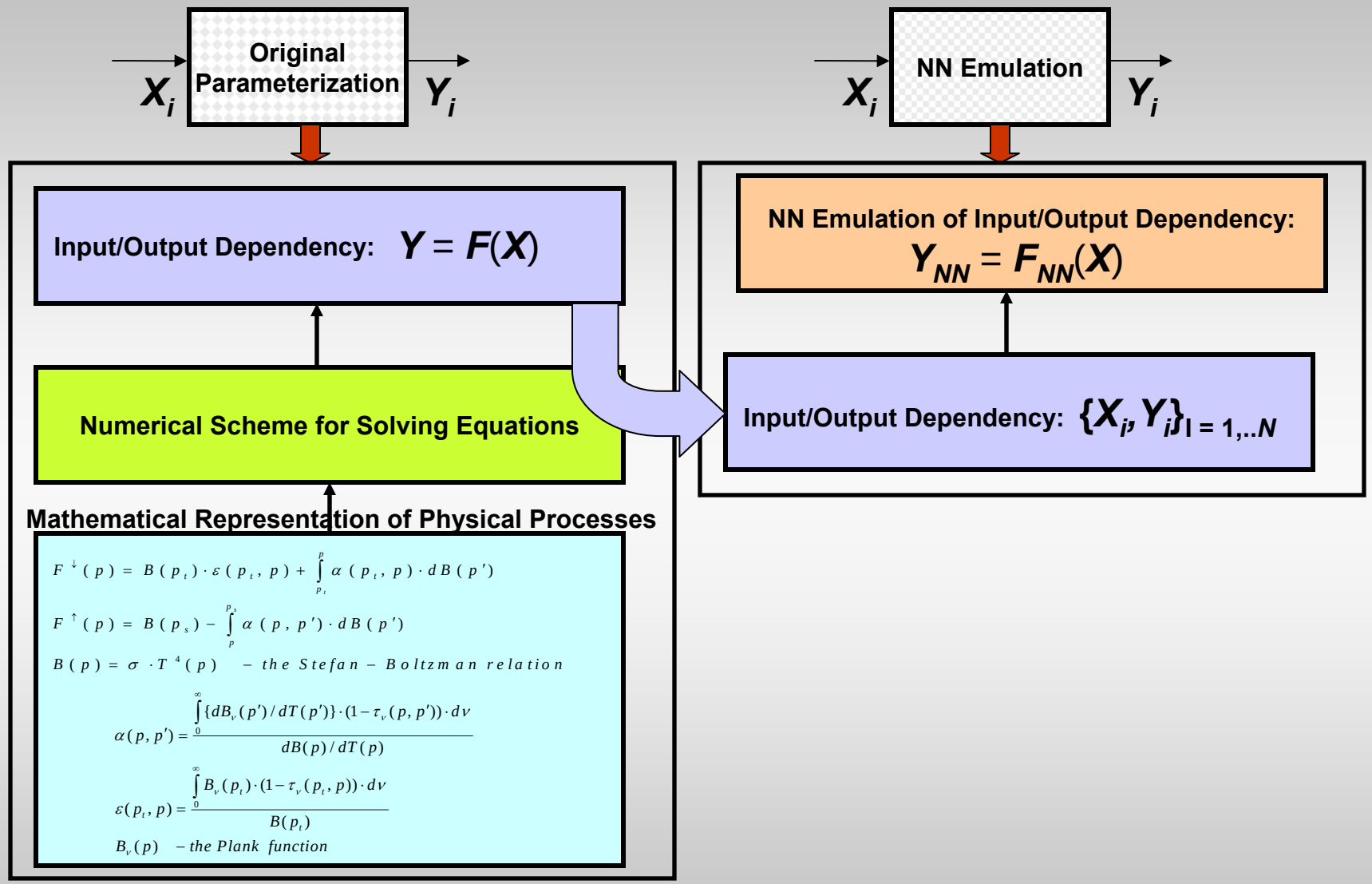


NN Emulations of Model Physics Parameterizations

Learning from Data



The Magic of NN Performance



NN for NCAR CAM Physics

CAM Long Wave Radiation

- Long Wave Radiative Transfer*:*

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

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

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

- Absorptivity & Emissivity (optical properties):

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

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

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

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

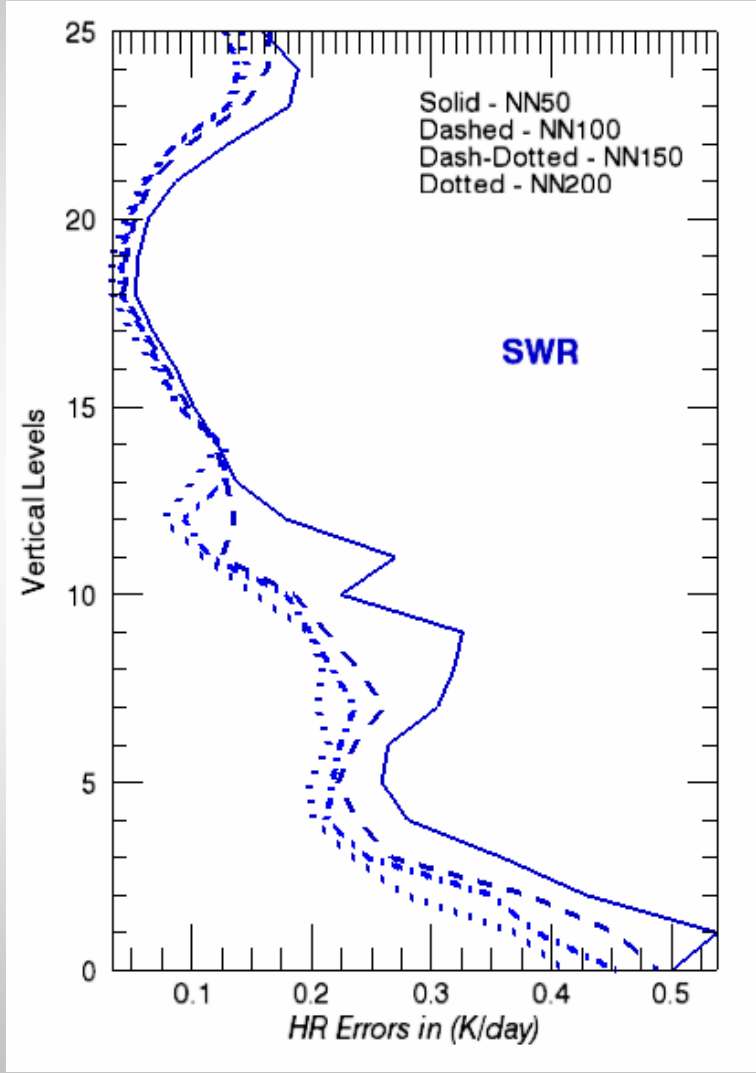
- 451 Inputs:
  - *21 Profiles:* **specific humidity, ozone concentration, pressure, cloudiness, aerosol mass mixing ratios, etc**
  - *7 Relevant surface characteristics*
- 33 Outputs:
  - **Profile of heating rates (26)**
  - **7 LW radiation fluxes: *fsns, fsnt, fsdc, sols, soll, solsd, solld***
- 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*)

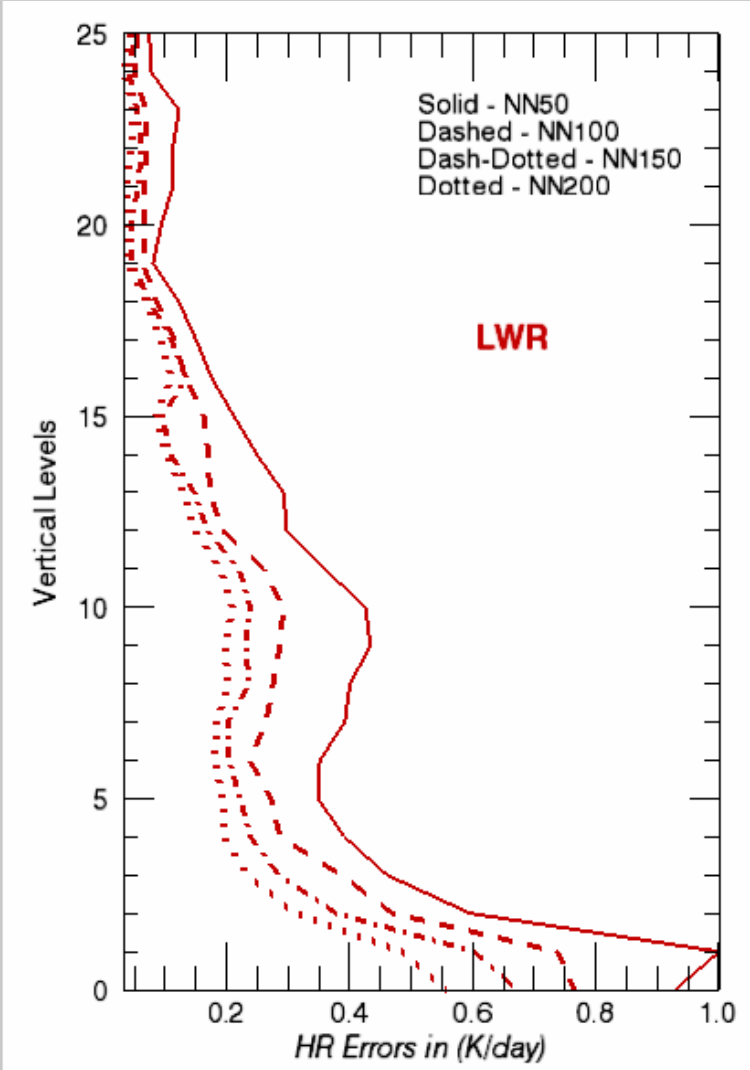
Parameter	Model	Bias	RMSE	Mean	σ	Performance
LWR (°*K/day*) NN150	NASA **M-D. Chou**	$1.\ 10^{-4}$	0.32	-1.52	1.46	
	NCEP **AER *rrtm***	$7.\ 10^{-5}$	0.40	-1.88	2.28	
	NCAR **W.D. Collins**	$3.\ 10^{-5}$	0.28	-1.40	1.98	~ 150 times faster
SWR (°*K/day*) NN150	NCAR	$6.\ 10^{-4}$	0.19	1.47	1.89	~ 20 times faster

