UNIVERSITY OF Southampton
School of Engineering Sciences

Using the NGS to study bi-stability of the Atlantic thermohaline circulation

NGS Roadshow - University of Exeter

Andrew Price
17 Sep 2009

GENIE

NATURAL ENVIRONMENT RESEARCH COUNCIL

Overview

- GENIE
- Example study of Thermohaline Circulation
- Use of the NGS
- Geodise and GENIELab software
- Study results

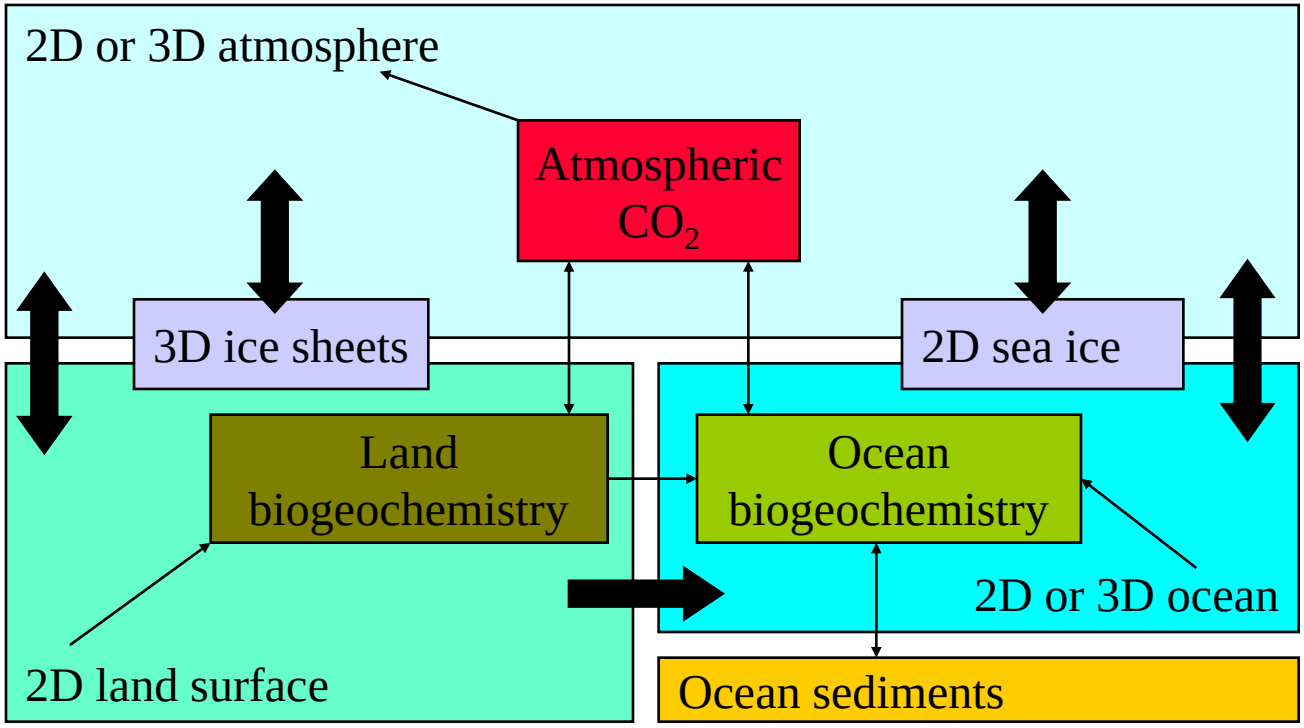
GENIE

- Grid ENabled Integrated Earth system model
- GENIE is a Grid-enabled computing framework enabling users to:
  - Flexibly couple together state-of-the-art components to form a unified Earth System Model (ESM)
  - Execute the resultant ESMs across computational Grid infrastructure
  - Share the resultant data produced by simulation runs
  - Provide high-level open access to the system, creating and supporting virtual organisations of Earth System modellers
- The central scientific goal of the GENIE project is to study the forcing and feedbacks driving the glacial-interglacial cycles that dominated the Earth's climate over the last 1 million years
- By better understanding the processes (physical and biogeochemical) which regulated these cycles in the past, we can be more confident about the predictions climate models make for the future

