Public Resource Distributed Modelling

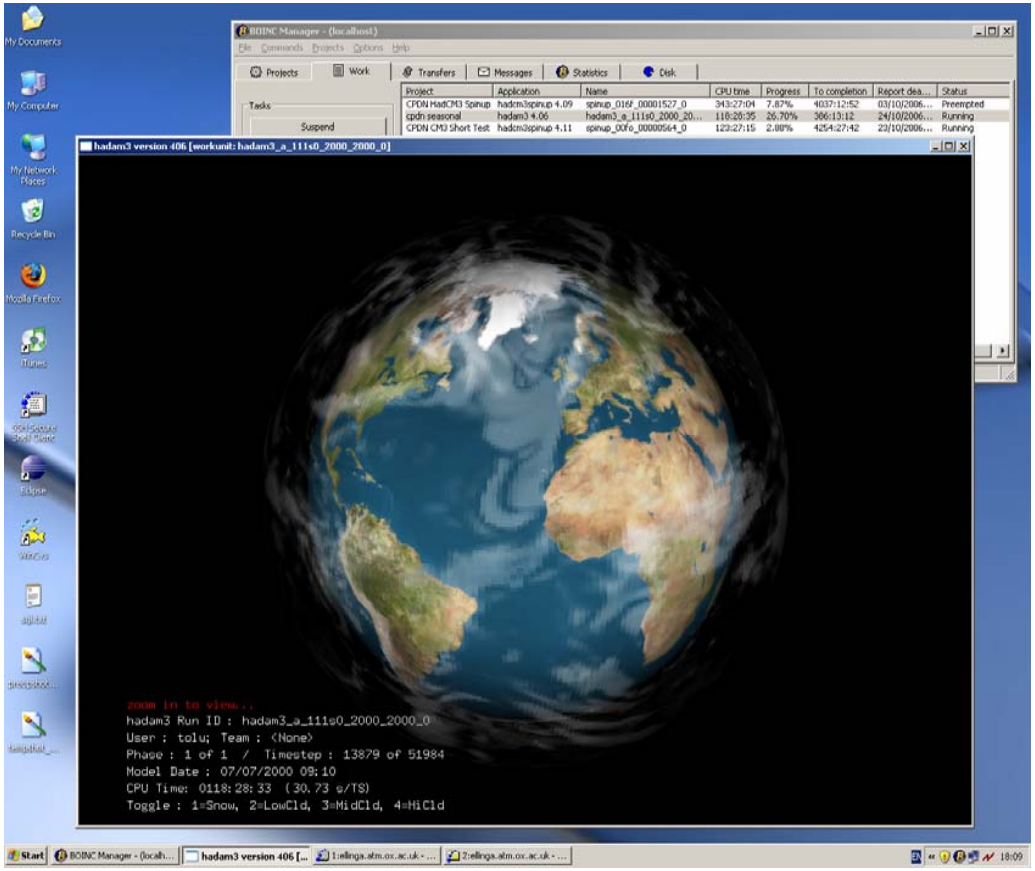
Dave Stainforth, Oxford University

Acknowledgements: Myles Allen, Dave Frame, Carl Christensen, Tolu Aina, Jamie Kettleborough, Mat Collins and many many others.