Error Vertical Variability Profiles

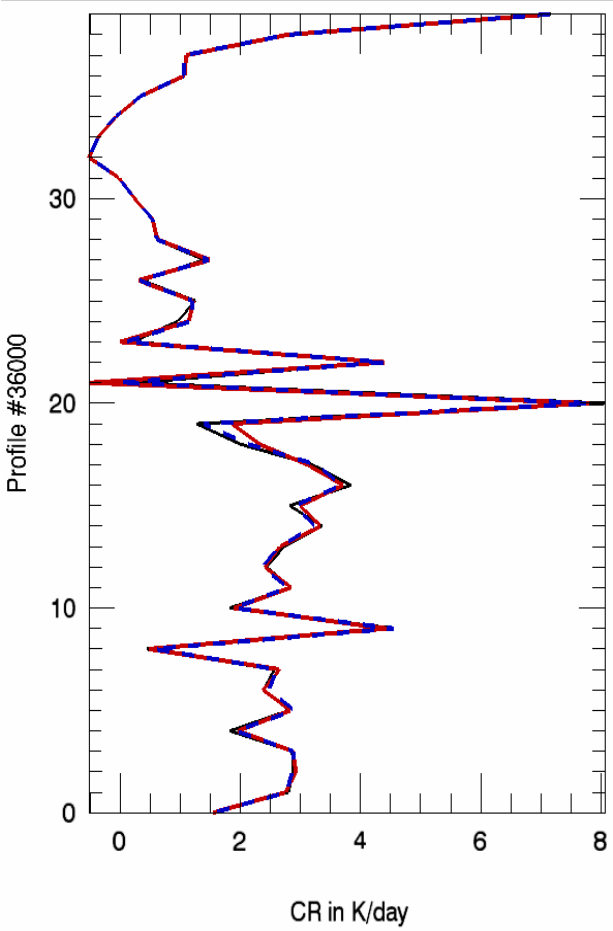
RMSE profiles in K/day

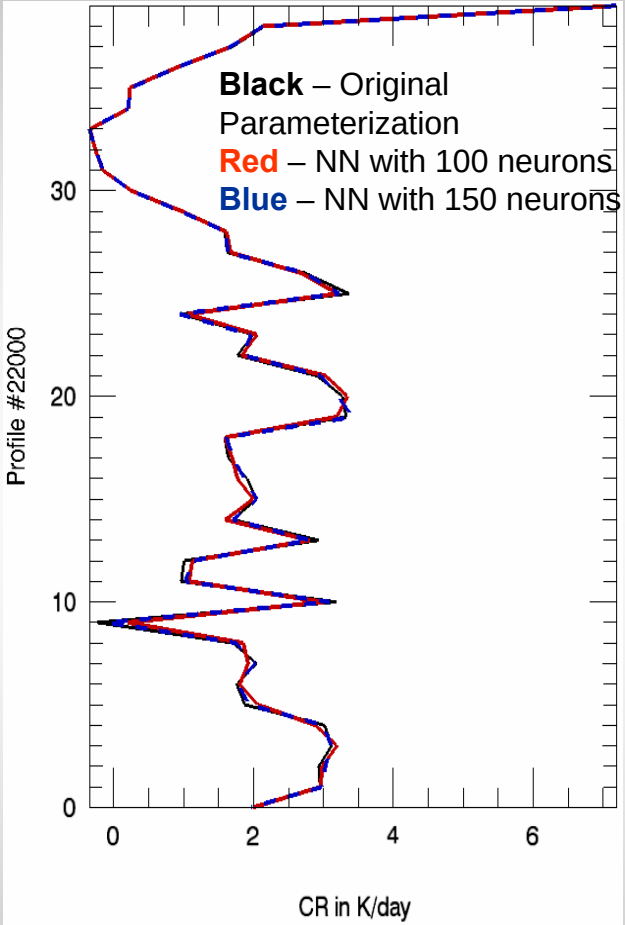


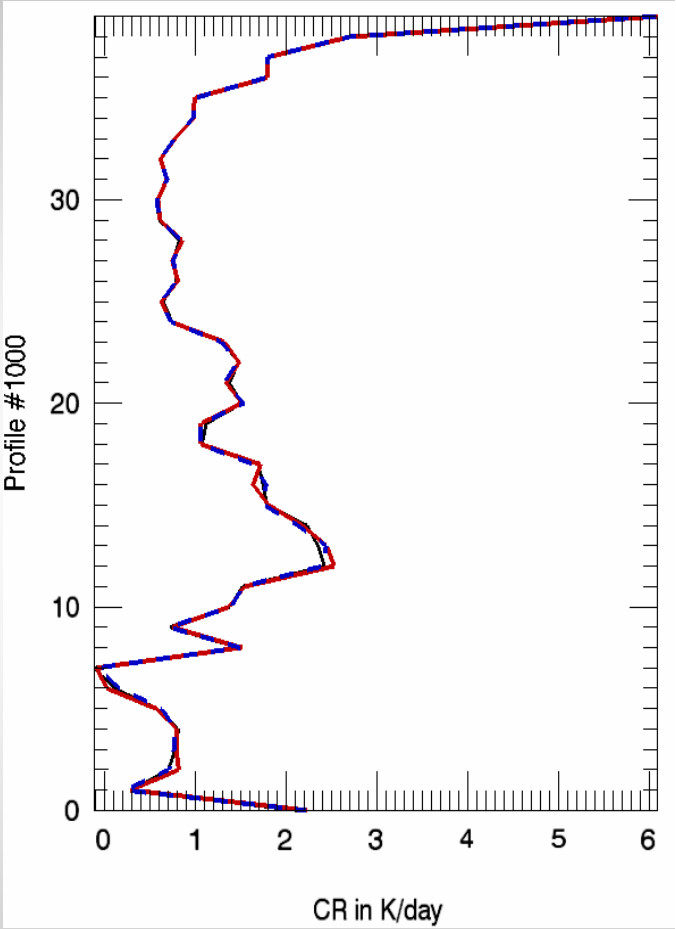
RMSE Profiles in K/day



Individual Profiles

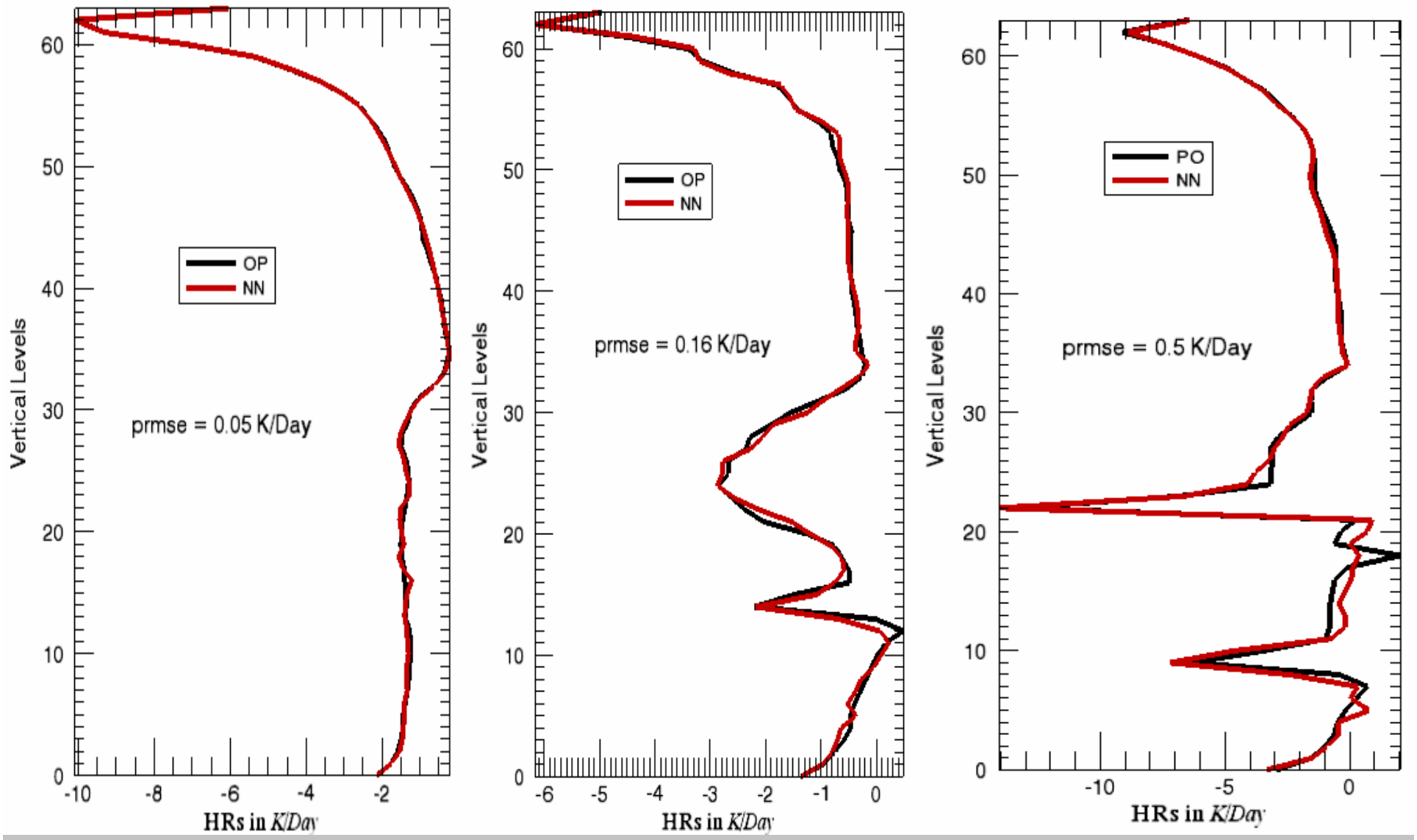






PRMSE = 0.18 & 0.10 K/day **PRMSE = 0.11 & 0.06 K/day** **PRMSE = 0.05 & 0.04 K/day**

Individual Profiles (NCEP)



