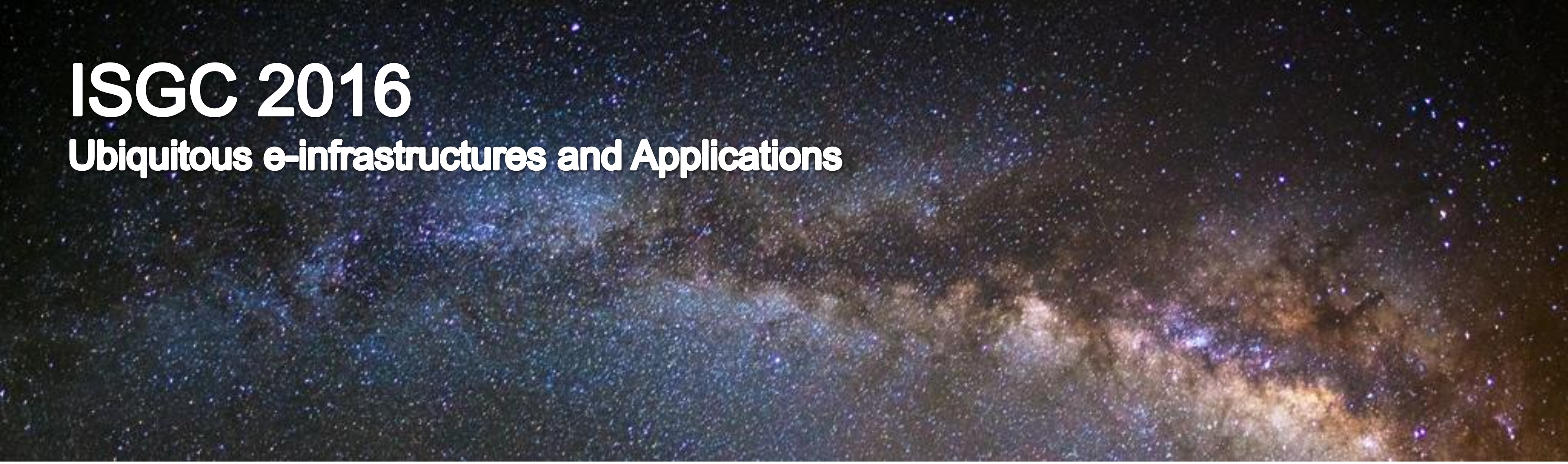


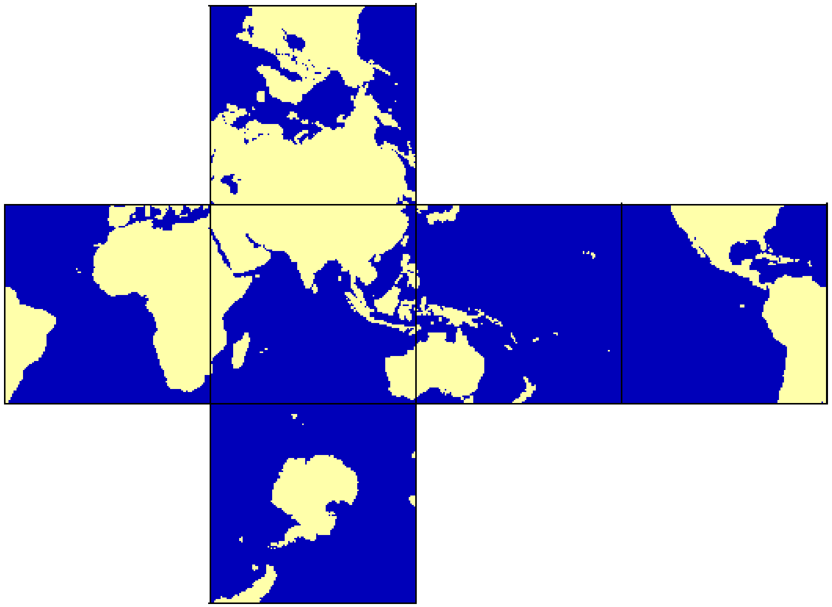
Data Processing and Visualization for Global High-Resolution Climate Simulations

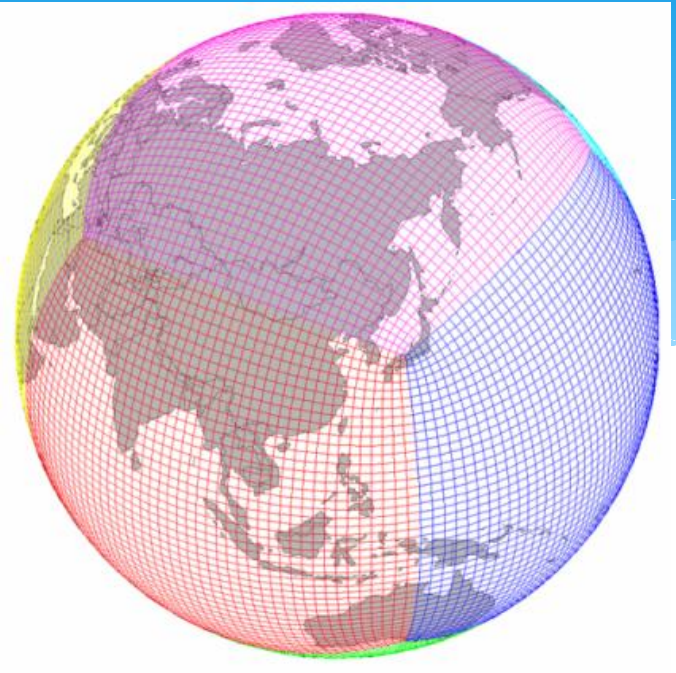
Chia-Ying Tu[R], Huang-Hsiung Hsu[R]
Lung-Cheng Lee[H], Ching-Yao Lin[H], Whey-Fone Tsai[H]

R - RCEC, Academia Sinica
H - NCHC, NARLabs

March 16, 2016 @ Academia Sinica, Taipei, Taiwan

FV3: Finite-Volume Dynamical Core on the Cubed-Sphere Globe



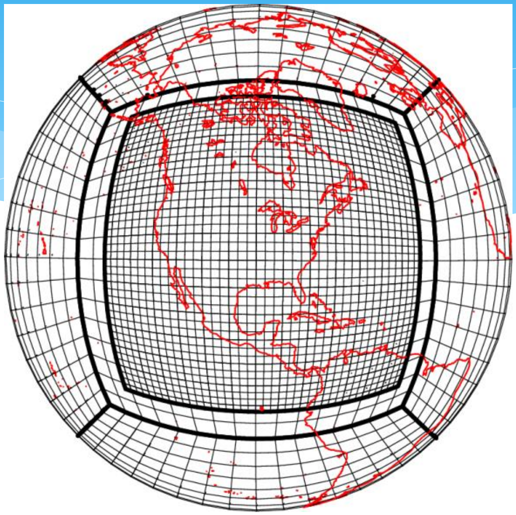


- 4th order Finite-Volume algorithms with a true Lagrangian vertical coordinate (Lin 2004)
- Designed for fine-grained parallelism (MPI, openMP, and GPU); scalable to ~million cores (IBM BG) at cloud-resolving resolution
- Hydrostatic or non-hydrostatic, regional or global (run time options)

Ultra-high resolution regional-global model

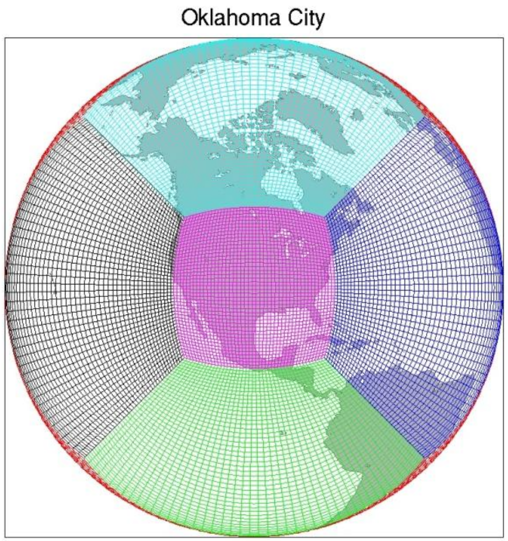
I. Nested regional-global model

- Regional model can be run independently or two-way nested with the global model
- Suitable for weather & climate