MISU, Stockholm

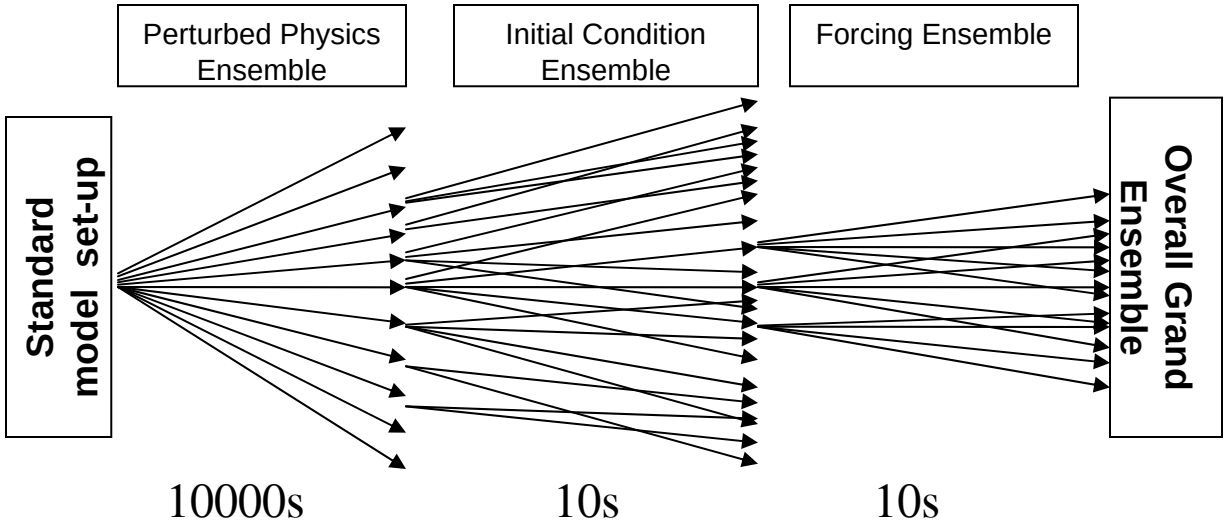
8th March 2006

1. Climate*prediction*.net
2. Public Resource Distributed Modelling c.f. Other Volunteer Computing Projects.



climateprediction.net

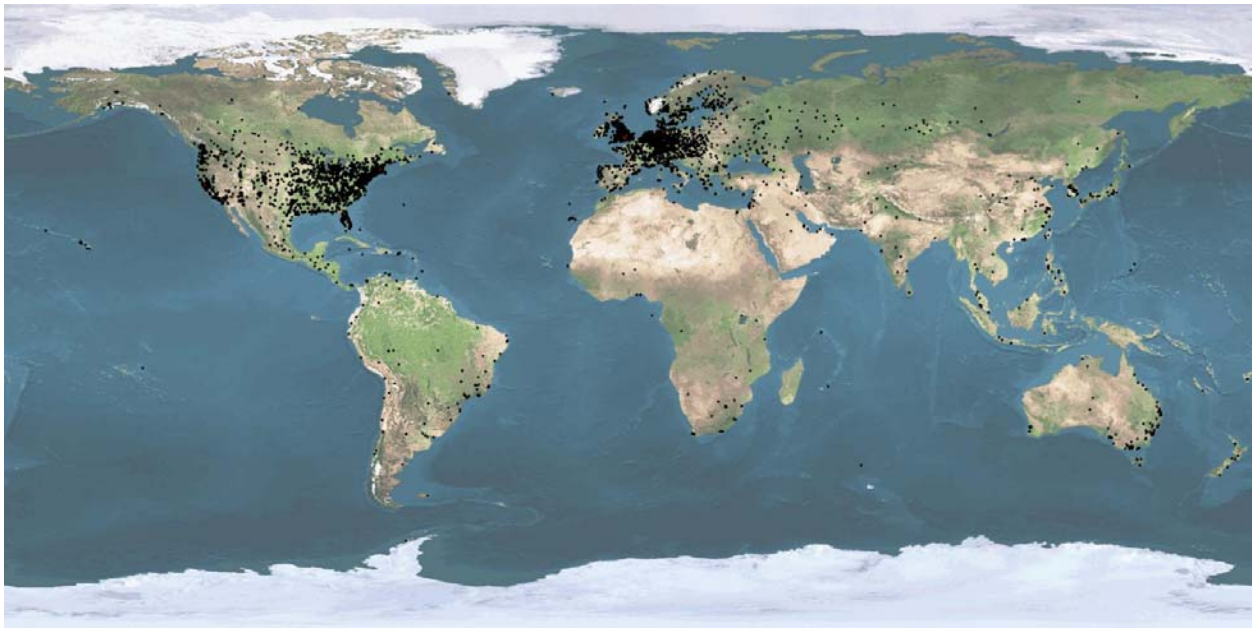
The Climate*prediction*.net Experiment



- To quantify uncertainty we need 100s of thousands of simulations.
- Impossible with super computers.
- But possible with distributed computing.
- Anyone can go to www.climateprediction.net and download the Hadley Centre model to run on their PC.

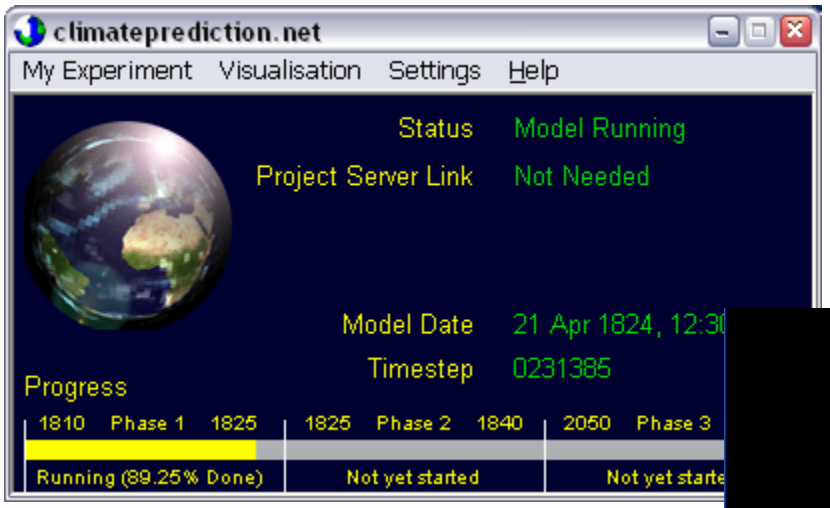
Statistics

- > 110,000 participants.
- > 24M years simulated.
- > 110,000 completed simulations. (Each 45years of model time)
- 10000 years of computing time.

