ARM Meeting, November 19, 2008, Princeton, NJ

Development of ensemble neural network convection parameterizations for climate models using ARM data: Initiation of the project

Project Participants: PI: Michael Fox-Rabinovitz (UMD), Co-PI: Vladimir Krasnopolsky (UMD and NCEP), Co-I: Philip Rasch (DOE PNNL and NCAR), Collaborators: Yefim Kogan (OU), and Alexei Belochitski (UMD)

Acknowledgments: Marat Khairoutdinov (SUNY) for providing SAM and consultations, Peter Blossey (UWA) for consultations on SAM

Outline

- Background: Hybrid Modeling: Synergetic Combination of Deterministic and Statistical Learning Model Components
- ARM and other data to be used for development of NN (Neural Network) convection parameterizations
- Initial SAM (System for Atmospheric Modeling)/CRM simulations for creating representative NN training data sets, and upscaling/averaging
- Initial development of NN convection
- Conclusions

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*,

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

- 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

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

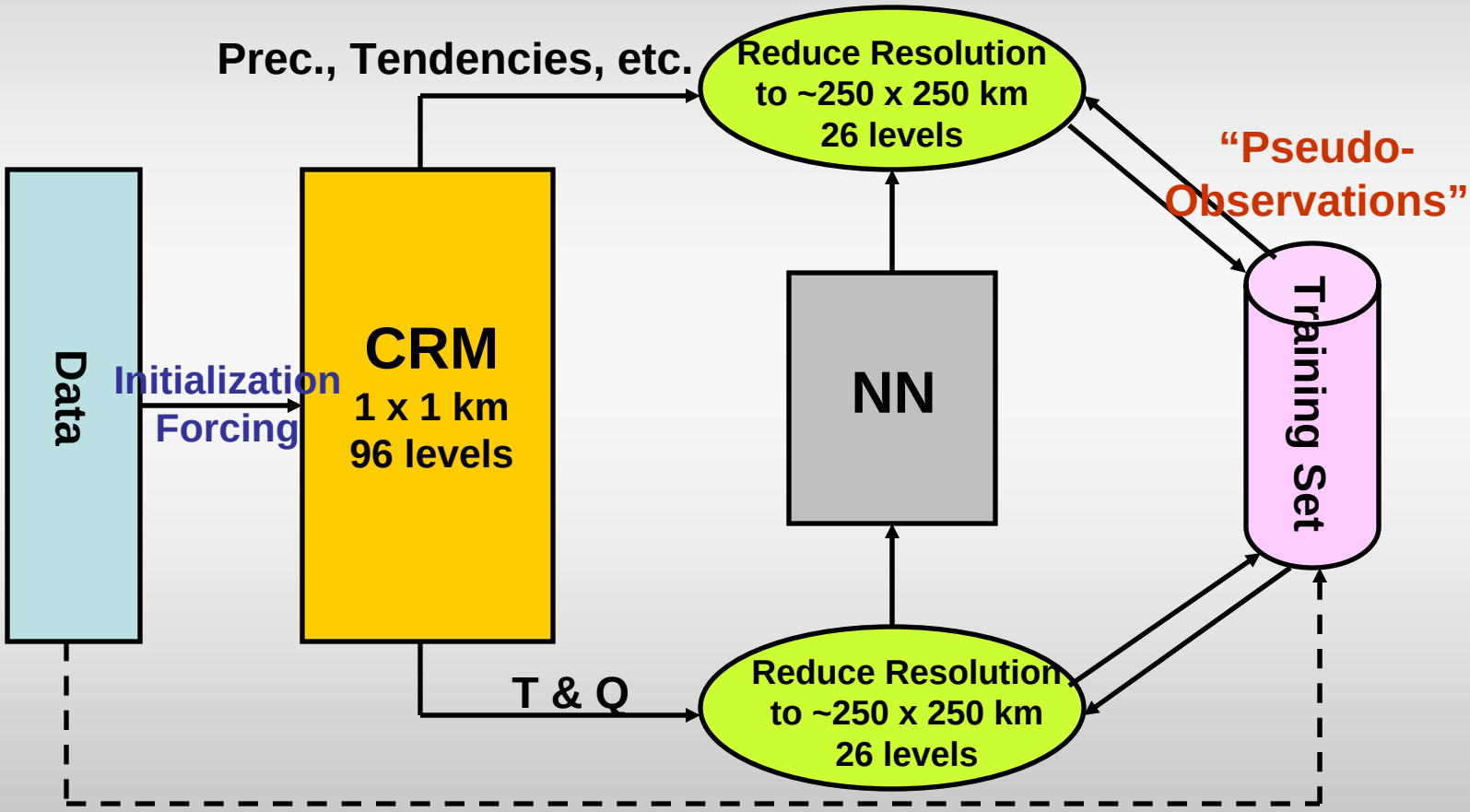
$$y_q = a_{q0} + \sum_{j=1}^{k} a_{qj} \cdot t_j$$

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

NN Parameterizations

- **New NN parameterizations** of model physics can be developed based on:
  - Data simulated by first principles models (cloud resolving models (CRM) (e. g. Khairoutdinov and Randall 2003)).
  - Observations (e.g. ARM)
- Our research is aimed at developing ***new more sophisticated and fast model convection schemes*** based on the NN approach, i.e. on **direct learning cloud physics from simulated CRM** (driven by ARM and other similar forcing) and **ARM** data.
- 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



Major stages of NN convection development

- **Developing an integrated/fused data sets**
- **Developing and testing NN convection parameterizations based on learning from data**
- **Developing and testing the NN ensemble approach**
- **Validating developed NN convection parameterizations in climate simulations using the NCAR CAM SCM (Single Column Model), and their validation using ARM data**
- **Exploratory development and testing of NN convection parameterizations for climate change scenarios**

ARM and other observational data available for development of NN convection

- Observational data available for initializing, driving/forcing and validating CRM:

  ARM-1997, ARM-2000, TOGA-COARE Long, TOGA-COARE, KWAJEX, GATE, ASTEX

The total period of observations = 264 days

- Other 1-day observational data (case studies)