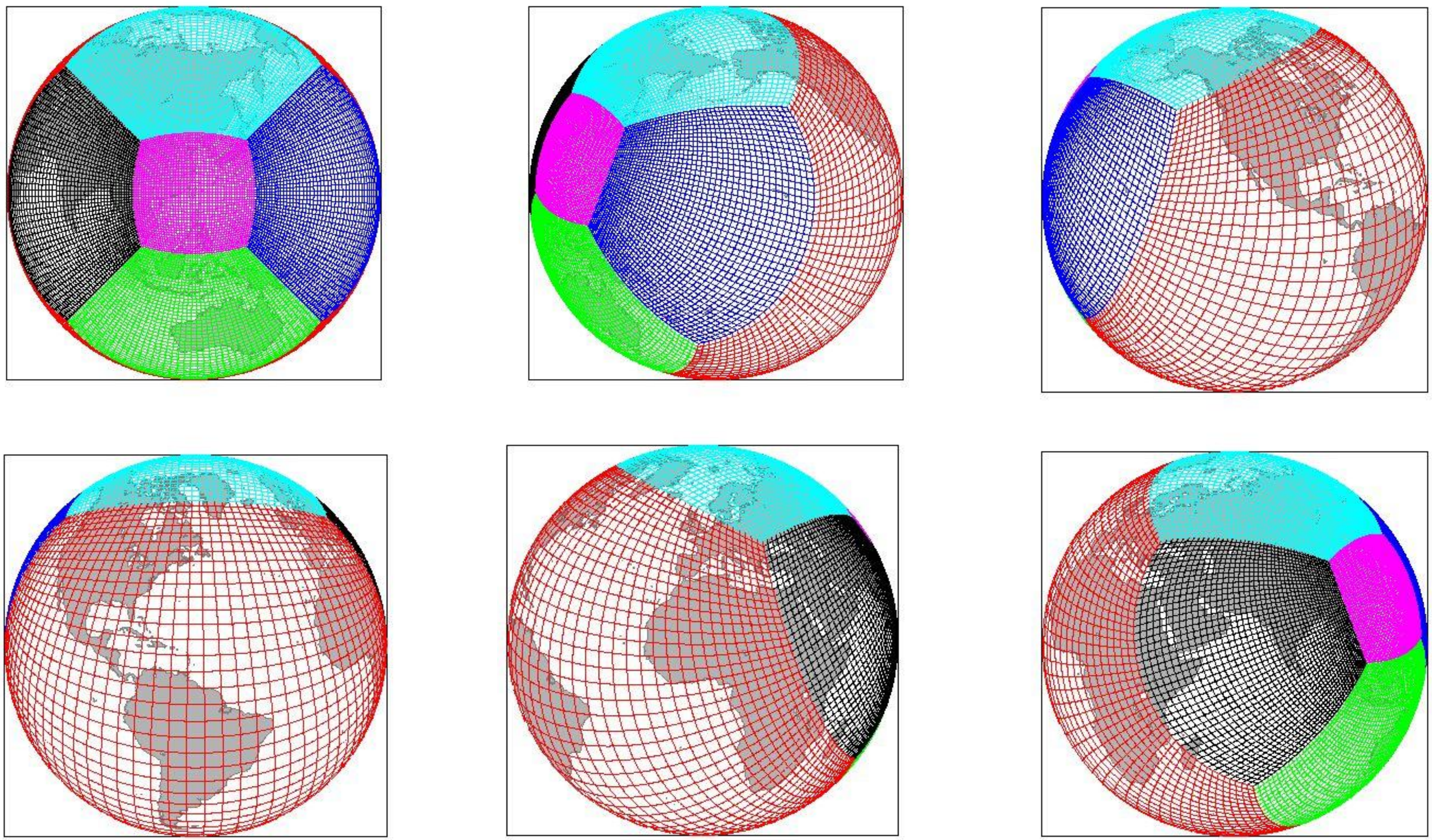


II. Variable resolution global model

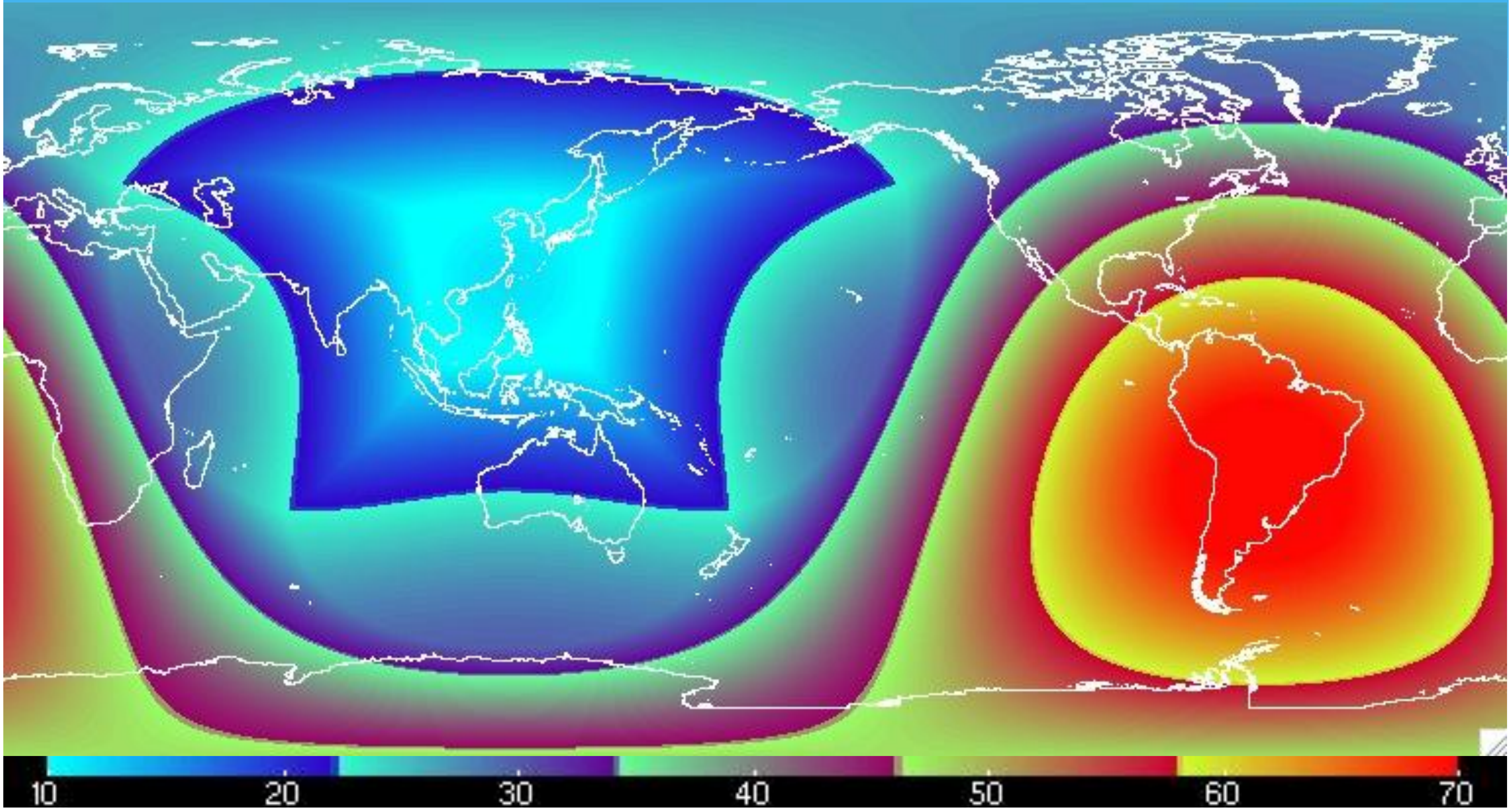
- Single model framework with smooth transition in resolution (e.g., 10 km over Taiwan stretched to 68 km over Andes Mountains)
- Suitable for weather & climate



Stretched Cube Sphere Globe (center @ TW) C384 Stretched Factor 2.5 ($\Delta x \approx$ 10km~68km)



Grid Size of Stretched Cube Sphere Globe C384R2.5 center @ Taiwan



- horizontal grid spacing for C384R2.5 ranges from around 10km to 68km.

Computing & Storage Demand for CMIP6 (ALPS: 177 T-Flops, 25,600 CPUs, 225M SU/yr)

MIPs	Model	Computing	Storage
CMIP6	TaiESM 1º 991 model years	4.31M SU (8*544 SU/yr)	60TB (60GB/yr)
HighResMIP	HiRAM C192 & C384 66 model years (Tier 1 only, no DECK)	4.84M SU (73,344 SU/yr)	370TB ((4.6+1)TB/yr)
PMIP	TaiESM 1º 200 model years	0.87M SU	12TB
CFMIP	TaiESM 1º 174 model years	0.76M SU	10.4TB
AerChemMIP	TaiESM 1º 600 yrs AMIP+1274 yrs coupled	8.11M SU	112.4TB
GMMIP	HiRAM C192 80 yrs extra from HighResMIP	0.7M SU	80TB (1TB/yr)
Total		**~20M SU**	**~650 TB**

Experimental Platform for Weather/Climate Data Processing and Visualization

ALPS Acer AR585 F1 Cluster
-177 TFLOPS
-25,600 compute cores

Storage Total : 160 TB
NAS Server TS-EC2480U-RP (24bay) - 96TB
QNAP REXP-1600U-RP (16bay) – 64TB

18 Slave Nodes
ACER Veriton M2611G
CPU:Intel i7-3770@3.40GHz(4Coresx2Threads)
RAM:16GB
HDD: 4TB

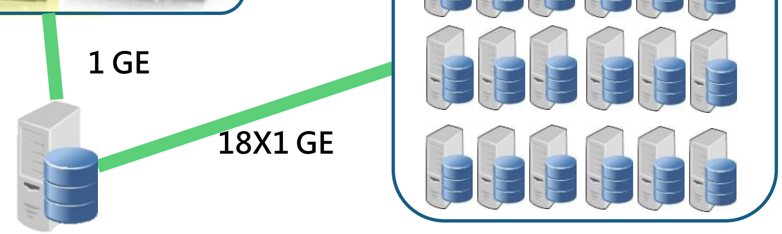
Hardware

10GE

1 GE

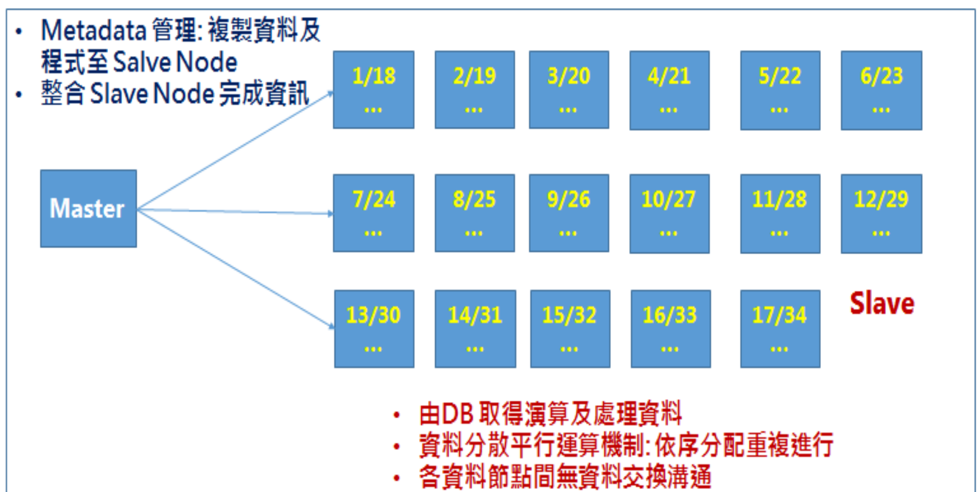
18X1 GE

Master Node



Software

Linux + Mysql



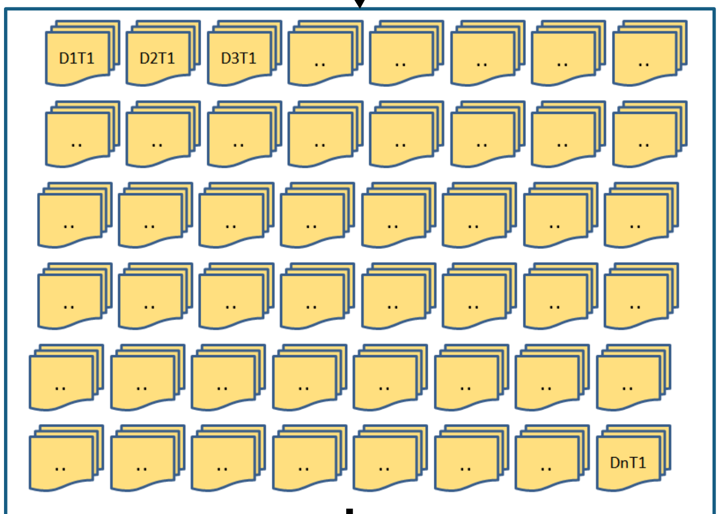
NetCDF File Partitioning

netCDF Library/nc_reader.f

1. **Install netCDF library (C /Fortran)& cdi to BigDataLab cluster**
2. **Use netCDF Fortran API to develop general netCDF reader (nc_reader.f)**
3. **Read netCDF file (nc) and partition the variables to multi blocks (by time)**
4. **Distribute these blocks to DataNodes and parallel load blocks into local MariaDB**
5. **Save dimension, unit, lon, lat, lev, time, varBlock tables to MariaDB**
6. **Save varBlockList tables, HostTable to master MariaDB**
7. **Will try to split variables data by time & level**

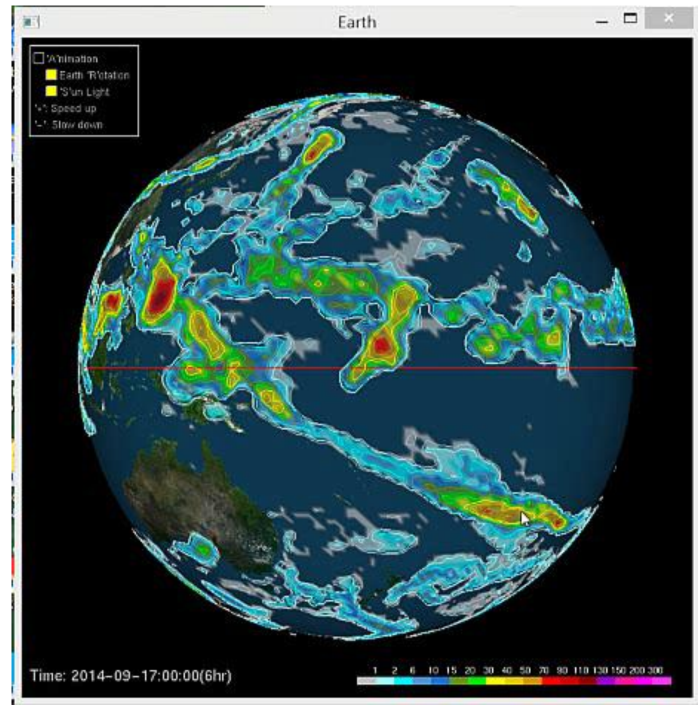
gfortran nc_reader.f -o nc_reader -I${NETCDFDIR}/include –L {NETCDFDIR}/lib -lnetcdff