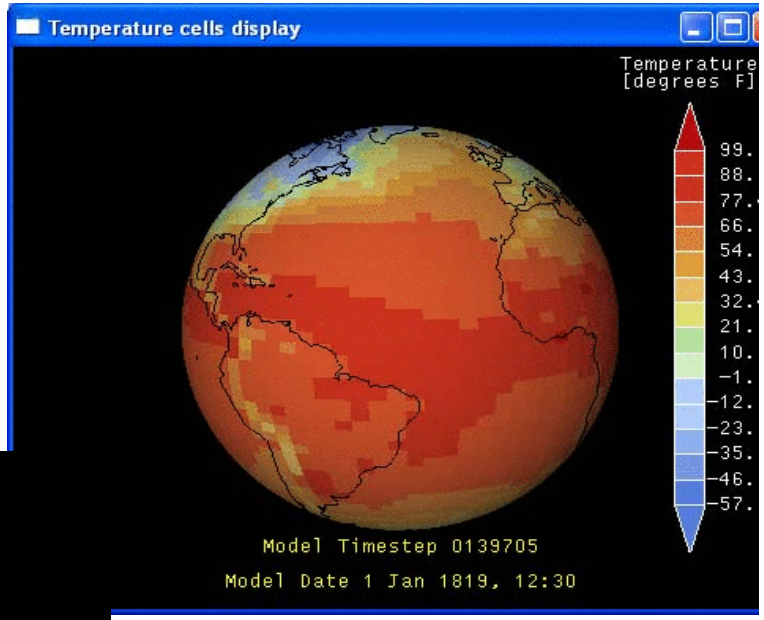


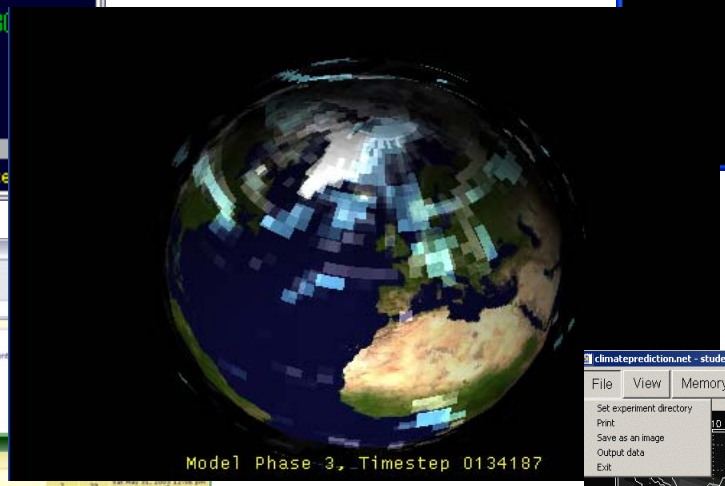
climateprediction.net

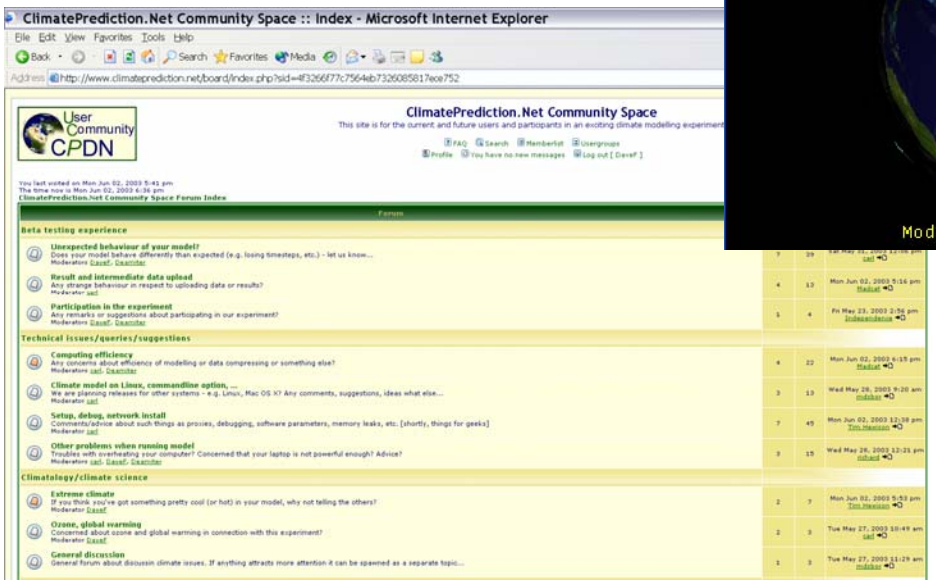
Climate*Prediction*.net : What it looks like.