Observational Data

Name	***Location***	Time Period	Domain	Resolution	Comment
TOGA-COARE, Long	**-2S, 155E**	120 days, Nov 1, 1992 -Mar 1, 1993	**32x32x128**	Horiz: 4000m x4000m Vert: 75-500m Time: 10s	
TOGA-COARE	**-2S, 155E**	21 days, Dec 18, 1992 – Jan 8, 1993	**32x32x128**	Horiz: 4000m x4000m Vert: 75-500m Time: 10s	
KWAJEX	**8.6N, 167.4E,** Marshall Islands	53 days, 23 Jul- 14 Sep 1999	**128x128x64**	Horiz: 1000m x1000m Vert: 50-500m	TRMM Ground Validation
GATE	**8.5N, -23.5W**	19 days, Sept. 1-18 1974		Horiz: 2000m x2000m Vert: 110-500m Time: 10s	GATE Phase III
ARM9707	**36.5N, -97.5W** (Oklahoma ARM site)	19 days	**128x128x64**	Horiz: 1000m x1000m Vert: 75-500m, Time: 10 s	ARM SGP July 1997 IOP
ARM0003	**36.5N, -97.5W** (Oklahoma ARM site)	22 days	**128x128x64**	Horiz: 2000m x2000m Vert: 50-500m Time: 10 s	ARM SGP March 2000 IOP dataset
ASTEX209	**Off the CA coast**	**10 days**		Horiz: 100m x100m Vert: 25m, Time: 2 s	Stratocumulus case based on ASTEX flight 209 soundings

NOTE: The total period of IOP observation data = 264 days

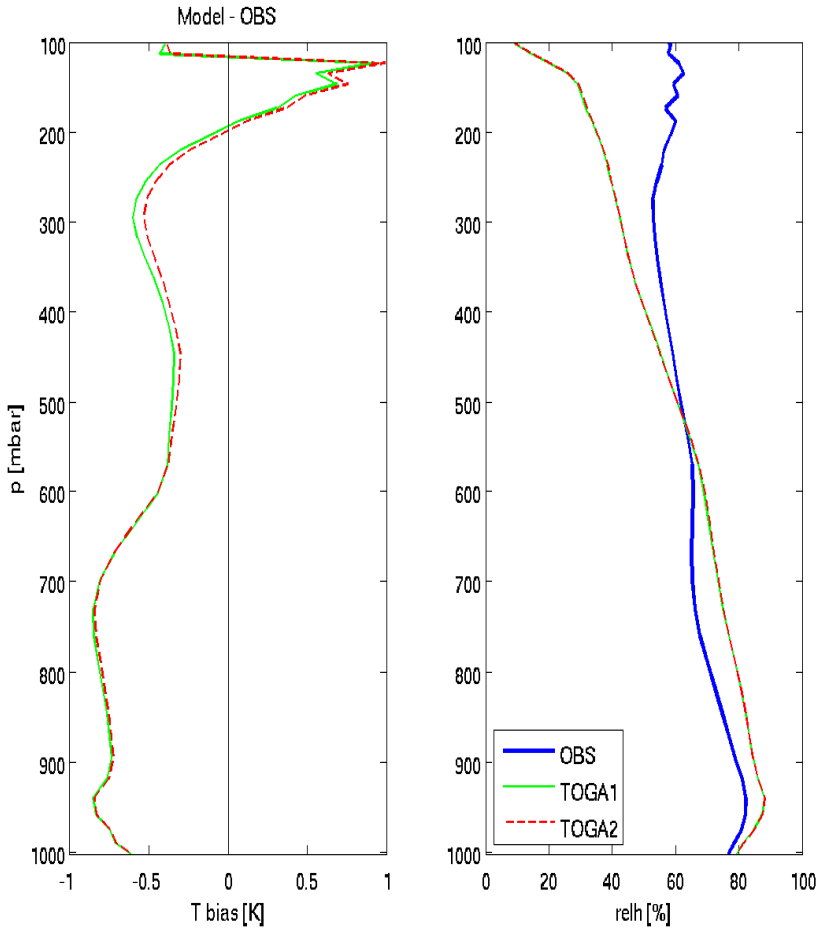
SAM/CRM runs for creating NN training data sets

- **Data (for CRM initialization and forcing):** ARM, TOGA-COARE,
- **CRM:** SAM/CRM (Khairoutdinov and Randall, 2003).
  - SAM/CRM simulated data
  - Hourly data: for 22 days for ARM and 120 days for TOGA-COARE
  - Resolution: 1 km over the domain of 256 x 256 km
  - 96 vertical layers (0 – 28 km)

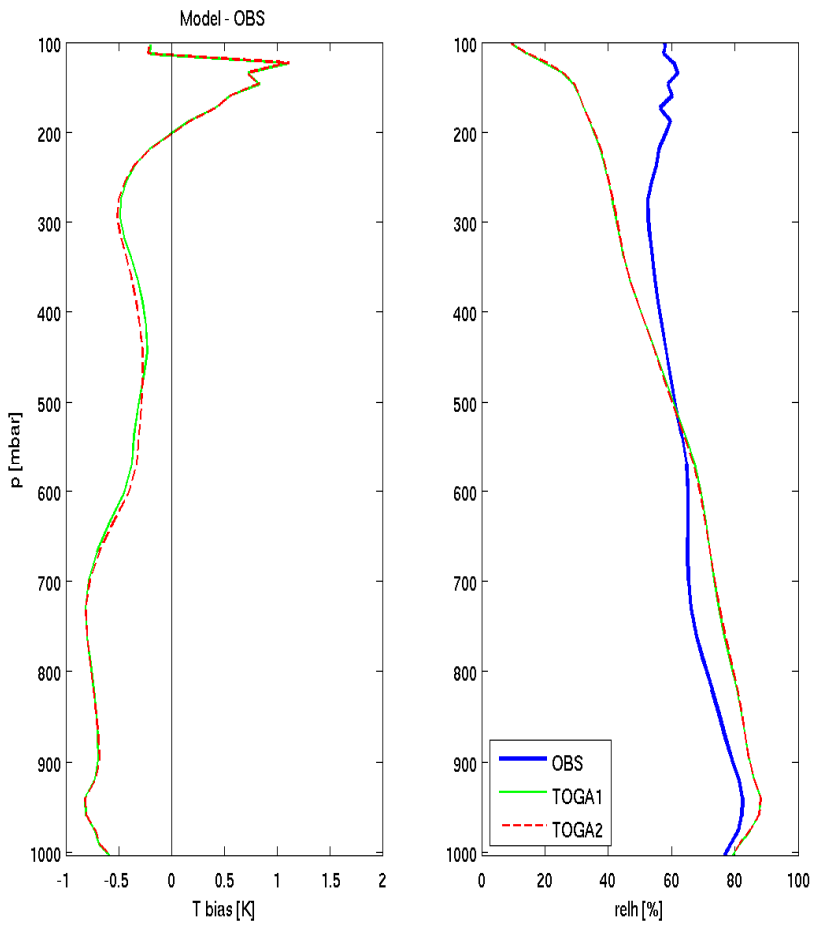
SAM/CRM runs with Perturbed IC

Two perturbed IC run (random, decreasing with height

perturbation to lower 5 levels + cloudy levels of T profile with the 0.1 K amplitude)



