Development of ensemble neural network convection parameterizations for climate models using ARM data

Michael Fox-Rabinovitz[1], Vladimir Krasnopolsky[1,2], Phil Rasch[3], Yefim Kogan[4], and Alexei Belochitski[1]

[1]University of Maryland – College Park, [2]NOAA NCEP, [3]DOE PNNL, [4]University of Oklahoma/UCSD-Scripps

Acknowledgements: Marat Khairoutdinov (SUNY – Stony Brook), Peter Blossey (UW), DOE NERSC supercomputing facilities

Objectives

1. Contributing to the major ARM goals: "developing new… cloud parameterizations for global models using ARM observations."
2. Developing novel, more sophisticated, and fast convection parameterizations for climate models based on applying neural networks (NN) for direct learning cloud physics from SAM (System for Atmospheric Modeling, Khairoutdinov and Randall, 2003) simulated data.
3. Using ARM (or TOGA-COARE) data for:
   - initializing and forcing SAM simulations,
   - creating NN training data sets for developing NN convection parameterizations, and
   - validating model simulations.
4. Designing and testing different NN architectures.

NN Parameterizations

1. New NN convection parameterizations are developed through learning from data using:
   - Observations
   - Data simulated by first principle process models (cloud resolving models).
2. Here NN serves as an interface transferring information about sub-grid scale processes from fine scale data or models (SAM/CRM) into GCM (up-scaling).

Data and NN Training

1. Data for initialization, forcing, and validation (?) of SAM:
   - ARM and TOGA-COARE meteorological conditions,
   - Hourly data over ~60 to 120 days,
   - 1 km resolution, 256x256 km domain, 96 layers (0–28 km)
2. Resolution of averaged SAM output:
   - Horizontal: 256 x 256 km (up-scaling to GCM resolution)
   - Vertical: 26 vertical layers (0 – 28 km)
   - Temporal: 1 hr
   - First 96 days are for NN training, last 24 -- for validation.
3. Averaged SAM output is "projected" on the space of the NN mapping, or in other words, only a subset of relevant SAM variables available in climate model (CAM) is selected, resulting in creating an NN training data set or "pseudo-observations".

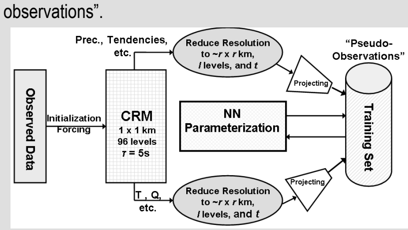

Fig. 1: Generation of the NN training data set. Horizontal resolution: 1 km < *r* ≤ *R*; vertical resolution: 96 layers > *l* ≥ *L*; averaging time *t*.

NN Architectures

Table 1: **T** - temperature, **QV** – the atmospheric moisture vapor mixing ratio, **W** – vertical velocity, **U** and **V** are horizontal components of the velocity vector, **RelH** is the relative humidity, **Rad** – the radiative heating/cooling rates, **Q1C** – the "apparent heat source", **Q2** – the "apparent moist sink", **Prec** – the precipitation rate, and **CLD** – the cloudiness. Numbers in the table show the dimensionality/number of the corresponding input and output parameters. **In:Out** stand for inputs and outputs and show their corresponding numbers.

NN Architecture In:Out	NN Inputs							NN Outputs			
	T	QV	W	U	V	Rel H	Rad	Q1C	Q2	PREC	CLD
{2} – 47:40	26	21	-	-	-	-	-	21	18	1	-
{3} – 47:59	26	21	-	-	-	-	-	21	18	1	19
{4} – 87:66	26	15	-	23	23	-	-	26	19	1	20
{5} – 58:66	26	15	17	-	-	-	-	26	19	1	20
{6} – 81:66	26	15	17	-	-	23	-	26	19	1	20
{7} – 66:66	26	-	17	-	-	23	-	26	19	1	20
{9} – 84:66	26	15	17	-	-	-	26	26	19	1	20

Accuracy of NN Approximation

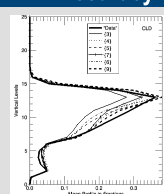

Fig. 2: Average CLD profiles for different NN architectures vs. validation (SAM) data

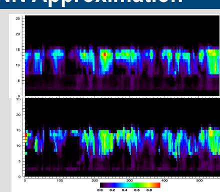

Fig. 3: Hovmöller diagrams for CLD profile time series; top validation data, bottom – NN.

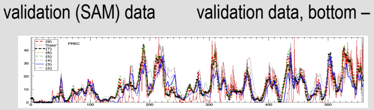

Fig. 4: Precipitation time series for different NN architectures vs. validation data, mm/day.

NN Convection in CAM

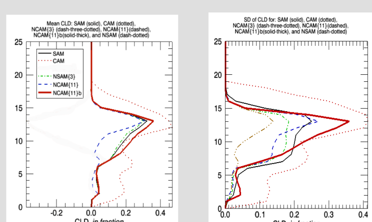

Fig. 5 : Vertical profiles of time-mean CLD (left) and its standard deviation (SD) (right) for NN architectures 3 and 11.

NN Convection in CAM (continued)

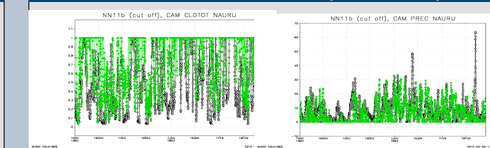

Fig. 6: Time series of total cloudiness (in fractions, left) and precipitation (mm/day, right) for the control CAM and CAM-NN-11b (the off-line/diagnostic CAM run with NN-11b) runs. Green: the control CAM run, black: the CAM-NN-11b run.

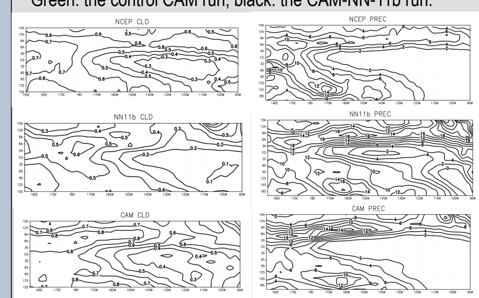

Fig. 7: Total cloudiness (in fractions, left) and precipitation (mm/day, right) for NCEP (top), CAM-NN11b (middle), and CAM (bottom) for the Eastern Tropical Pacific Ocean for 25 S to 15 N, 150 E to 90 W. The contour interval for CLD is 0.1, and for PREC is 2 mm/day.

Conclusions

1. A novel approach based on using neural networks (NNs) is formulated and used for development of NN ensemble convection parameterizations for climate models.
2. SAM/CRM simulations initialized with and driven/forced by TOGA-COARE and ARM data have been averaged and projected onto the GCM space of atmospheric states and used to derive very fast NN convection parameterizations with different architectures, and their accuracy is estimated.
3. Developed NN convection parameterizations have been tested in an off-line/diagnostic CAM (CAM-NN) run vs. the control CAM run. The initial results are encouraging: Total precipitation and cloudiness time series and tropical distributions for CAM-NN and CAM are realistic and consistent.

Future Plans

1. Using SAM simulations driven by CAM forcing for longer times, more geographic locations, and more diverse weather conditions so that NNs can be used globally and for all seasons.
2. Testing NN convection, trained using these new data, in CAM.
3. In-depth analysis of the results aimed at specifying the potential and challenges of the NN approach for representation of convective processes in CAM.