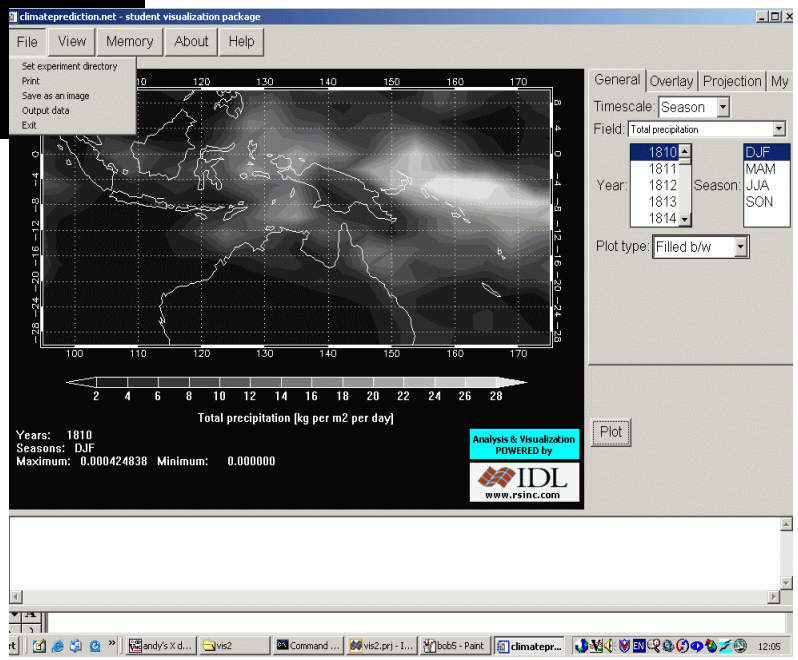


Teams.
P2P?









climateprediction.net

Distributed Computing is much more than climate*prediction*.net

climate*prediction*.net

Other Public Resource Distributed Computing (PRDC) Projects

GIMPS	SETI@home	Folding@home
LHC@home	Einstein@home	Lifemapper
Find-a-drug	FightAIDS@home	Evolution@home
Eon	Compute Against Cancer	Drug Design Online
Muon1	Seventeen of Bust	

- There are 100-200M PCs connected to the internet.
- <1% are involved.
- Project development is getting easier thanks to generic platforms.
  e.g. the Berkeley Open Infrastructure for Network Computing (BOINC)

climateprediction.net

Running an Ensemble of GCM Simulations is not a Typical PRDC Application

- A typical GCM simulation takes weeks or months, not hours.
- GCMs use more memory.
- GCM simulations **produce** data for further analysis. Most PRDC projects analyse data or test an hypothesis so data transfer is not a problem.
- Potentially **all** the simulations are useful, not just those which find a result.
- The research is of direct social relevance so there is a remit to engage the participants in the research itself.

climateprediction.net

Climate Specific Design Issues

- Security for the participants and the experiment.
- Model stability and error catching.
- Implementation of legacy code and specific model versions e.g. HadSM3 + sulphur cycle, HadAM3+stratosphere.
- Flexible mechanism for simulation perturbation.
- Distributed (pre-data-collection) analysis and choice of diagnostics (dynamic updating of analysis code?)
- Reliable data collection to a network of international servers.
- Timing of data collection and management of bandwidth. (Trickle-back)
- Flexibility to include other models.
- Maintaining participant interest – visualization packages etc.

climateprediction.net

Comparison with Other “Classic” Volunteer Computing Projects

Execution times on an Intel Pentium 4 / 1.6GHz laptop

Application	*Memory (MB)*	*Disk (MB)*	*Time (hours)*	*Total Users*
CPDN	50	600	840	90,000
SETI@home	20	3	6	5,000,000
Folding@home	5	3	12	500,000
Distributed.net	5	1	8	250,000