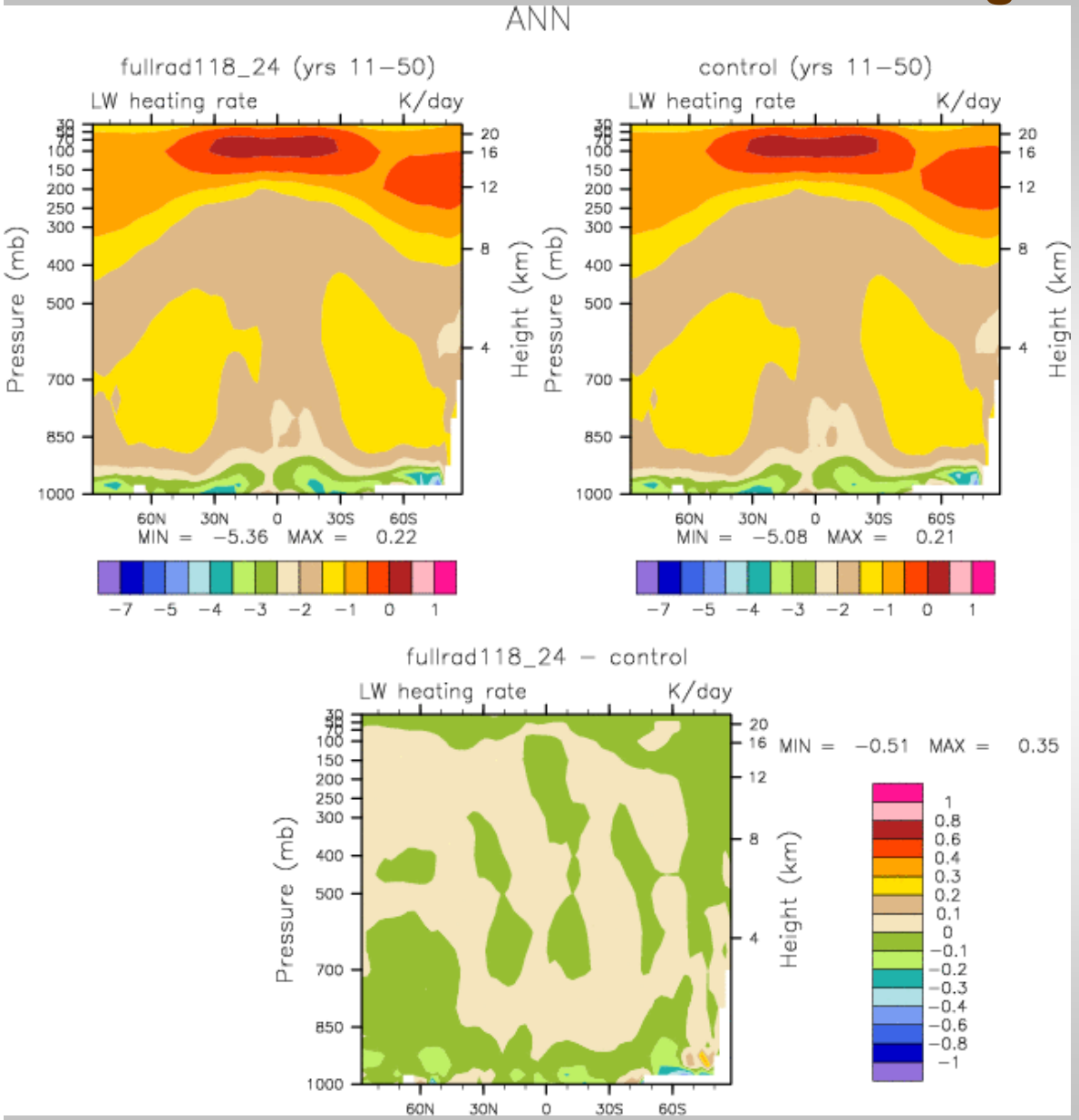
Hybrid Numerical Models (HNM)

- 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

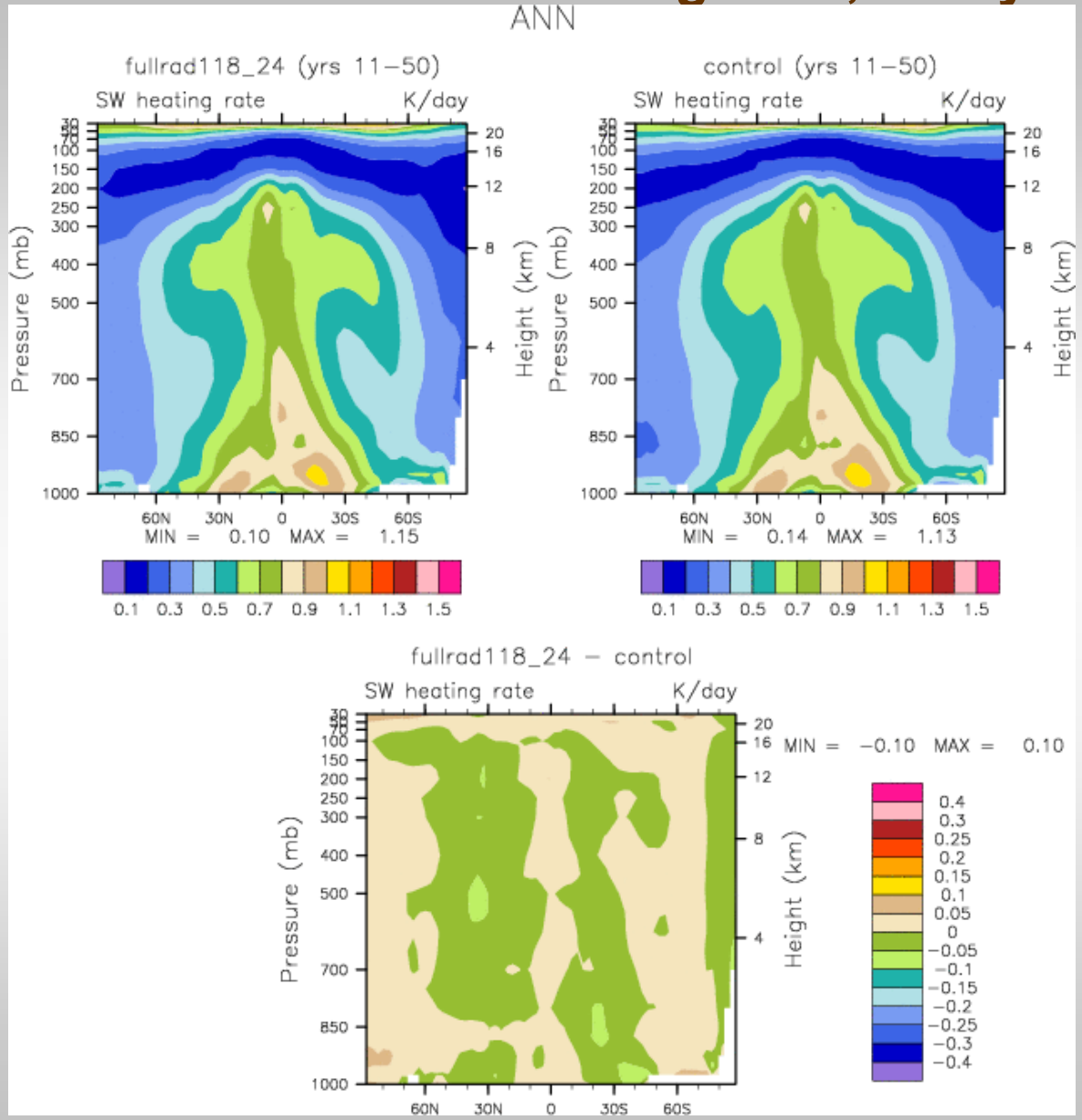
- The control NCAR CAM with the original LWR and SWR parameterizations is run for 40 years (in the 50 year run first the 10 years are skipped to account for spin-up effects).
- The Hybrid NCAR CAM with LWR and SWR NN emulations is run for 40 years.
- The validation of the Hybrid NCAR CAM with NN emulation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

Zonal Mean LW Heating Rate, K/day.

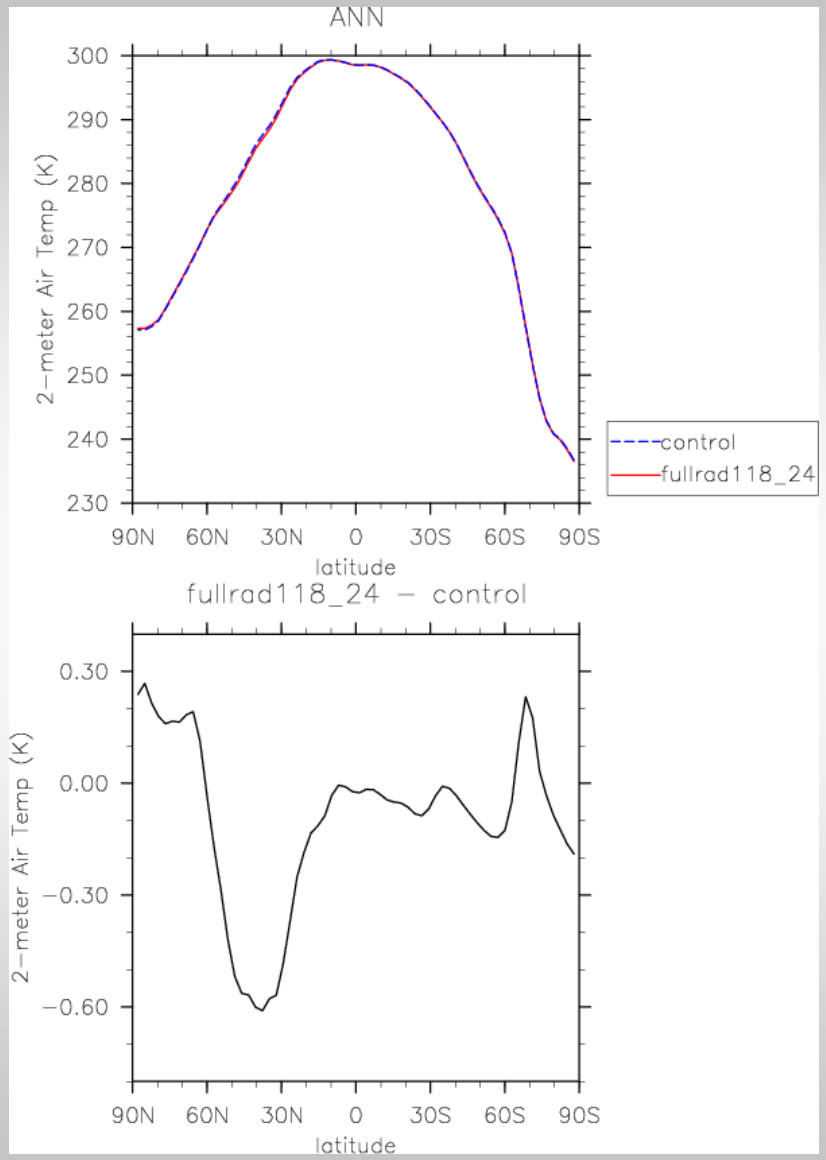


Note: light green and light beige colors correspond to the near 0 values!

