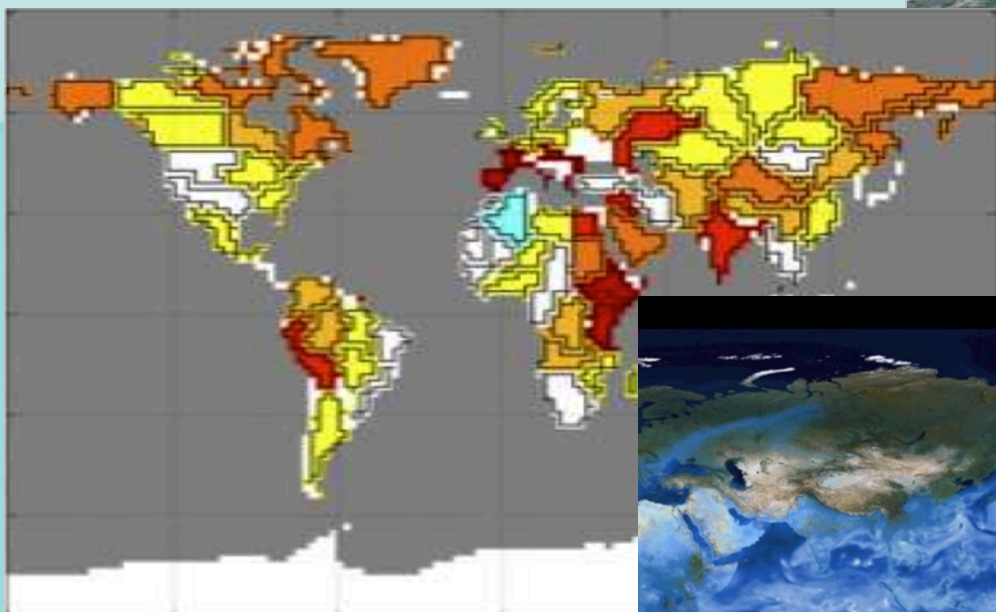
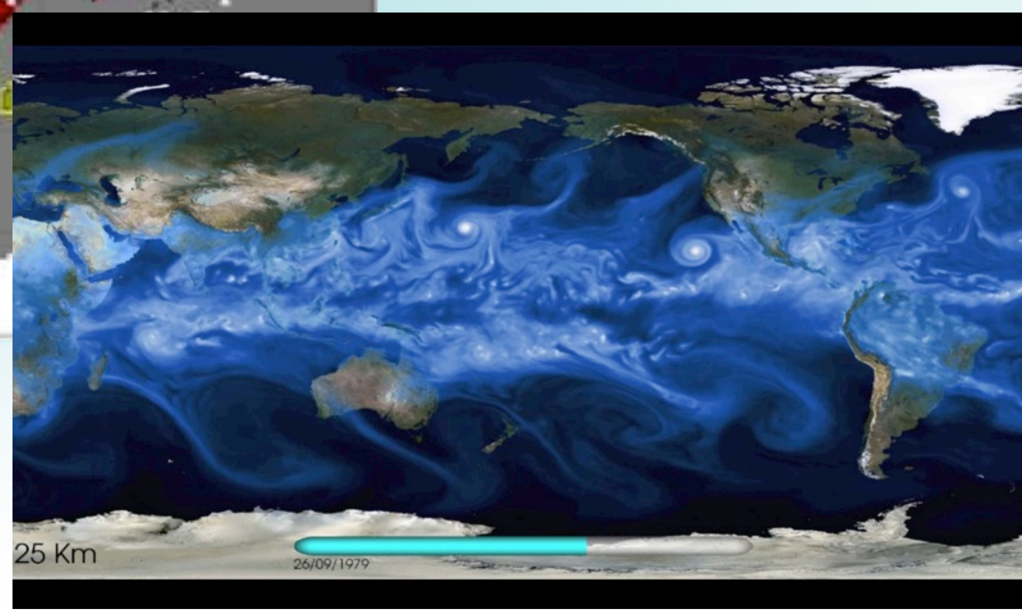
Large climate data sets in Probabilistic Weather-Event Attribution

Pardeep Pall
Michael Wehner
Daithi Stone





Motivation:
'Big data' in the climate community

- **The amount of data available for analysis within the climate community has increased substantially over recent years**

- Due to increased resolution and sophistication of climate models and observational products (~100km global to ~1km regional)

- [Show 25km resolution tropical cyclone movie]

- Also due to larger ensembles of simulations afforded by increased computing power (e.g. CMIP5, HadCRUT4)

⇒ ***Here we show some examples of big data in the context of 'Probabilistic Event Attribution' of weather-related events to anthropogenic climate change***

Motivation:
Probabilistic Event attribution

- **Can't attribute any specific event** because it could also have happened by chance in a non-anthropogenic climate

- **Can attribute *loading* in event *risk*, *if*:**

  1) define climate as encompassing all weather states consistent with initial condition uncertainty

  2) do this both with and without anthropogenic drivers

Loading the climate of the 'weather dice'

⇒ ***Probabilistic Event Attribution***

⇒ *"What fraction of event risk was attributable anthropogenic drivers?"*

An example:
UK Autumn 2000 floods

River Trent, Nottingham
Autumn 2000

October & November 2000 floods occurred during the wettest autumn since record began in 1766

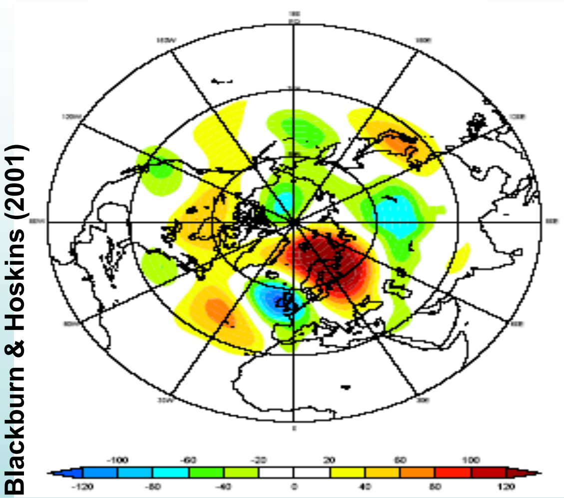
Nationwide impact resulting in £1.3 billion of insured losses

Worst floods since the snowmelt-driven floods of 1947

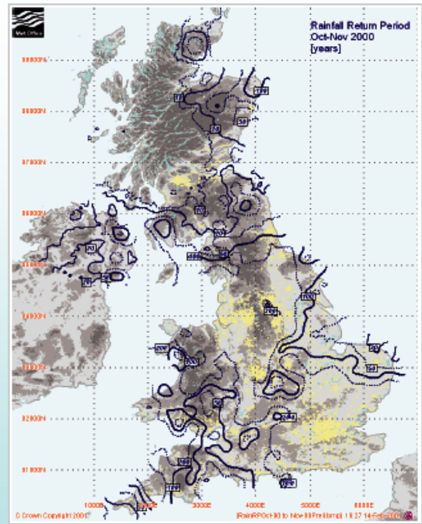
UK autumn 2000 floods:
Hydrometeorology

Geopotential-height anomaly at 300hPa

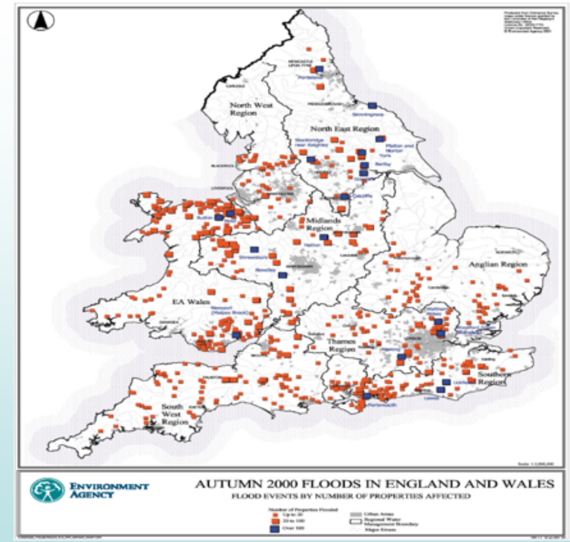
Blackburn & Hoskins (2001)



