TECA, 13TB, 80,000 processors

Or: Characterizing extreme weather in a changing climate

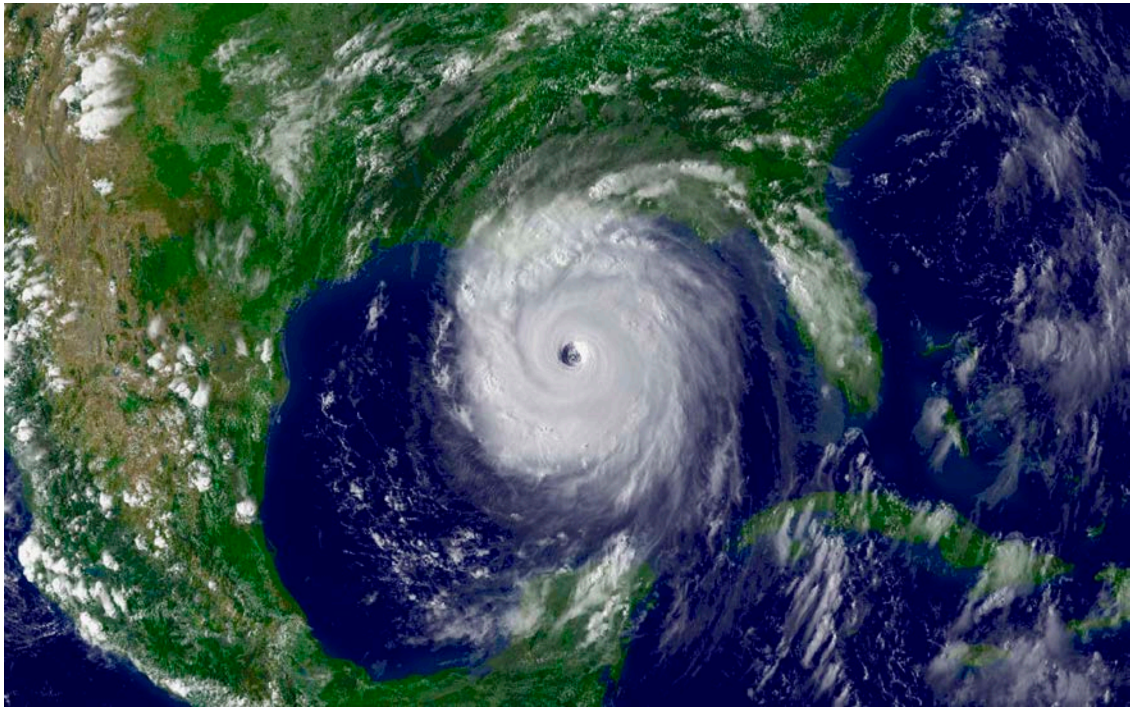
Second Workshop on Understanding Climate Change from Data
University of Minnesota

Michael F. Wehner, Prabhat, Surendra Byna, Fuyu Li, Erin LeDell, Thomas Yopes, Gunther Weber, Wes Bethel, William D. Collins

Lawrence Berkeley National Laboratory
mfwehner@lbl.gov

Extreme weather in a changing climate

- Severe storms
  - Hurricanes (Tropical Cyclones)
  - Extra Tropical Cyclones
  - Atmospheric Rivers
  - Mesoscale Convective Systems
- Blocking events
  - Heat waves and droughts
  - Cold snaps

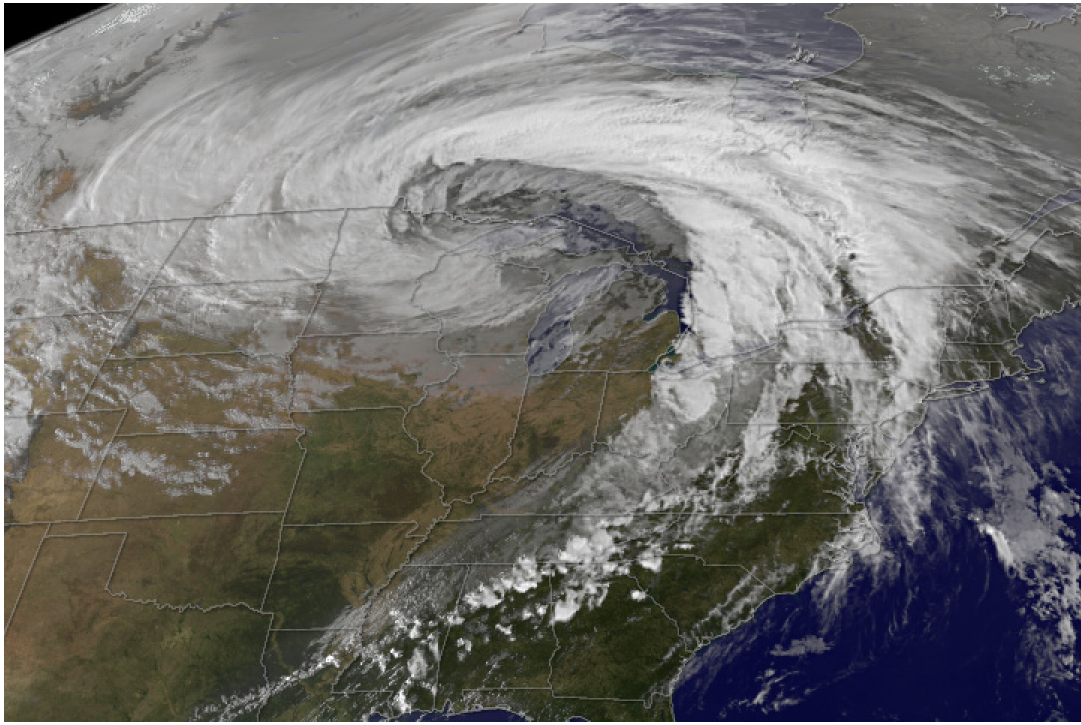


NASA/NOAA

Extreme weather in a changing climate

- **Severe storms**
  - **Hurricanes (Tropical Cyclones)**
  - **Extra Tropical Cyclones**
  - **Atmospheric Rivers**
  - **Mesoscale Convective Systems**

- **Blocking events**
  - **Heat waves and droughts**
  - **Cold snaps**



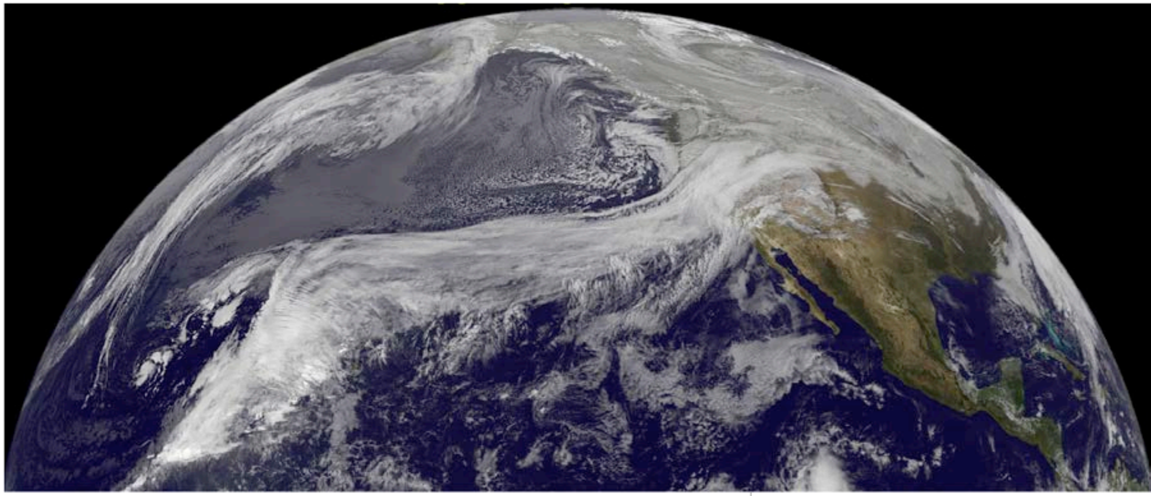
NASA GOES

Extreme weather in a changing climate

- **Severe storms**
  - **Hurricanes (Tropical Cyclones)**
  - **Extra Tropical Cyclones**
  - **Atmospheric Rivers**
  - **Mesoscale Convective Systems**