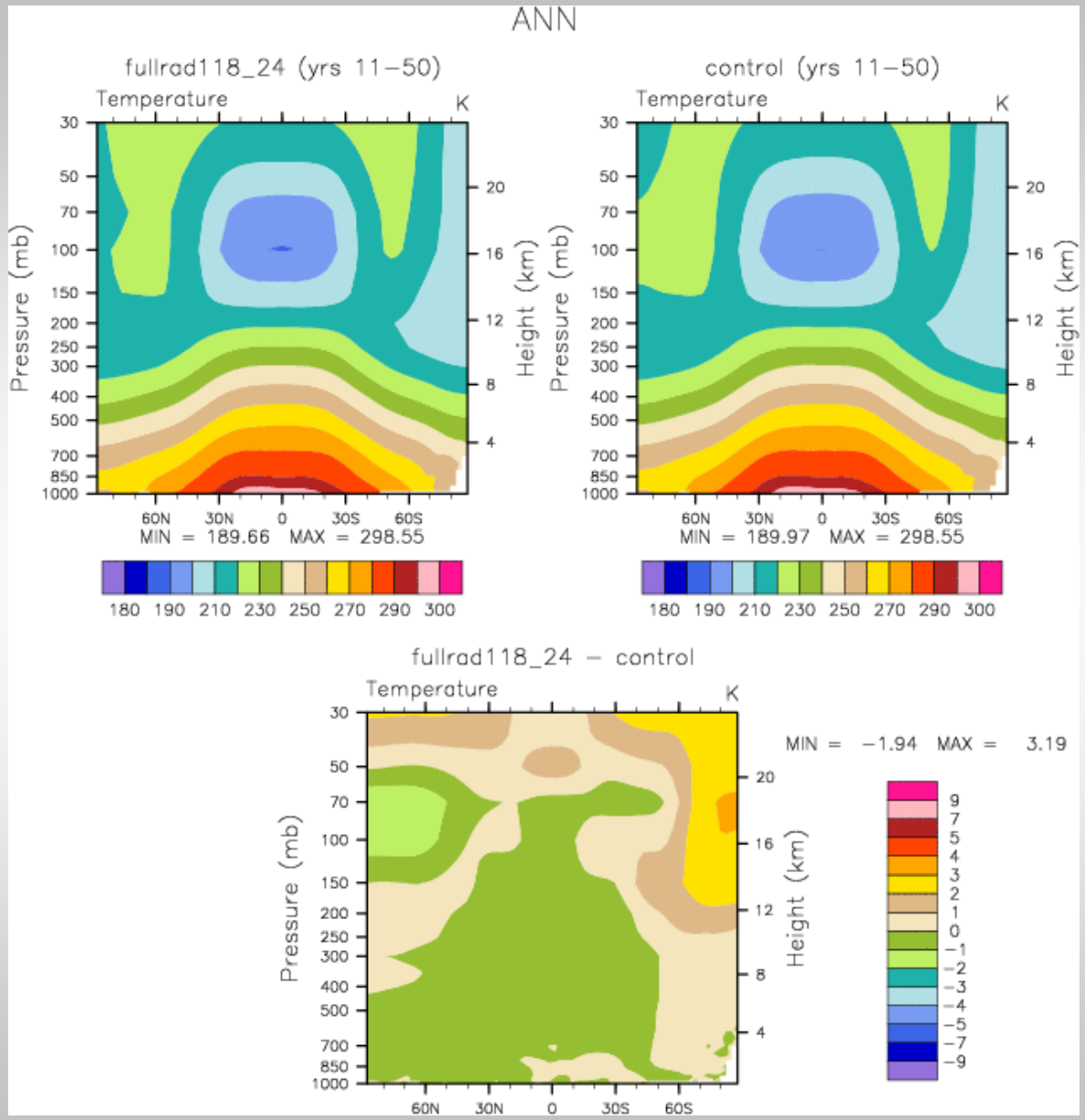
Zonal Mean SW Heating Rate, K/day.



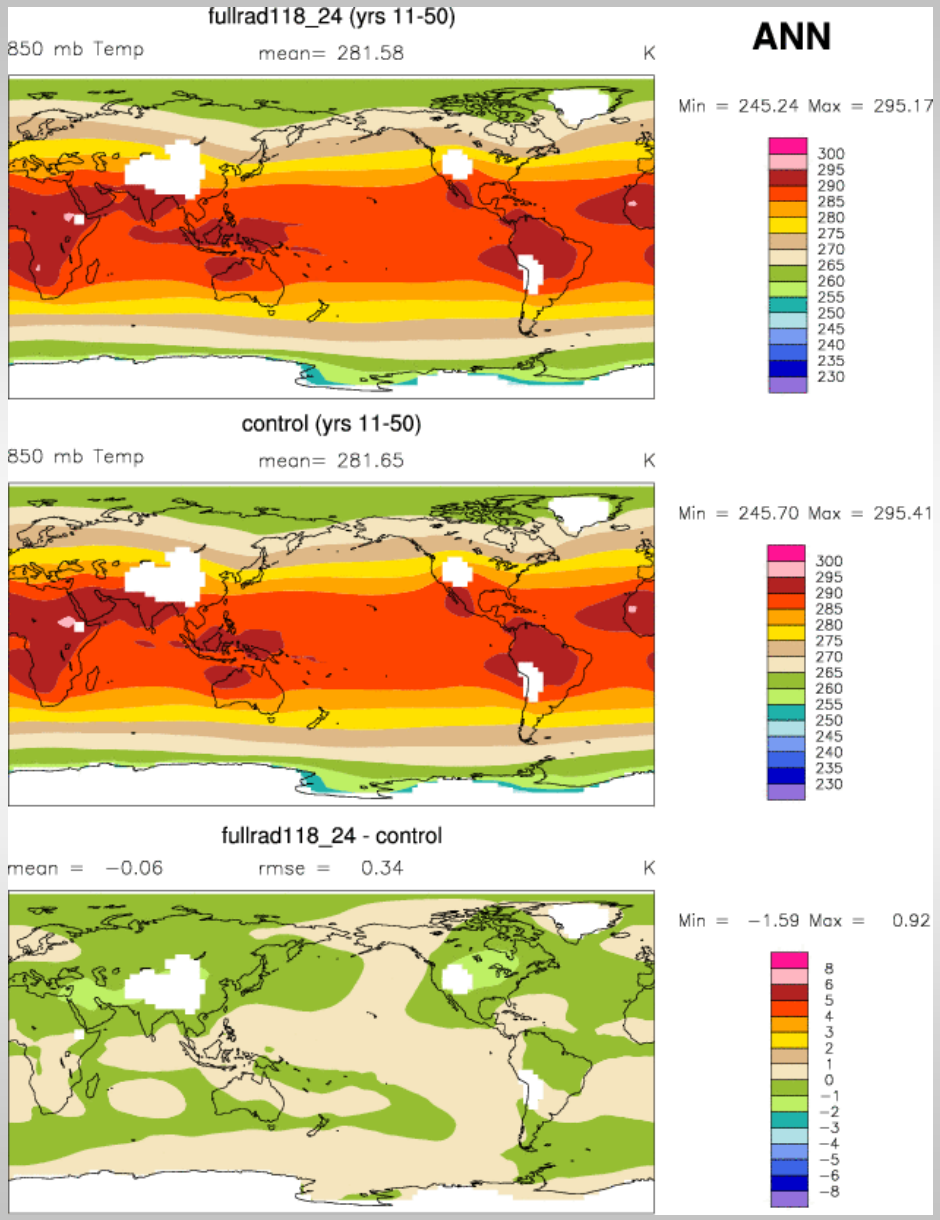
Zonal Mean 2-meter T, K.



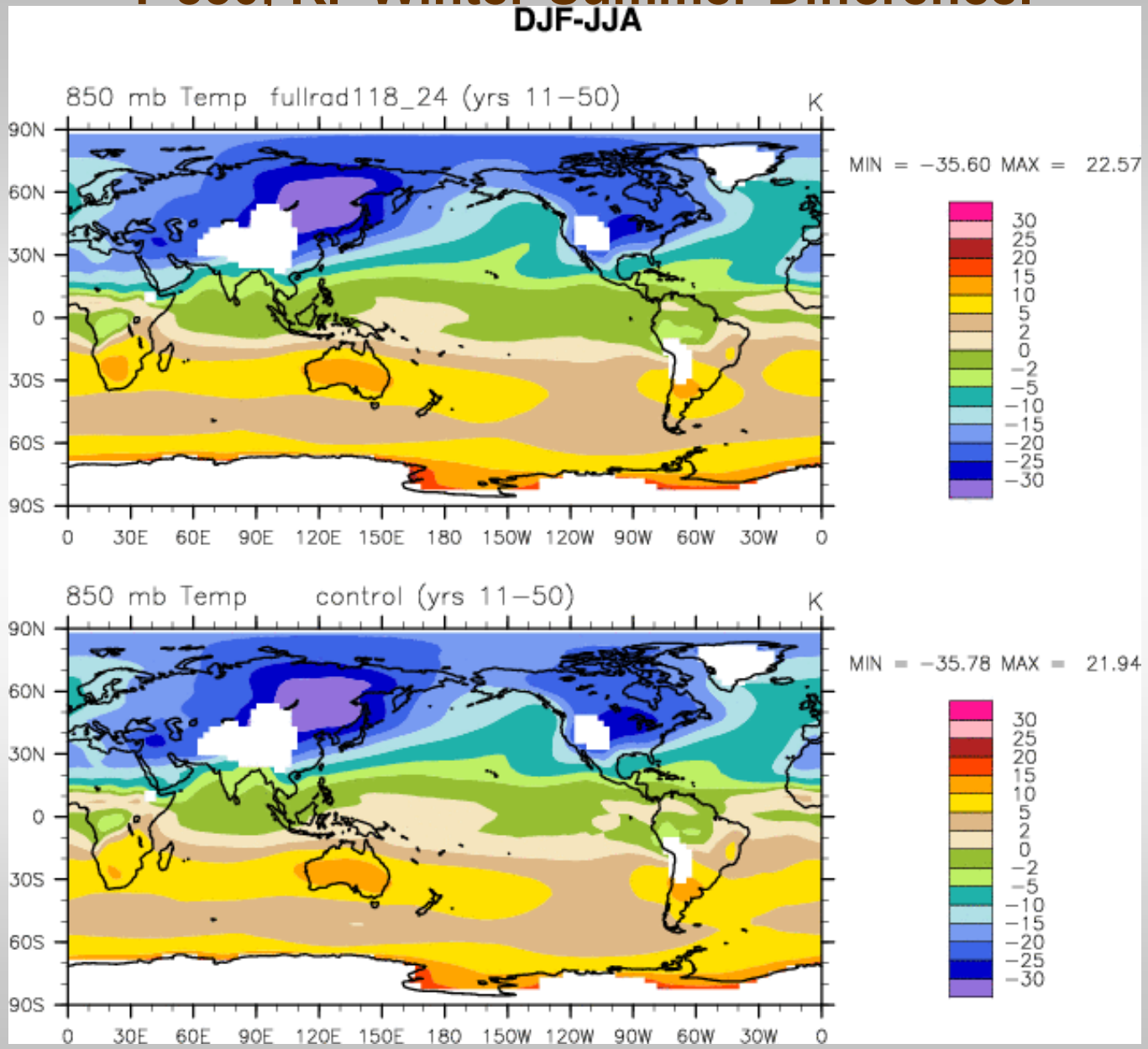
Zonal Mean T, K.