Precipitation anomaly



Properties damaged



1. Common but anomalously strong **'Scandinavia' circulation pattern** [Rossby wave-train]

   → Eastward + southward displacement of NH jet stream
   [More storms]

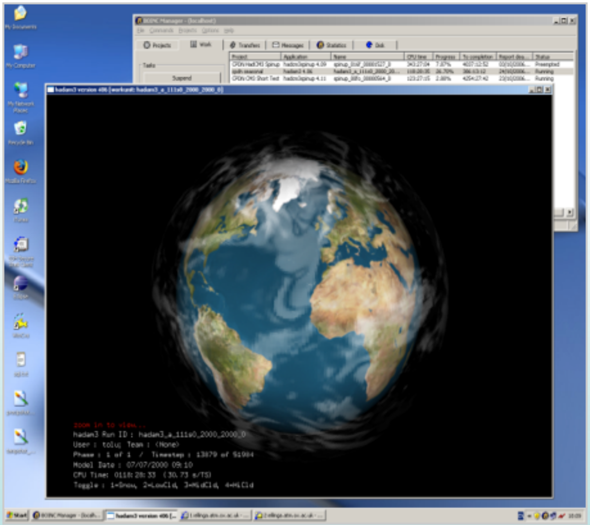
2. Persistent and repeated frontal depressions
   Not strong localised events

   → Three **main precip events, over several day, triggering saturation and floods:**
   9-12th October
   29-30th October
   5-9th November

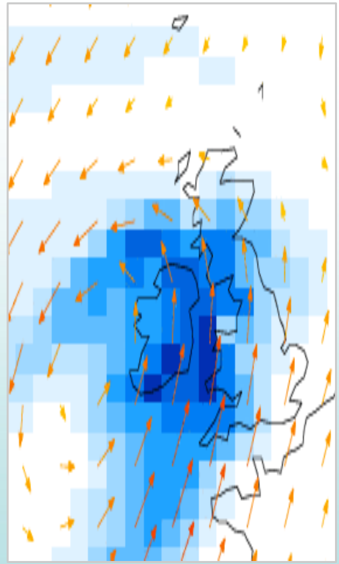
- 3. Nationwide impact
  10,000 properties flooded
  Transport and power disrupted

  → £1.3 billion insured losses

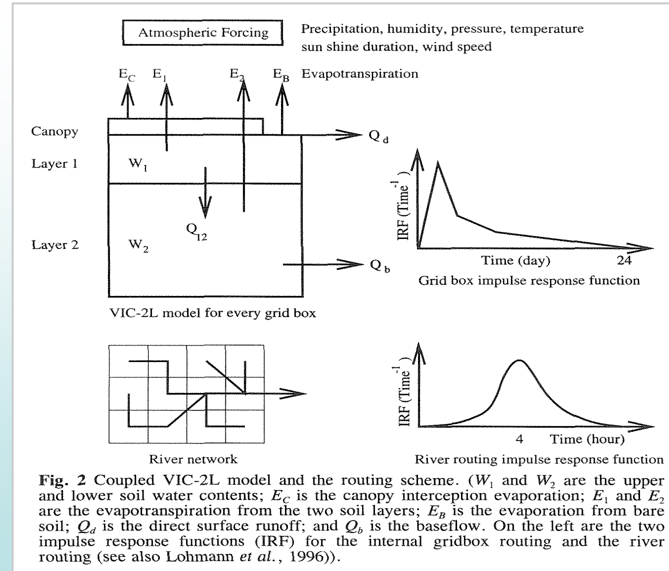
UK autumn 2000 floods:
Experiment design



1. Simulate autumn 2000 weather using a detailed climate model, **both with & without 20thC GHGs**



2. Extract UK precipitation



3. Feed into to a UK flood model

[Use climate*prediction*.net

HadAM3-N144 model;

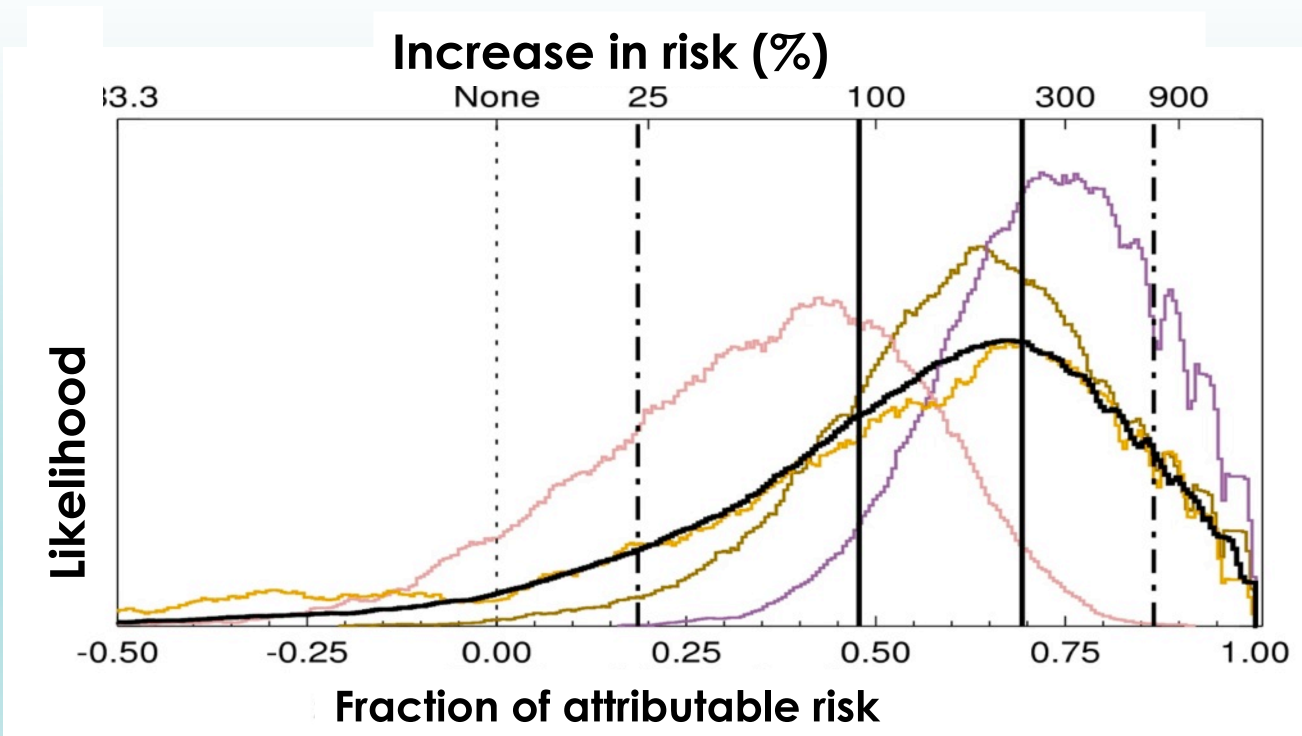
~95km resolution;

>10K simulations.]