- **Blocking events**
  - **Heat waves and droughts**
  - **Cold snaps**

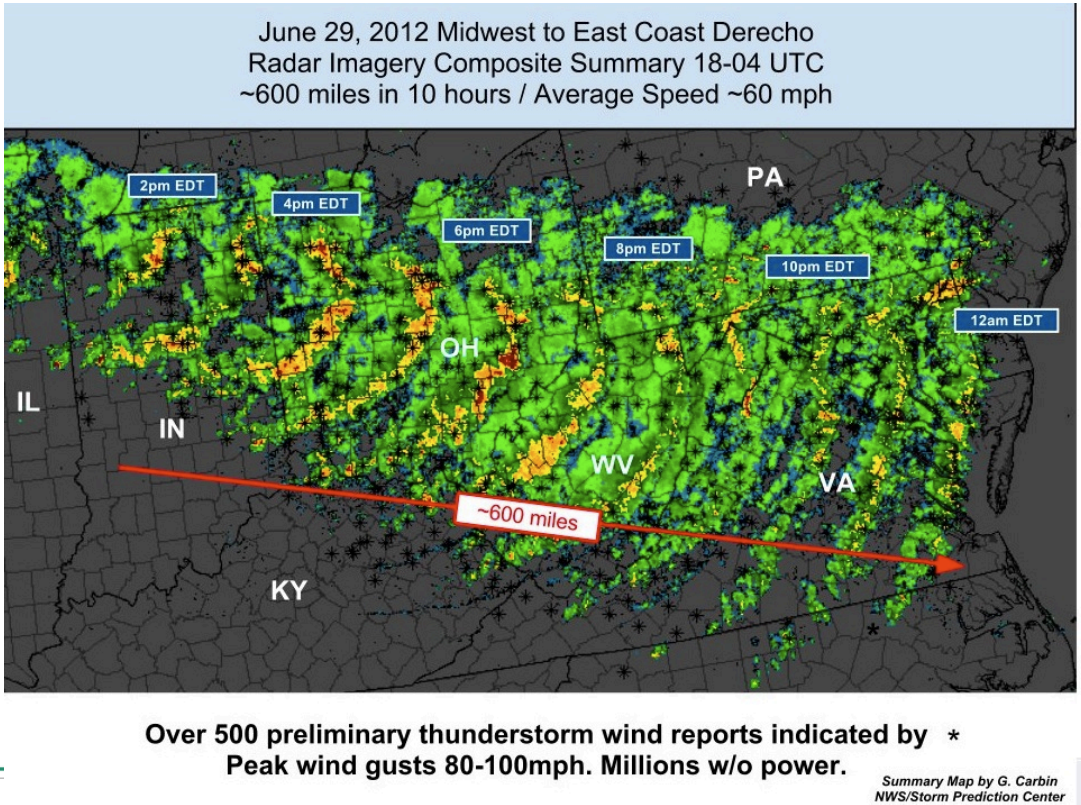
NASA



Extreme weather in a changing climate

- **Severe storms**
  - Hurricanes (Tropical Cyclones)
  - Extra Tropical Cyclones
  - Atmospheric Rivers
  - **Mesoscale Convective Systems**

- Blocking events
  - Heat waves and droughts
  - Cold snaps

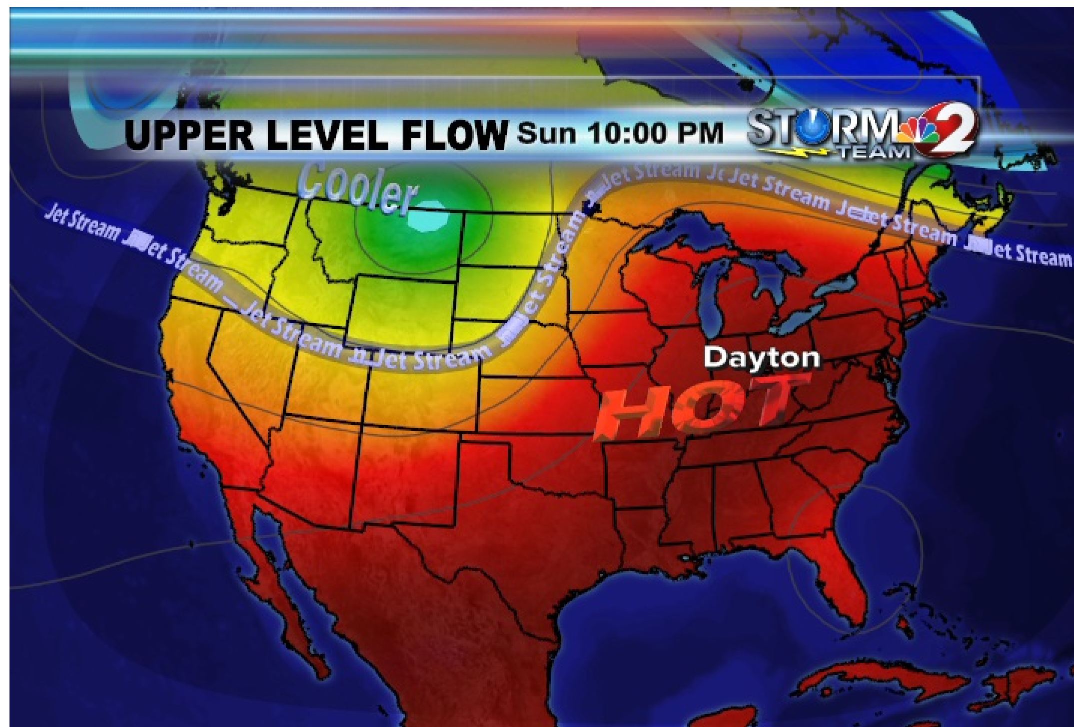


NOAA NWS

Extreme weather in a changing climate

- **Severe storms**
  - Hurricanes (Tropical Cyclones)
  - Extra Tropical Cyclones
  - Atmospheric Rivers
  - Mesoscale Convective Systems

- Blocking events
  - **Heat waves and droughts**
  - Cold snaps

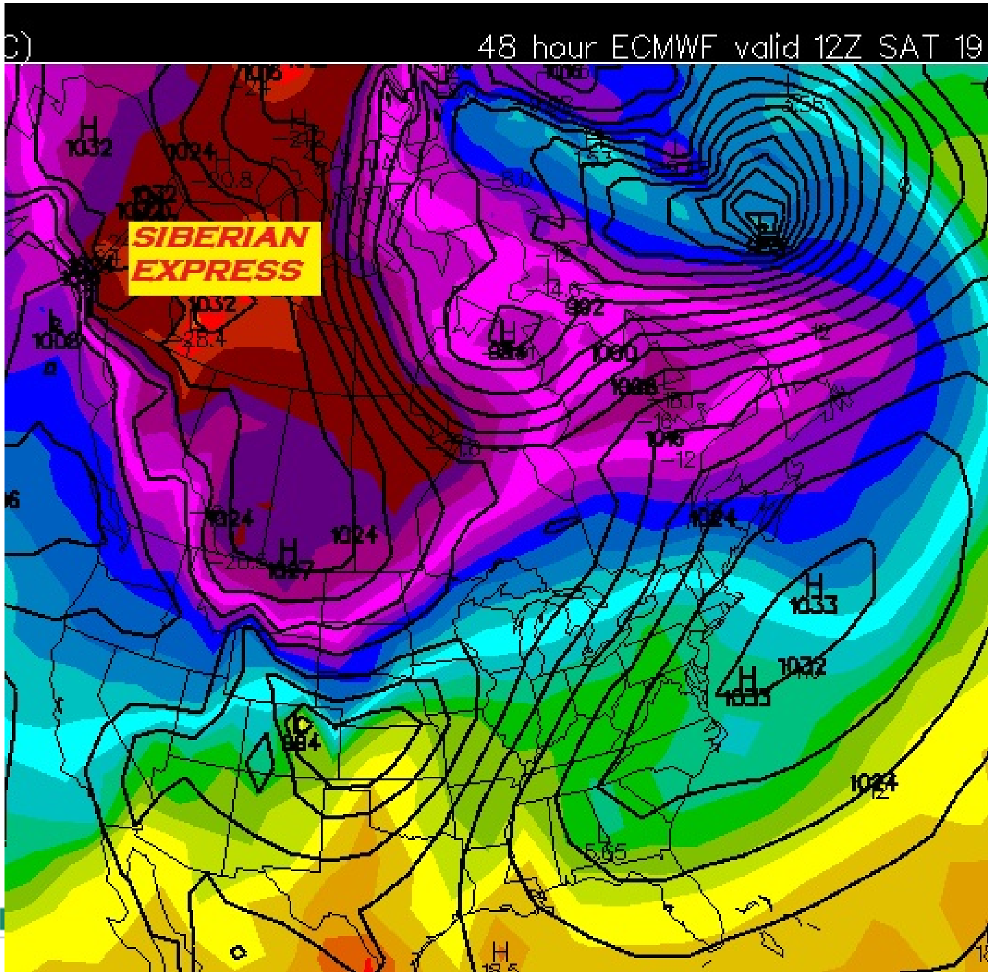


WDTN
Dayton, Ohio

Extreme weather in a changing climate

- **Severe storms**
  - Hurricanes (Tropical Cyclones)
  - Extra Tropical Cyclones
  - Atmospheric Rivers
  - Mesoscale Convective Systems

- Blocking events
  - Heat waves and droughts
  - **Cold snaps**



ECMWF
Weather underground

Challenges to indentify extreme events

- State of the art high resolution climate models can inform us about extreme weather changes, but
  - Unprecedented volumes of data need be generated.
    - We generated 100TB of output in a 26 year integration of a ~25km global atmospheric model (NCAR CAM5.1).
- Tracking extreme weather events is data intensive.
  - Scales poorly with resolution. In some cases as n^4.
    - Often high frequency (3 or 6 hourly)
  - Can be I/O bound on the input side.
  - Parallel processor tools are essential.
    - But not widely available to the climate model analyst community.

n=number of points on a horizontal direction

The abstraction:

- Identifying extreme weather events in high frequency climate model output involves two steps:
  1. Search through the data for candidate events at each individual time step that meet some defined criteria.
  2. Stitch together candidate events at multiple time steps, rejecting candidates that fail continuity criteria.
- Step 1 can be very computationally intensive. But is embarrassingly parallel across time steps
- Step 2 is cheap.

TECA: A Toolkit for Extreme Climate Analysis

- It is and must be highly parallel!

- Currently, we have implemented:
  - Hurricane tracking
  - Extratropical cyclone tracking
  - Atmospheric river identification
  - Atmospheric blocking

- Other extreme event opportunities exist as well as other variants of the present set of feature tracking algorithms.