NetCDF File → NetCDF File Partition



Shared File System

Distributed Local DataBases → Parallel Reader / BigData Analysis / Visualization

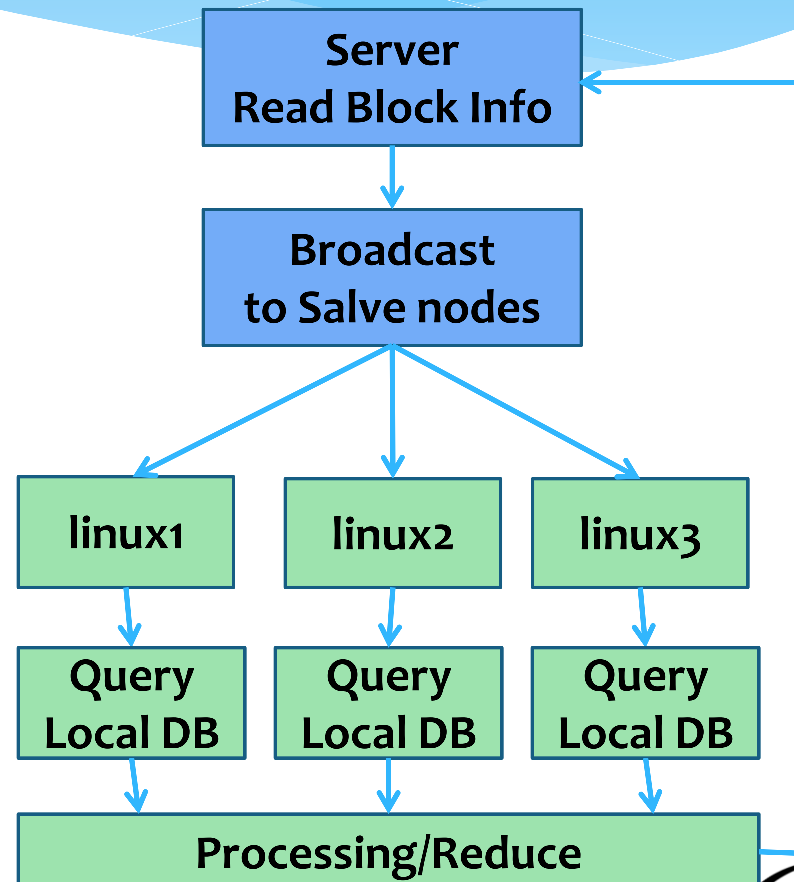


Multi-DB Query & Post Processing

```
|     132 | st20140824750000blk | 20140824.75 | linux14 |
|     133 | st20140825000000blk |    20140825 | linux15 |
|     134 | st20140825250000blk | 20140825.25 | linux16 |
|     135 | st20140825500000blk |  20140825.5 | linux17 |
|     136 | st20140825750000blk | 20140825.75 | linux18 |
|     137 | st20140826000000blk |    20140826 | linux1  |
|     138 | st20140826250000blk | 20140826.25 | linux2  |
|     139 | st20140826500000blk |  20140826.5 | linux3  |
|     140 | st20140826750000blk | 20140826.75 | linux4  |
|     141 | st20140827000000blk |    20140827 | linux5  |
|     142 | st20140827250000blk | 20140827.25 | linux6  |
|     143 | st20140827500000blk |  20140827.5 | linux7  |
|     144 | st20140827750000blk | 20140827.75 | linux8  |
|     145 | st20140828000000blk |    20140828 | linux10 |
|     146 | st20140828250000blk | 20140828.25 | linux11 |
|     147 | st20140828500000blk |  20140828.5 | linux12 |
|     148 | st20140828750000blk | 20140828.75 | linux13 |
|     149 | st20140829000000blk |    20140829 | linux14 |
|     150 | st20140829250000blk | 20140829.25 | linux15 |
|     151 | st20140829500000blk |  20140829.5 | linux16 |
```

```
cluster: linux1, size: 11
cluster linux1 got 20140723.000000, st20140723000000blk
cluster linux1 got 20140727.250000, st20140727250000blk
cluster linux1 got 20140809.000000, st20140809000000blk
cluster linux1 got 20140731.500000, st20140731500000blk
cluster linux1 got 20140821.750000, st20140821750000blk
cluster linux1 got 20140826.000000, st20140826000000blk
cluster linux1 got 20140817.500000, st20140817500000blk
cluster linux1 got 20140830.250000, st20140830250000blk
cluster linux1 got 20140804.750000, st20140804750000blk
cluster linux1 got 20140903.500000, st20140903500000blk
cluster linux1 got 20140813.250000, st20140813250000blk
```

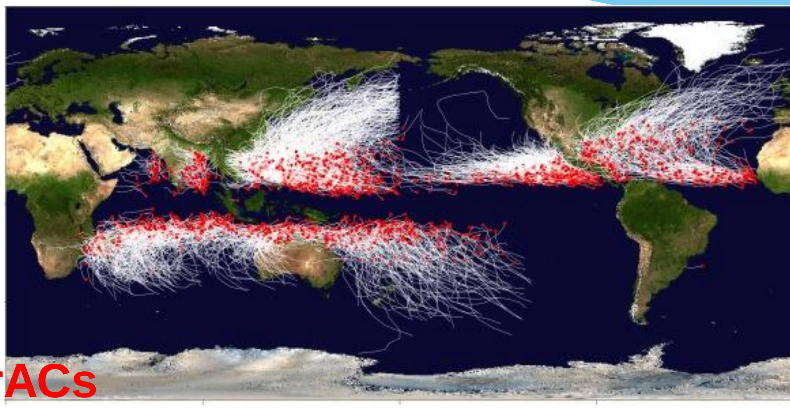
* Using MPI



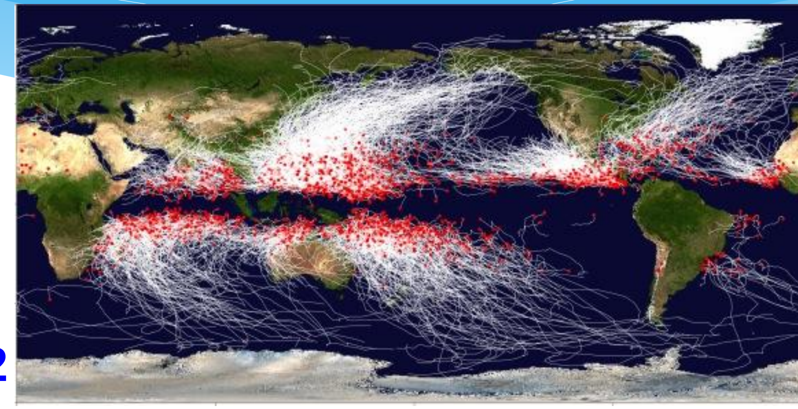
Tropical Cyclones and Extreme Precipitation in HiRAM

Typhoon and Hurricanes

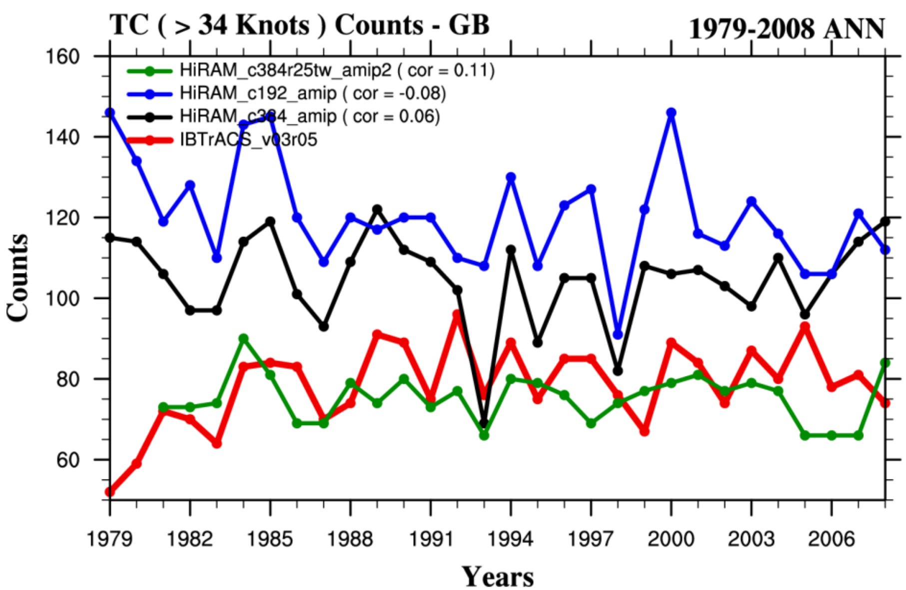
Typhoon Tracks (1979~2008)