[Statistical precip-runoff model;
Parameters derived from a more
comprehensive hydrological-hydraulic model, based on

UK autumn 2000 floods:
Result

20thC anthropogenic GHG emissions significantly increased the risk of floods occurring in England & Wales in Autumn 2000



Risk increases *very likely* (9-out-of-10 cases) by more than 20%,
Risk increases *likely* (2-out-of-3 cases) by more than 90%

UK autumn 2000 floods:
Experiment design

- Flood event was rare, so tried to capture it by generating several thousand weather simulations
- Beyond conventional comput

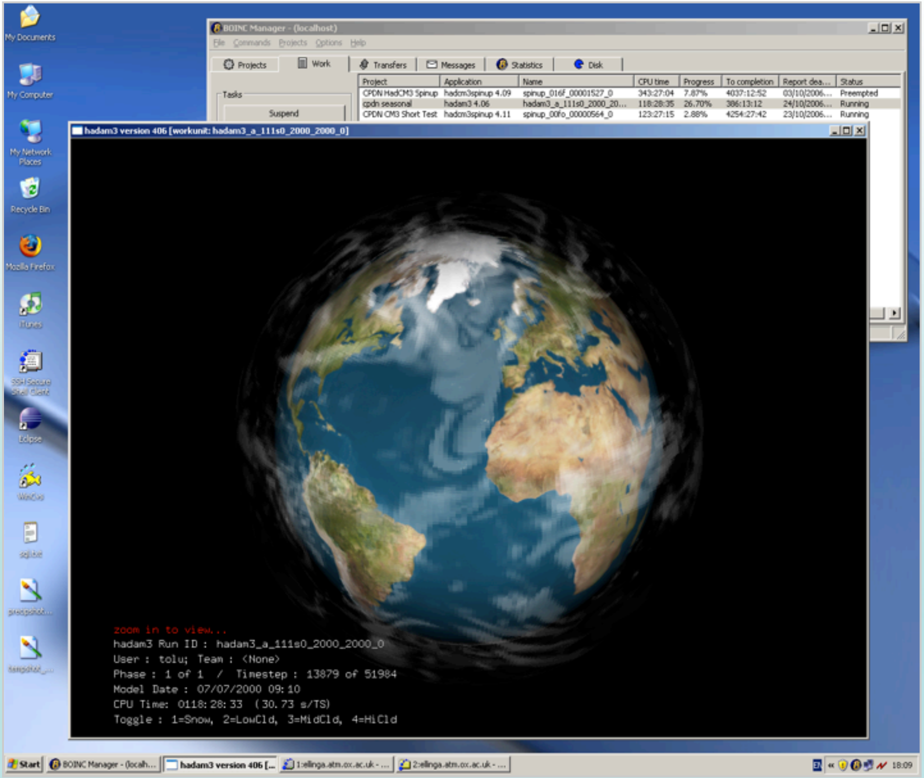
climateprediction.net

- ~150 – 450 MB RAM

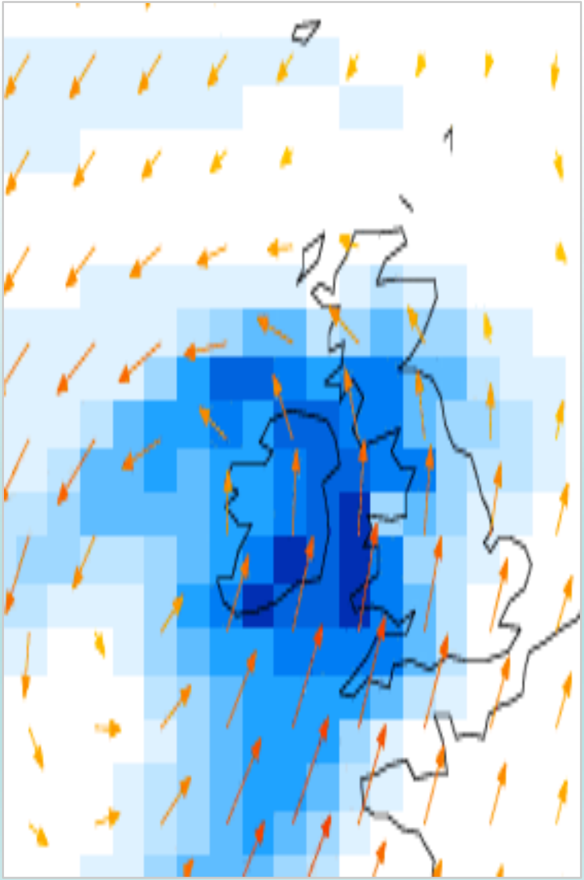
  ~ 3 weeks to simulate 1yr

- >10,000 simulations

100s of GB



UK autumn 2000 floods:
Analysis



Data analysis performed on a single processor desktop computer!

Manageable for a small region precipitation such as the UK

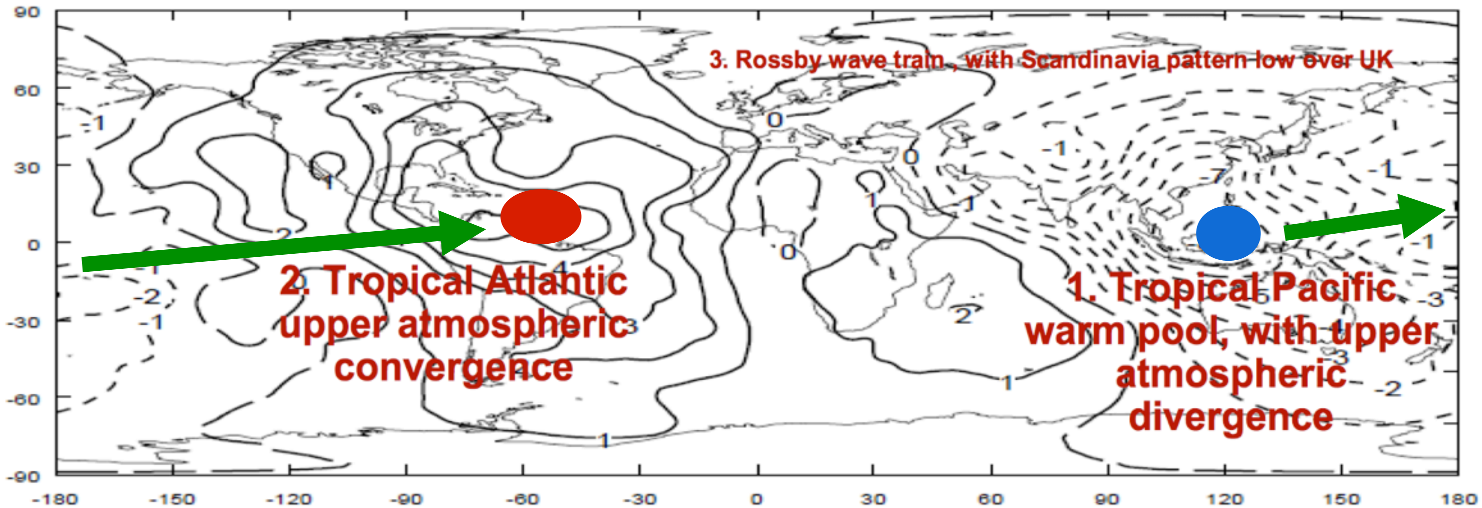
What about analysing and visualising larger regions and more complicated metrics?