GFDL hurricane tracking algorithm

1. Candidate detection:

– Find local vorticity maxima at 850 hPa exceeding $1.6*10^{-4}$ /s
– Find closest local minima in sea-level pressure (storm center) and maxima in 300-500 hPA temperature (warm-core center)
– Surface pressure should increase by 4hPa from the storm center within a ~400km radius
– Distance of warm-core center from storm center should not exceed ~200km.

2. Stitching:

– Search candidates within 400km radius over 6-hr window
– Look closer, westward and poleward candidates first
– Trajectory should last for 2 days, and have max surface wind velocity >17m/s during at least 2 days (but not necessarily consecutively)

Documented in Knutson, et al. (2008) BAMS

Tropical Storm

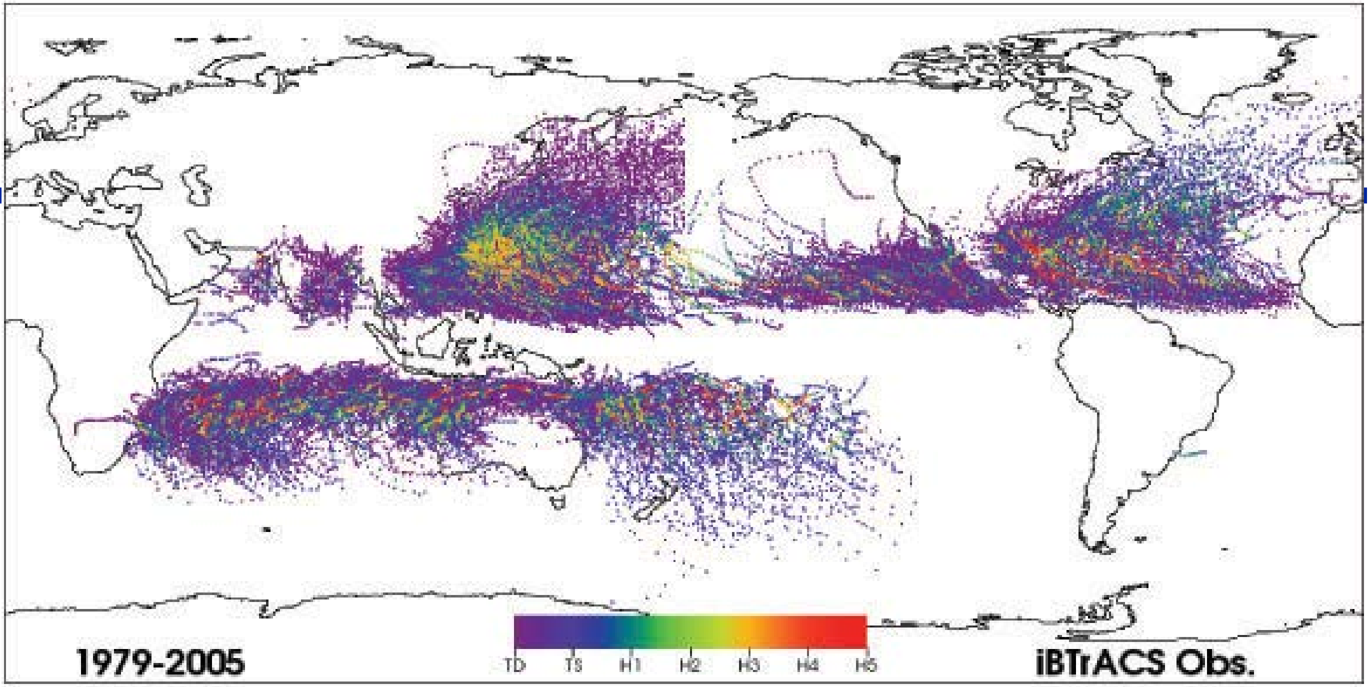
Cat1

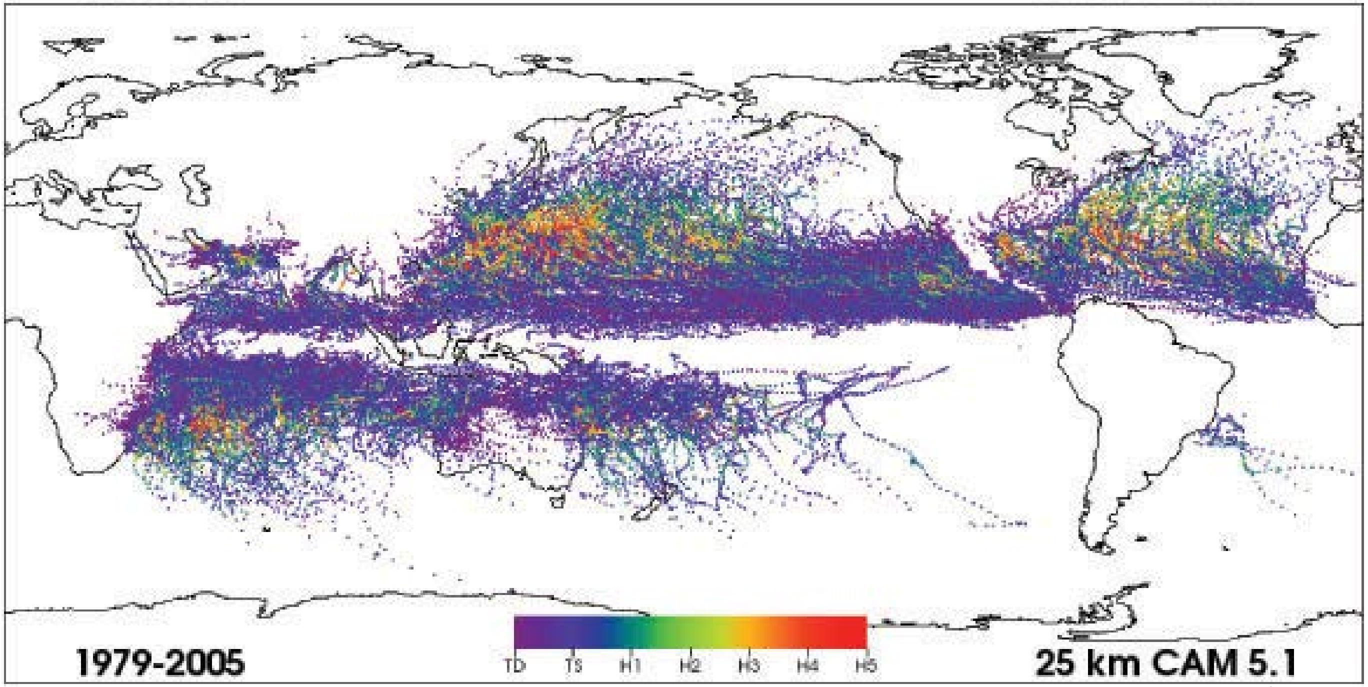
Cat2

Cat3

Cat4

Cat5





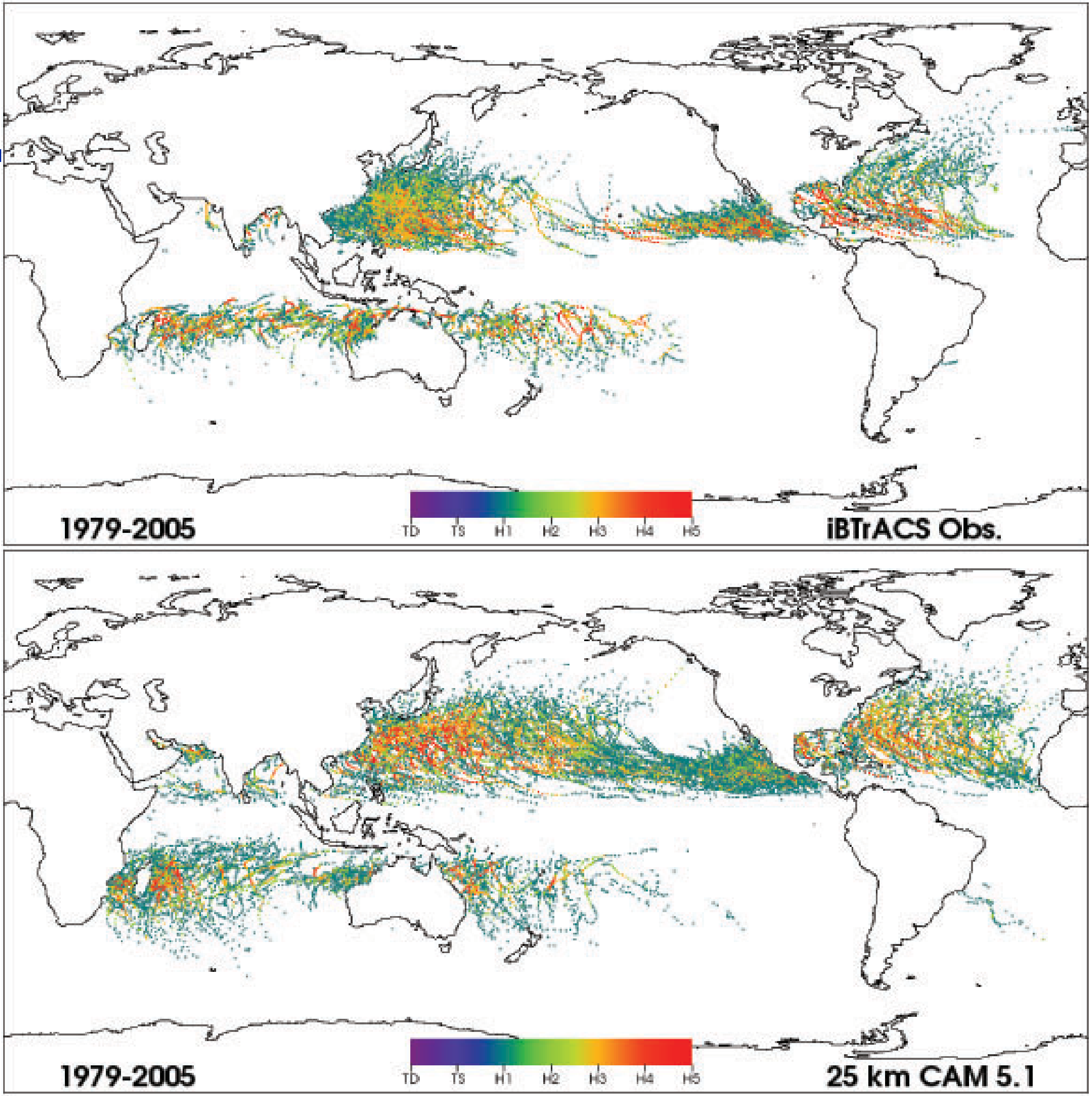
Cat1

Cat2

Cat3

Cat4

Cat5

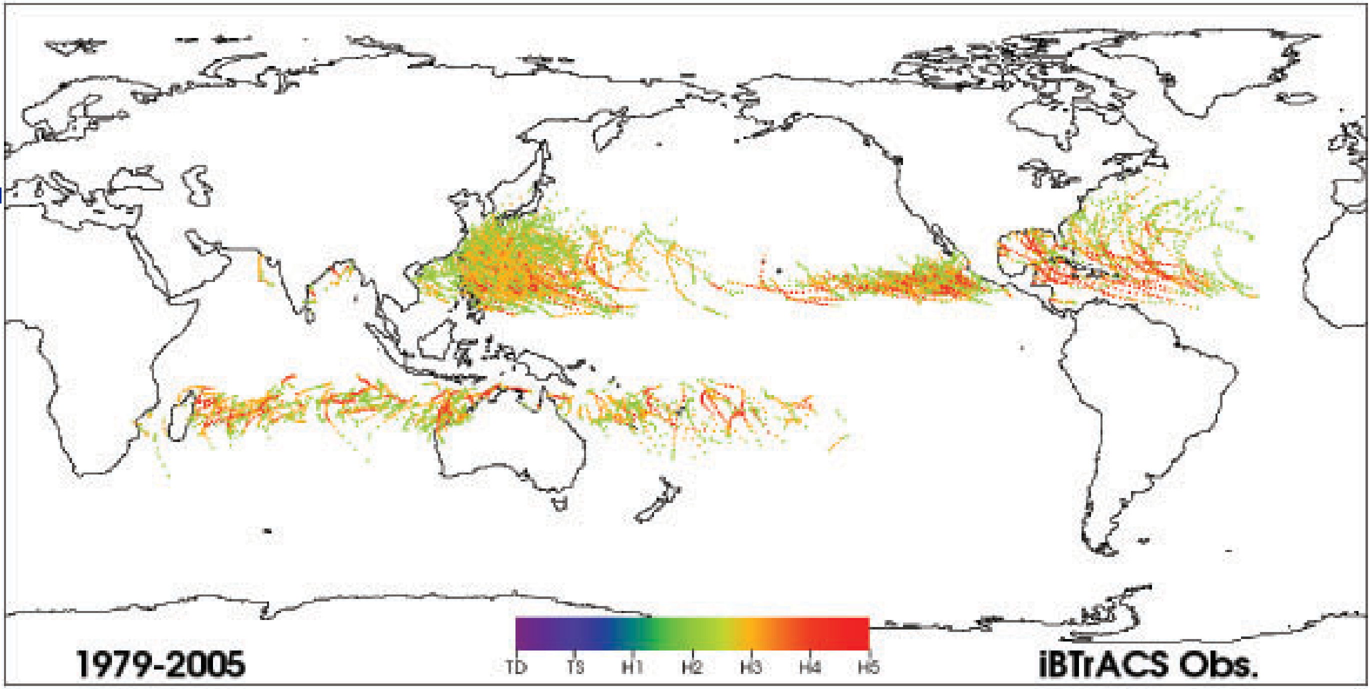


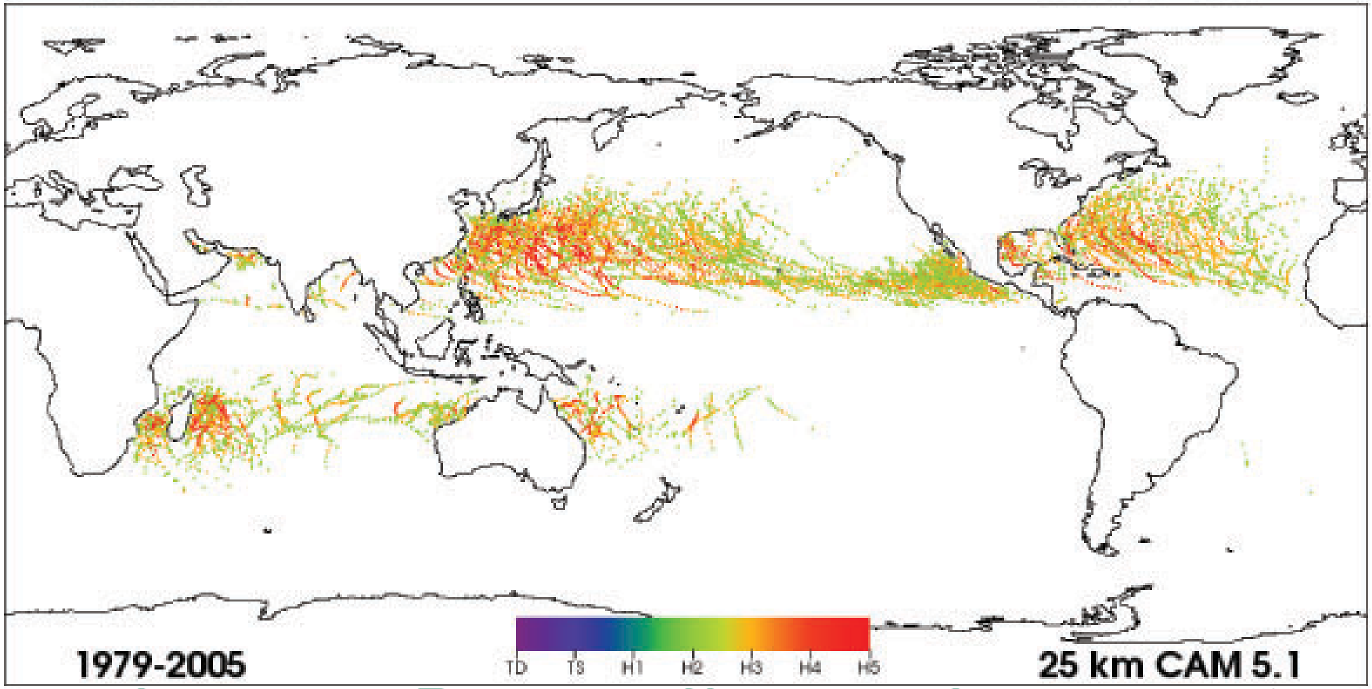
Cat2

Cat3

Cat4

Cat5

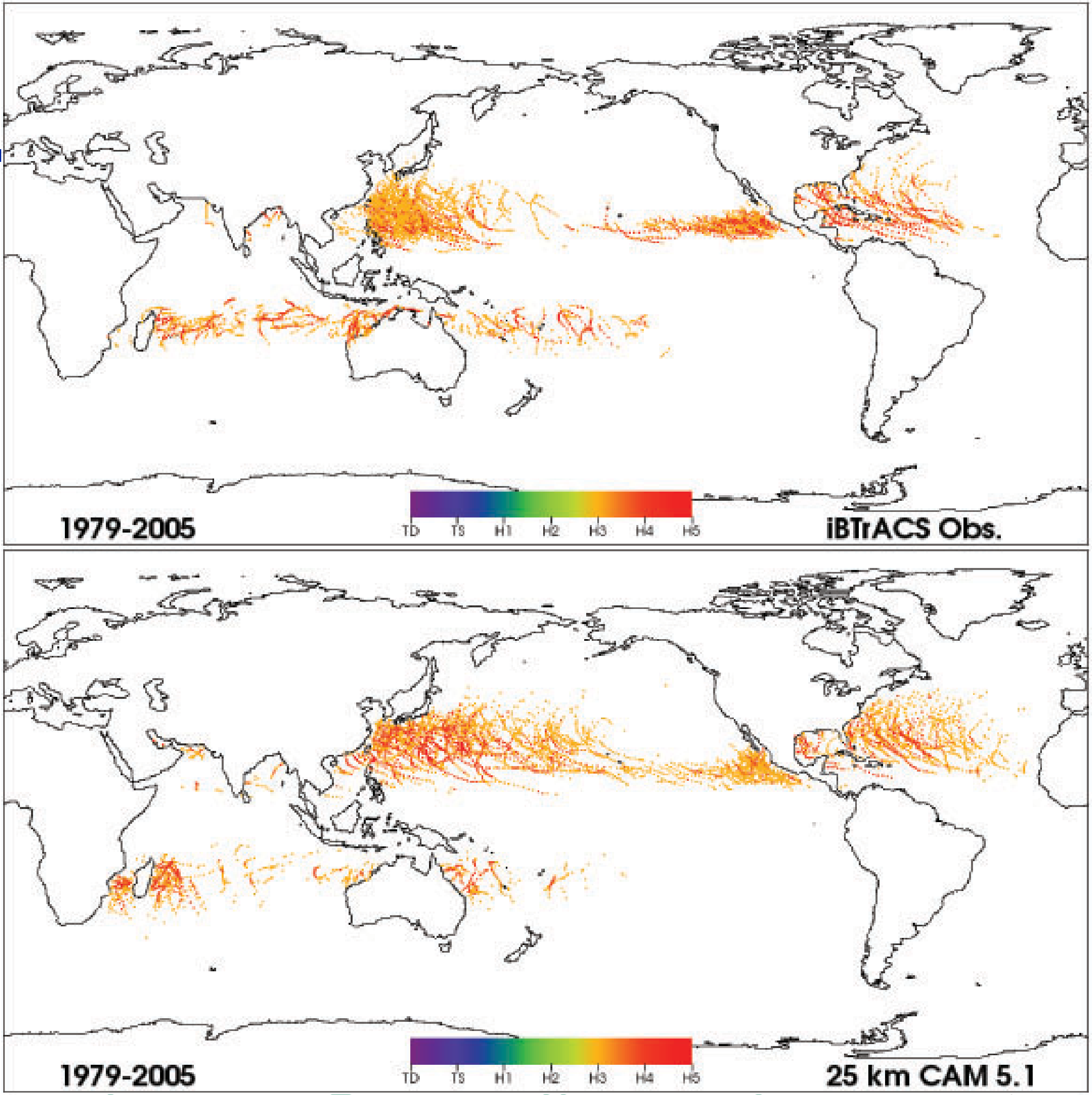


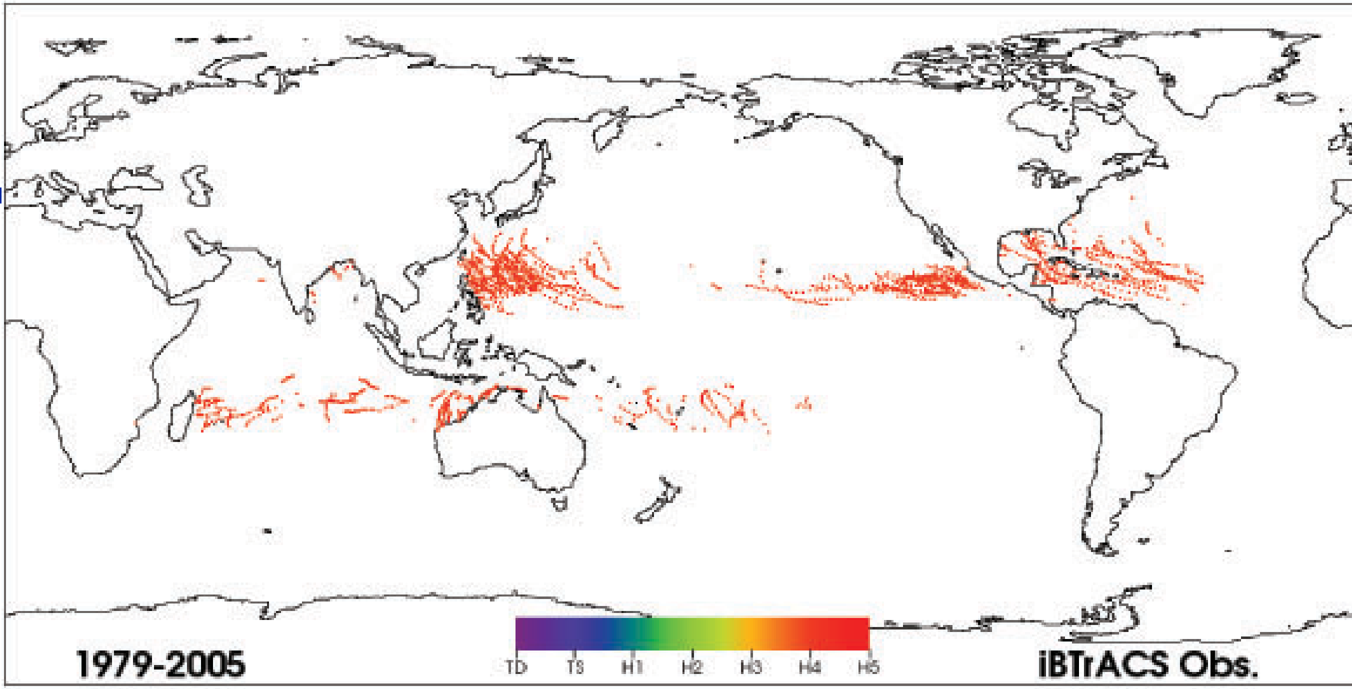


Cat3

Cat4

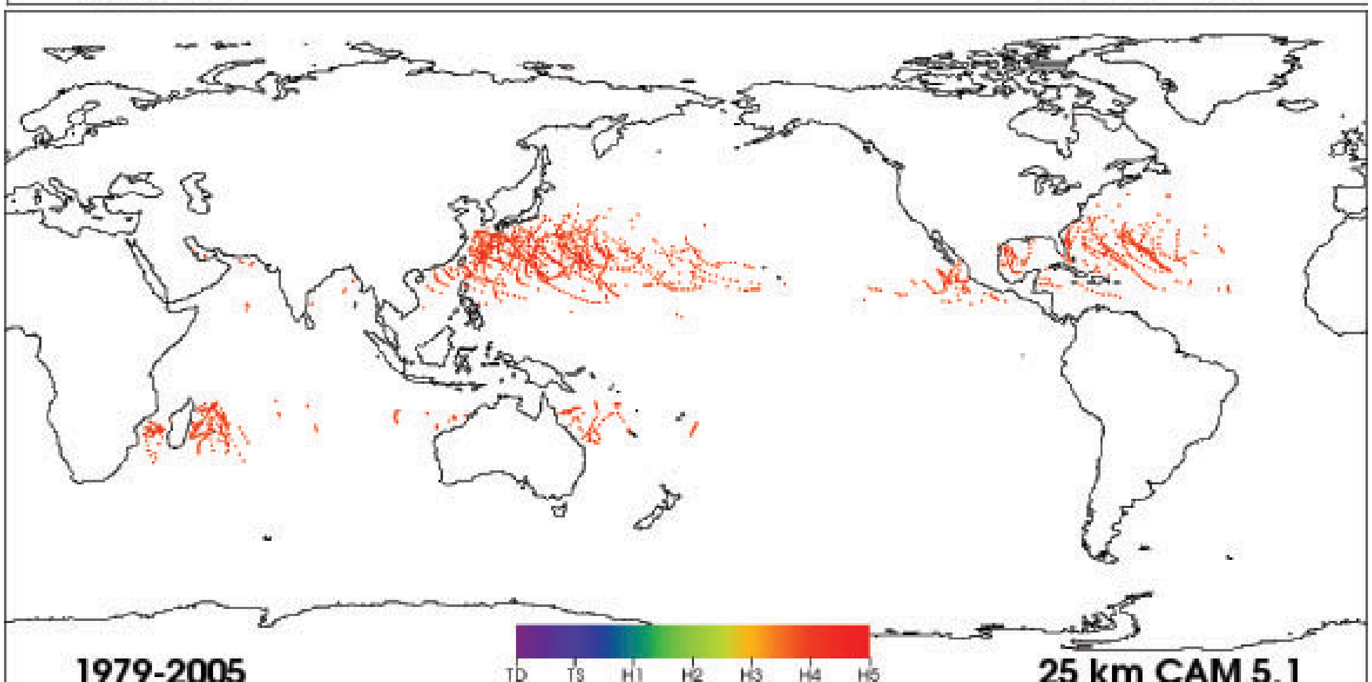
Cat5