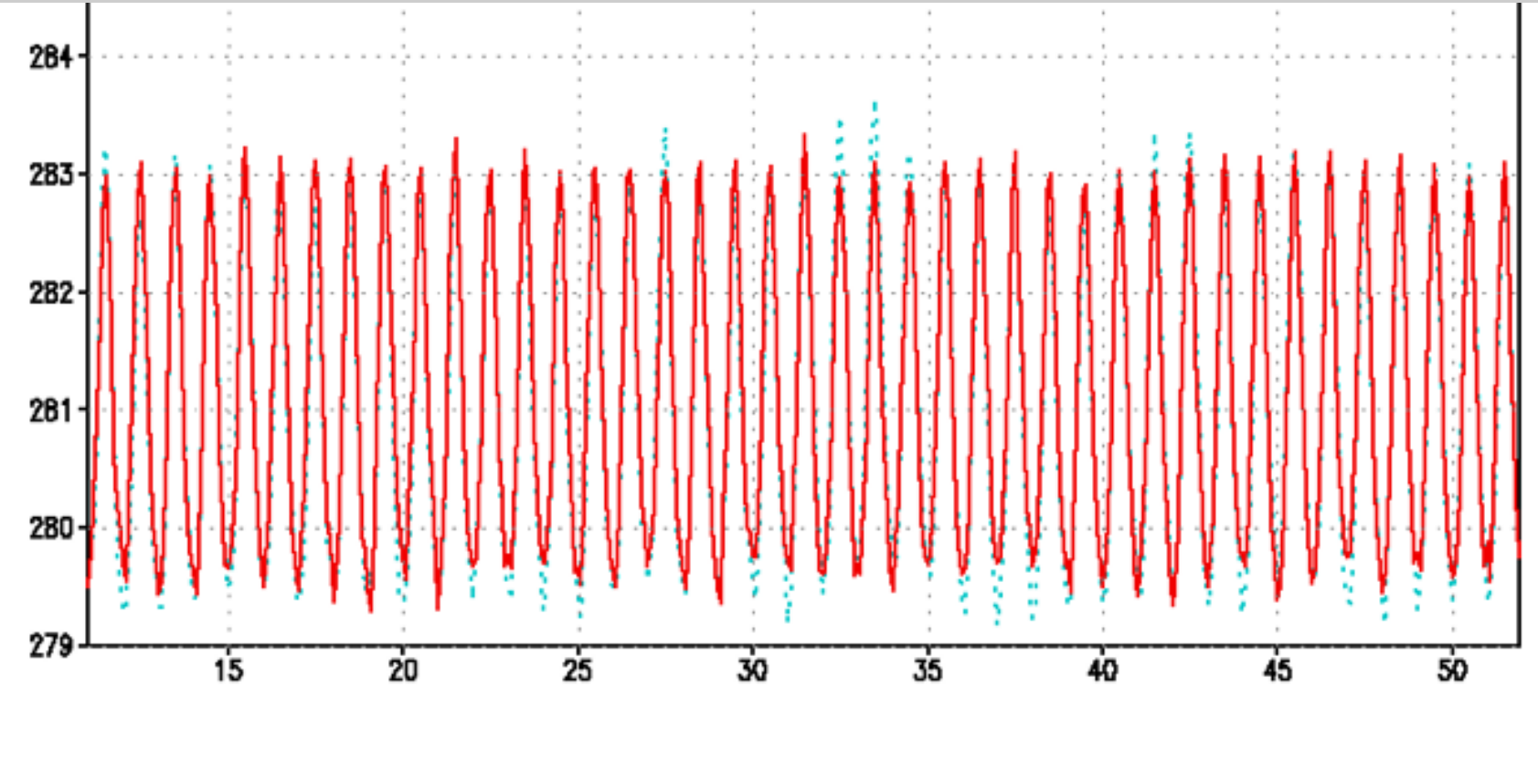
T-850, K.


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, K. Winter-Summer Difference.

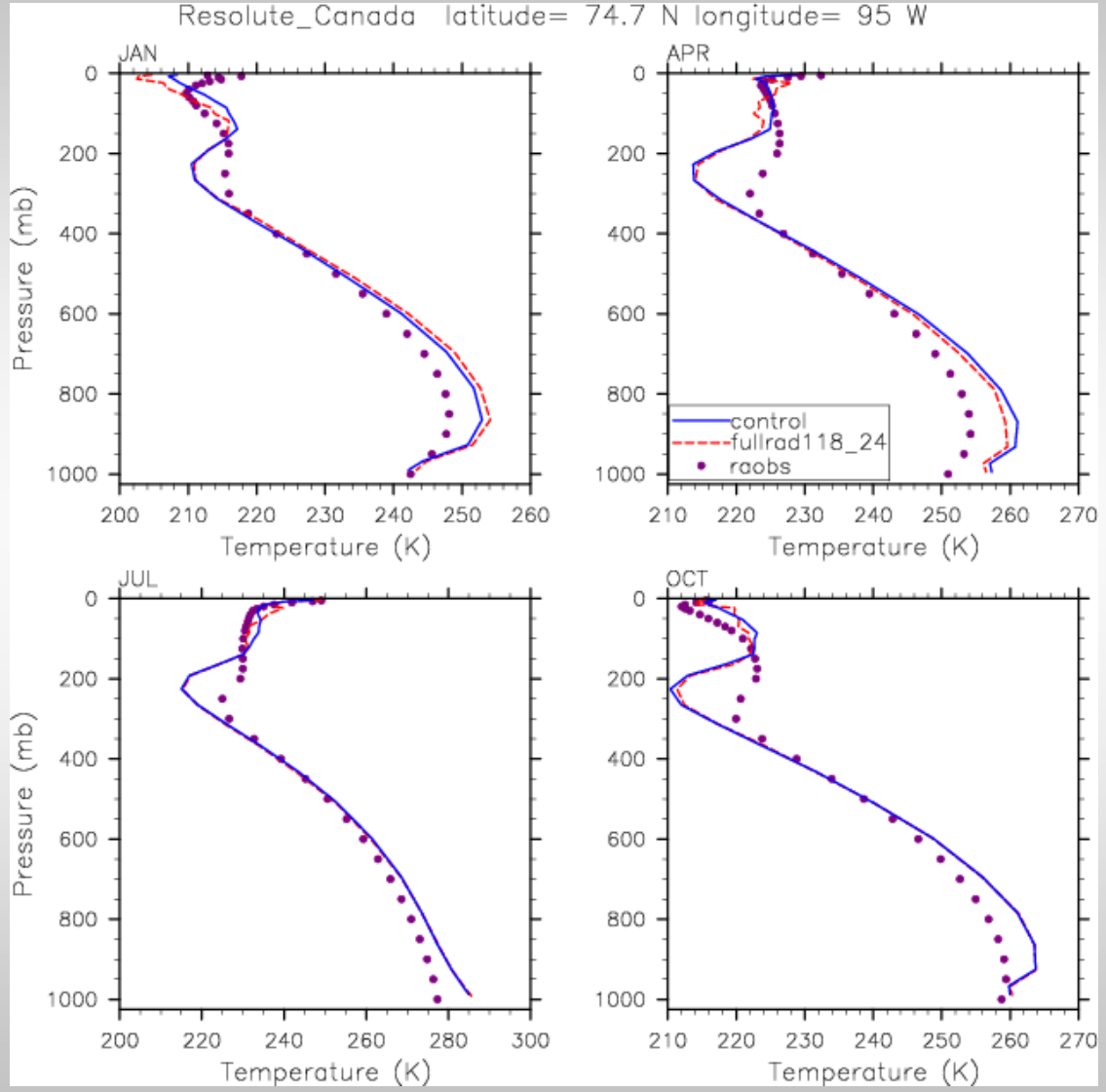


T-850, K. Global Time Series.



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

Vertical Profile of T (Resolute, Canada), K

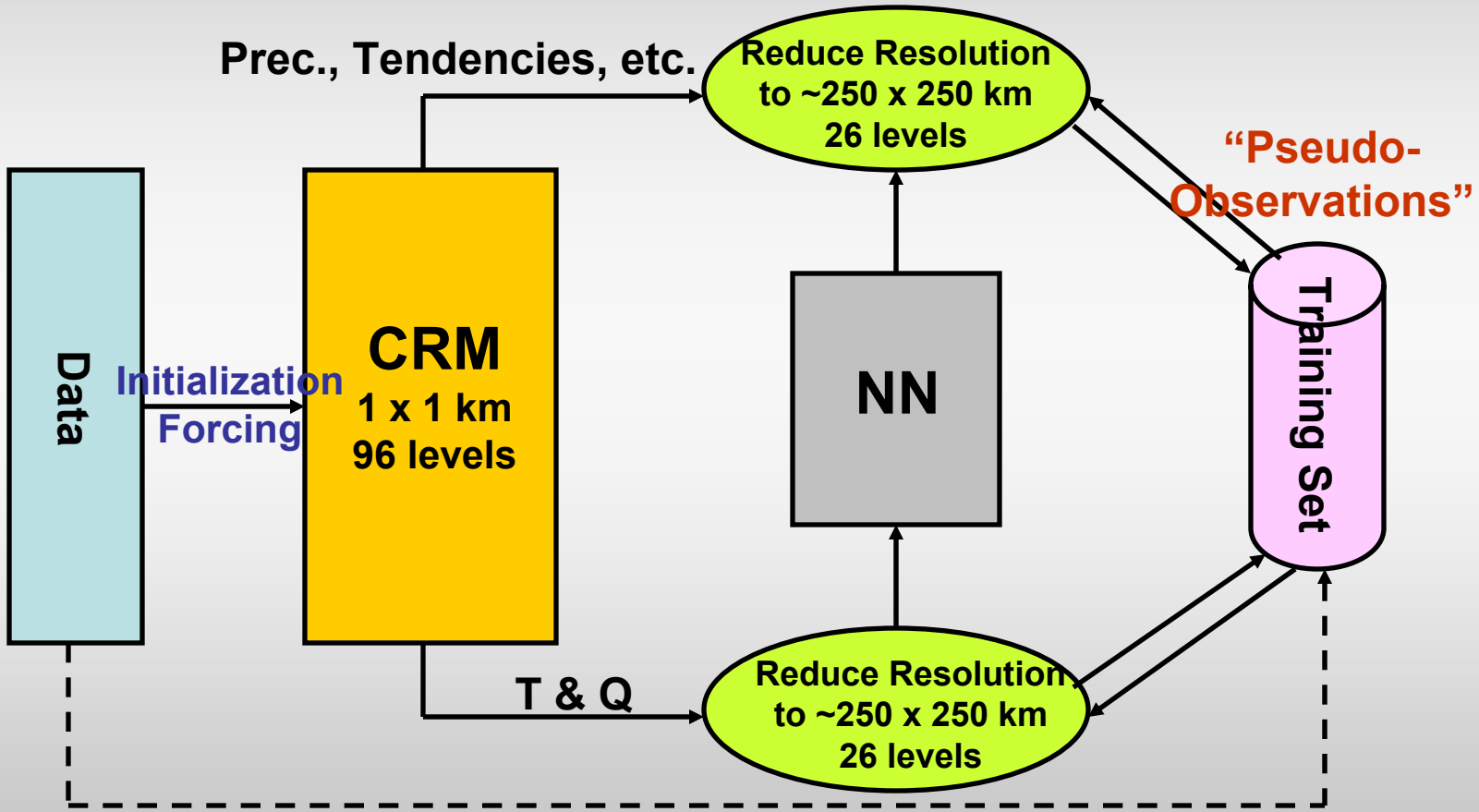


NN Parameterizations

- **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.