UK autumn 2000 floods:
Atmospheric teleconnection?

A proposed tropical catalyst for UK floods → **teleconnection**:

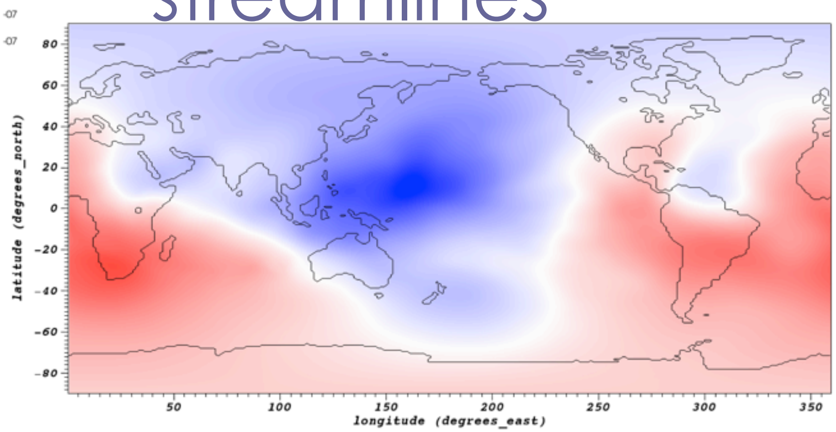
1. Eastward propagating upper-atmospheric flow from Pacific

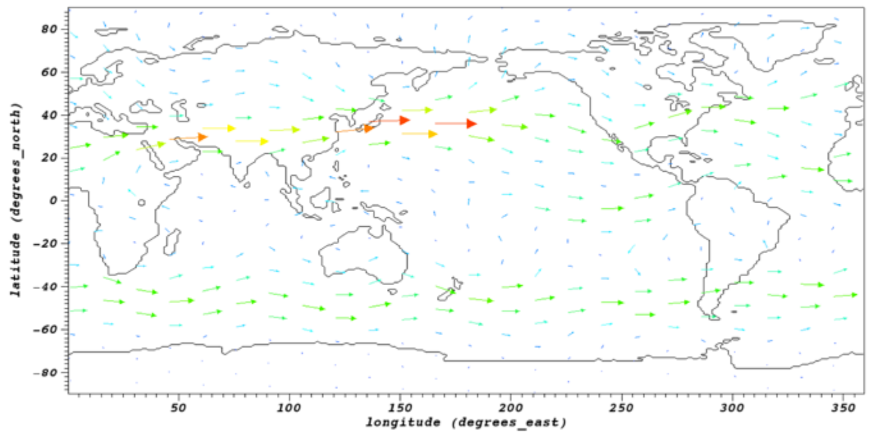
200hPa velocity potential anomaly in October 2000, relative to ECMWF re-analysis climatology



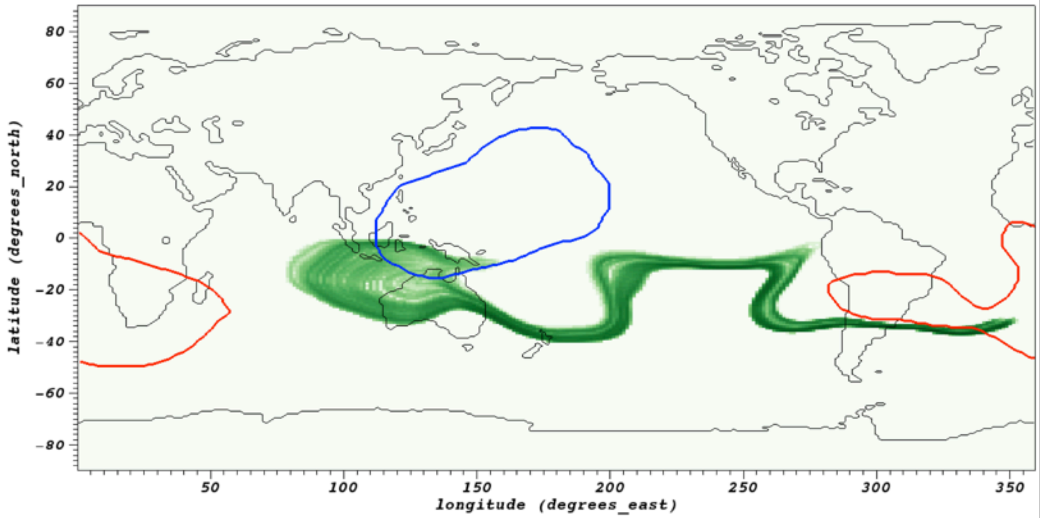
Evidenced by observed:
- Divergence of air over Indonesian Pacific warm pool region
- Convergence of air over tropical South America

UK autumn 2000 floods:
Identifying teleconnection nodes and streamlines





Velocity Potential



1) Search for minima/maxima (nodes) of upper-air velocity potential by comparing nearest neighbour cells

2) Seed particles in minima region

3) Advect particles through the wind field (streamlines) to see which reach a maxima

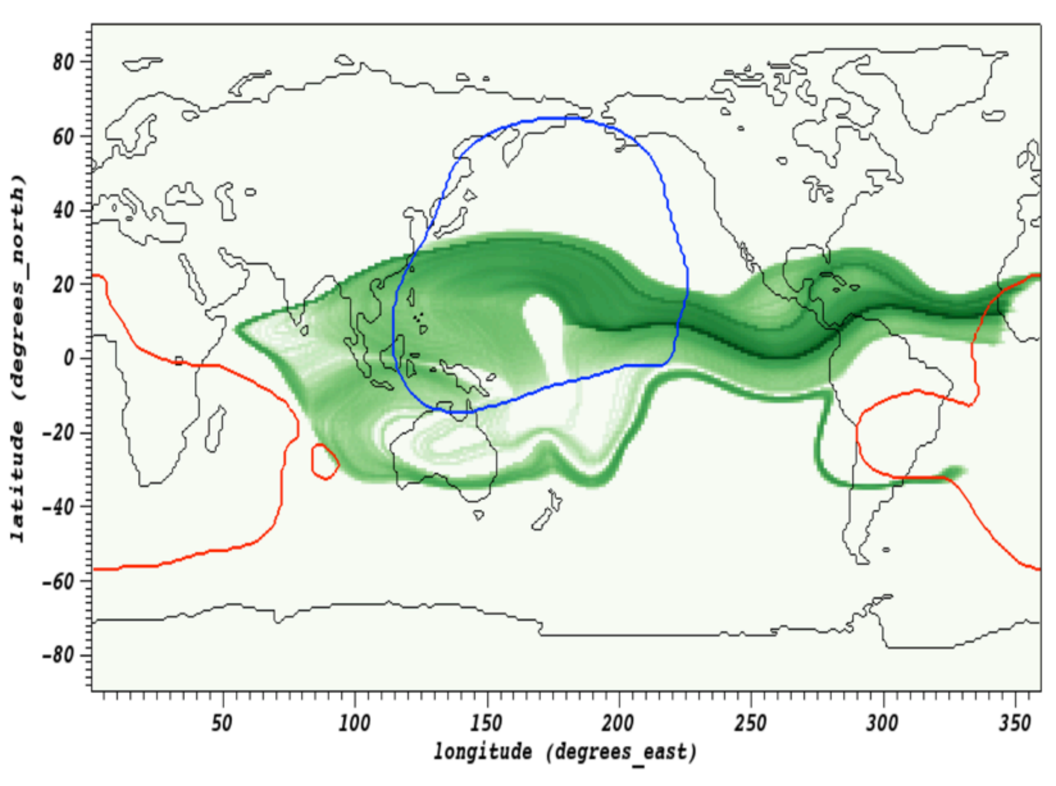
UK autumn 2000 floods:
Parallelisation of teleconnection analysis