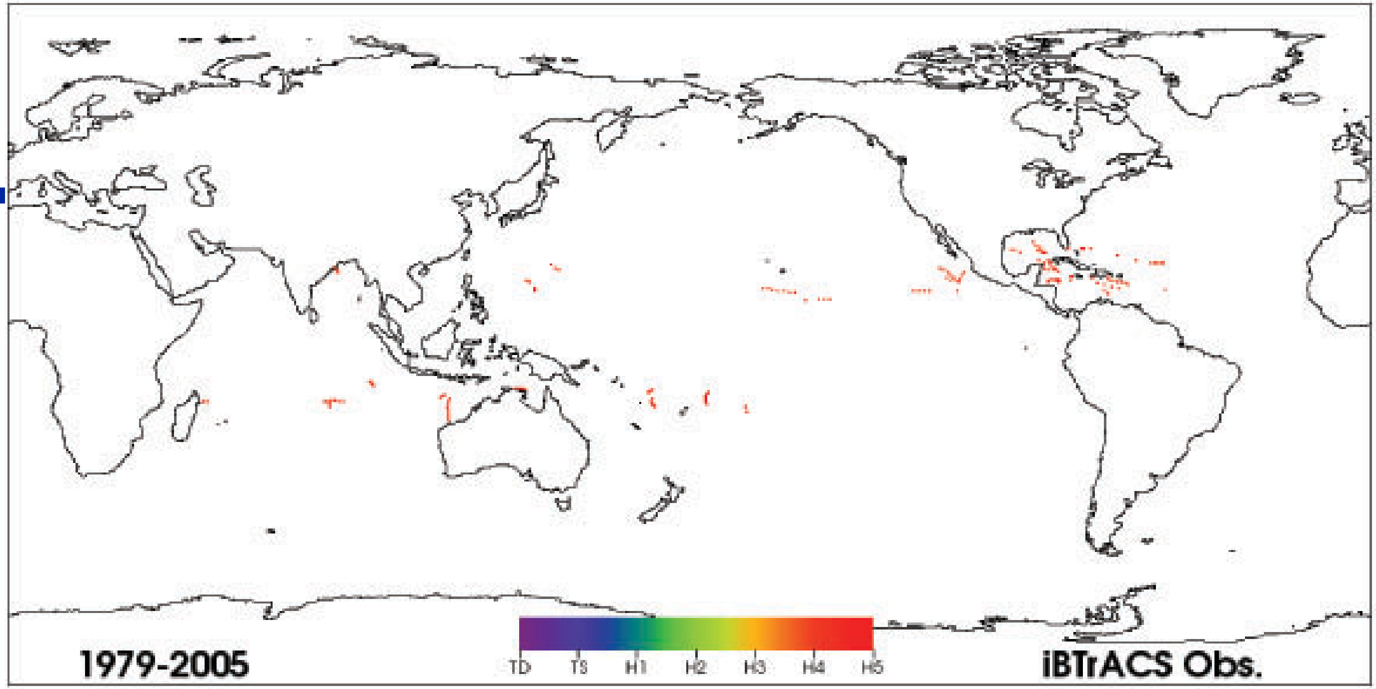

1979-2005
TD
TS
H1
H2
H3
H4
H5
iBTrACS Obs.

Cat4

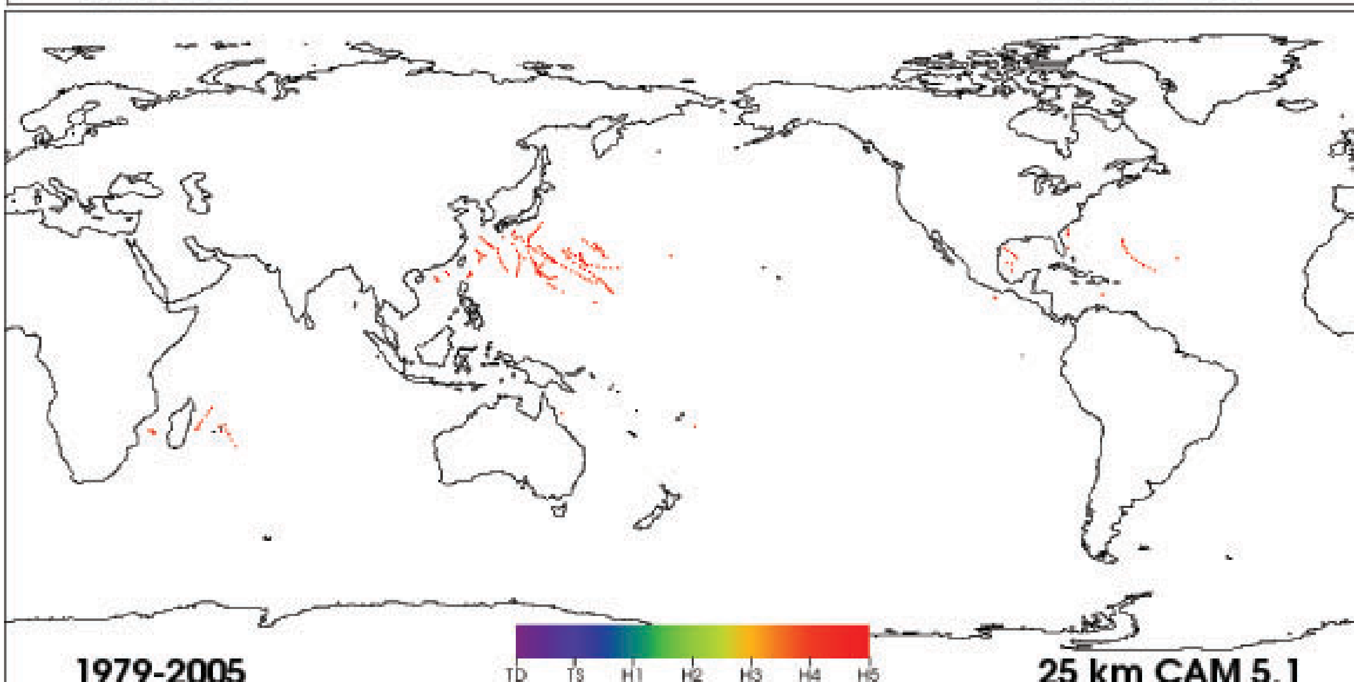
Cat5


1979-2005
TD
TS
H1
H2
H3
H4
H5
25 km CAM 5.1

Cat5



1979-2005
TD TS H1 H2 H3 H4 H5
iBTrACS Obs.

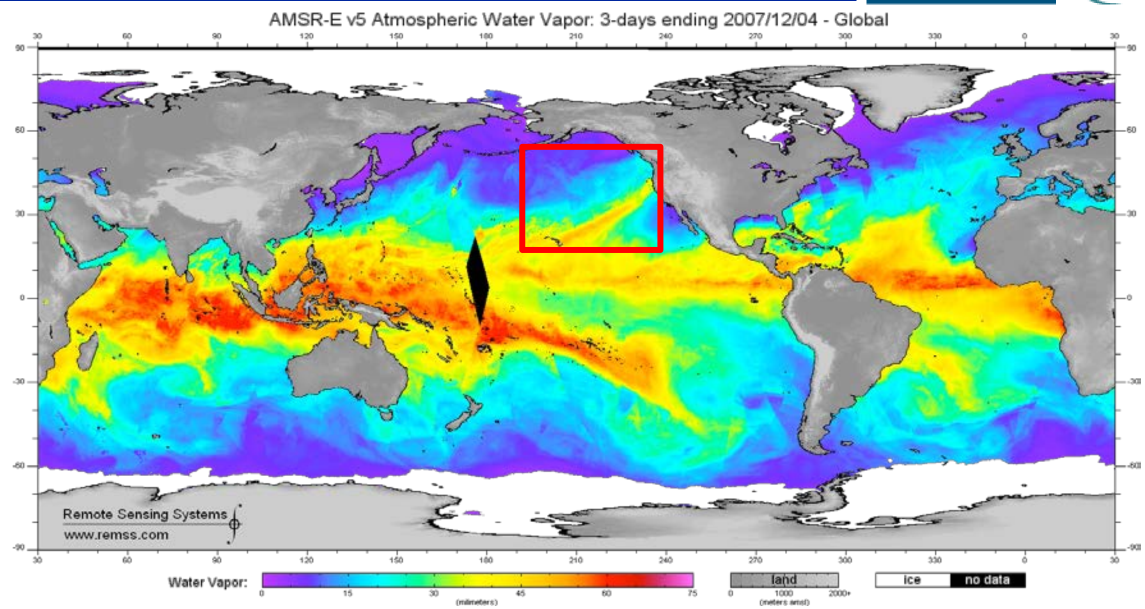


1979-2005
TD TS H1 H2 H3 H4 H5
25 km CAM 5.1

TC detection: Computational Performance

- fvCAM5 ~0.25 degree 3 hourly output
  - 1979-2005 simulated period
  - 13 TB dataset (subset of 100TB)
- Ran detection step on up to 80,000 Hopper cores
  - Hopper is a NERSC Cray XE6 system with a high performance Lustre filesystem