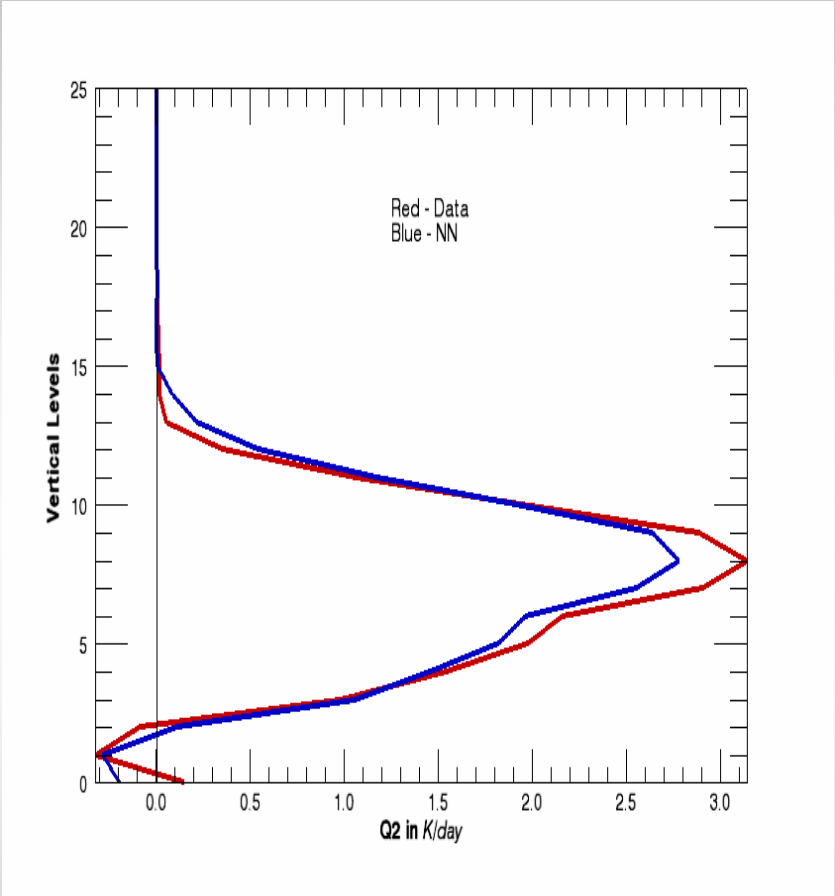
Proof of Concept (POC) -1.



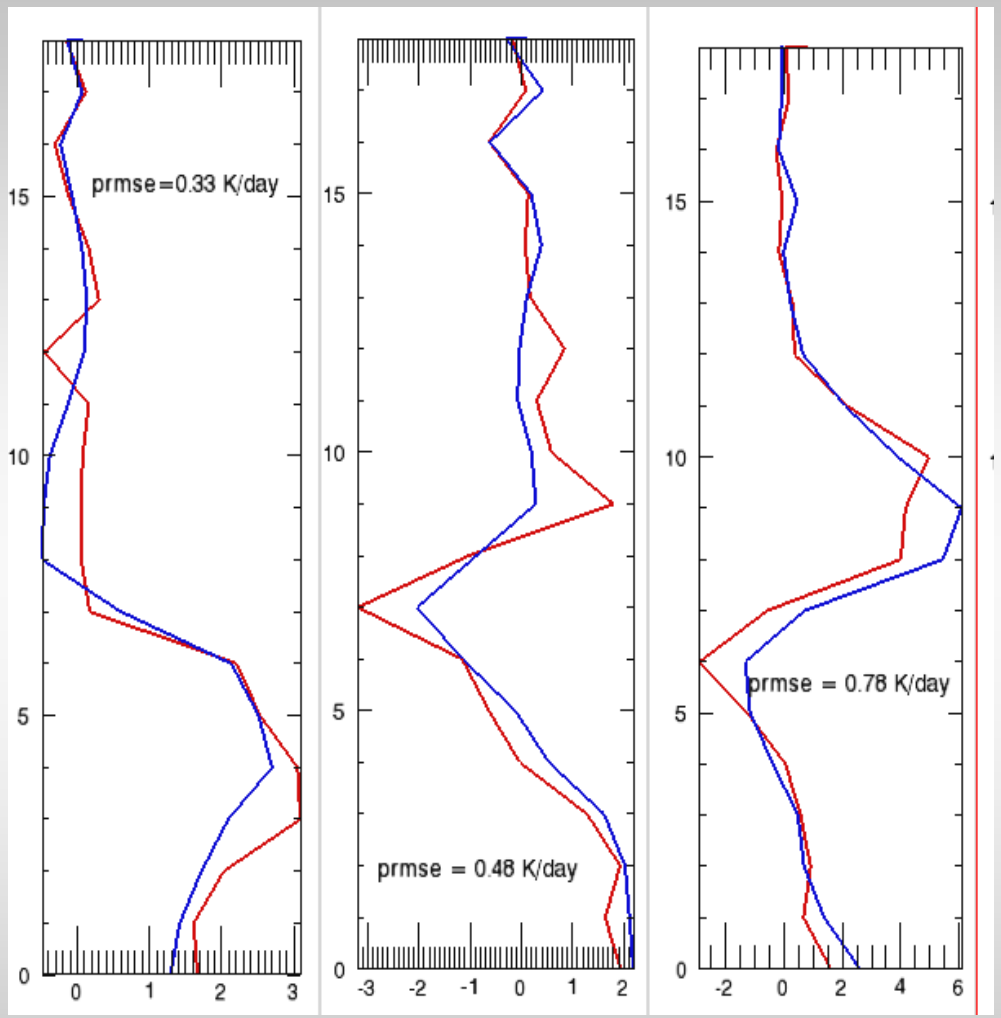
Proof of Concept - 2

- **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

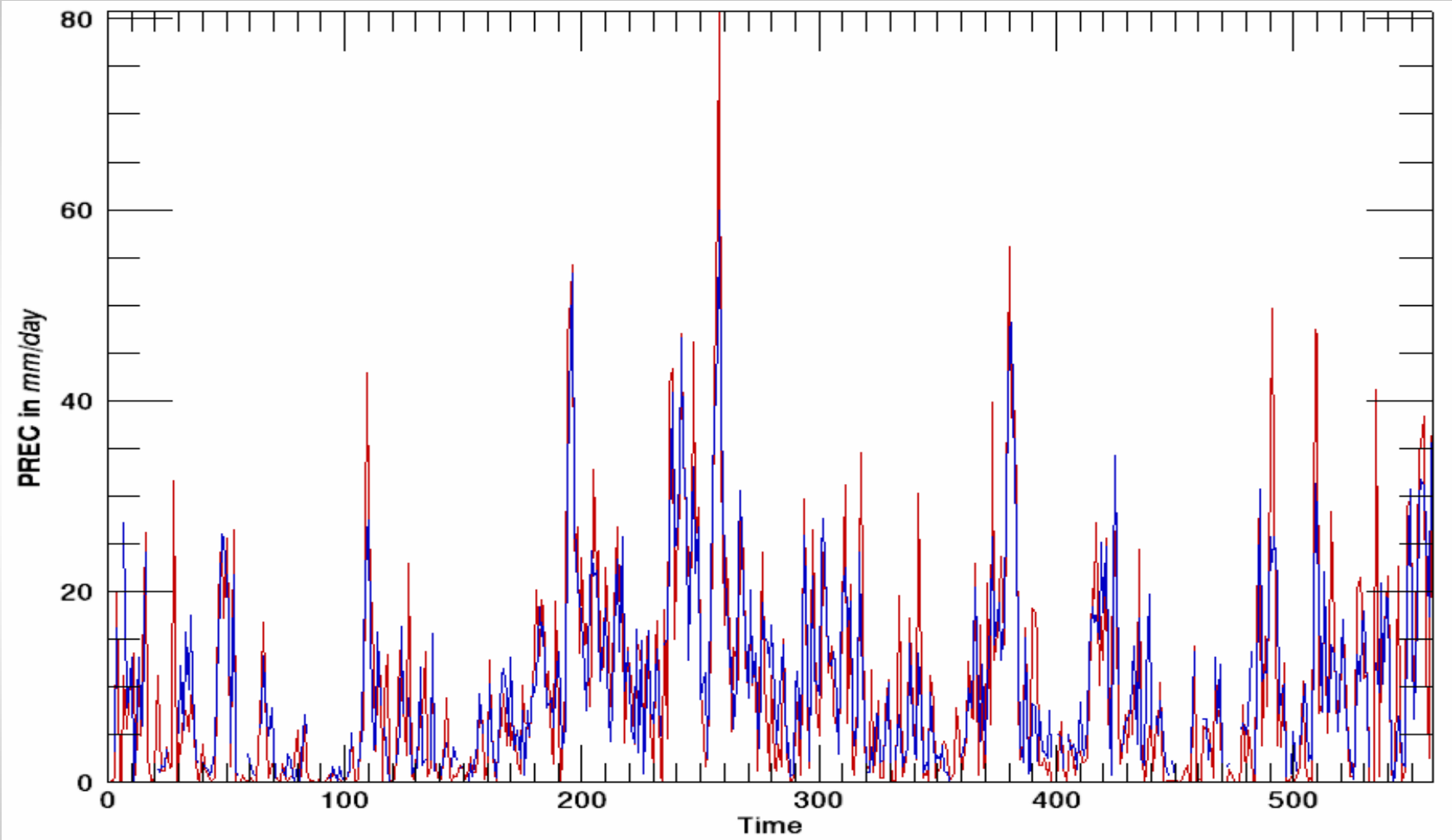


Time averaged water vapor tendency (expressed as the equivalent heating) for the validation dataset.