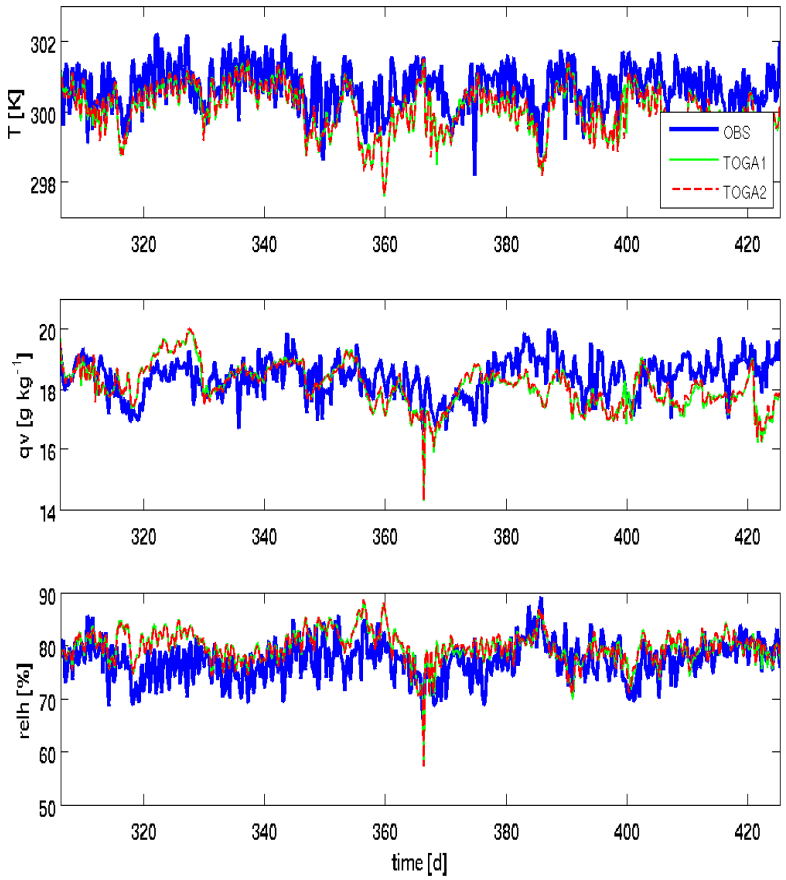
Two perturbed IC runs (random perturbation to the entire moisture and temperature profiles with amplitude of 0.5 K)



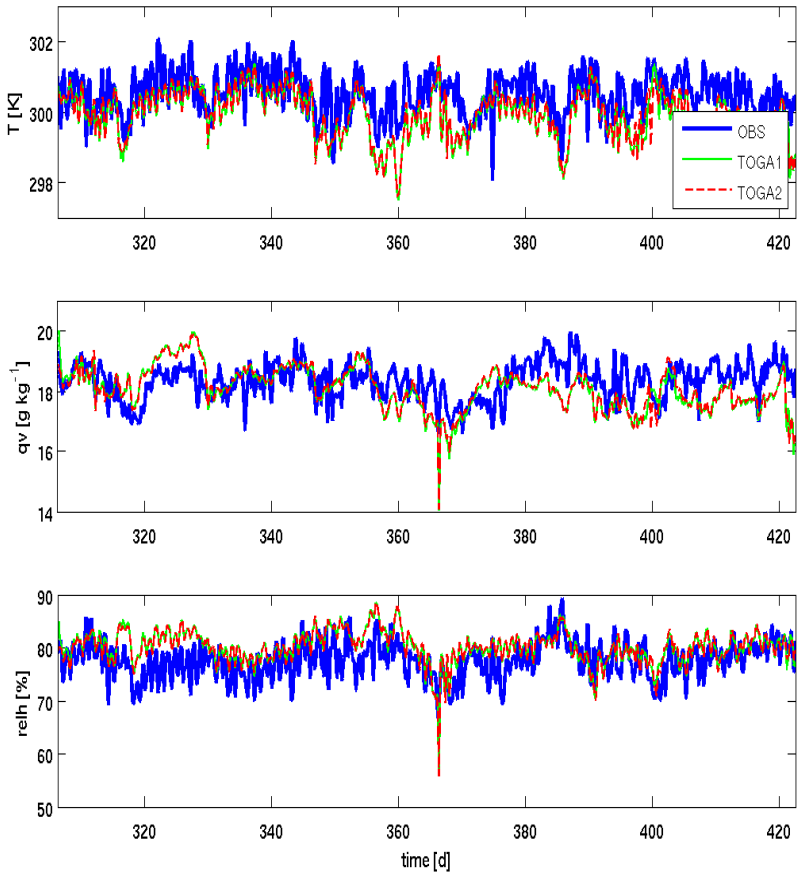
SAM/CRM runs with Perturbed IC

Two perturbed IC run (random, decreasing with height

perturbation to lower 5 levels + cloudy levels of T profile with the 0.1 K amplitude)