Each teleconnection identification calculation takes ??? on a ???

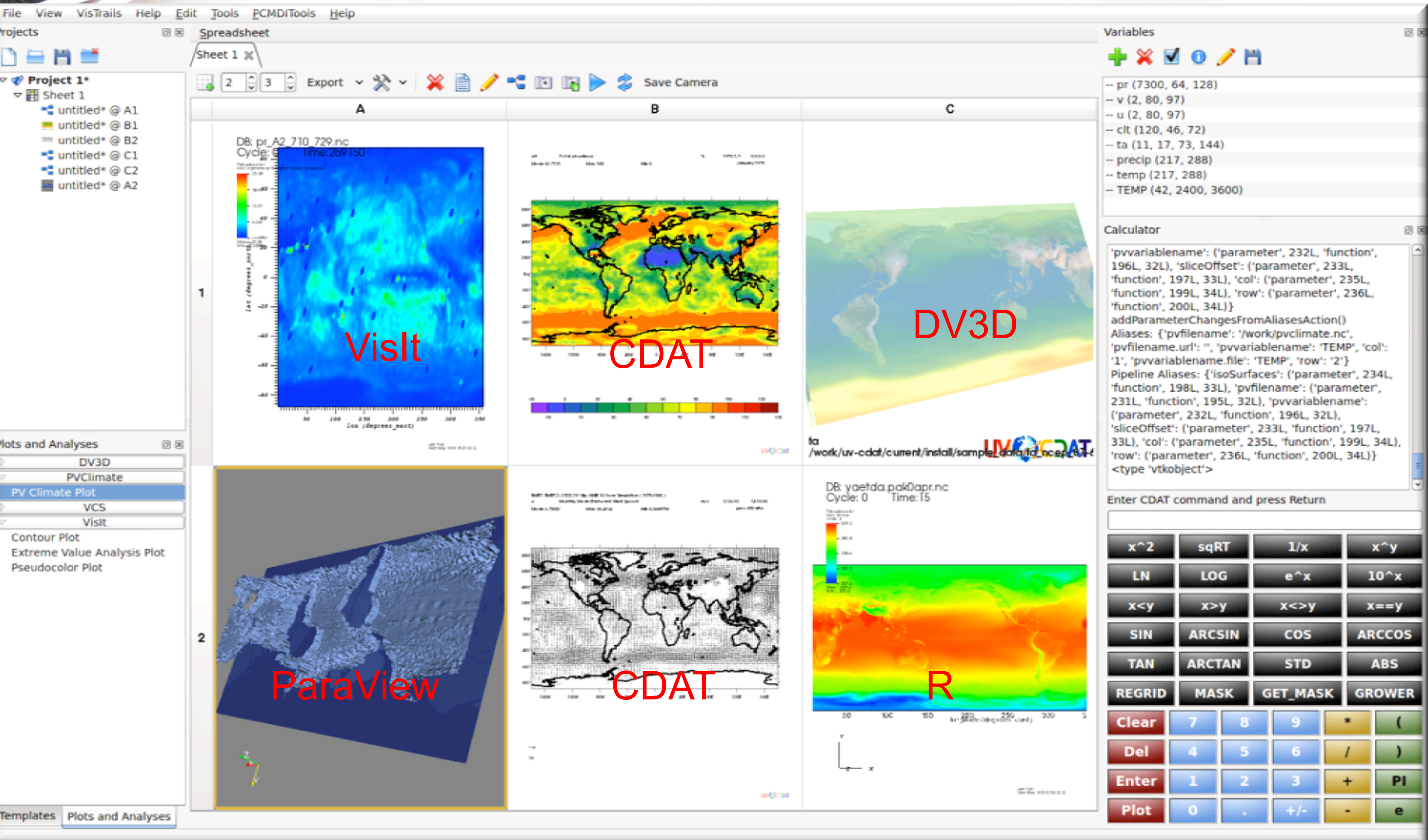
-> ~??? for ~10K simulations

Embarrassingly parallel problem

[Needed for calculating anthropogenic change in teleconnection frequency]



UV-CDAT Integrating CDAT, DV3D, ParaView, VisIt, and R



TECA: A Toolkit for Extreme Climate Analysis

Concept:

Two-step identification of extreme weather events in high frequency climate model data:

1. Search the data for candidate events at each individual time step that meet some defined criteria.

Computationally intensive, but embarrassingly parallel across time steps

1. Stitch together candidate events at multiple time steps, rejecting candidates that fail continuity criteria.

Computationally cheap

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
   - Warm-core and storm center should be co-located within ~200km

2. Stitching
   - Search candidates within 400km radius over 6-hr window
   - Closer, westward, and poleward candidates are preferred
   - 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