- Detection step completed in ~1 hour, Stitching in ~10 seconds.
  - But 3 days in queue. Plus 3 days for tape transfer from archive.
- Comparable serial job would have taken ~9 years to execute on single core

Atmospheric Rivers



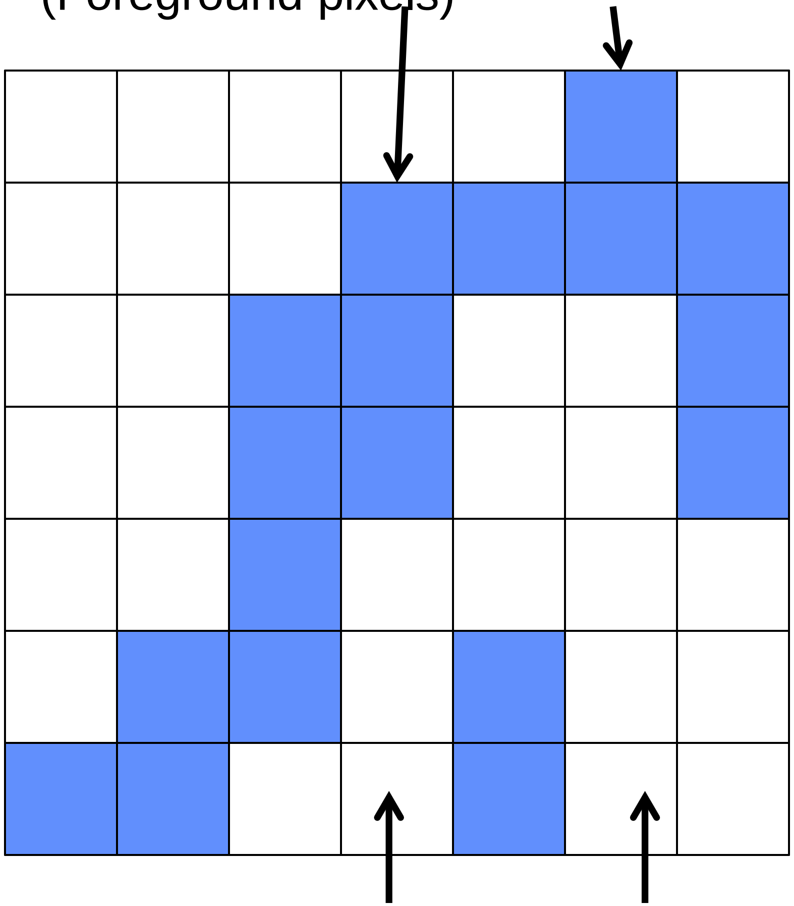
- **Atmospheric Rivers are long filamentary structures that can transport copious amounts of water vapor from the tropics to the mid-latitudes**
  - –**Along the West coast of North America, AR can cause severe flooding, wind damage if they reach land.**
  - –**AR happen in all ocean basins.**