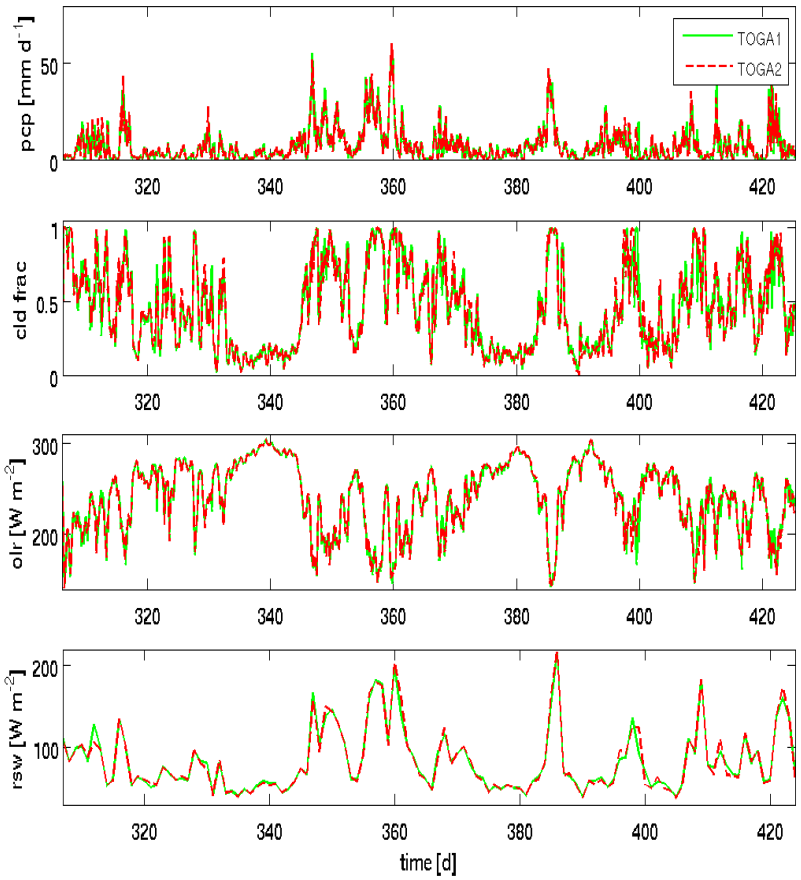
Two perturbed IC runs (random perturbation to the entire moisture and temperature profiles with amplitude of 0.5 K)



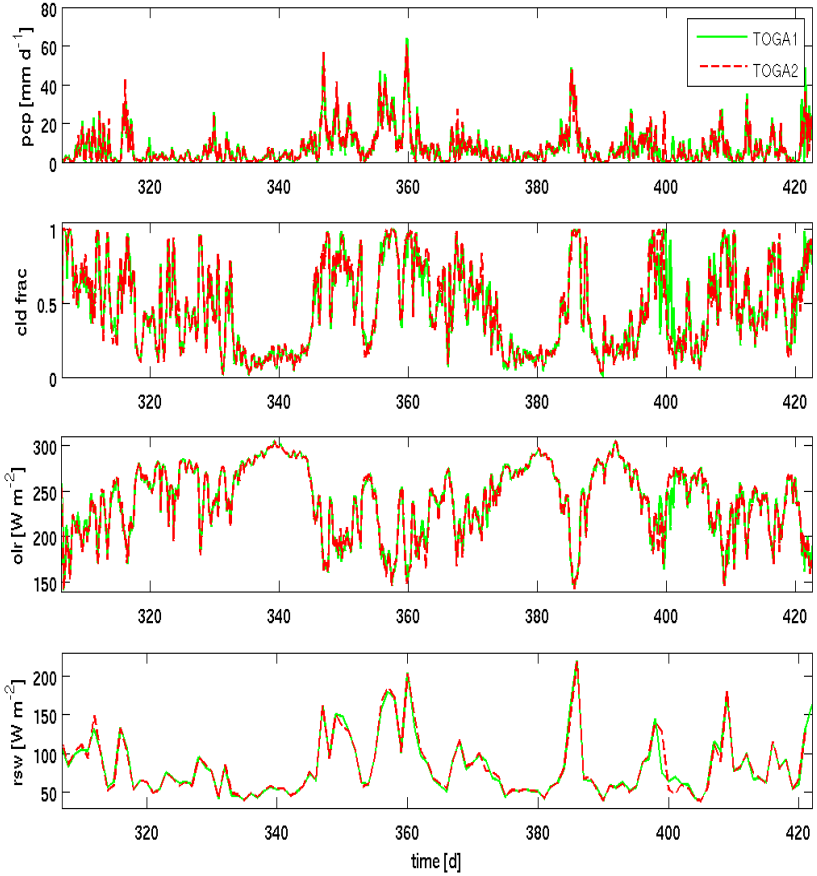
SAM/CRM runs with Perturbed IC

Two perturbed IC run (random, decreasing with height

perturbation to lower 5 levels + cloudy levels of T profile with the 0.1 K amplitude)



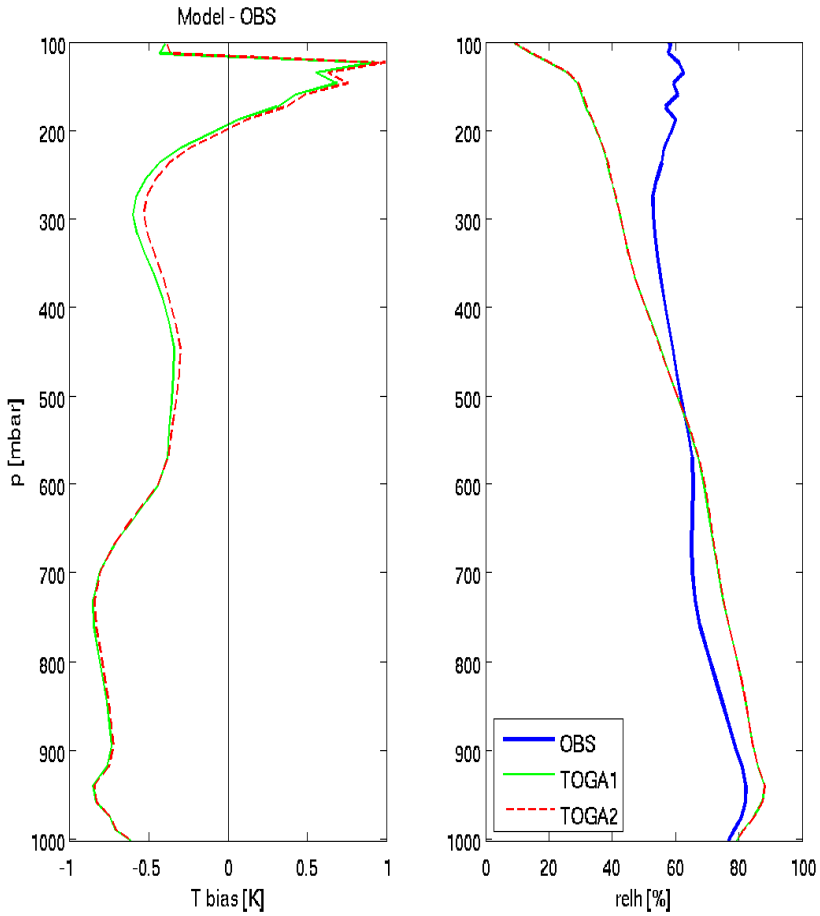
Two perturbed IC runs (random perturbation to the entire moisture and temperature profiles with amplitude of 0.5 K)