UNIVERSITY OF Southampton
School of Engineering Sciences

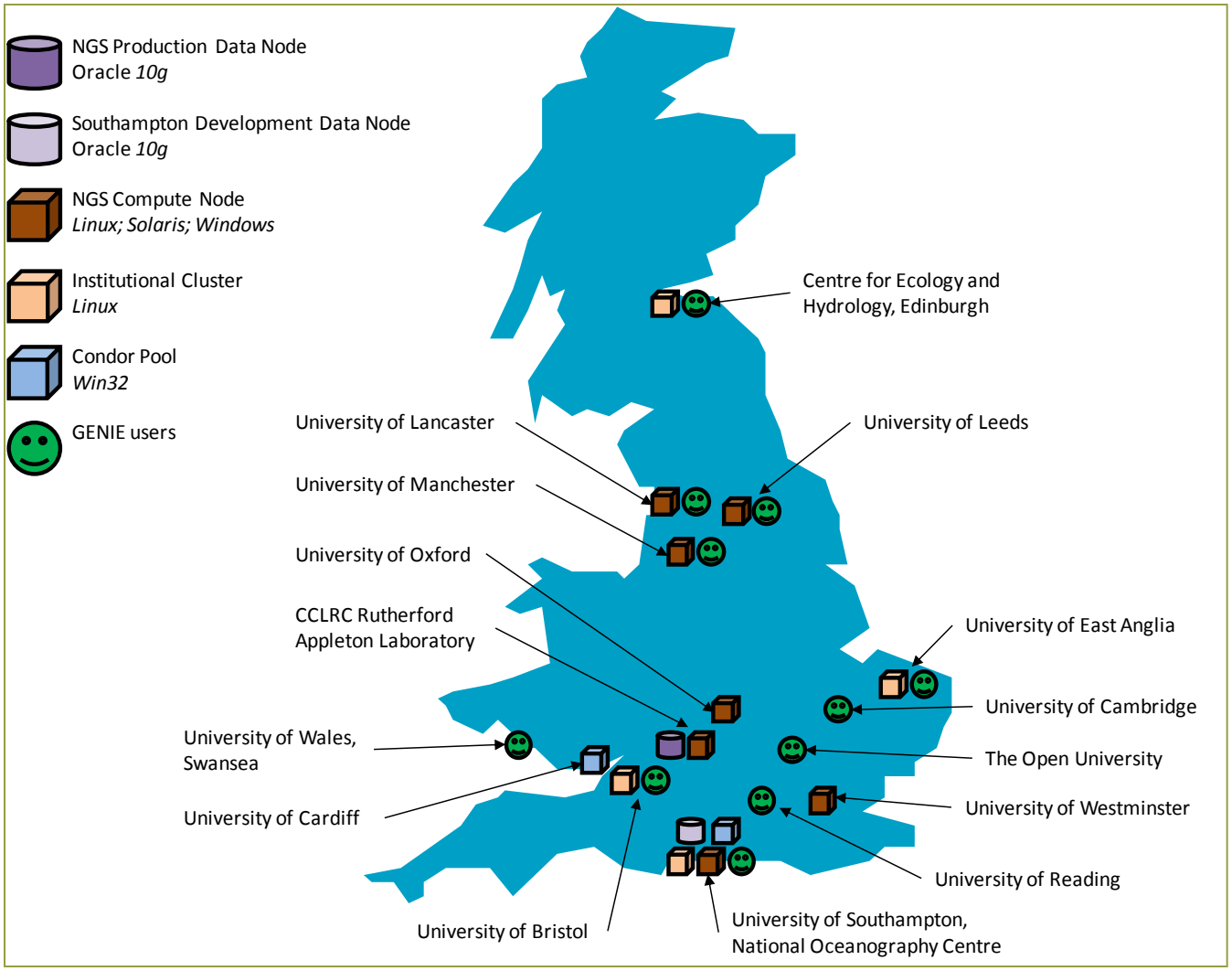
GENIE Modelling Framework

- Integrates component models of varying resolution, dimensionality & complexity
- Modular, scalable and traceable to more complex models



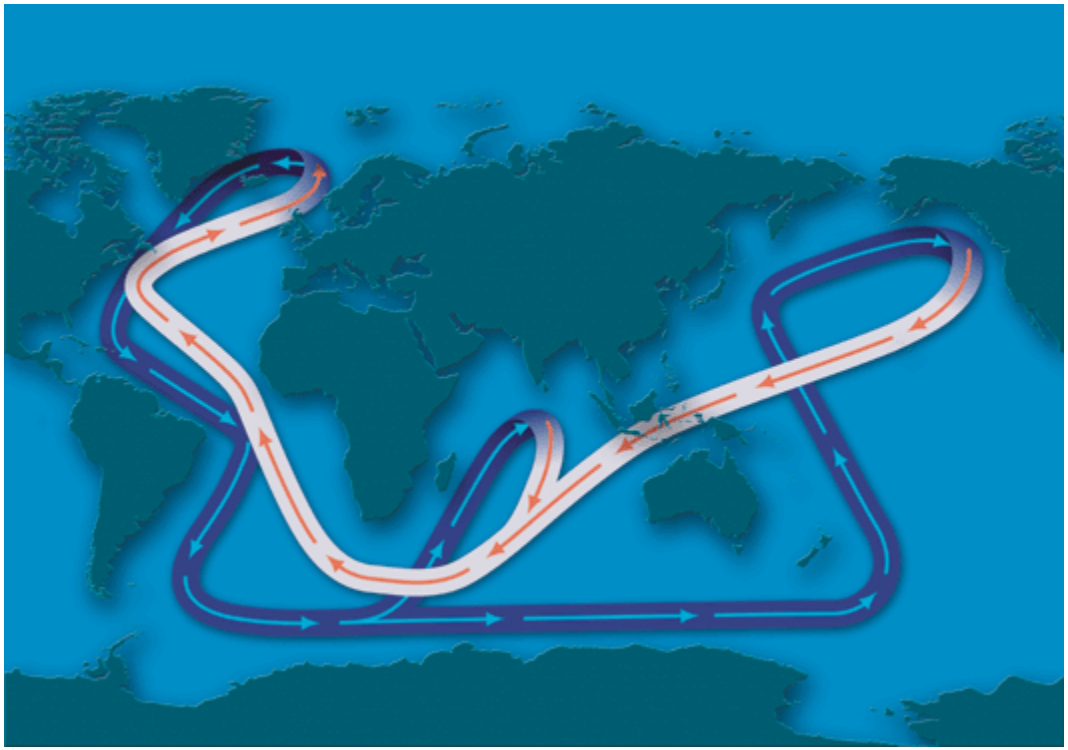
Schematic of the major component codes that are integrated to form climate models in the GENIE framework.

GENIE UK Community



The GENIE user base is primarily UK based with a growing number of international collaborators and users. The Grid computing software is used to exploit compute and data resources on the UK National Grid Service and institutional clusters and Condor pools.

Thermohaline Circulation

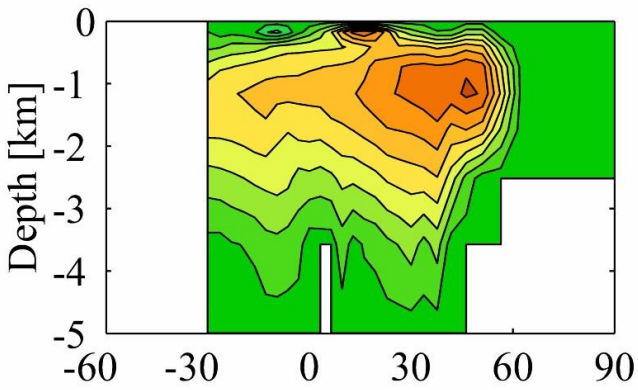


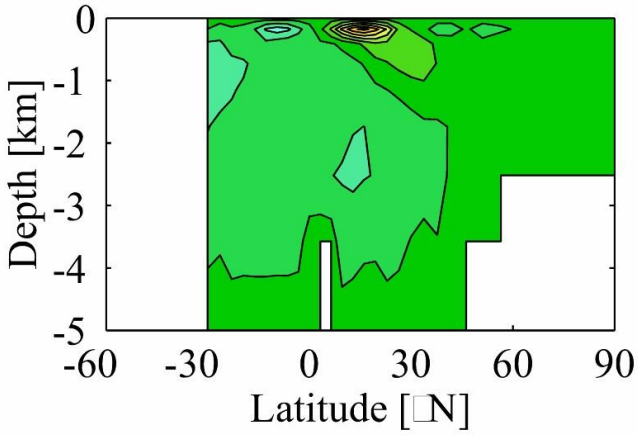
The Earth's oceans are linked by a system of currents, driven by temperature and salinity, collectively known as the Global Conveyor Belt.