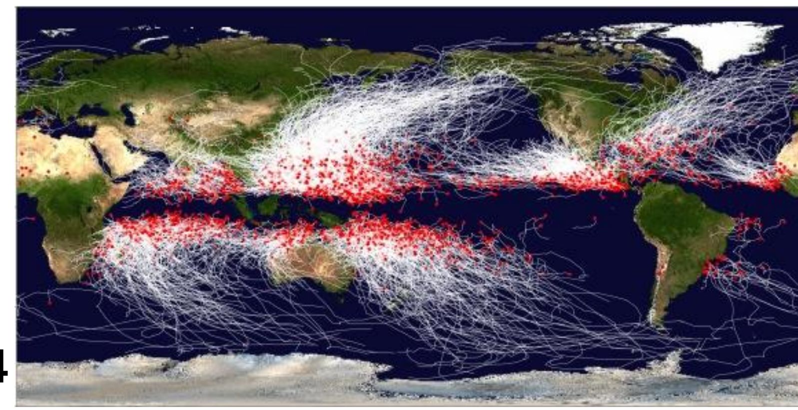


IBTrACs

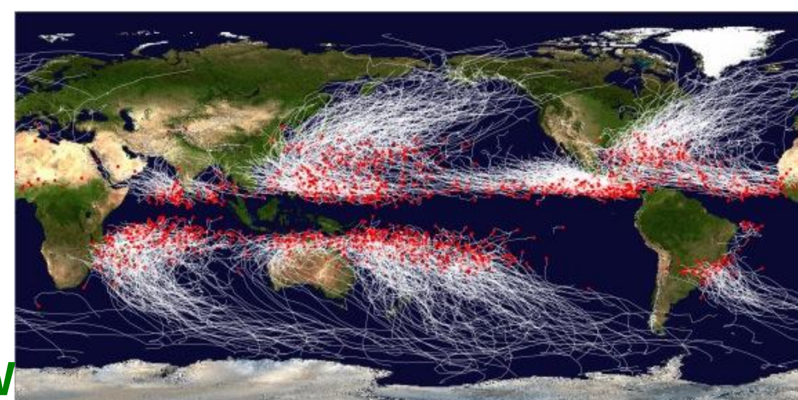


C192



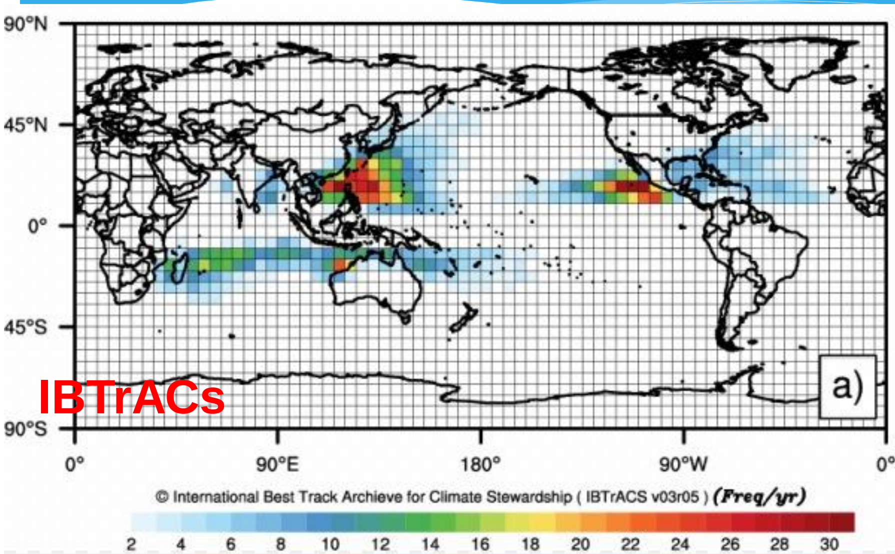


C384

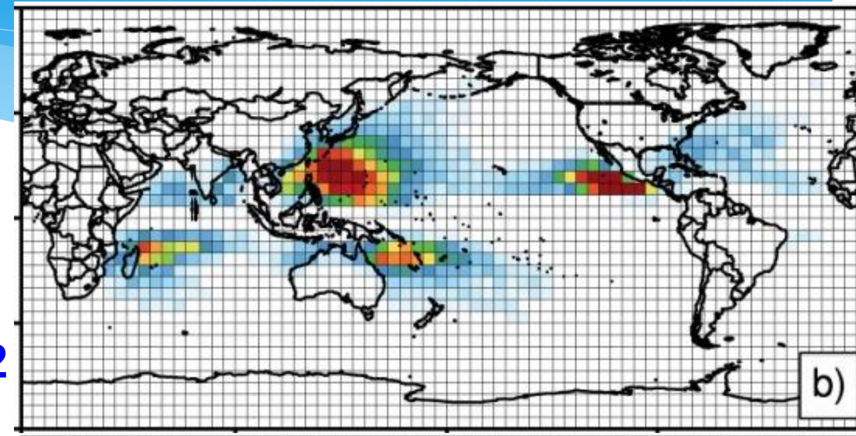


C384R2.5TW

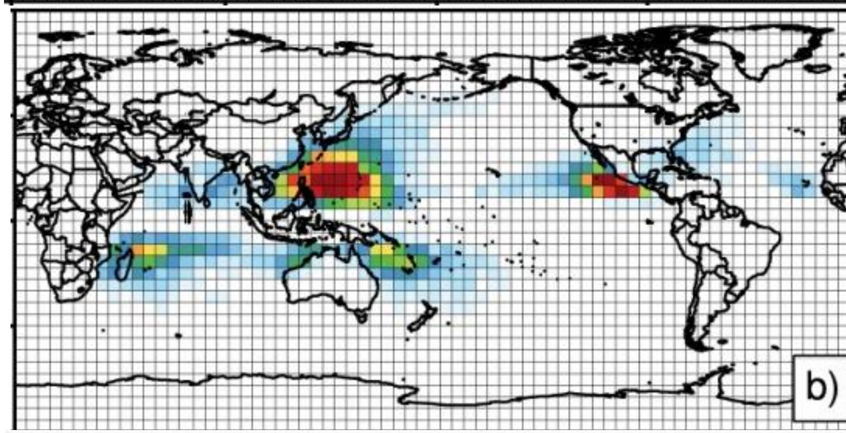
Typhoon Density (1979~2008)

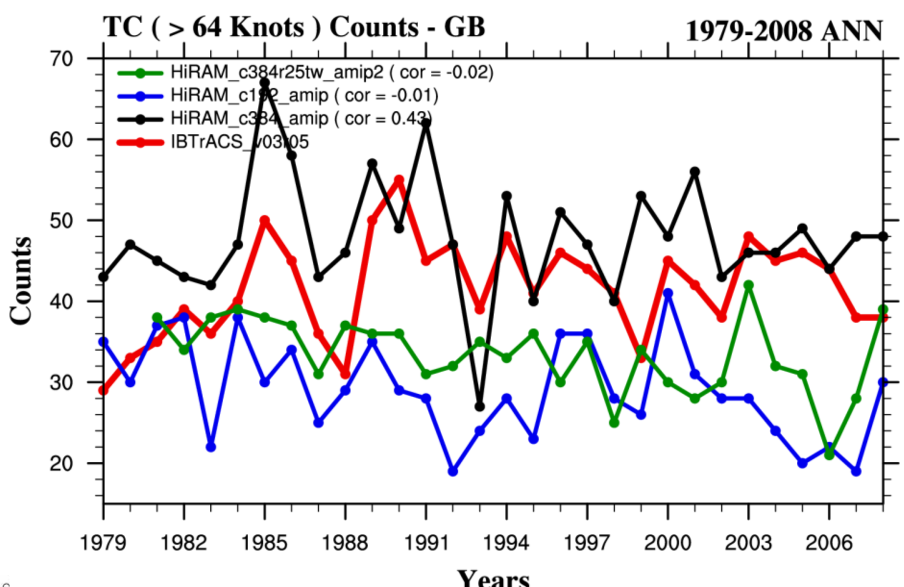


C192

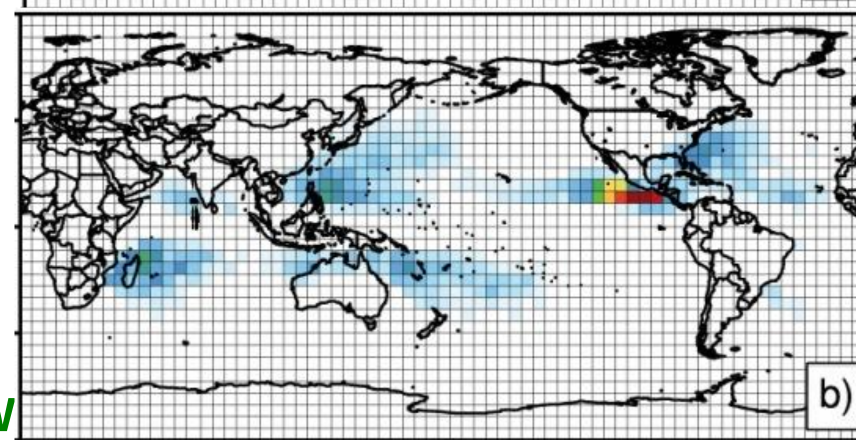


C384



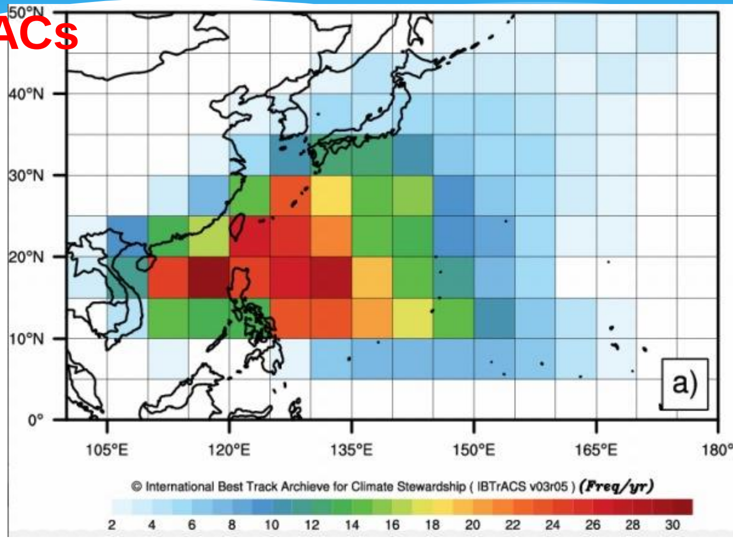


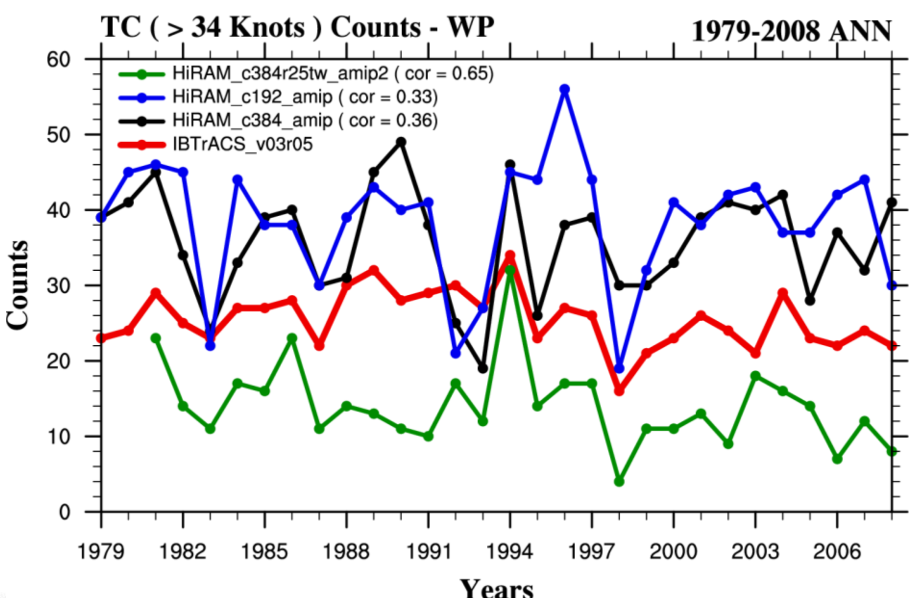
C384R2.5TW

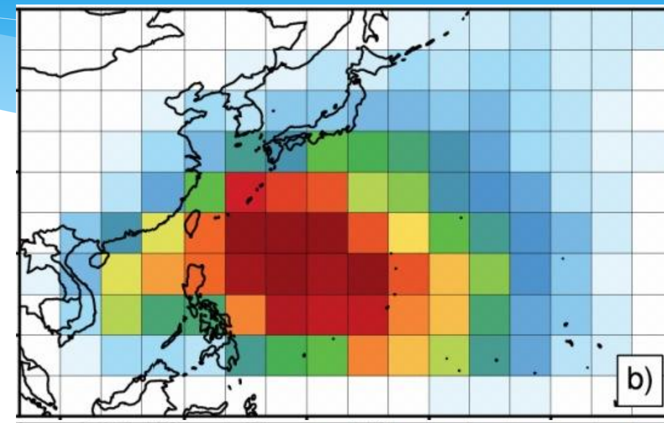


Western North Pacific - Typhoon Density (1979~2008)