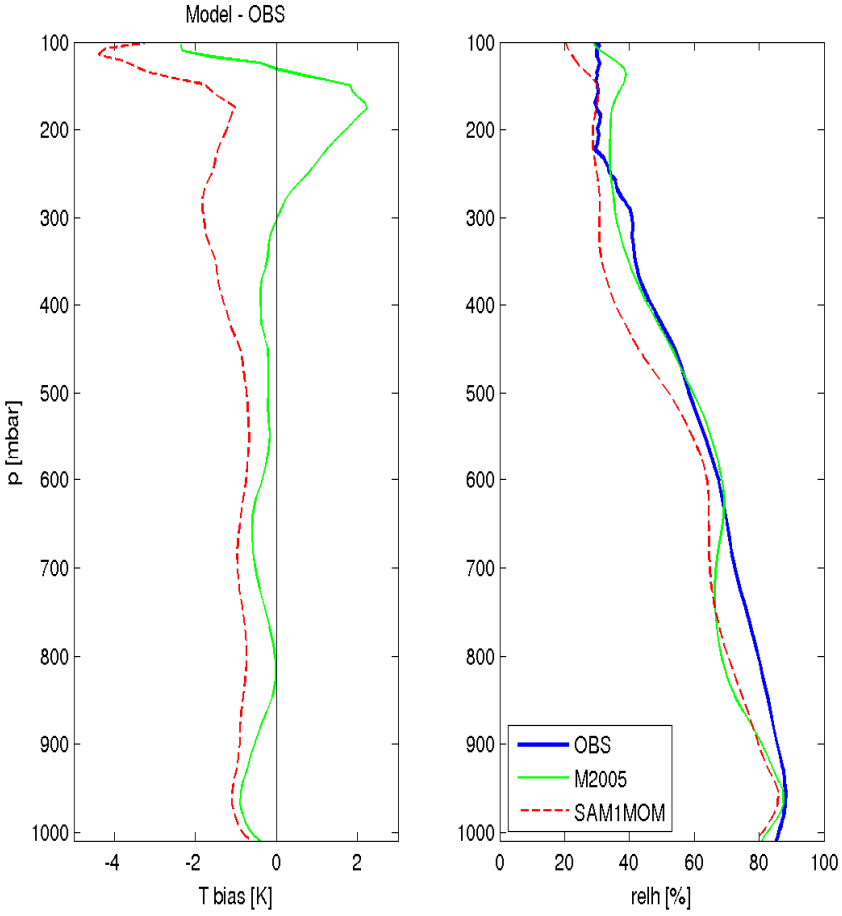


SAM/CRM runs with perturbed IC and with different microphysics

Perturbed IC run (random, decreasing with height perturbation to lower 5 levels + cloudy levels of T profile with the 0.1 K amplitude) vs. the unperturbed IC run

KWAJEX runs performed by Peter Blossey for two different microphysics (Morrison-2005 and the SAM original)