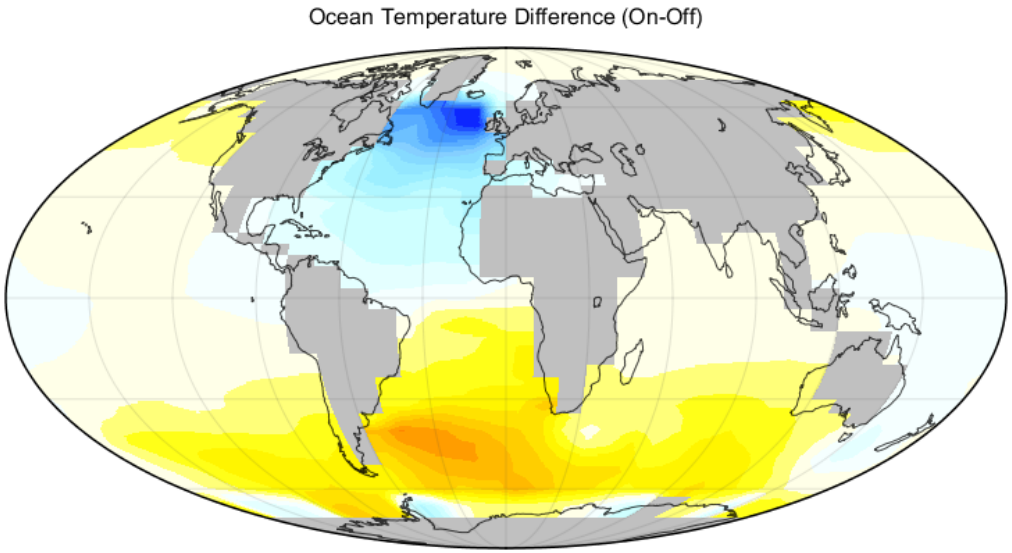
UNIVERSITY OF Southampton
School of Engineering Sciences

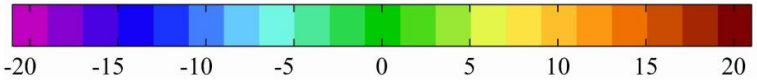
THC shutdown in the model

- THC “On” (top left)
- THC “Off” (bottom left)
- Temperature consequences (below)









Study design

- Two parameters of the model are identified affecting the freshwater budget of the Atlantic …
  - FW_x Atlantic-to-Pacific zonal transport
  - K_q Atmospheric diffusivity (meridional transport)
- Simulations performed using a systematic sweep of the values of these parameters
- The results of these simulations (boundary conditions) were then used to feed new simulations to examine “classic bi-stability”

Use of the NGS

- Individual model runs require ~10 hrs CPU time to simulate 5000 years and reach equilibrium
- Initial study requires a minimum of (11x11x10x3)/24 = 151.25 CPU days
- Local resource can prove insufficient to meet project needs
  - Find that institutional HPC clusters can limit an individual's resource usage
  - University Condor pools have practical limits for the runtime of compute tasks
  - The NGS provides the means to access a greater pool of resource
- Parametric sweeps are more practical if employing a collaborative task-farming approach
- The NGS database service provides the means to collaboratively study GENIE models
- Other NGS/Grid use cases
  - Building Response Surface Models for parameter estimation exercises
  - Use of Condor pools for large sampling of parameter space

Geodise Toolboxes

Geodise Compute Toolbox

- Grid access from the Desktop
- Matlab and Jython interfaces
- Globus and Condor support

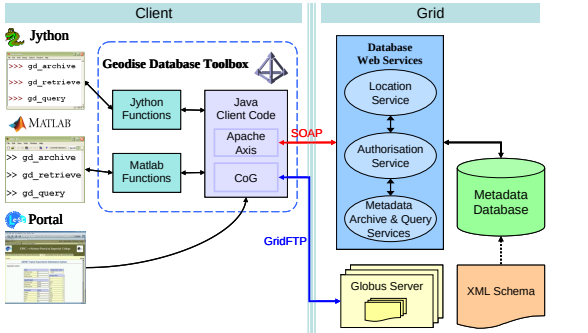
Geodise Database Toolbox

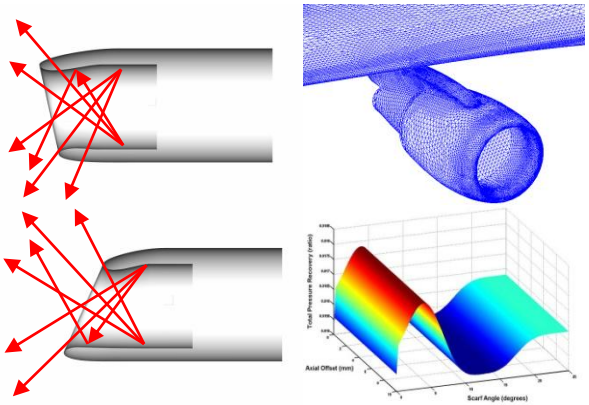
- Associate metadata with data
- Programmatic and GUI access

OptionsMatlab

- Engineering Design Optimisation
- Suite of multi-dimensional optimisation algorithms
- Parameter estimation

Geodise Compute Toolbox	gd_createproxy.m	Creates a Globus proxy certificate for the user's credentials
	gd_destroyproxy.m	Destroys the local copy of the user's Globus proxy certificate
	gd_jobsubmit.m	Submits a compute job to a Globus GRAM job manager
	gd_jobstatus.m	Gets the status of a Globus GRAM job
	gd_putfile.m	Puts a remote file using GridFTP
	gd_getfile.m	Retrieves a remote file using GridFTP
	gd_rmfile.m	Deletes a remote file using GridFTP
	gd_makedir.m	Creates a remote directory using GridFTP
	gd_rmdir.m	Deletes a remote directory using GridFTP
Geodise Database Toolbox	gd_archive.m	Archives a file or data structure to the database
	gd_query.m	Query the database for data matching specified criteria.
	gd_retrieve.m	Retrieves a file or data structure from the database