climateprediction.net

Comparison with Other BOINC (new) Volunteer Computing Projects

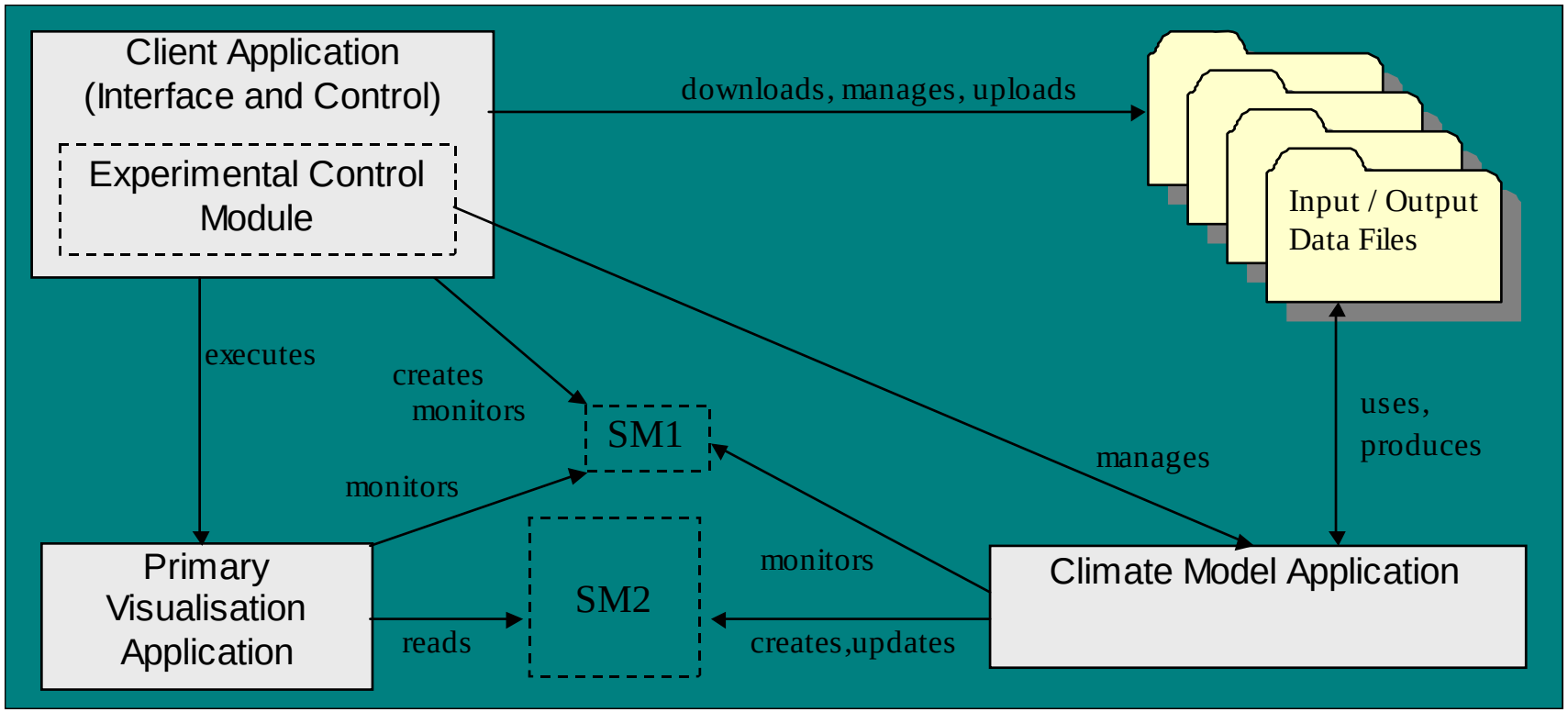
Execution times on an Intel Xeon / 3.4GHz

NB: Time given as days : hours : minutes : seconds

Application	*Min CPU Time*	*Avg CPU Time*	*Max CPU Time*	*Count*
CPDN	20:12:06:38	21:06:49:08	21:22:50:45	3
SETI@home	(did not finish)	00:02:28:48	00:06:40:43	751
Predictor@home	00:01:09:32	00:01:09:58	00:04:43:49	1521
Einstein@home	00:09:47:17	00:10:37:07	00:16:26:38	189
LHC@home	(did not finish)	00:01:07:35	00:12:13:03	589