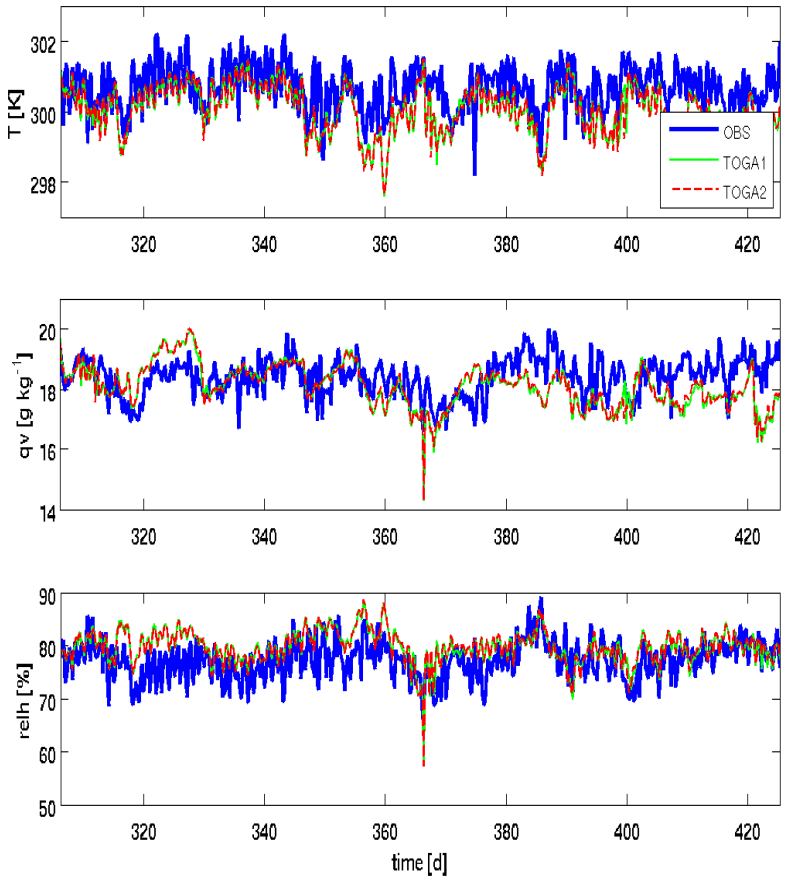


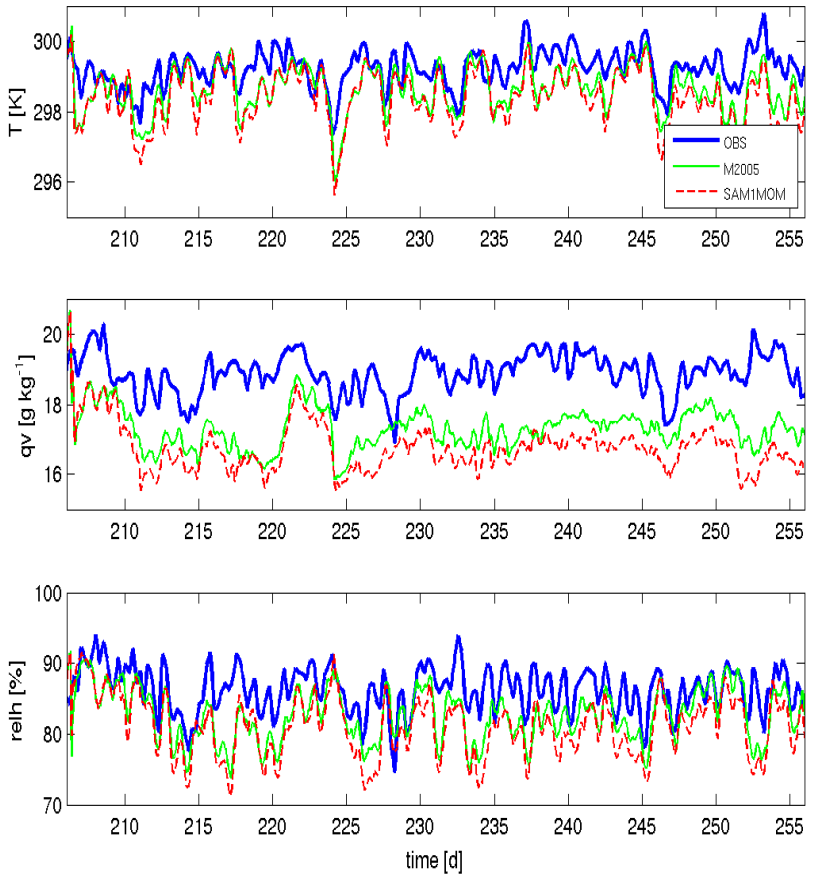
SAM/CRM runs with perturbed IC and with different microphysics

Perturbed IC run (random, decreasing with height perturbation to lower 5 levels + cloudy levels of T profile

with the 0.1 K amplitude) vs. the unperturbed IC run

KWAJEX runs performed by Peter Blossey for two different microphysics (Morrison-2005 and the SAM original)