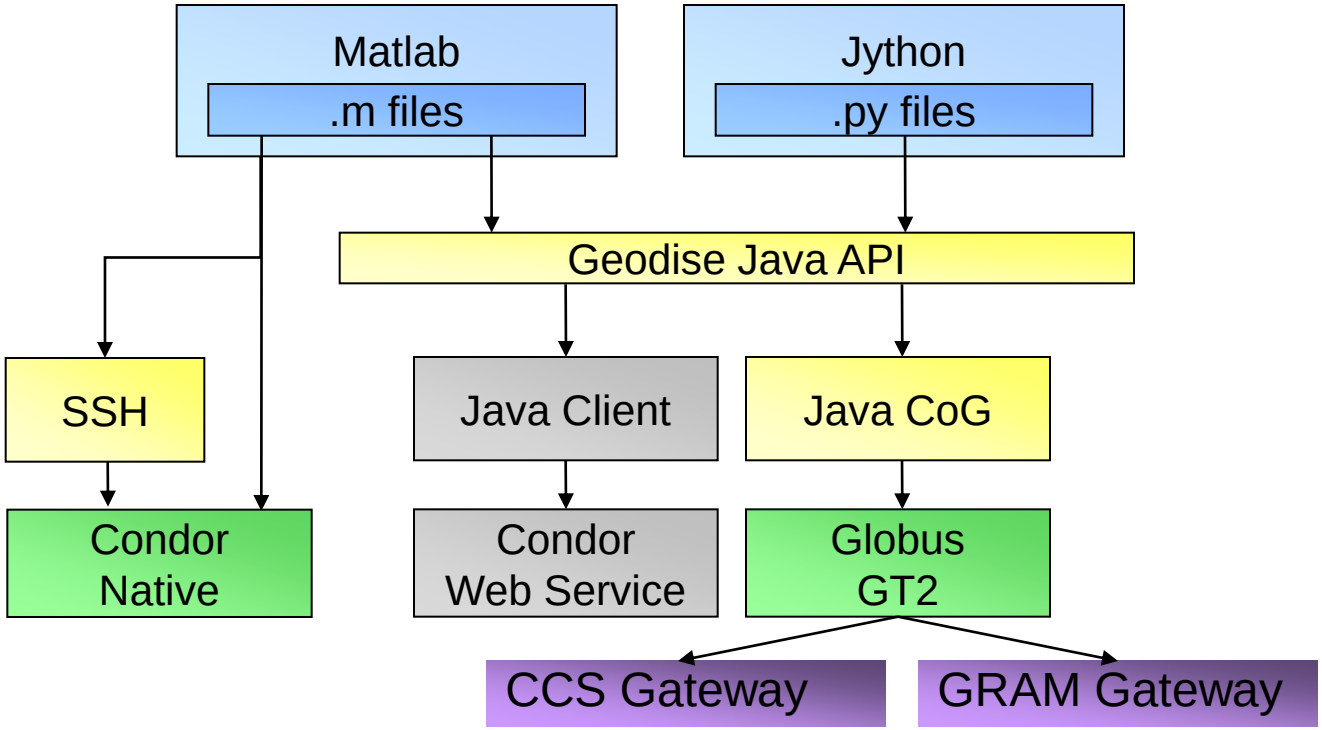
UNIVERSITY OF Southampton
School of Engineering Sciences

Geodise Compute Toolbox

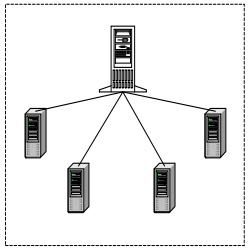
`gd_createproxy.m`	Creates a Globus proxy certificate for the user's credentials
`gd_destroyproxy.m`	Destroys the local copy of the user's Globus proxy certificate
`gd_certinfo.m`	Returns information about the user's certificate
`gd_proxyinfo.m`	Returns information about the user's proxy certificate
`gd_proxyquery.m`	Queries whether a valid proxy certificate exists
`gd_jobsubmit.m`	Submits a compute job to a Globus GRAM job manager
`gd_jobstatus.m`	Gets the status of a Globus GRAM job
`gd_jobpoll.m`	Queries the status of a Globus GRAM job until complete
`gd_jobkill.m`	Kills a Globus GRAM specified by job handle
`gd_putfile.m`	Puts a remote file using GridFTP
`gd_getfile.m`	Retrieves a remote file using GridFTP
`gd_rmfile.m`	Deletes a remote file using GridFTP
`gd_makedir.m`	Creates a remote directory using GridFTP
`gd_rmdir.m`	Deletes a remote directory using GridFTP

Grid Computation

UNIVERSITY OF Southampton
School of Engineering Sciences



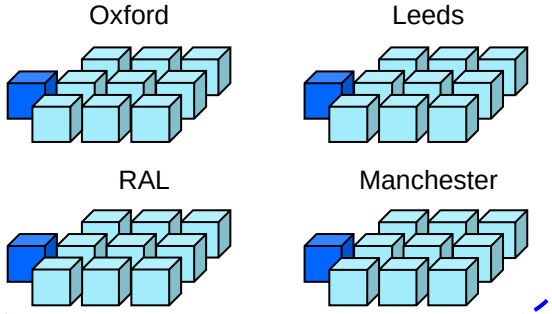
Condor Pool



Microsoft Compute Cluster Server

Institutional Resources (GT2)

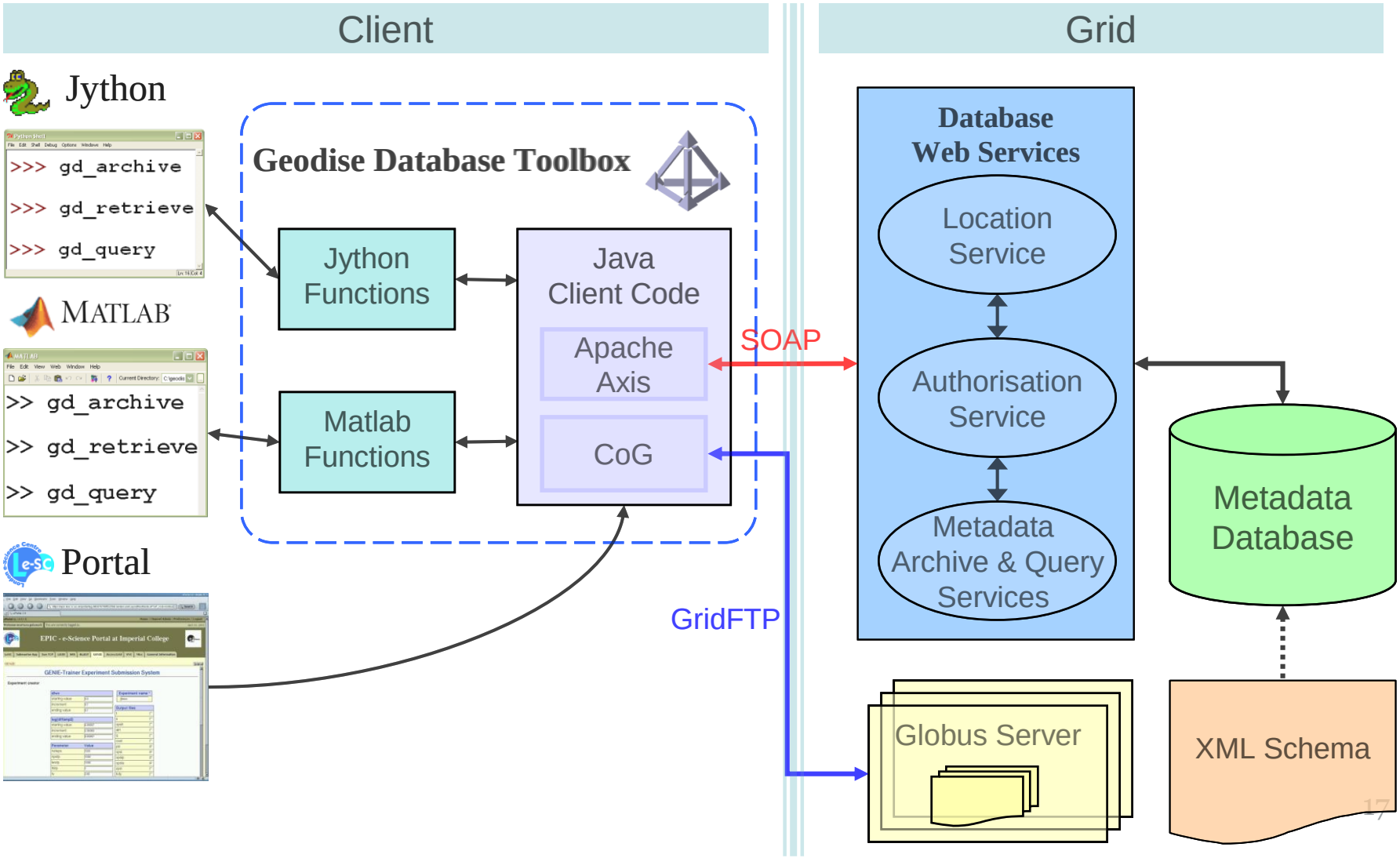
UK National Grid Service (GT2)