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

Proof of Concept - 3

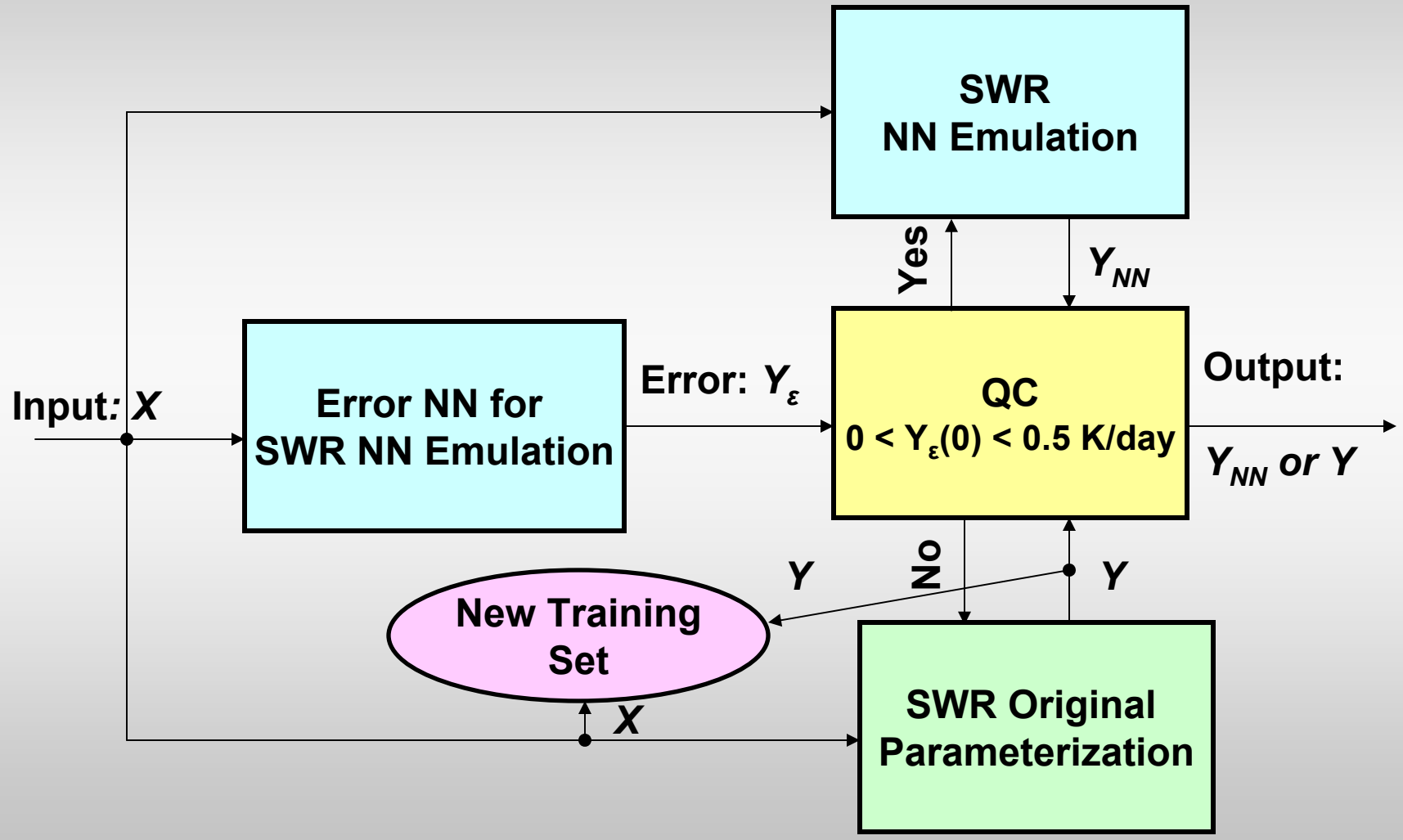


Precipitation rates for the validation dataset. Red – data, blue - NN

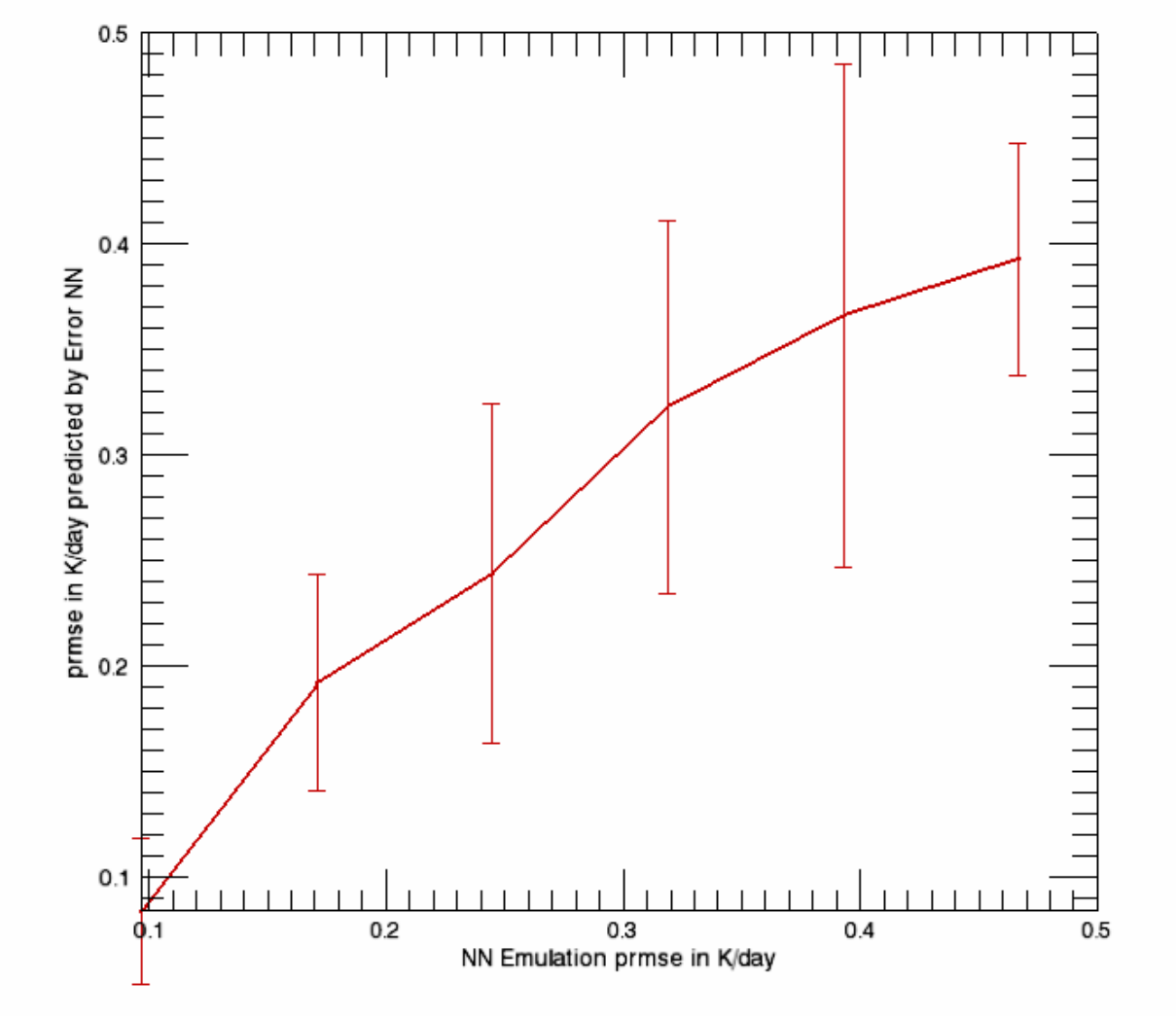
Compound Parameterization (CP): Quality Control (QC) for Outliers and Adaptivity to Climate Change

- 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

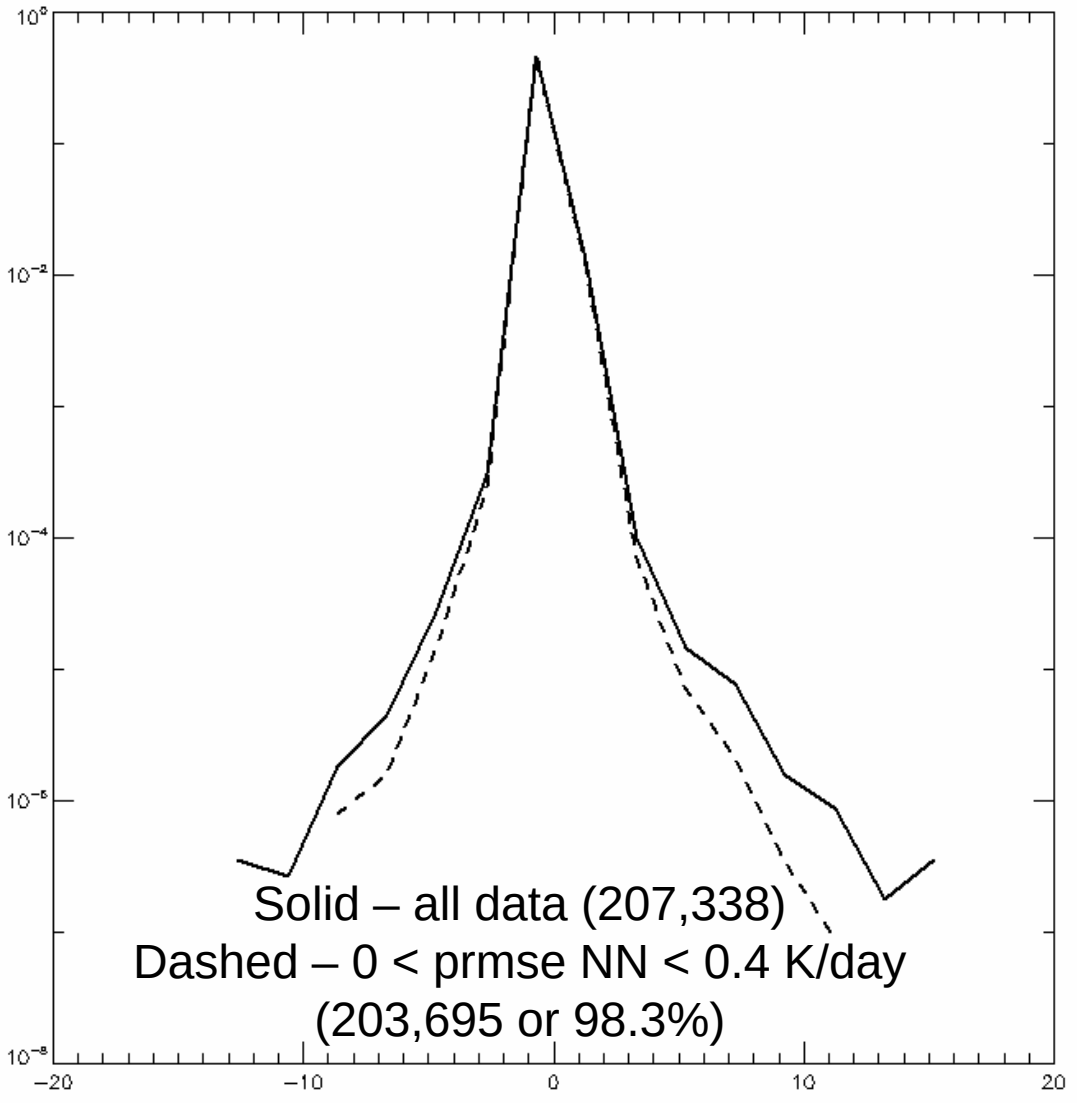
Compound Parameterization Based on Error NN