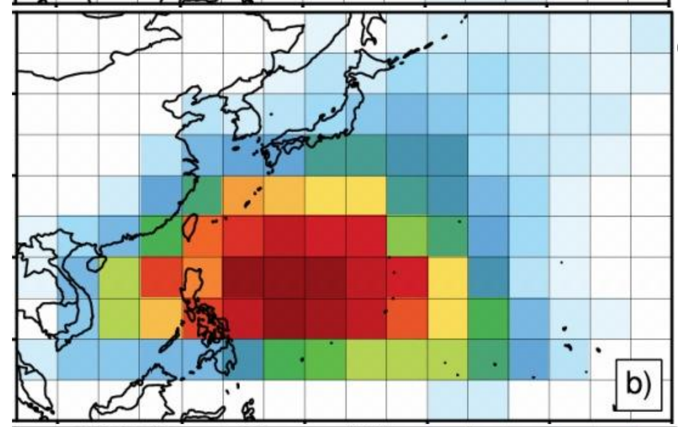
IBTrACs



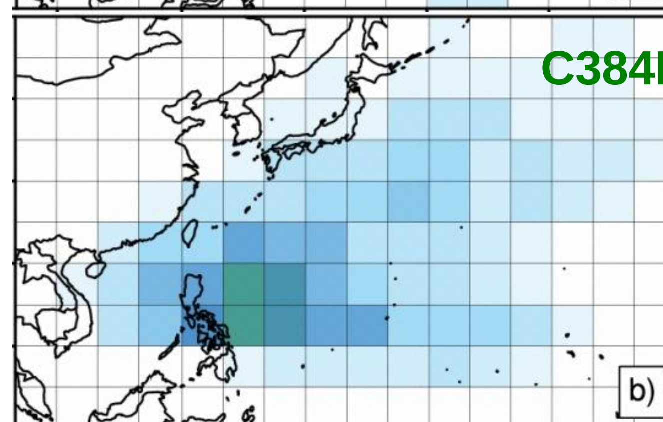




C192

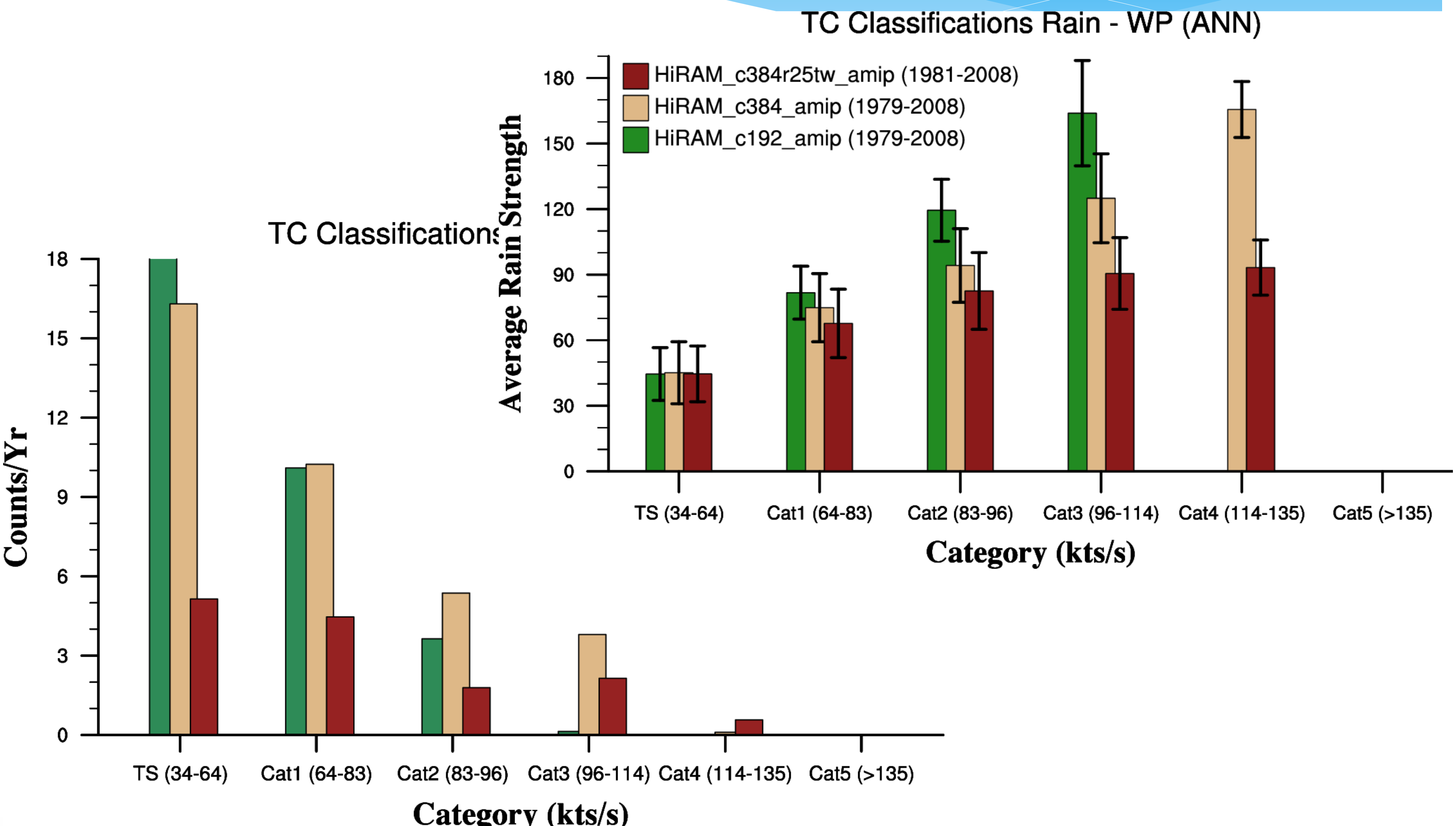


C384

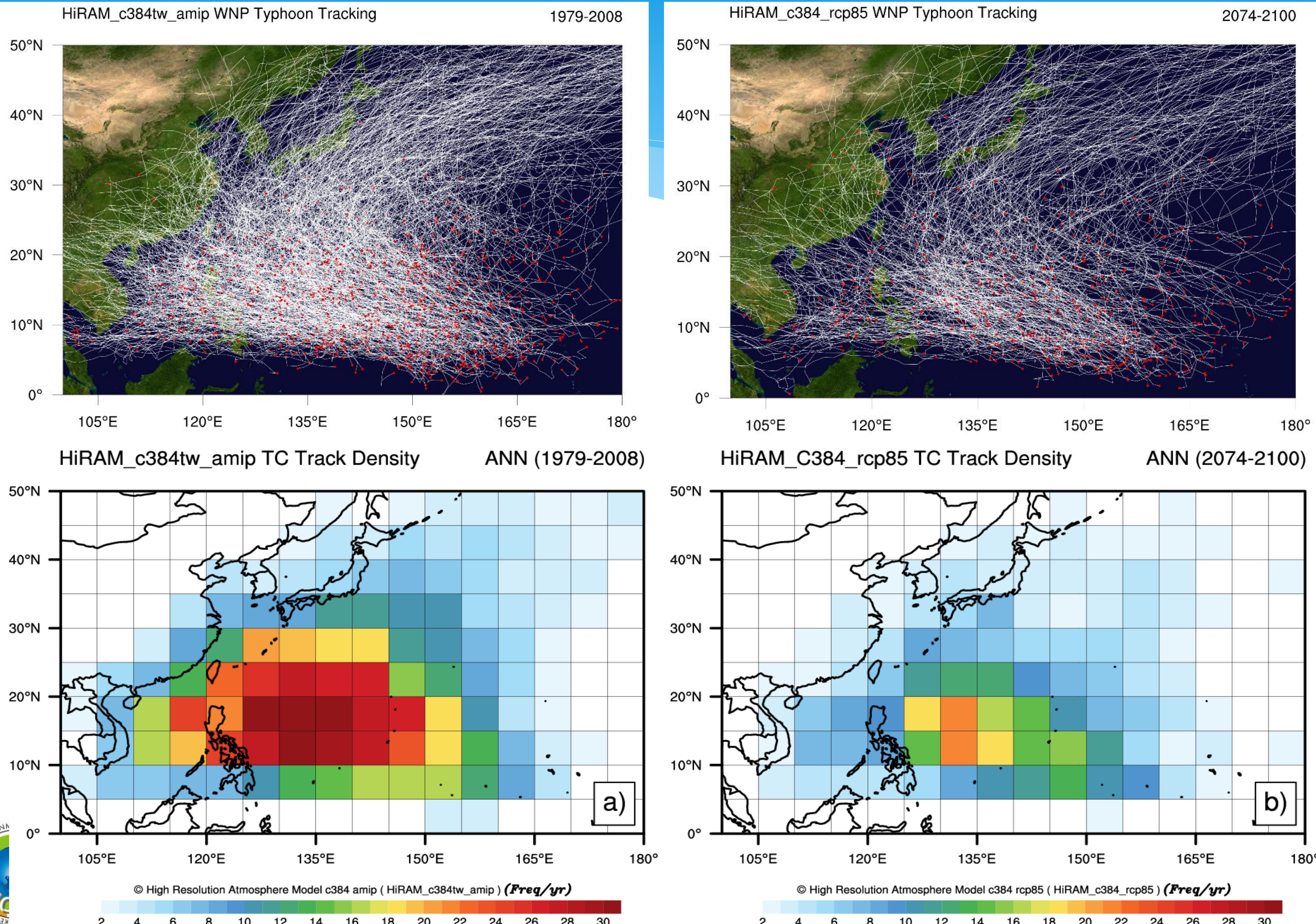


C384R2.5TW

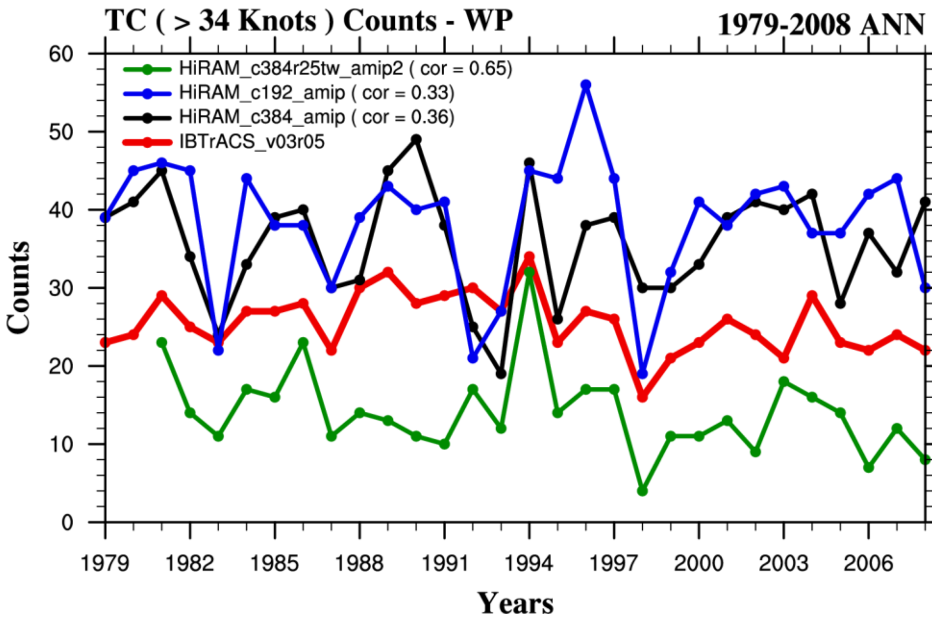
Western North Pacific - Typhoon Counts & Precipitation