Geodise Database Toolbox

`gd_archive`	Stores a file or structure with metadata into an archive
`gd_query`	Queries over metadata or structures in the archive
`gd_retrieve`	Retrieves a file or structure from the archive to the local machine
`gd_addusers`	Grants users permission to access some data
`gd_datagroup`	Creates a new datagroup to aggregate files and structures
`gd_datagroupadd`	Adds a file or structure to a datagroup
`gd_display`	Displays the results of a query

Data Management System

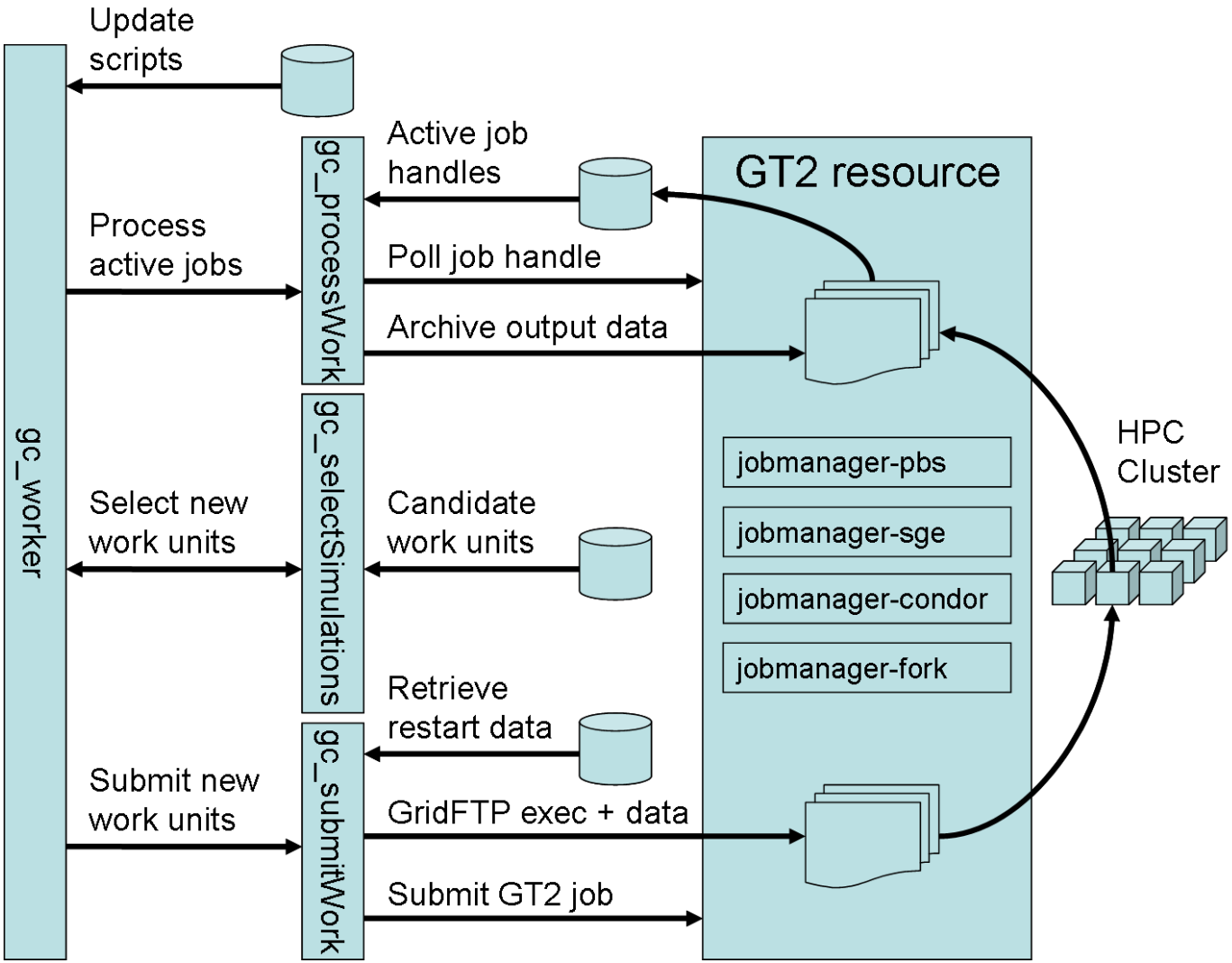


UNIVERSITY OF Southampton
School of Engineering Sciences

GENIELab

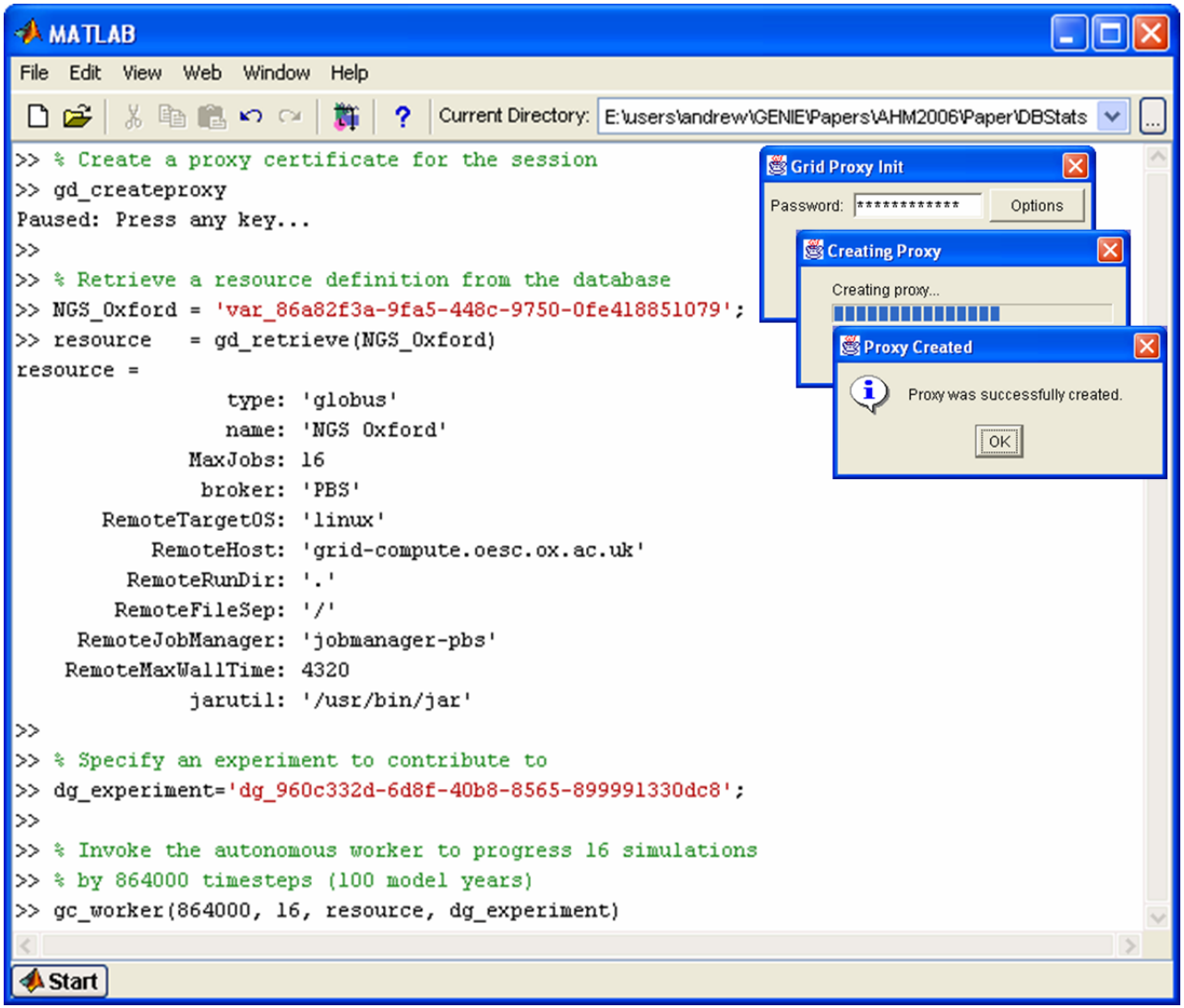
- GENIELab provides a set of Matlab functions to manage GENIE models on the Grid
- Built upon the Geodise toolboxes