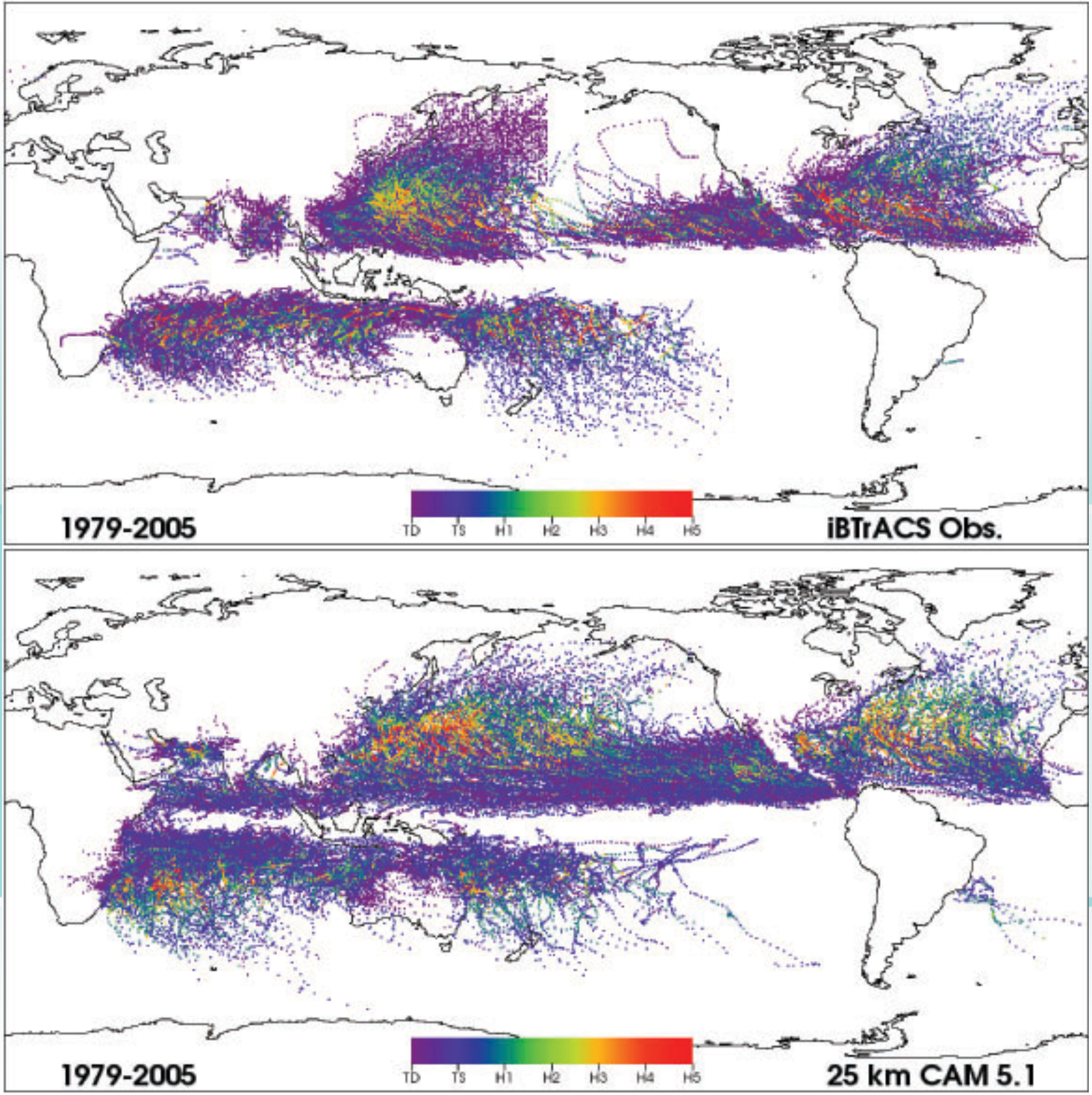
Cat1

Cat2

Cat3

Cat4

Cat5



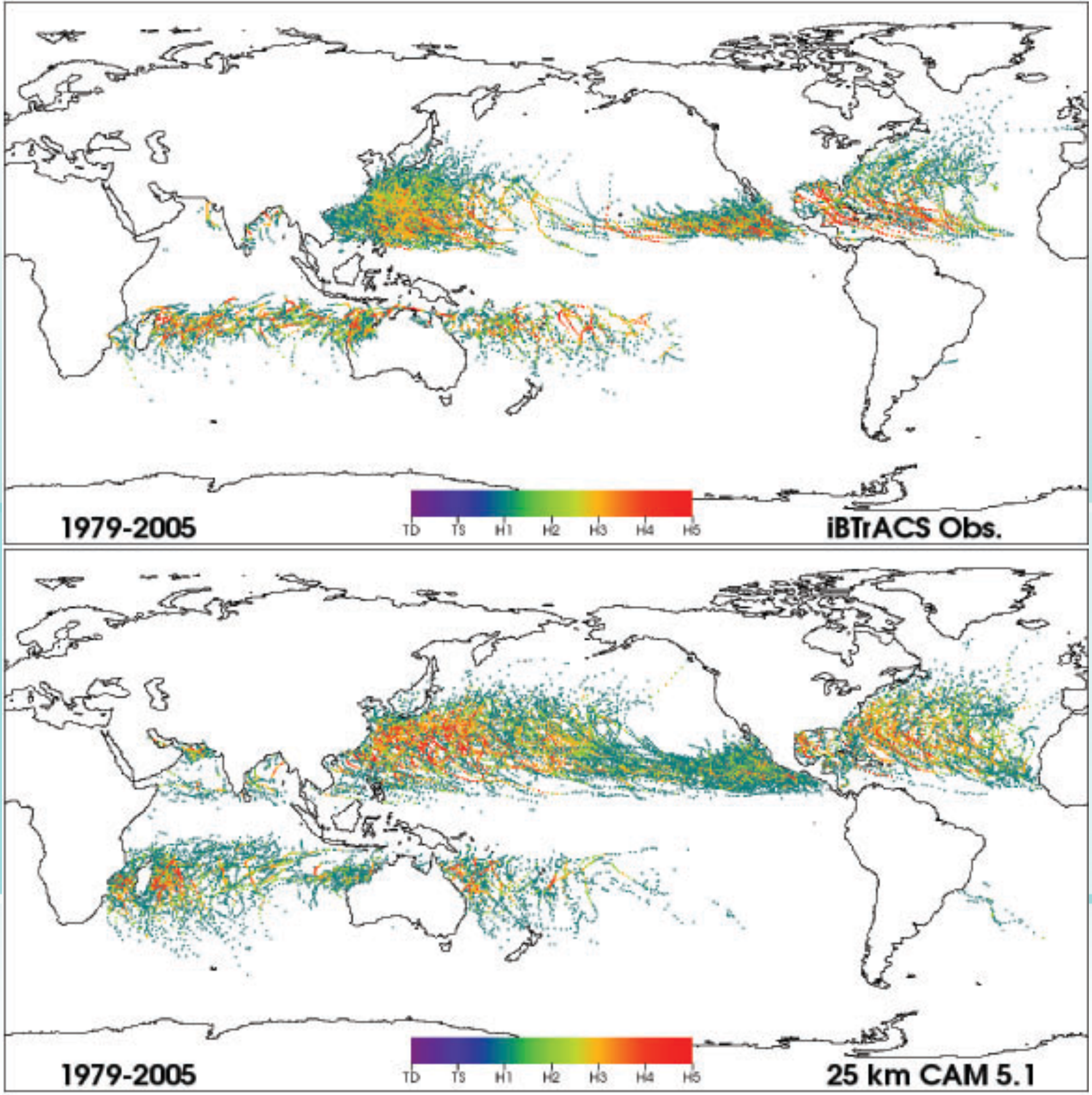
Cat1

Cat2

Cat3

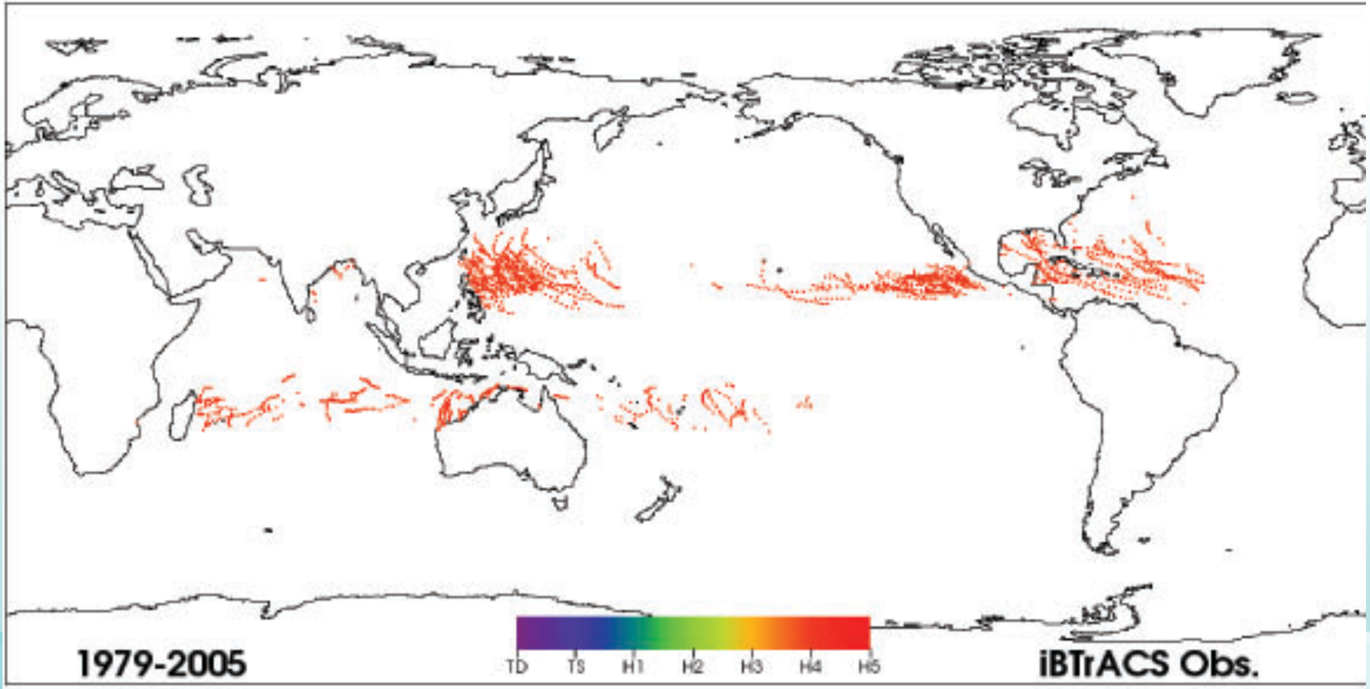
Cat4

Cat5

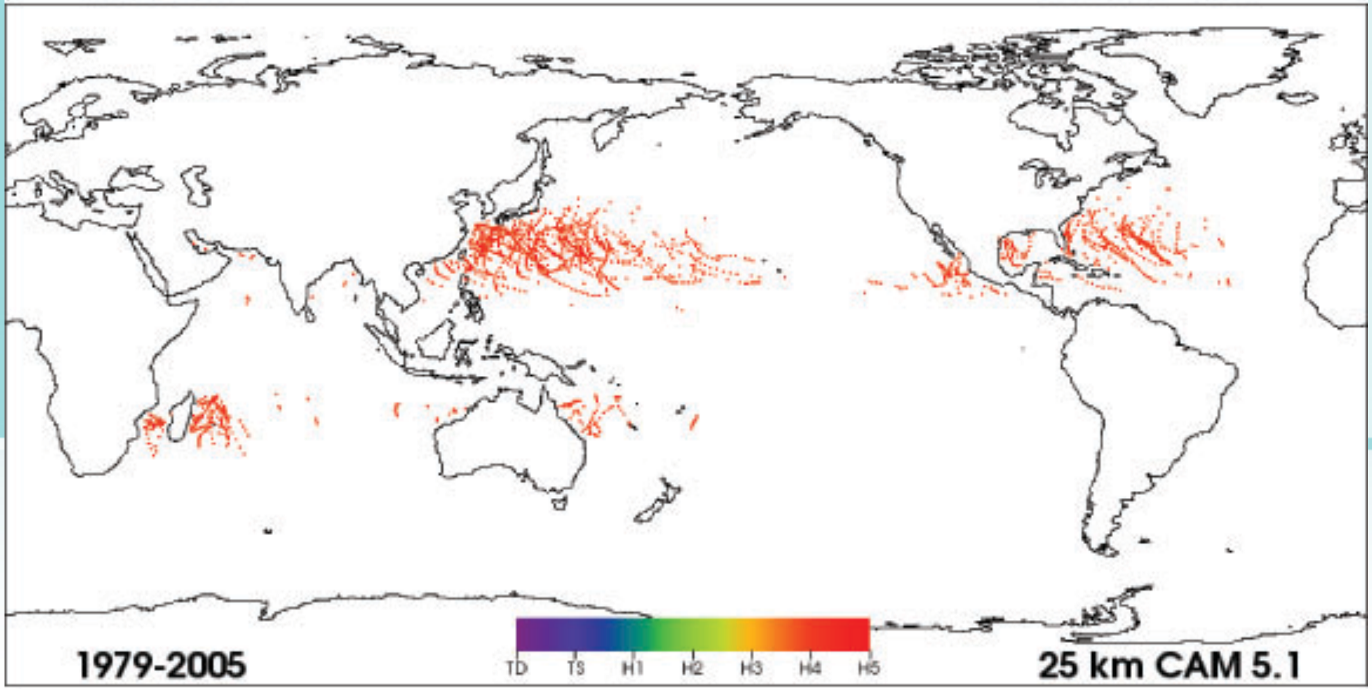


Cat4

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 simulation 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
  - Assigned 1 time step per processor core
- 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!

TECA: A Toolkit for Extreme Climate Analysis

Tracking 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. 100TB of output in a 26 year integration of a ~25km global atmospheric model (fvCAM5.1) → effectively a few MB

- 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.

Designed by Prabhat, LBNL

The C20C Detection and Attribution Project

How about a more systematic event attribution product?

→ “Core” project of the WCRP CLIVAR C20C project.

Scientific aims:

– to characterise historical trends and variability in the probabilities of damaging weather events, including the differences across climate models

– to estimate the fraction of the historical, present, and future probabilities of damaging weather events that is attributable to anthropogenic emissions, and to characterise underlying uncertainties in these estimates.

The C20C Detection and Attribution Project

Or in other words:

– to get O(10) modelling groups around the world to run >O(20) atmospheric models in a semi-coordinated weather-risk-attribution framework

– 50-member ensemble simulations per climate forcing, including 3-D daily output and 2-D 3-hourly output

– Any analysis must be conducted on a per-ensemble basis (to estimate probabilities/uncertainties etc.), so amenable to parallel computing

The C20C Detection and Attribution Project

The “core” project

Scenario	Description	SST and SIC	Period
All-Hist/est1	Including changes in “all” known external forcings (anthropogenic and natural)	Observed	1960-2012
Nat-Hist/CMIP5-est1	Including changes in natural external forcings only	Observed minus CMIP5 estimate of anthropogenic signal	2000-2012
Nat-Hist/?	Including changes in natural external forcings only	Observed minus another estimate of anthropogenic signal	2000-2012
Nat-Hist/?	Including changes in natural external forcings only	Observed minus another estimate of anthropogenic signal	2000-2012