climateprediction.net

Client Package Design



climateprediction.net

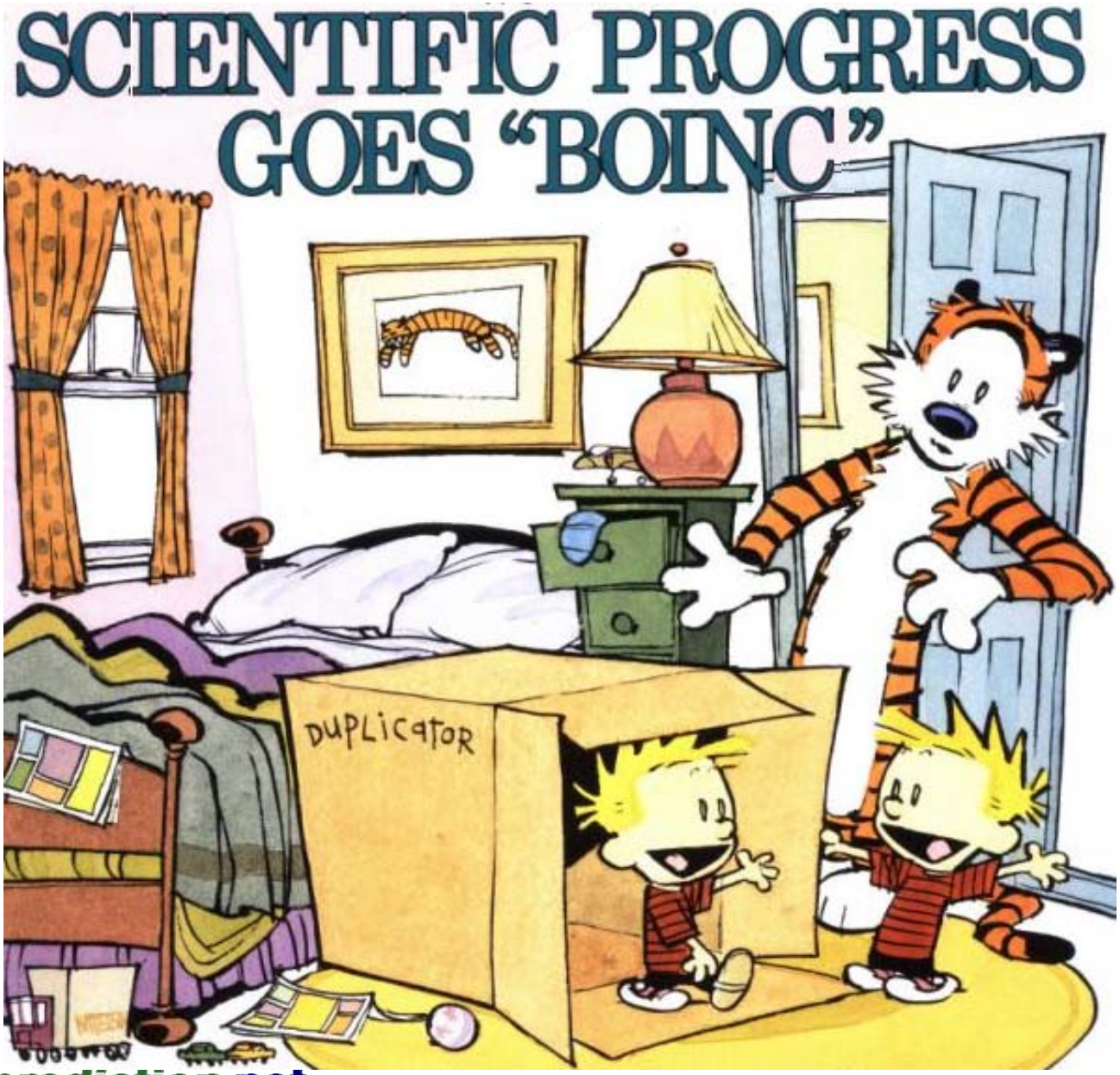
Security Issues

- Threats to participants (unexpected costs of participation):
  - Software package is digitally signed.
  - Communications are always be initiated by the client.
  - HTTP over a secure socket layer will be used where necessary to protect participant details and guarantee reliable data collection.
  - Digitally signed files can be used where necessary.

- Threats to the experiment (falsified data):
  - Two types of run replication:
    - Small number of repeated identical runs.
    - Large numbers of initial condition ensembles.
  - Checksum tracking of client package files to discourage casual tampering.
  - Opportunity to repeat runs as necessary.
  - Server security management and frequent backups.

climateprediction.net

CPDN and BOINC Integration



climateprediction.net

Apologies to Bill Watterson

Distributed Computing and Distributed Modelling.

