Collaborative study of GENIE*fy* Earth System Models using scripted database workflows in a Grid-enabled PSE

A. R. Price[1], Z. Jiao[1], I. I. Voutchkov[1], T. M. Lenton[2], G. Williams[3], D. J. Lunt[3], R. Marsh[4], P. J. Valdes[3], S. J. Cox[1] and the GENIE team

[1]School of Engineering Sciences, University of Southampton, Southampton, UK
[2]School of Environmental Sciences, University of East Anglia, Norwich, UK
[3]School of Geographical Sciences, University of Bristol, Bristol, UK
[4]National Oceanography Centre, University of Southampton, Southampton, UK

Abstract

The integration of computational grids and data grids into a common problem solving environment enables collaboration between members of the GENIE*fy* project. In addition, state-of-the-art optimisation algorithms complement the component framework to provide a comprehensive toolset for Earth system modelling. In this paper, we present for the first time, the application of the non-dominated sorting genetic algorithm (NSGA-II) to perform a multiobjective tuning of a 3D atmosphere model. We then demonstrate how scripted database workflows enable the collective pooling of available resource at distributed client sites to collaboratively progress ensembles of simulations through to completion on the computational grid. A systematic study of the oceanic thermohaline circulation in a hierarchy of 3D atmosphere-ocean-sea-ice models is presented providing evidence for bi-stability in the Atlantic Meridional Overturning Circulation.

1 Introduction

The GENIE project (Grid ENabled Integrated Earth system model [1]) has created a Grid enabled component framework for the composition, execution and management of Earth System Models (ESMs). Mature simulation codes that model constituents of the Earth system (e.g. ocean, atmosphere, land surface, etc.) at varying resolution, complexity and dimensionality can be coupled together to provide a comprehensive hierarchy of climate models. The focus of the project is to create computationally efficient models for the study of the Earth system over millennial timescales and in particular to study ice age cycles and long-term human induced climate change.

Grid technology is a key enabler for the collaborative flexible coupling of constituent models, subsequent execution of the resulting ESMs and the management of the data that they generate. In this paper, we demonstrate how the flexible construction of ESMs in the GENIE framework in conjunction with the Grid software deployed for the project is exploited to tune a constituent component and then execute large ensemble simulations of computationally expensive models to study bi-stability in the oceanic "thermohaline circulation". We demonstrate how scripted database workflows enable the collective pooling of available resource at distributed client sites to progress an ensemble of simulations through to completion.

In this paper, we present the results of recent work extending the study of Marsh *et al* [2]. Section 2 presents the scientific problem that this work addresses. We discuss the Grid-enabled Problem Solving Environment (PSE) that we have exploited to perform these studies in Section 3. The results from a comprehensive study of bi-stability in a 3D atmosphere-ocean-sea-ice model are presented in Section 4. We discuss the merits of our approach in Section 5 and conclude in Section 6.

2 Scientific Challenge

A significant aspect of the climate system is the thermohaline circulation (THC), the name given to the system of large currents that connect and flow through the major oceans of the world. A principal component of the present day THC is the Gulf Stream in the North Atlantic which brings warm surface waters from the Gulf of Mexico to northern Europe. In the middle and high latitudes, heat and moisture from the ocean are lost to the atmosphere giving rise to the temperate climate in this region. As a consequence of this transfer the ocean waters become both cooler and more saline and hence increase in density. By a process of deep convection cool water sinks to the ocean floor and flows southwards until it reaches the

Southern Ocean. The north flowing surface current and the south bound deep current form a major part of the global ocean "Conveyor Belt" and are often referred to collectively as the Atlantic Meridional Overturning Circulation (MOC). The vast currents that circulate the globe because of the MOC are responsible for transport of heat energy and salinity and play a major role in determining the Earth's climate.

The strength of the Atlantic MOC is sensitive to the global hydrological cycle. In particular, variation in the atmospheric fresh water transport between the Atlantic and Pacific basins could lead to systematic changes in the THC. Since moisture transports by the atmosphere increase significantly under global warming scenarios, a key concern for climate change predictions is an understanding of how the THC may be affected under increased greenhouse gas forcing. It is possible that the system will react in a non-linear fashion to increased fresh water transports and the Atlantic MOC could collapse from its current "on" state into an "off" state, where no warm conveyor belt exists, and a colder, drier (more continental) winter climate may be expected in northern Europe. Studies using box models [3] and models of intermediate complexity (EMICs) such as GENIE-1 (C-GOLDSTEIN) [2], have found bi-stability in the properties of the THC; the "on" and "off" states can exist under virtually the same fresh water flux properties of the atmosphere depending on the initial conditions of the model (starting from an "on" or "off" state). However, the most comprehensive type of climate model, the coupled Atmosphere-Ocean General Circulation Models (AOGCMs), have yet to find any conclusive evidence for this bi-stability. In this paper, we extend the work performed with the GENIE-1 model (comprising a 3D ocean, 2D atmosphere and sea-ice) and present the first study of THC in the GENIE-2 model, a fully 3D ocean-atmosphere-sea-ice EMIC model from the GENIE framework. We thus take a step up the ladder of model complexity to study the behaviour of the same ocean model but now under a 3D dynamical atmosphere in place of the simple 2D energy moisture balance code (EMBM) of GENIE-1.