The C20C Detection and Attribution Project

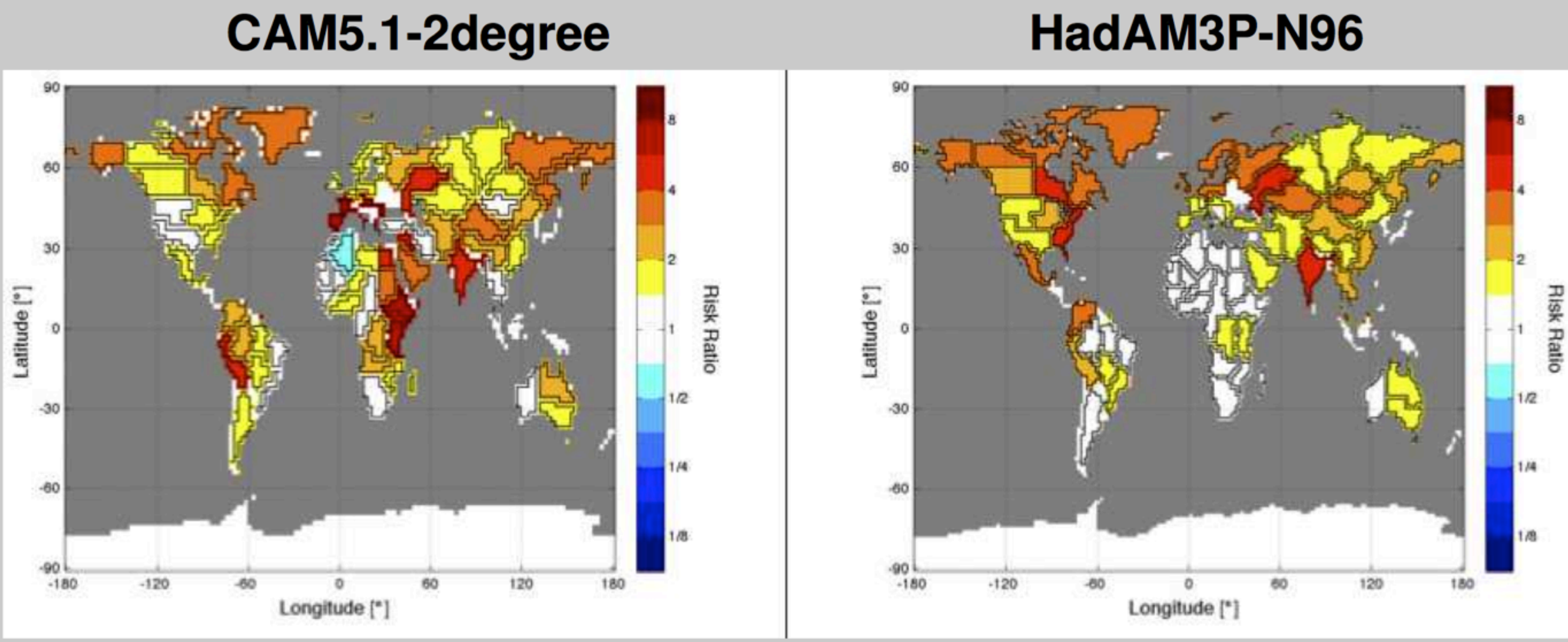
Groups that have already started simulations...

Institution	Model
LBNL (USA)	CAM5.1-2degree, CAM5.1-1degree, CAM5.1-0.25degree
MOHC (UK)	HadGEM-3A-N96, (HadGEM-3A-N216?)
NIES (Japan)	MIROC5
UCT (South Africa)	HadAM3-N48, HadAM3P-N96
UC Davis (USA)	WRF-CAM5.1-1degree

The C20C Detection and Attribution Project

Some pilot results...

5-day regional precipitation, 1-in-1-year event, 2009-2011

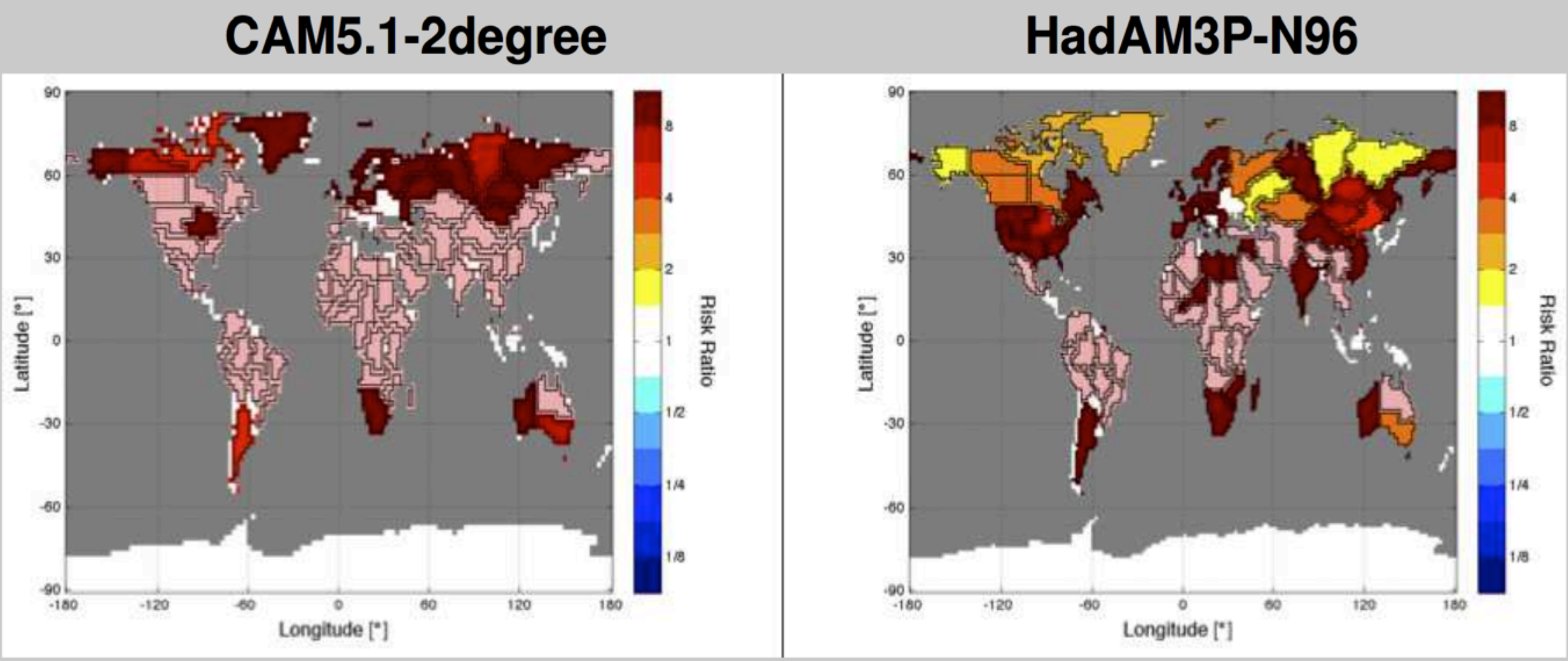


Courtesy Oliver Angélil

The C20C Detection and Attribution Project

Some pilot results...

5-day regional temperature, 1-in-1-year event, 2009-2011



Courtesy Oliver Angélil

Thank you

Please feel free to use the C20C data...

- Documentation:
  - http://portal.nersc.gov/c20c/
- Data:
  - http://esg.nersc.gov
    - Search for "c20c"
    - Or contact Dstone@lbl.gov / mfwehner@lbl.gov