- Provides routines to automatically manage metadata creation for archival of simulation output
- Scripted workflows available for
  - collaborative study of ensembles of simulations
  - parameter estimation studies exploiting OptionsMatlab
  - visualisation of simulation and ensemble output

Collaborative Ensemble Study Workflow



Workflow schematic for the activities of the gc_worker script. Simulations are advanced through to completion by the automated submission of compute jobs

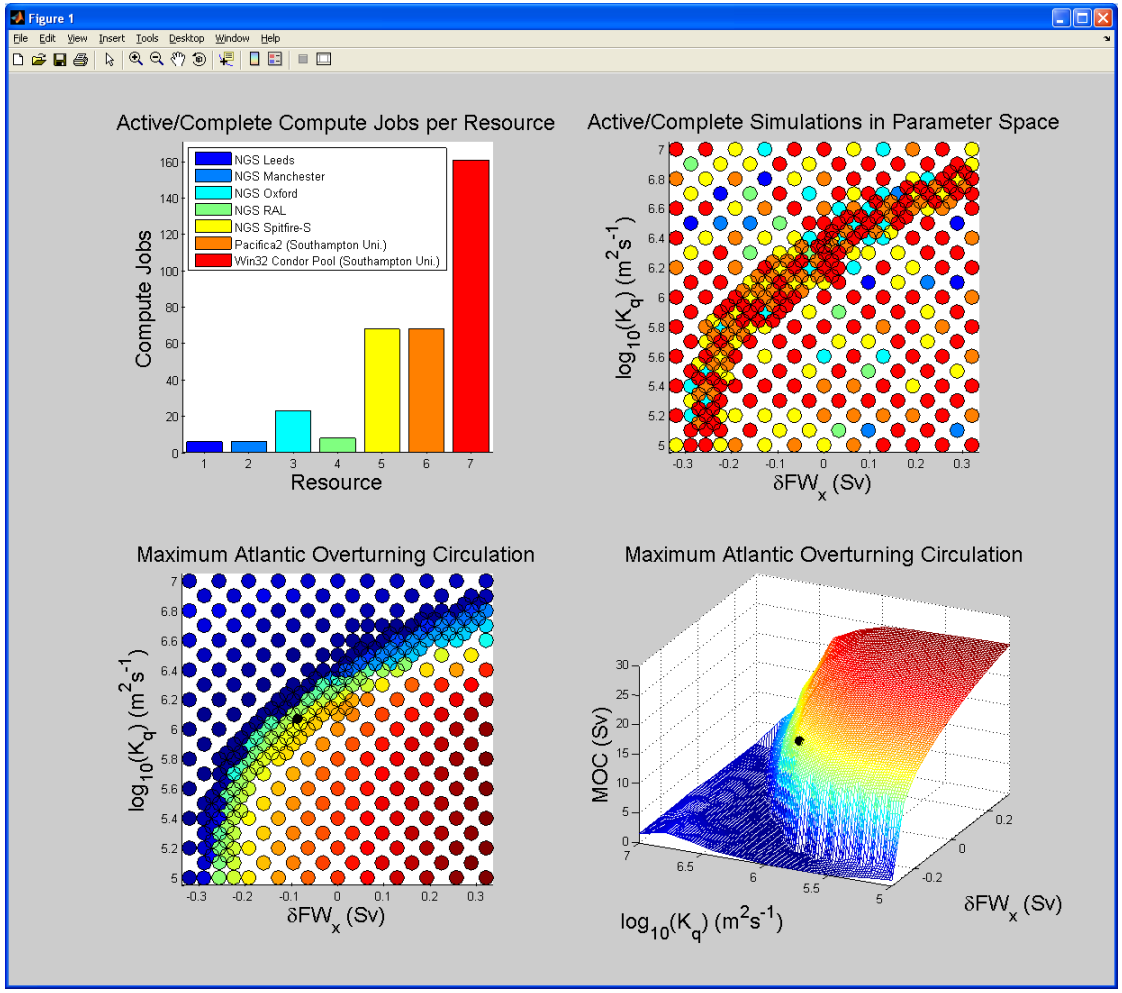
Client Session



Typical client session used to contribute resource for the advancement of a specified number of simulations. In practice, such sessions are typically invoked at regular intervals with automated job scheduling.