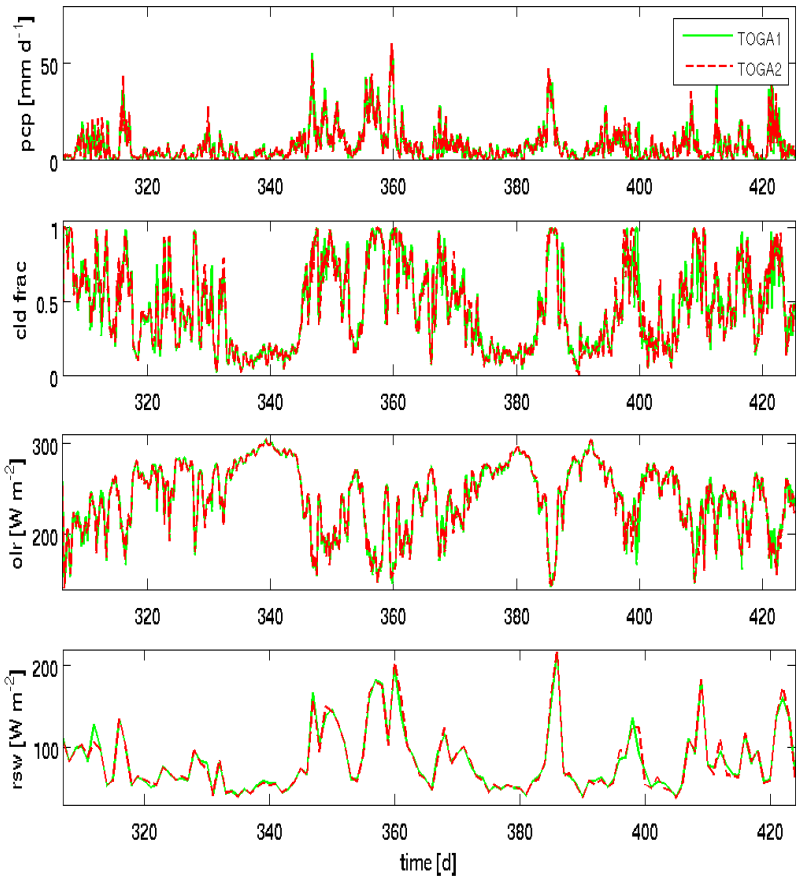


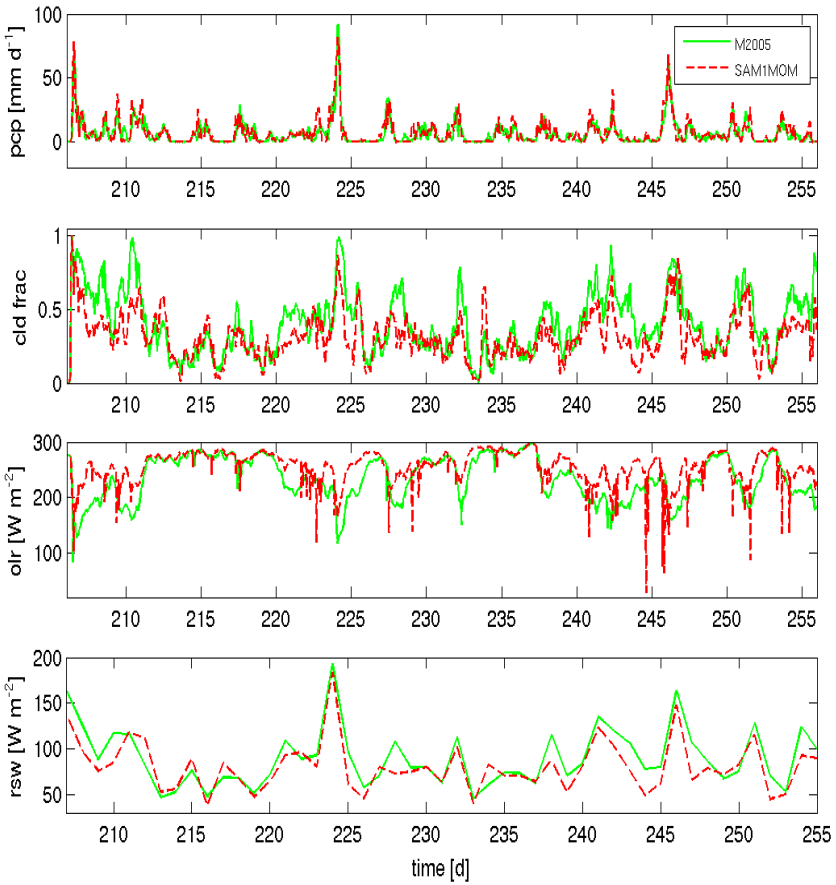
SAM/CRM runs with perturbed IC and with different microphysics

Perturbed IC run (random, decreasing with height perturbation to lower 5 levels + cloudy levels of T profile

with the 0.1 K amplitude) vs. the unperturbed IC run

KWAJEX runs performed by Peter Blossey for two different microphysics (Morrison-2005 and the SAM original)