AR detection scheme

1. **Candidate detection:**
   – Read 3 day average Integrated Water Vapor (IWV)
   – Threshold and label the areas where IWV > 2cm
   – Find all the connected components via labeling
   – Verify the origin and landing spots for all connected components
   – If a connected component satisfies origin and landing criteria, measure length and width of the component
   – If length and width criteria meet, then an AR exists
2. **Stitching:**
   – Simpler than for the other events we have considered as these events are so rare
   – Post-detection analysis of storm statistics.

Connected Component Labeling

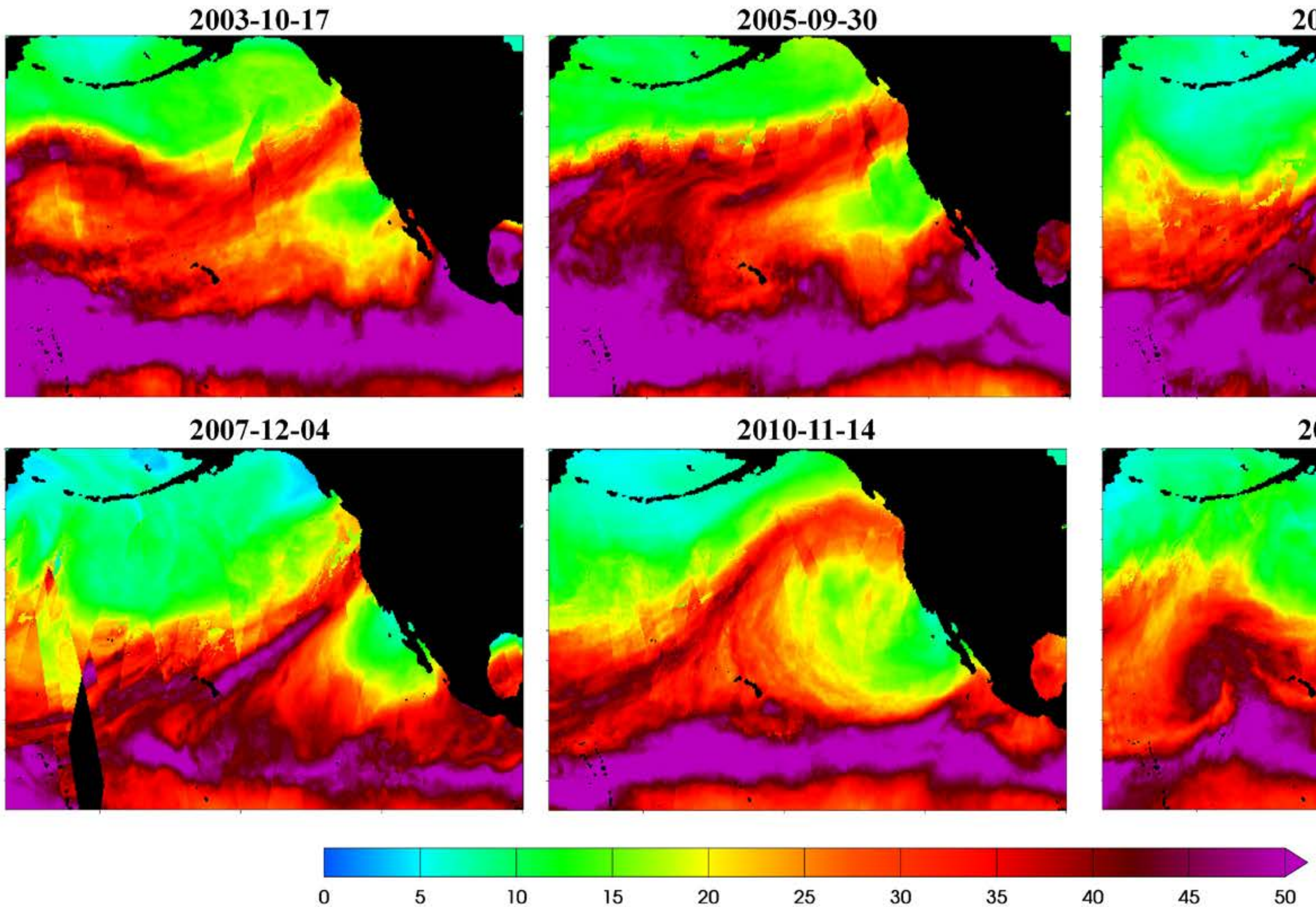
Exceed threshold
(Foreground pixels)



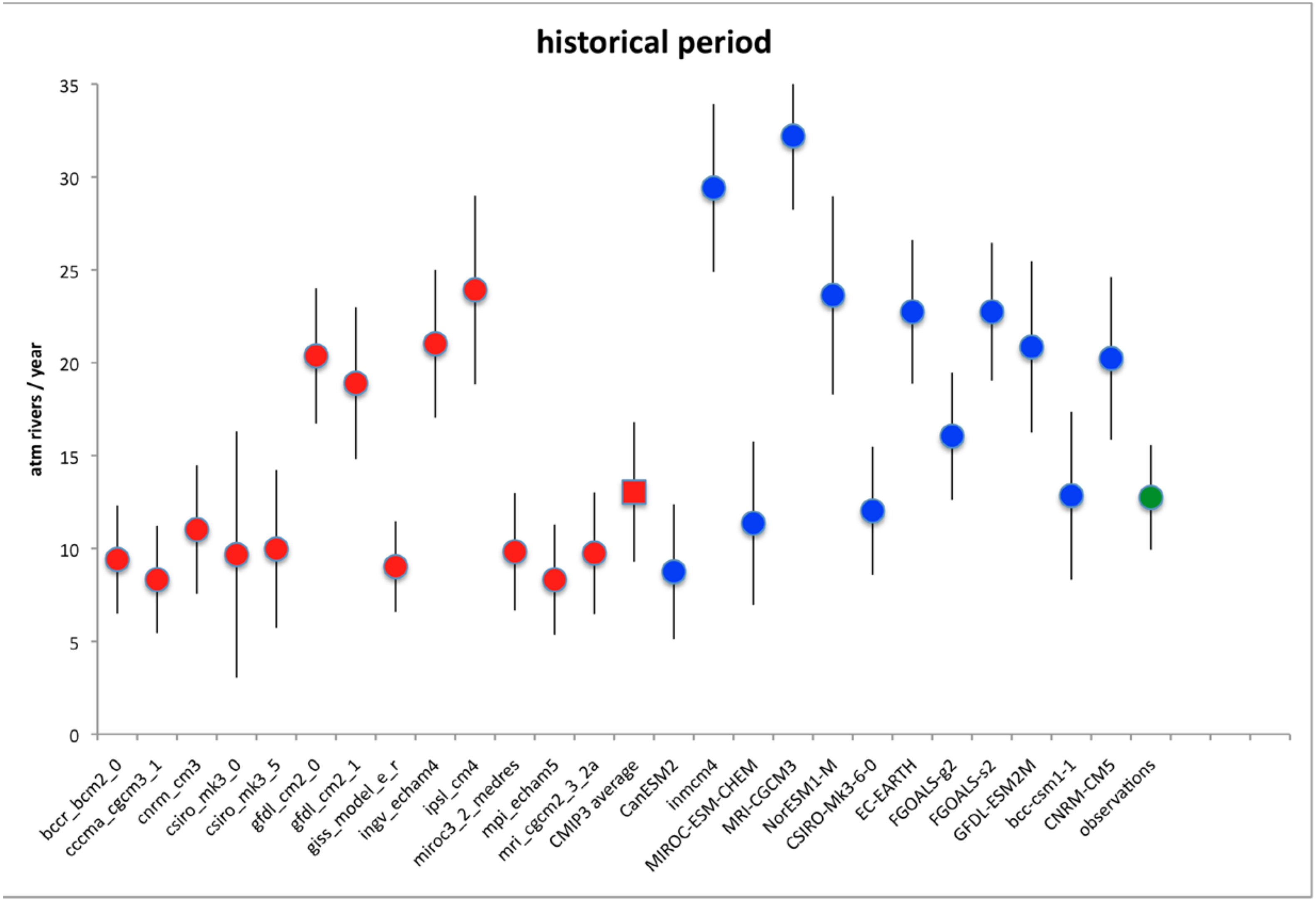
Don’t exceed threshold
(Background pixels)