UNIVERSITY OF Southampton
School of Engineering Sciences

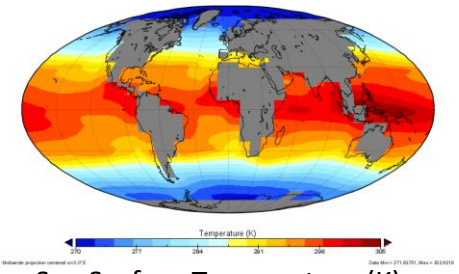
Interactive Monitoring GUI



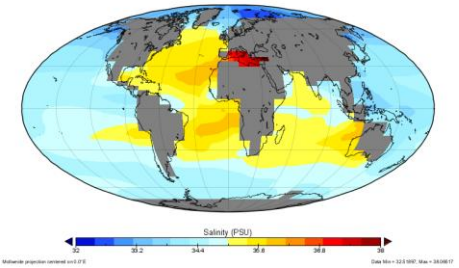
The existing Matlab GUI is also used for interactive analysis and post-processing of the experiment data sets.

Atlantic Overturning Circulation "ON"

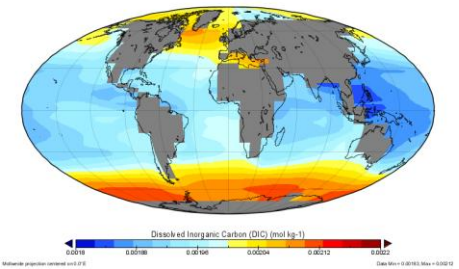
Difference: "ON" - "OFF"



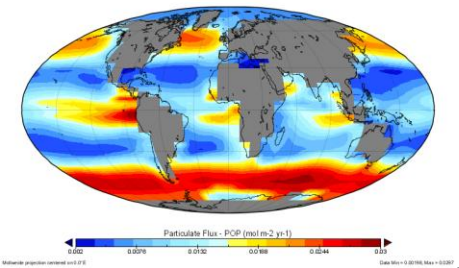
Sea Surface Temperature (K)



Sea Surface Salinity (PSU)



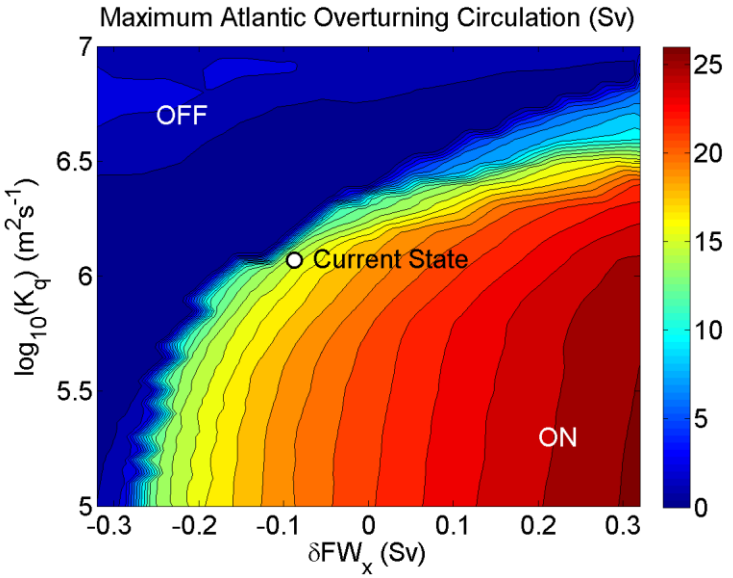
Dissolved Inorganic Carbon (mol kg^{-1})

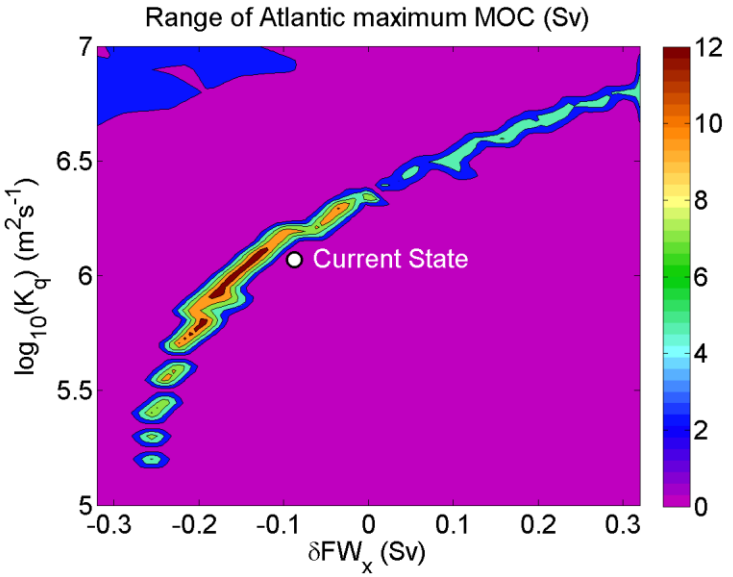


Biological Export Flux (mol m^{-2} yr^{-1})

Atmosphere **more** diffusive

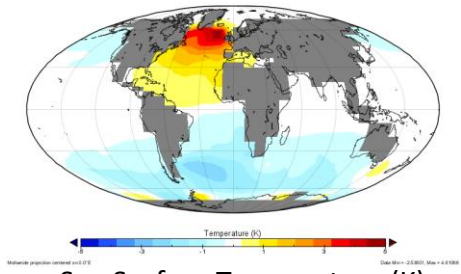
Atmosphere **less** diffusive



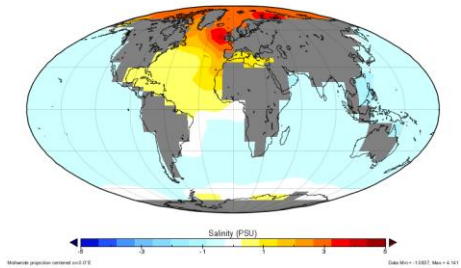


Atlantic **wetter**

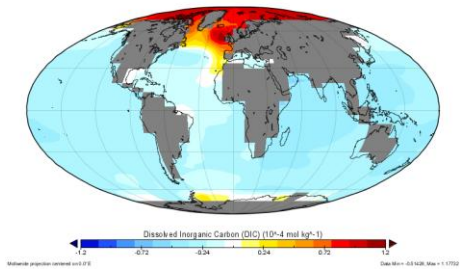
Atlantic **drier**



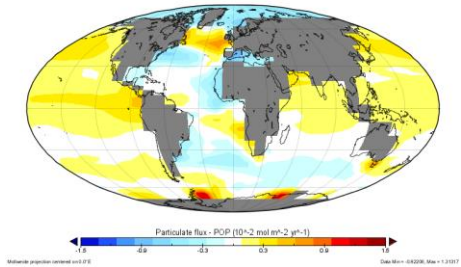
Sea Surface Temperature (K)