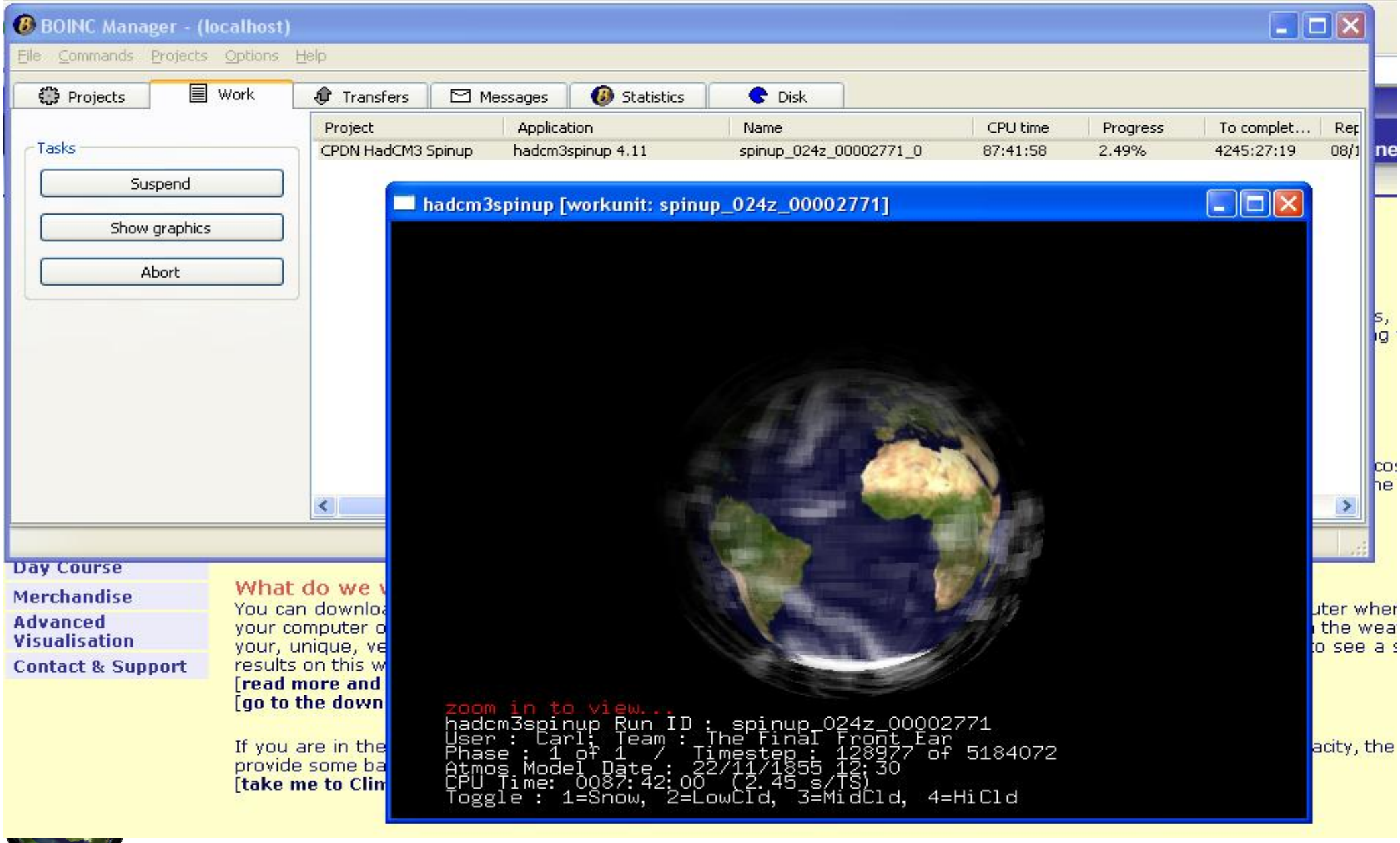
Climate*prediction*.net and BOINC

- Initially climate*prediction*.net developed all its own software but collaboration with the University of California at Berkeley has benefited both teams.
- The SETI@home / BOINC team at Berkeley have been helpful from the point climate*prediction*.net was conceived in 1999.
- David Anderson has provided very valuable guidance and we have now moved to the BOINC software.
- BOINC had to undergo considerable development to cope with the type of modelling application required.
- Differences of data volumes, memory usage and participant commitment are significantly different from most distributed (or volunteer) computing.
- Having solved these problems the cyberspace is open for other modelling projects presenting the opportunity for scientists to access vast amounts of computing power and to engage with the public worldwide.

climateprediction.net

BOINC!

Screenshot


BOINC Manager - (localhost)
File Commands Projects Options Help
Projects Work Transfers Messages Statistics Disk
Tasks
Suspend
Show graphics
Abort
Project Application Name CPU time Progress To complet... Rep
CPDN HadCM3 Spinup hadcm3spinup 4.11 spinup_024z_00002771_0 87:41:58 2.49% 4245:27:19 08/1
hadcm3spinup [workunit: spinup_024z_00002771]
zoom in to view...
hadcm3spinup Run ID : spinup_024z_00002771
User : Carl; Team : The Final Front Ear
Phase : 1 of 1 / Timestep : 128977 of 5184072
Atmos Model Date : 22/11/1855 12:30
CPU Time: 0087:42:00 (2.45 s/TS)
Toggle : 1=Snow, 2=LowCld, 3=MidCld, 4=HiCld
Day Course
Merchandise
Advanced Visualisation
Contact & Support
What do we
You can downlo
your computer
your, unique, ve
results on this w
[read more and
[go to the down
If you are in the
provide some ba
[take me to Clim
climateprediction.net

Beyond Climate*prediction*.net

Distributed modelling has the potential to be a major resource for the climate and meteorological research communities.

And for many other areas of scientific research.

It is likely to be particularly valuable in understanding uncertainty in model development.

It can also play a major role in engaging the public in scientific research. Particularly important in an era when scientific research and model results are directly guiding social and business policy.

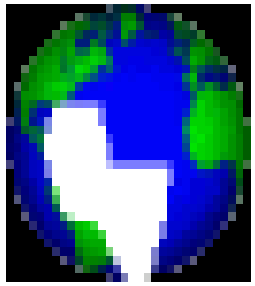
climate*prediction*.net

Climate*prediction*.net Experiments

- Experiment 1: Explore sensitivity of climate in a grand ensemble.
- Sub-experiment: Implications of THC shutdown.
- Experiment 2: Explore uncertainty in the effects of sulphate aerosols.
- Sub experiment: Spin-up of oceans for use in experiment 3.
- Experiment 3: Transient simulation of 1920 to 2080 with HadCM3L exploring:
  - Model uncertainty in the atmosphere.
  - Model uncertainty in the ocean.
  - Uncertainty in historic forcing.
  - Some uncertainty in future forcing.
  - Natural variability.
- Detection and attribution experiment: Quantify the change in risk of specific events such as the autumn 2000 floods.
- *Sub experiments: Mini-ensembles (a few hundred members) with embedded regional models.*
- *Arctic model ensemble? Impact model ensembles? Ocean shelf ensemble? Paleo-ensemble …*