Predicted vs. Real Errors
Mean CC = 0.87

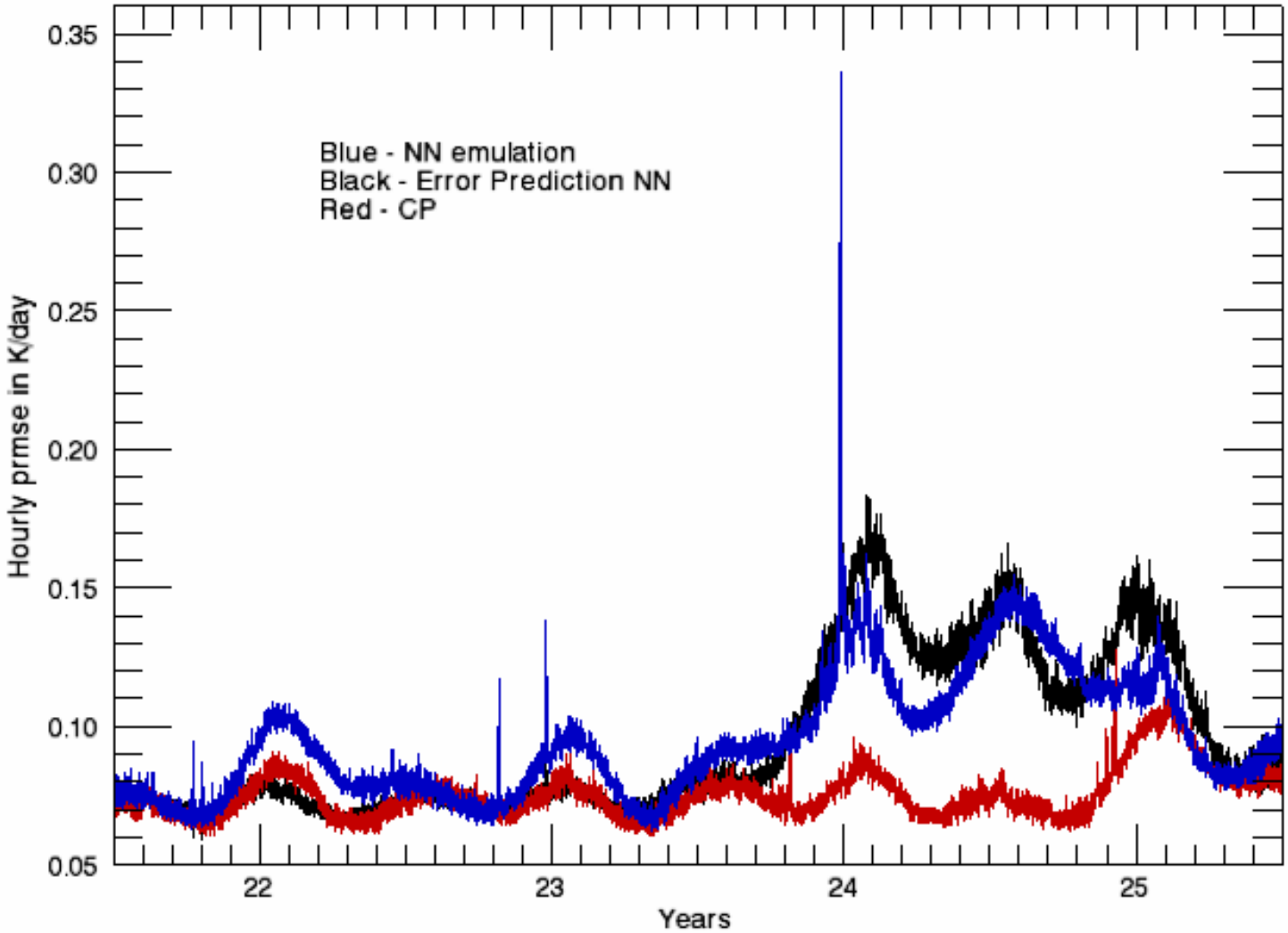


Error PDF



HR Errors in K/day

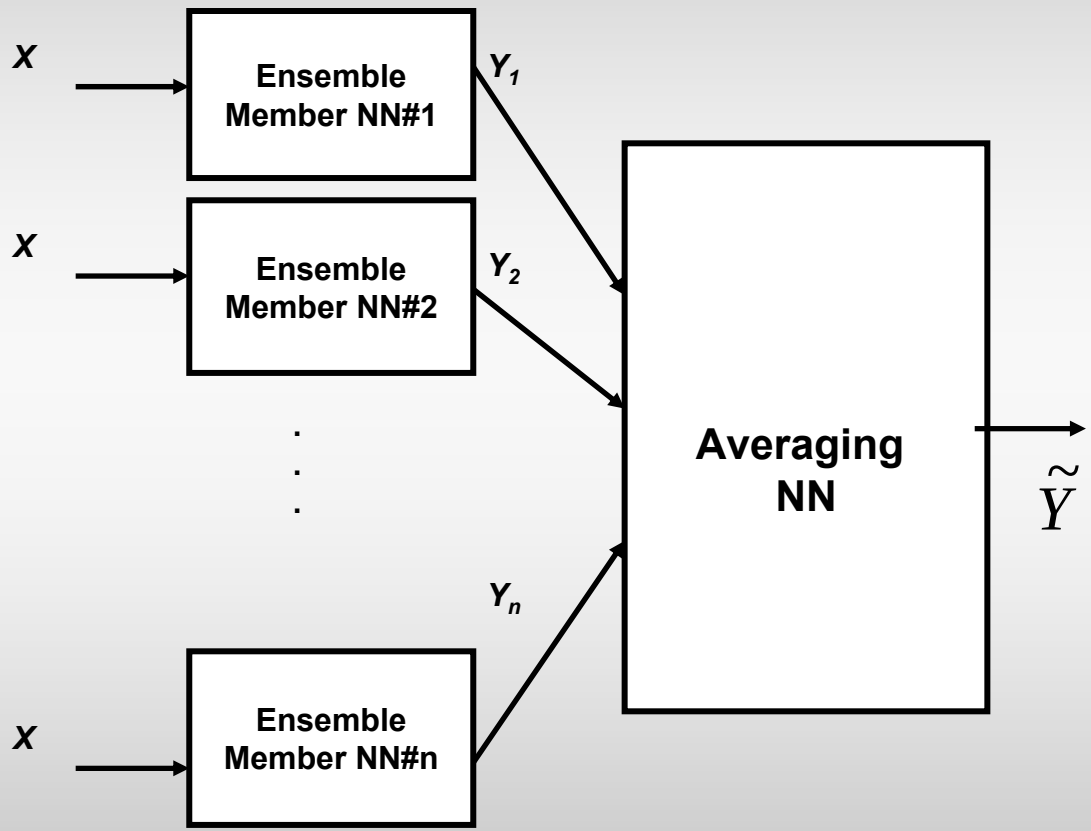
Hourly prmse



NN Ensembles

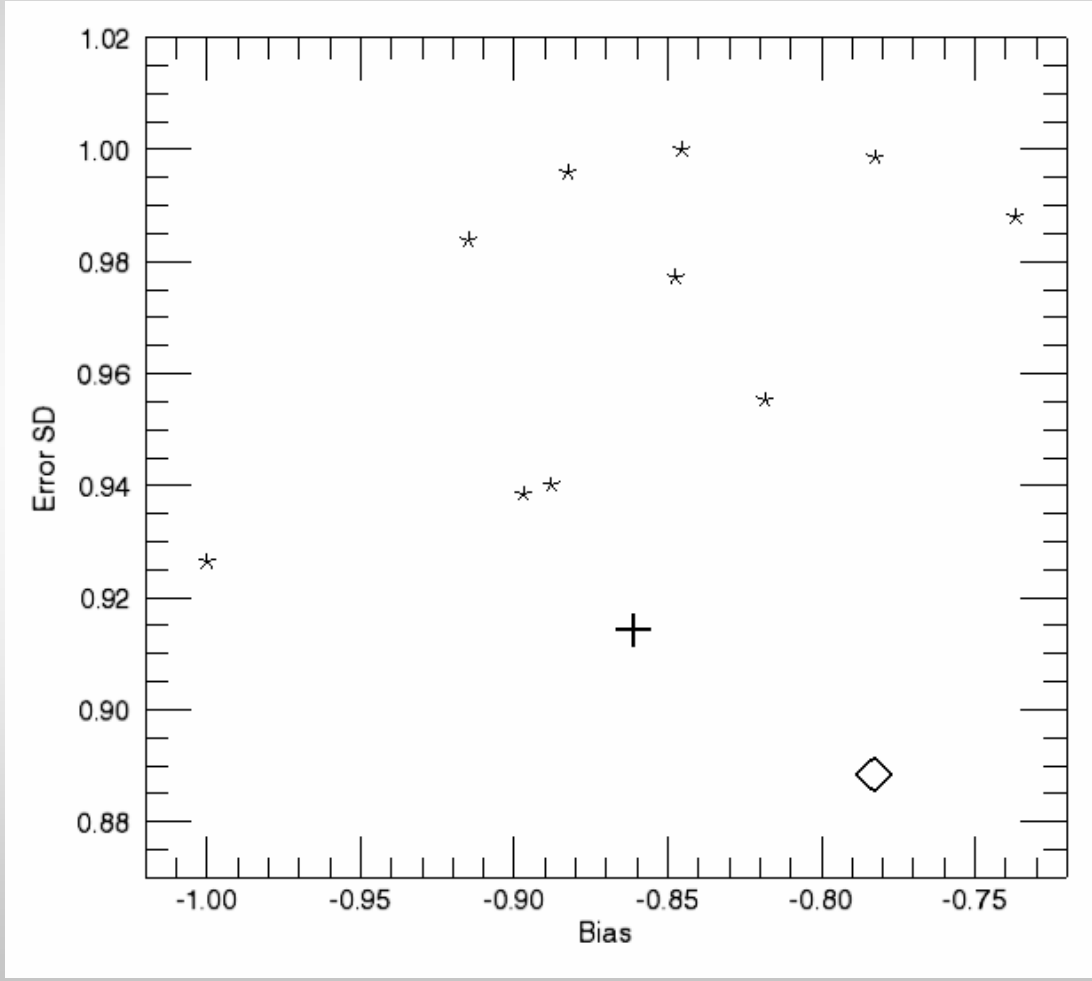
- **Ensemble approach is easily and naturally compatible with NN techniques**
  - Many ways of generating NN ensembles
    - Different initial conditions for NN training
    - Different NN architectures
    - Different partitioning of training set and/or domain
  - NNs can be applied for a nonlinear 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.

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 - 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