Examples of atmospheric rivers

- **Satellite total column water vapor**

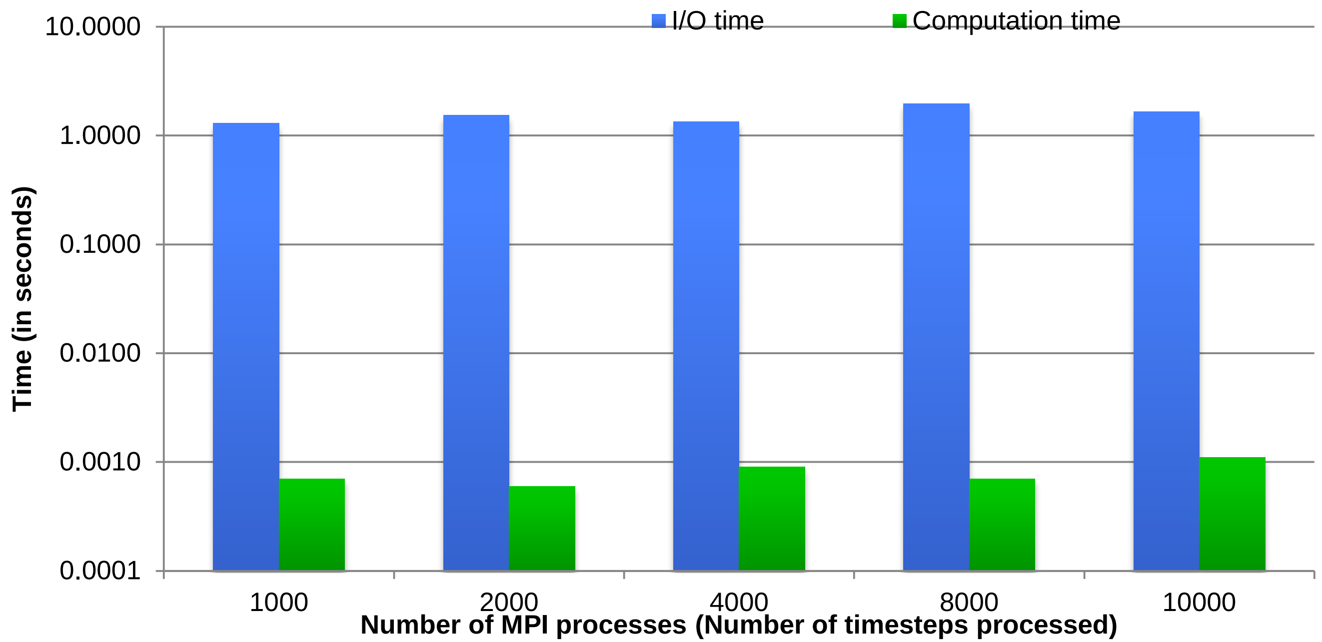


Results from CMIP3/5



AR performance

- **Fully parallel implementation → Each MPI task processes one day's worth of data. I/O bound.**
- **Weak scaling: The number of MPI processes increased as the number of time steps processed increases**



- **Post processing step to generate statistics (number of AR events per year, duration of each AR, etc.) takes ~0.5 second**

Tracking Extratropical Cyclones

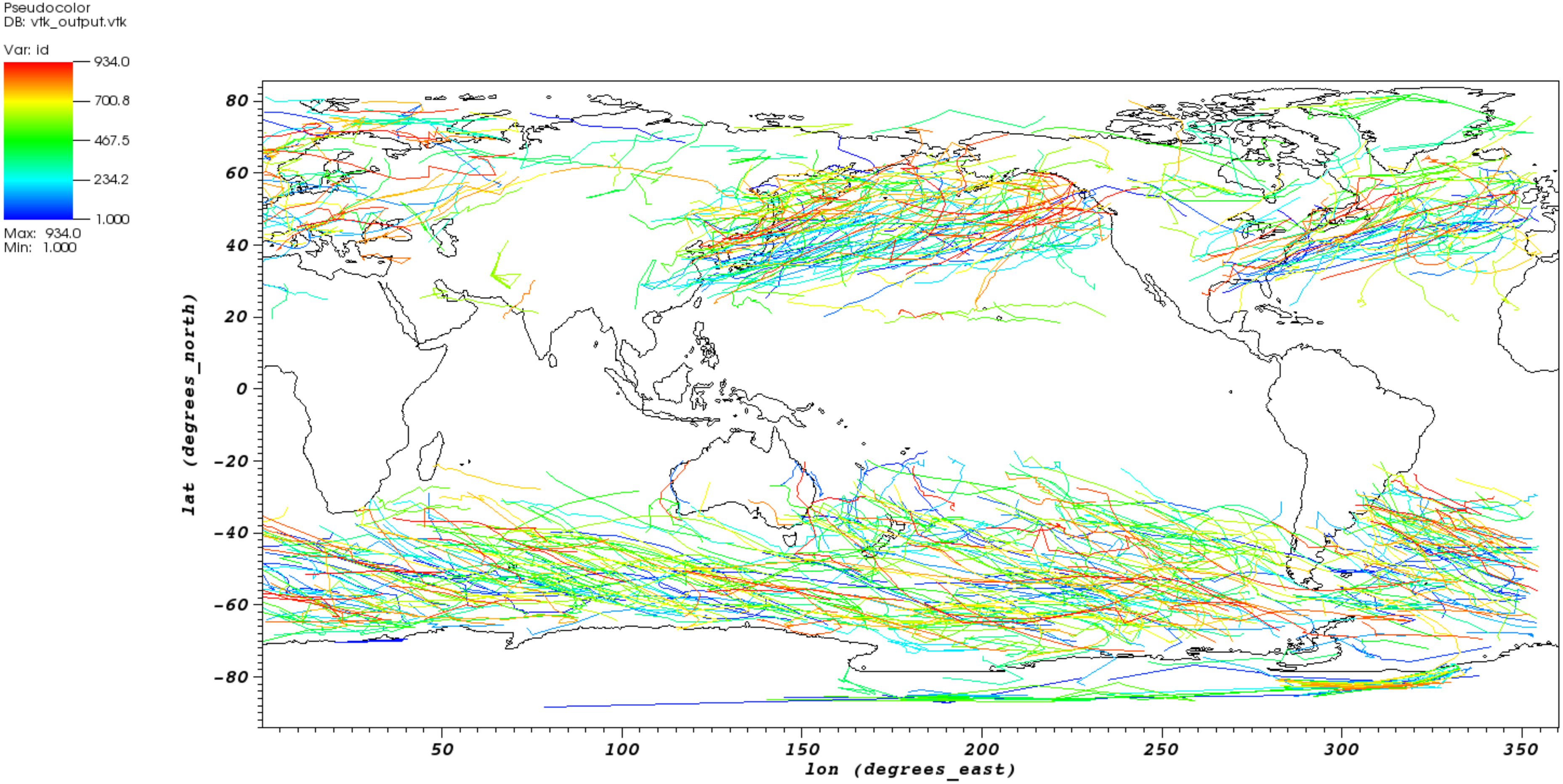
1. Candidate detection step:

- Calculate 850 hPa relative vorticity (Bengtsson et al. 2006; 2009).
- Search within a 800km x 800km region to identify the local maxima vorticity as a candidate storm.
- storm center is the defined at the interpolated pressure minimum within 5° radius of this maximum vorticity

2. Stitching step

- At 6 hour intervals candidate storms must be within a distance of 800 km during the following 6-h time period to be stitched together. Otherwise, the storm track has stopped.
- An event must last longer than 2 days and travel further than 1000km to be considered as an extratropical storm.

1º CAM5 test problem



Identifying Blocking Events

- **Two different types of blocks are identified as a persistent and significant quasi-isolated features**
  - **Low potential mid-tropospheric average vorticity <PV> (rare)**
  - **Negative <PV> anomaly.**

1.Candidate detection step:

- **Calculate 6 hourly mid-tropospheric average <PV> and its anomaly from U, V, T (expensive, done in parallel)**
- **Scan for closed contours of low (negative) <PV> (anomaly)**
- **Thresholds:**
  - **Magnitude<1.0PVU (-1.2PVU)**
  - **Area>1.0e^{6}km^{2}(1.8e^{6}km^{2})**

2.Stitching step

- **Areal overlap of 50%(70%) every 6 hours**
- **Must last more than 5 days**

Lessons learned

- Workflow patterns on large data sets should be carefully planned.
  - Serial extraction of time series from raw output on tapes and its conversion from model (σ) to standard (p) coordinates can take weeks.
  - It can be far more efficient to duplicate some output variables for specific purposes.
    - Don't put lots of variables into one class of file.
    - Do break up variables into groups destined for specific analyses.
    - Do calculate as much as you can prior to archival.
  - Due to queue delays, modest parallelism (1000s) running 10s of hours, provides faster turnaround than large parallelism (10000s) running in minutes.

TECA:conclusion

1. Embarrassingly parallel in time pattern recognition.
2. Serial stitching of candidate events.

- This abstraction is very general to feature tracking in climate data.
- Easy to be efficient.

- Codes will be made publically available in FY13.

Available online at www.sciencedirect.com
ScienceDirect

ELSEVIER

Procedia Computer Science 00 (2012) 1–10

Procedia Computer Science

International Conference on Computational Science, ICCS 2012

TECA: A Parallel Toolkit for Extreme Climate Analysis

Prabhat, Oliver Rübel, Surendra Byna, Kesheng Wu, Fuyu Li, Michael Wehner and Wes Bethel

Lawrence Berkeley National Laboratory

Thank you!

mfwehner@lbl.gov