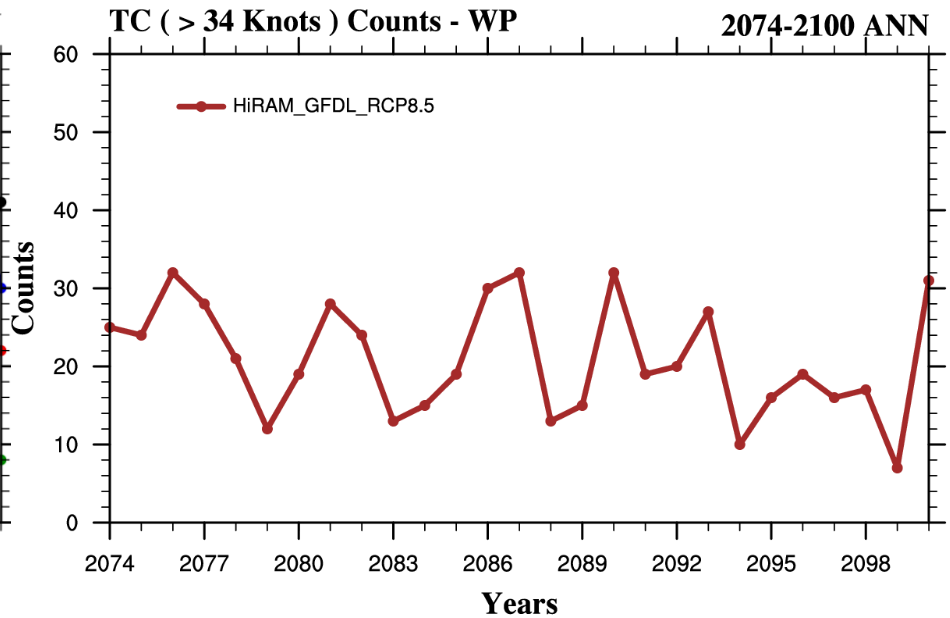


Western North Pacific – Typhoon Future Trend



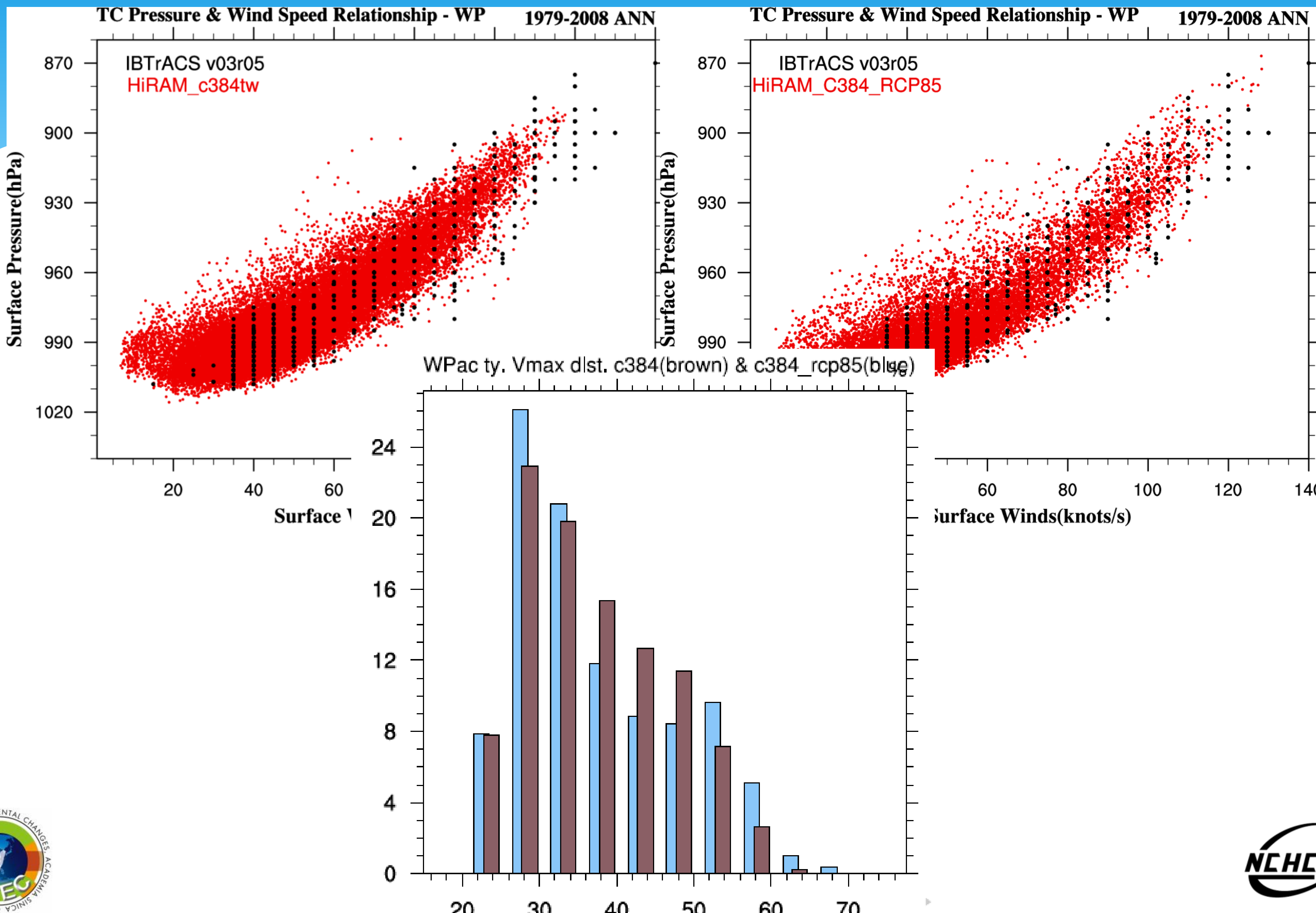
Western North Pacific – Typhoon Future Trend



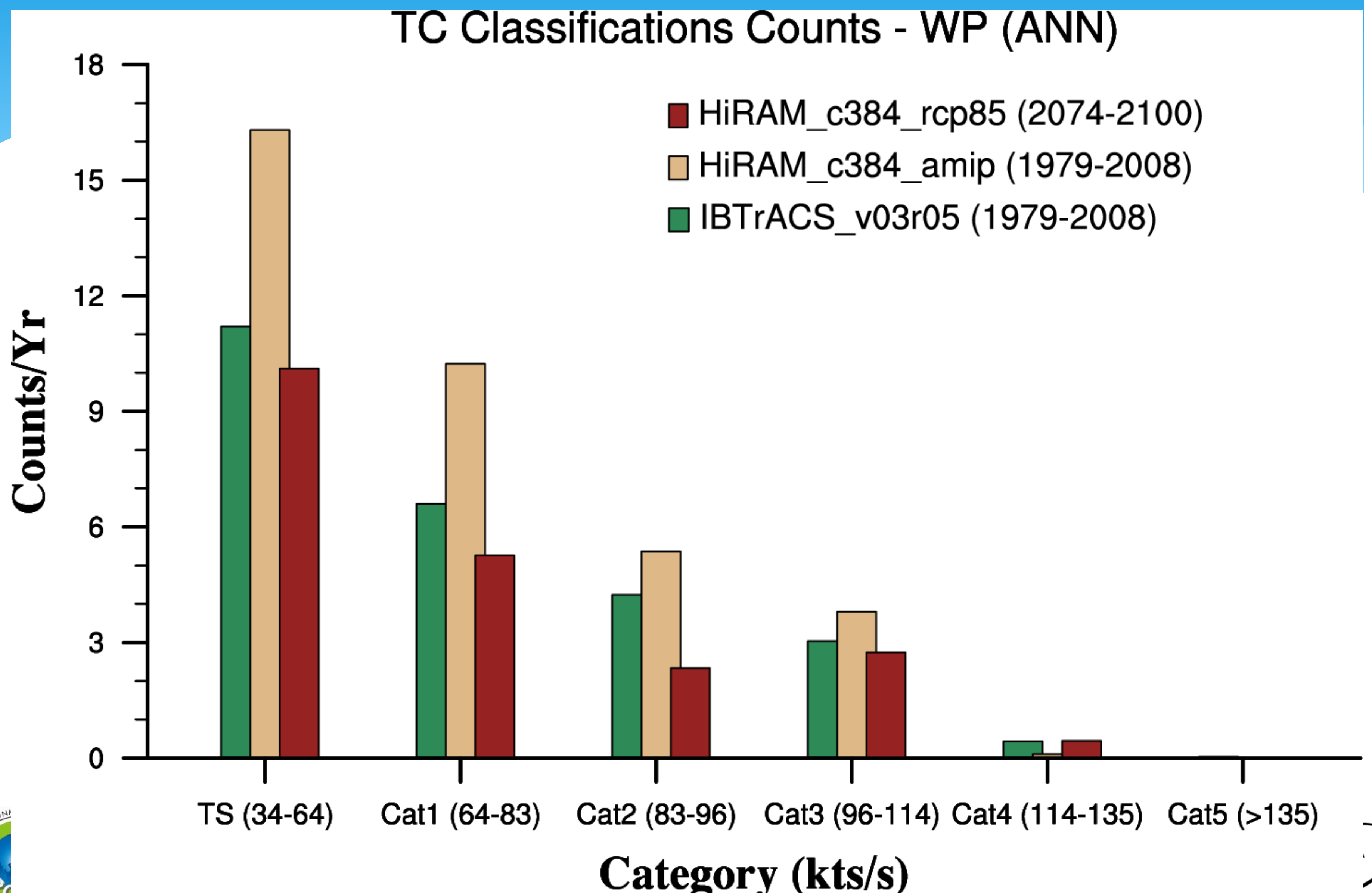


	> 34 knots (TS)	> 64 Knots (Cat1)
IBTrACs_v03r05	25.5	14.3
HiRAM_c384tw_amip	35.8	19.5
HiRAM_c384_rcp85	20.89	10.78

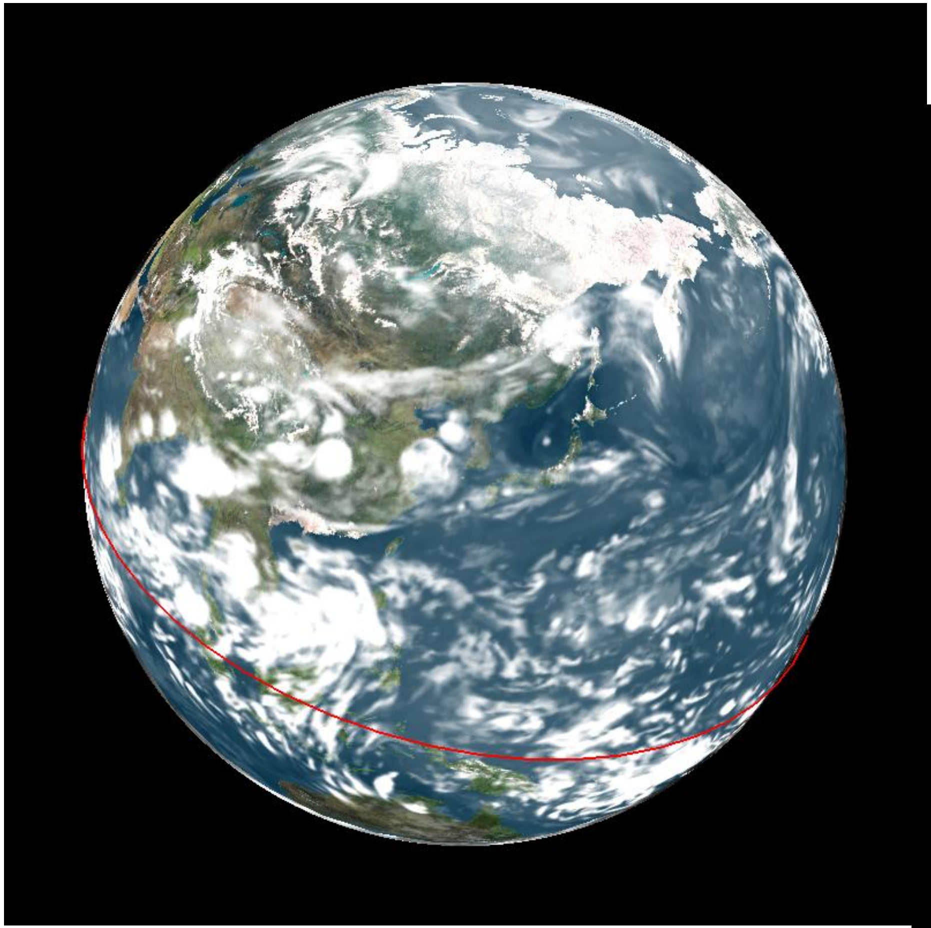
Western North Pacific – Typhoon Future Trend