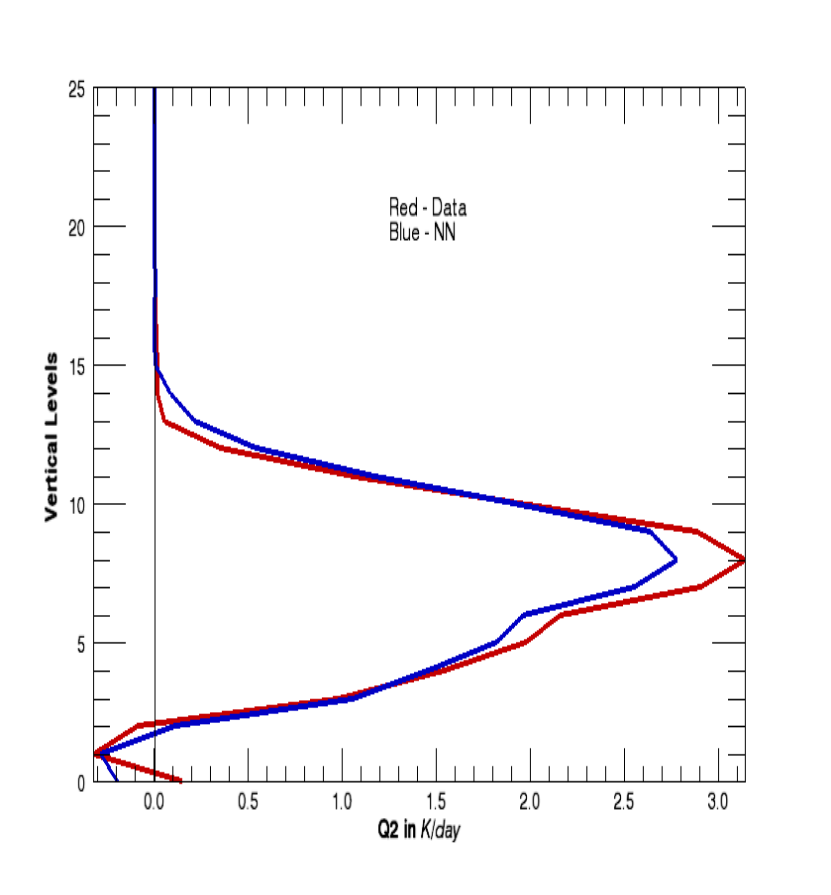




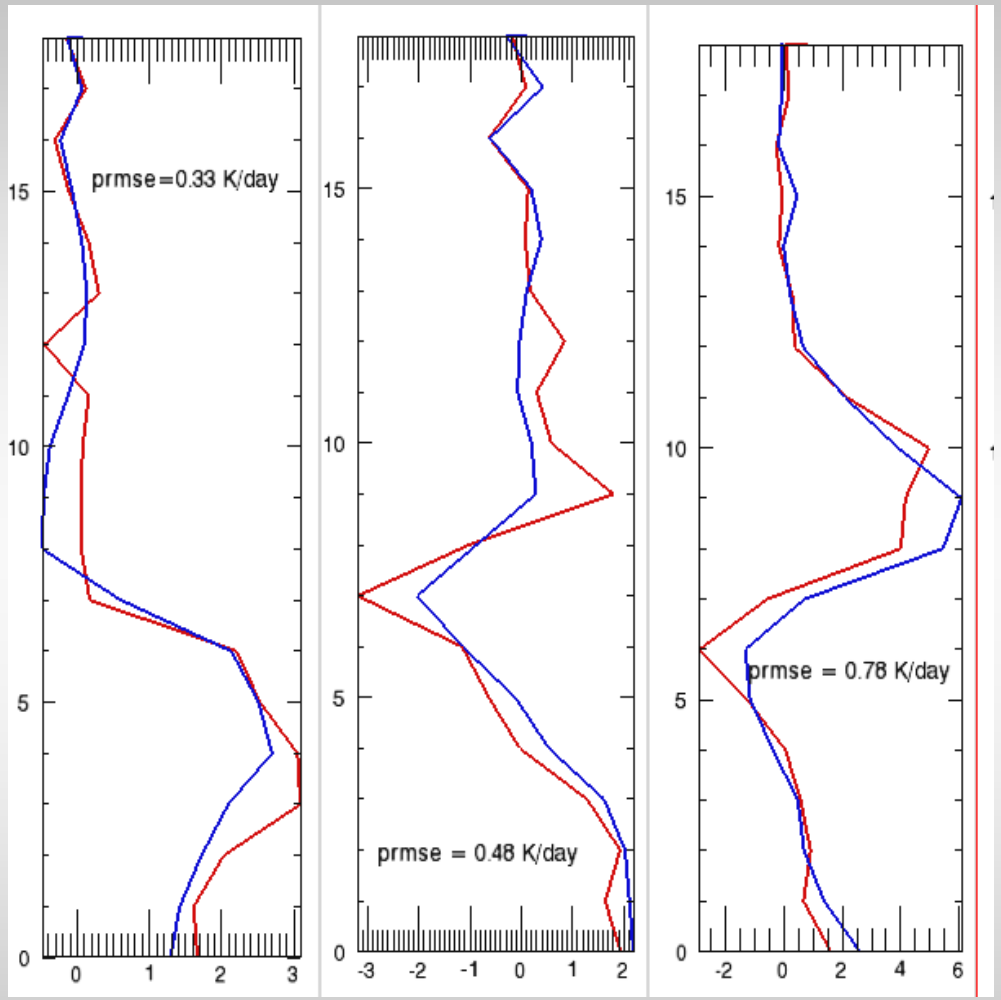
Initial NN convection

- **Data (for CRM initialization and forcing):** ARM, TOGA-COARE
- **CRM:** SAM CRM (Khairoutdinov and Randall, 2003).
  - Data from SAM CRM simulations
  - Hourly data: over 22 days for ARM and 120 days for TOGA-COARE
  - 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, 128 x 128 km, 64 x 64 km, 32 x 32 km
  - Vertical: 26 vertical layers (as in CAM)
- **NN inputs:** temperature and water vapor fields; a limited training data set used for the initial development of NN convection
- **NN outputs:** precipitation & the tendencies T and q, i.e. "apparent heat source" (Q1), "apparent moist sink" (Q2), and cloud fractions (CLD)

Initial NN convection parameterization

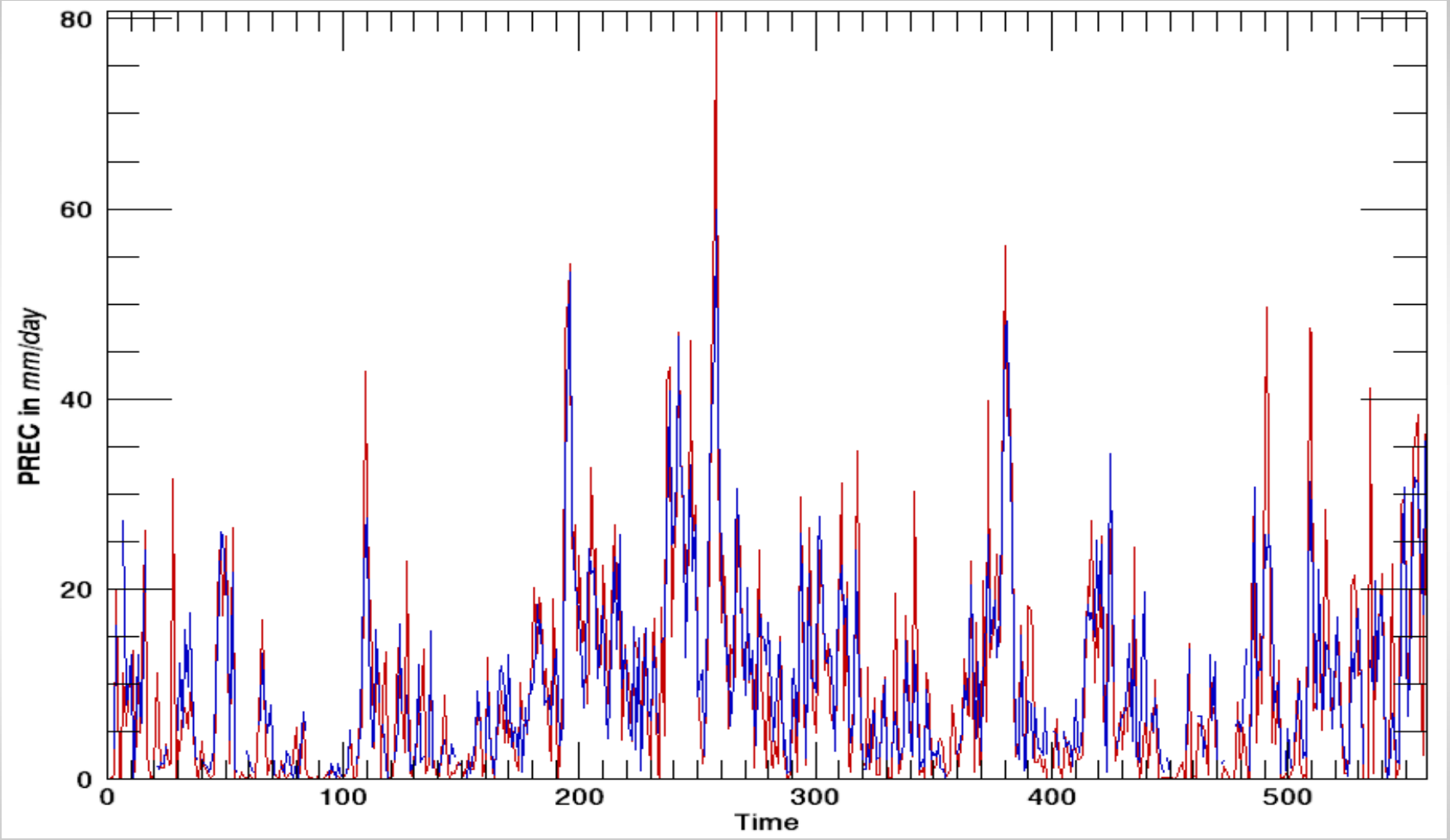


Time averaged water vapor tendency (expressed as the equivalent heating) for the validation dataset.



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

Initial NN convection parameterization - 2



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

Conclusions - Initial developments of NN convective parameterizations

- Approach has been conceptually formulated
- NN training data sets are being produced with averaging to horizontal resolution of 256 x 256 km, 128 x 128 km, 64 x 64 km, 32 x 32 km, and finer if necessary
- Analysis and questions regarding creating representative NN training data sets are being addressed
- Producing training data sets Initial NN convection parameterization has been produced
- Initial results: Even 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
- **The *NN Parameterization* approach is a promising tool for developing new NN convection parameterizations**