Sea Surface Salinity (PSU)



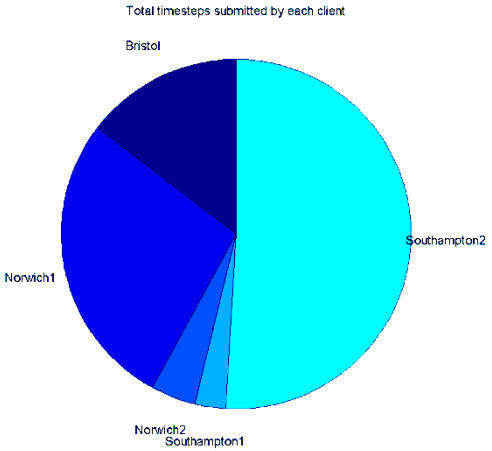
Dissolved Inorganic Carbon (mol kg^{-1})

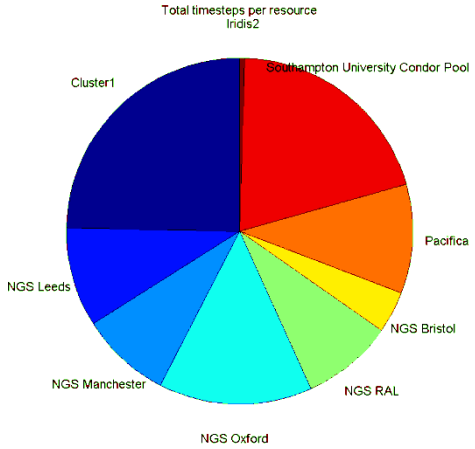


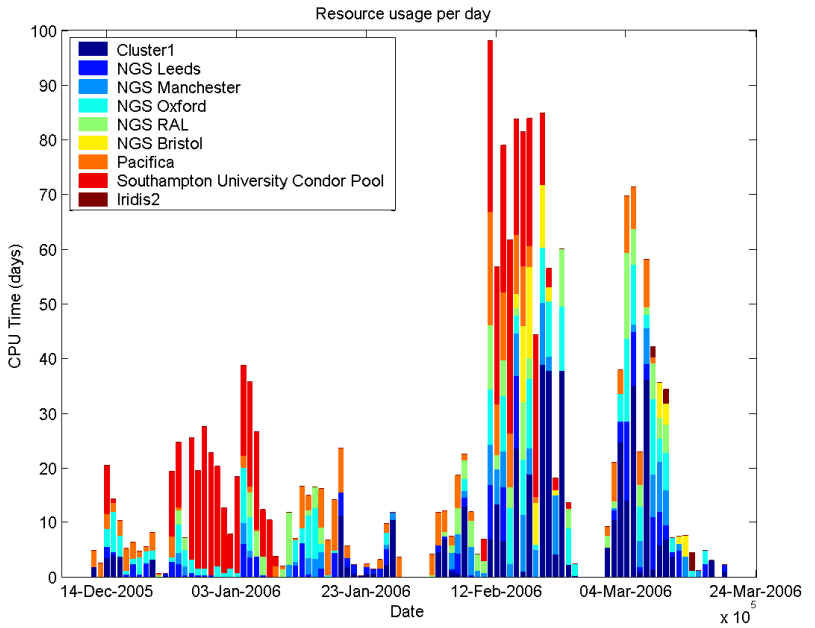
Biological Export Flux (mol m^{-2} yr^{-1})

Use of Grid Computing

- NGS usage in a more extensive previous study
- 12 ensemble studies
  - 362 simulations
  - 428,000 model years
  - 3,736 compute tasks
- 46,992 CPU hours
  - 1958 days = 5.36 years
- 9 compute resources
  - 5 nodes of UK National Grid Service
  - 3 Institutional HPC Clusters
  - Southampton University Condor pool
- Studies completed in just over 3 months







Summary

- GENIELab (built on Geodise) provides the means to manage GENIE models on the Grid from the desktop
- Users exploit NGS computational resources in conjunction with local institutional facilities
- NGS (RAL) Oracle 10g database service is used to manage and share simulation output and descriptive metadata
- NGS provides the services that GENIE members exploit for large collaborative study of climate models

University of Southampton
School of Engineering Sciences

The GENIE Team

Principal Investigator – GENIEfy (2005-2008):

Tim Lenton – *UEA Norwich*

Principal Investigator – GENIE (2002-2005):

Paul Valdes – *Bristol*

Research Team and Collaborators:

James Annan – *FRSGC, Japan*
Chris Armstrong – *Manchester*
Chris Brockwell – *UEA Norwich*
David Cameron – *CEH Edinburgh*
Peter Cox – *Hadley Centre (UKMO)*
Neil Edwards – *Open University*
Sudipta Goswami – *UEA Norwich*
Robin Hankin – *NOC Southampton*
Julia Hargreaves – *FRSGC, Japan*
Phil Harris – *CEH Wallingford*
Zhuoan Jiao – *Southampton e-Science Centre*
Martin Johnson – *UEA Norwich*
Eleftheria Katsiri – *London e-Science Centre*
Valerie Livina – *UEA Norwich*
Dan Lunt – *Bristol*
Richard Myerscough – *NOC Southampton*
Sofia Panagiotidi – *London e-Science Centre*
Andrew Price – *Southampton e-Science Centre*
Andy Ridgwell – *Bristol*
Ian Rutt – *Bristol*
Gethin Williams – *Bristol*
Mark Williamson – *UEA Norwich*
Gang Xue – *Southampton e-Science Centre*
Andrew Yool – *NOC Southampton*

Co-Investigators / Management team:

Peter Challenor – *NOC Southampton*
Trevor Cooper-Chadwick – *Southampton e-Science Centre*
Simon Cox – *Southampton e-Science Centre*
John Darlington – *London e-Science Centre*
Rupert Ford – *Manchester*
Eric Guilyardi – *Reading*
John Gurd – *Manchester*
Richard Harding – *CEH Wallingford*
Robert Marsh – *NOC Southampton*
Tony Payne – *Bristol*
Graham Riley – *Manchester*
John Shepherd – *NOC Southampton*
Rachel Warren – *UEA Norwich*
Andrew Watson – *UEA Norwich*

The Open University