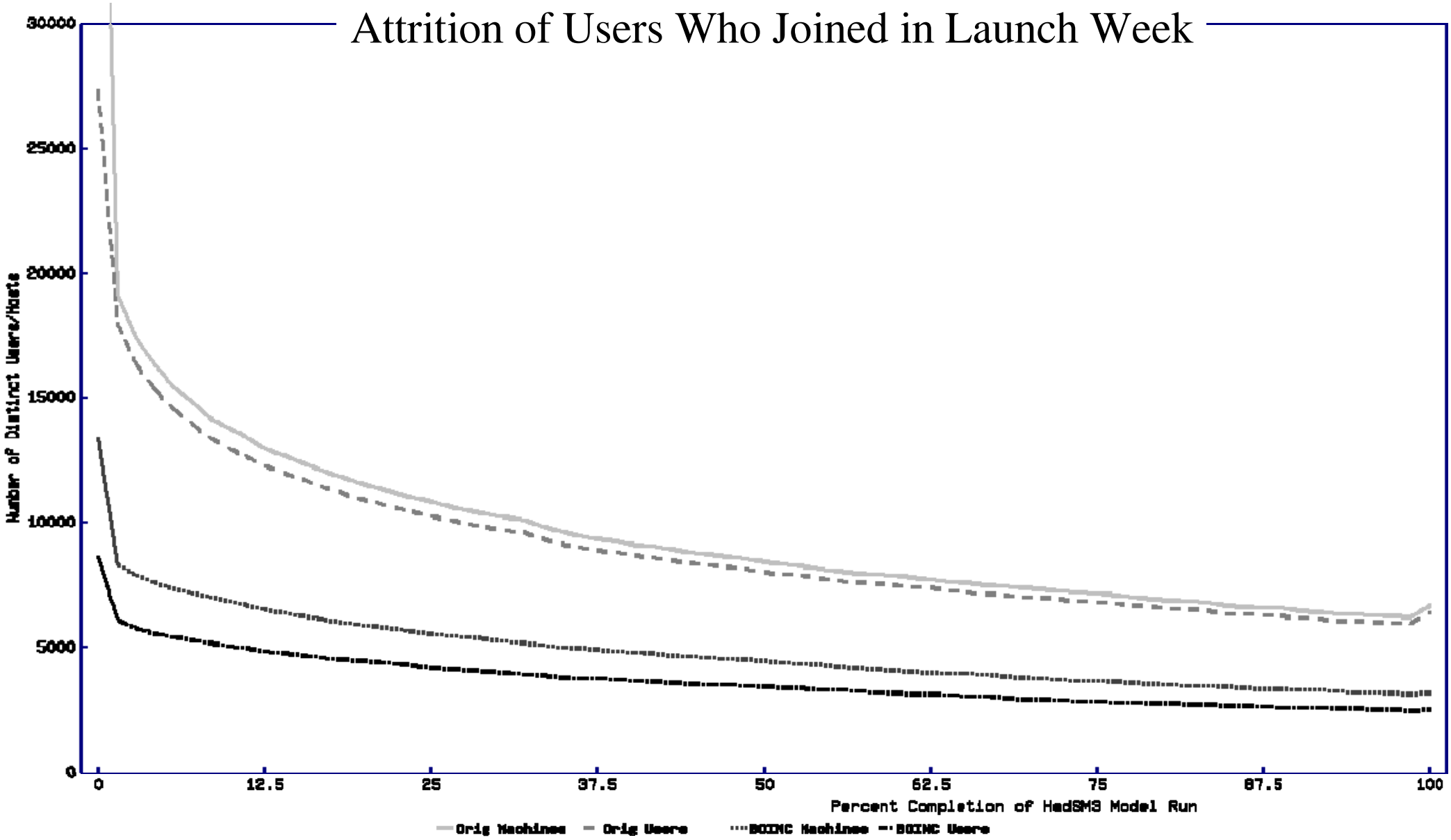
climateprediction.net

Visualization - Seasonal



Gfx1.exe

Attrition of Users Who Joined in Launch Week



climateprediction.net

Attrition Rates ("Classic" vs BOINC)

	Original CPDN	CPDN/BOINC
Launch Date	12/9/2003	29/08/2004
New Users Launch Week	27263	8570
1st Trickle Users	18373	6038
(made it to 1.4% completion)		
User Attrition	32.6%	29.5%
New Machines	51877	13301
1st Trickle Machines	19604	8285
Machine Attrition	62.2%	37.7%
Completed Runs	14895	8321
Completion Rate (per all user)	54.6%	97.1%
Completion Rate (per all mach)	28.7%	62.6%
Completion Rate (per 1st T user)	81.1%	137.8%
Completion Rate (per 1st T mach)	76.0%	100.4%
Active After 6 Mths		
Users	2101 (7.7%)	2894 (33.8%)
Machines	2195 (4.2%)	3748 (28.2%)

climateprediction.net

Conclusions

- We can't yet rule out the possibility of extreme responses to relatively modest stabilization levels. Even current levels could lead to dangerous climate change.
- Climate predictions need to include uncertainty analyses. It comes hand in hand with confidence.
- These uncertainty analyses will require very large numbers of simulations. Distributed (or volunteer) computing is currently the only way we can consider doing this.
- Future work will need to encompass inter-disciplinary and regional studies.

climateprediction.net