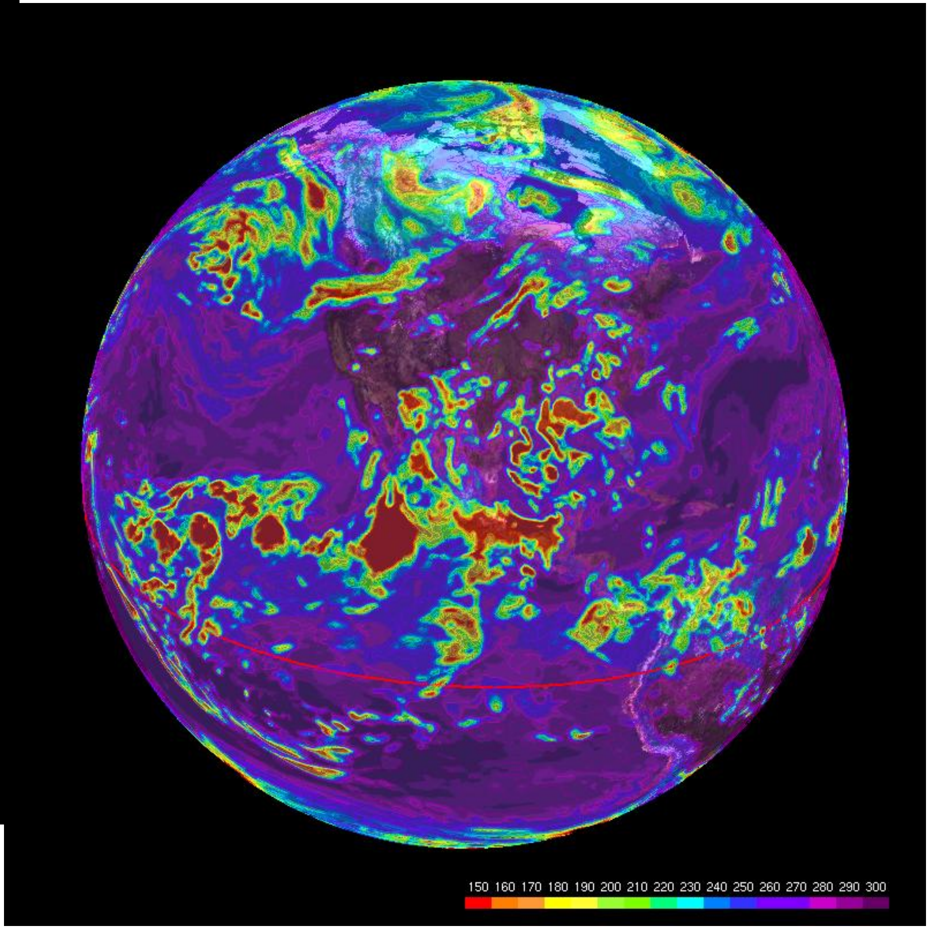


Western North Pacific – Typhoon Future Trend

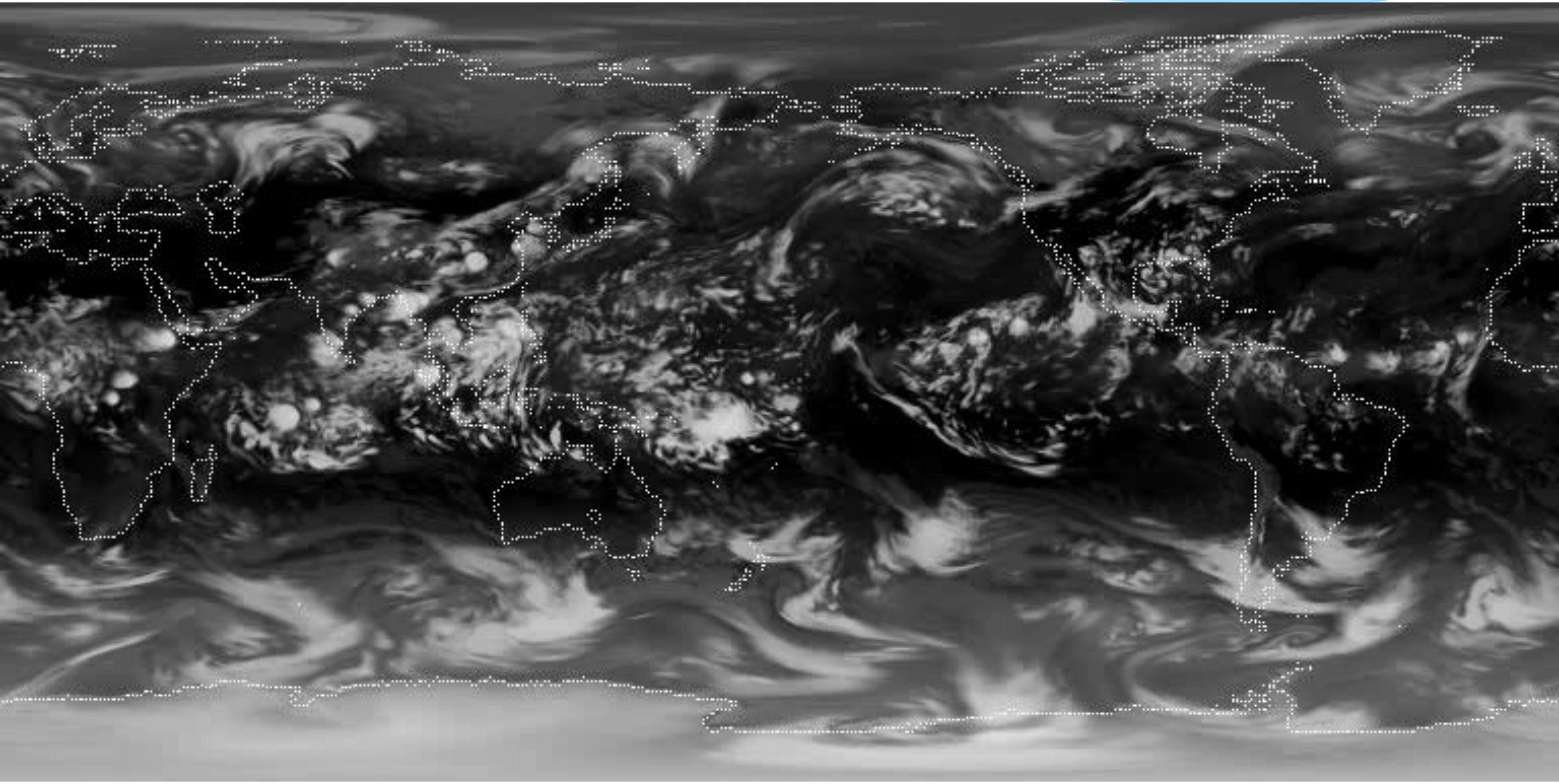


Visualization

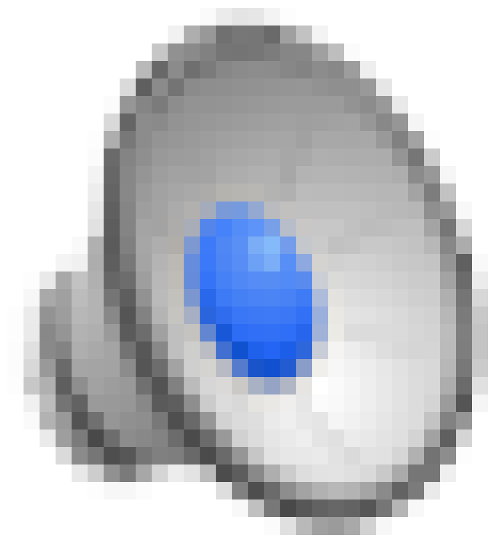




Visualization



Visualization



Feature Detection from Visual Animation

Formation of Typhoon eddies are detected through image pattern recognition

Concluding Remarks

- ***Here we present a big data experimental platform for climate data processing and visualization.***

- ***This platform links computing environment for global climate simulation and big data analysis system that improves data management and processing power.***

- ***Climate data analysis is more complicated and much larger in data size than what we have shown.***

- ***NetCDF is wide used in weather/climate data, but is only popular in atmospheric research community.***

Performance

Table 1、Experimental Data

	Atmos_20750801	Case20140723	Case20090801
Block Number	248	181	36
Block Size	1,179,648	819,200	18,874,368
Total Size(byte)	7,021,264,896	3,558,604,800	16,307,453,952

Table 2、Execution time (ms)

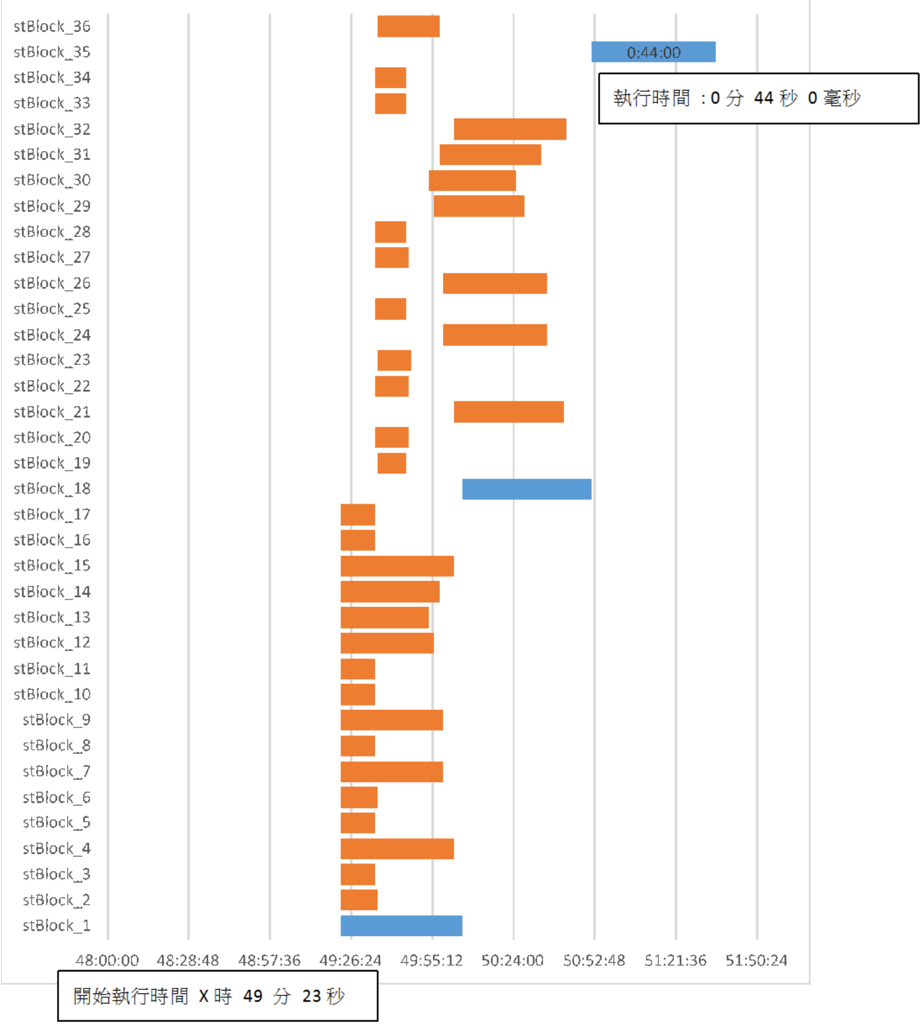
	Atmos_20750801	Case20140723	Case20090801
Sequence(ms)	364,360	278,759	1,154,360
1 (task/node)(ms)	46,396	23,846	152,326
2 (task/node)(ms)	42,516	22,969	138,805
4 (task/node)(ms)	39,126	20,958	118,652
8 (task/node)(ms)	33,173	19,982	x
16 (task/node)(ms)	27,666	16,661	x

Linux1-linux8, linux10-linux18
RAM:8G (600MB-1.0GB free)

Nb=L*N+M
Nb=36, N=17 => L=2, M=2

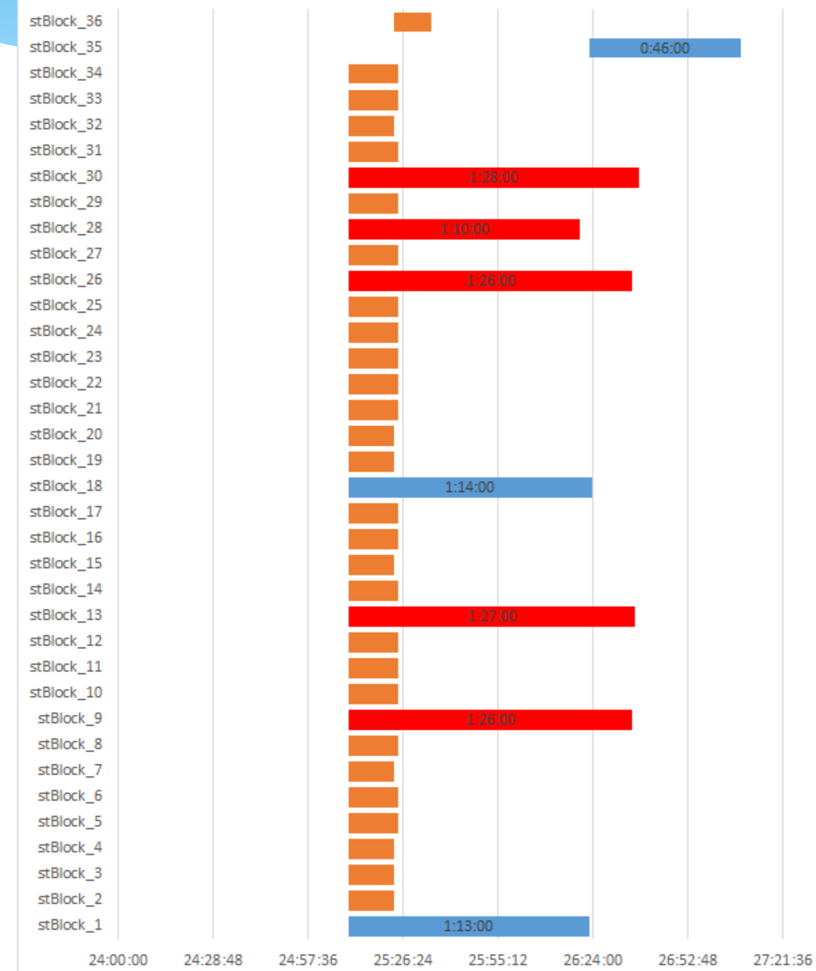
Performance

1 task/node



Nb=L*N+M=>Nb=36, N=17 => L=2, M=2

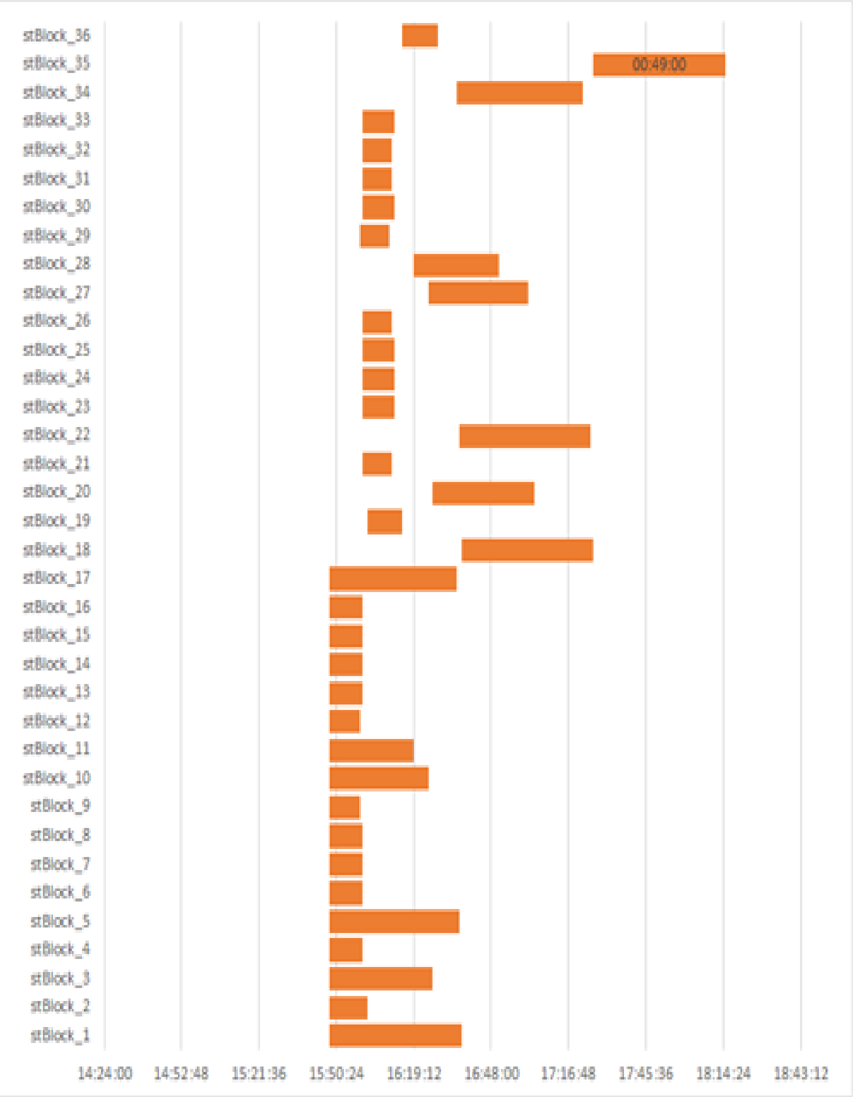
2 tasks/node

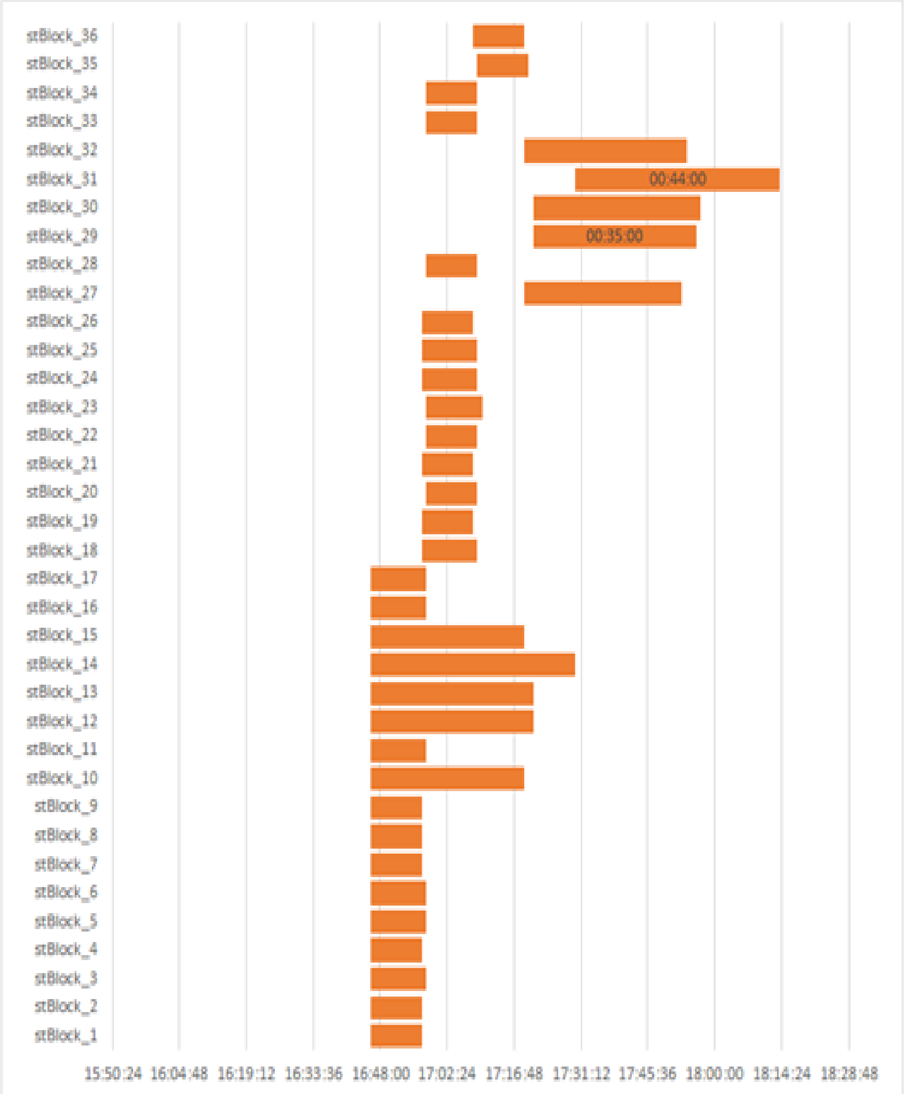


Nb=L*N+M=>Nb=36, N=34 => L=1, M=2

Performance

1 (task/node)





In Place Query Driven Post Processing for Global Climate Simulation

atmos_8xdaily_20750801.nc (1536x768x248)

1. On linux250 with **sequential** process : 670 sec

2. Distributed DB with local query & process (linux1-linux8, linux10-linux18) then send back to linux250: 58 sec

3. Nb=248=14(L)x17(N)+10(M)=15(L) => 58/15=3.87s

4. 670/248=2.7s