This study extends the work of Marsh [2] in two key areas. First, the new atmosphere model must be tuned to provide a reasonable climatology when coupled to the 3D ocean. We present, for the first time, the application of a multiobjective tuning algorithm to this problem. The second issue relates to computational complexity. Due to the need for shorter timesteps to handle atmospheric dynamics and the addition of a third (vertical) dimension in the atmosphere, the GENIE-2 model requires approximately two orders of magnitude more CPU time than GENIE-1 to simulate an equivalent time period. In Marsh [2] each GENIE-1 simulation required only a few hours of CPU time and the entire study was performed in about 3 days on a 200 node flocked Condor pool. However, for GENIE-2 models, individual simulations require ~5-10 days of continuous run time and a cycle stealing approach is no longer appropriate. Indeed, simply securing a sufficient number of CPUs over weekly timescales would be a significant challenge. In the remainder of this paper we show how the Grid computing software described in [4] is essential for the break down of large ensembles of lengthy simulations into manageable compute tasks. The GENIE data management system is used to mediate the execution of large ensemble studies enabling members of the project to pool their resource to perform these sub-tasks. Through a collaborative effort the more expensive study of GENIE-2 is enabled.

3 Collaborative PSE

GENIE has adopted software developed by the Geodise project [5] to build a distributed collaborative problem solving environment [4] for the study of new Earth system models. The Geodise Toolboxes integrate compute and data Grid functionality into the Matlab and Jython environments familiar to scientists and engineers. In particular, we have built upon the Geodise Database Toolbox to provide a shared Grid-enabled data management system, a central repository for output from GENIE Earth system models. An interface to the OPTIONS design search and optimisation package [6] is also exploited to provide the project with access to state-of-the-art optimisation methods which are used in model tuning.

3.1 Data Management System

The Geodise data model allows users to create descriptive metadata in Matlab / Jython data structures and associate that metadata with any file, variable or datagroup (logical aggregation of data) archived to the repository. The Geodise XML Toolbox is used to convert the metadata to a valid XML document which is then ingested and made accessible in the database. The GENIE database system augments this model by defining XML schemas that constrain the permissible metadata and improve subsequent query performance. The interface to the database is exposed through web services and all access is secured and authenticated using X.509 certificates. Through a common client members of the project can access the database both programmatically and with a GUI interface.

Recent work on the Geodise database toolbox has further enhanced its functionality by supporting aggregated queries (e.g. max, count), datagroup sharing (i.e. users working on the same task can add data into a shared data group), and improved efficiency of data access control. These enhancements enable the scientist to compose more powerful and flexible scripted workflows which have the ability to discover, based on the content of the database, the status of an experiment and make further contributions to the study.

3.2 Task Farming

We exploit the GENIE database and the restart capabilities of the models to break down ensembles of long simulations into manageable compute tasks. By staging the restart files in the shared repository all members of the project can access the ensemble study, query against the metadata in the database to find new work units, submit that work to compute resources available to them and upload new results as they become available. Using the common client software project members can target local institutional clusters, Condor pools or remote resources available to them on the grid (Globus 2.4).

The task farming paradigm is ideally suited to Grid computing where multiple heterogeneous resources can be harnessed for the concurrent execution of work units. Examples of such systems are Nimrod/G [7] and GridSolve [8] which provide task farming capabilities in a grid environment. Nimrod/G allows users to specify a parametric study using a bespoke description language and upload the study to a central agent. The agent mediates the execution of the work on resources that are available to it. The GridSolve system works in a similar fashion, exploiting an agent to maintain details about available servers and then selecting resource on behalf of the user. In contrast we exploit the GENIE database system, providing scriptable logic on the client side to coordinate the task farming based on point-in-time information obtained from the central database. This allows more dynamic use of resource as the client is more easily installed within institutional boundaries. The study is mediated in persistent storage, as with Nimrod/G and GridSolve, allowing progress to be monitored and output data to be shared.

3.3 Collaborative study

The programmatic interface to the shared data repository allows the database to become an active part of an experiment workflow.

To define an ensemble in the database the coordinator of the study creates an experiment "datagroup", a logical entity in the database describing the purpose of the ensemble and acting as parent to a set of simulation definitions. The simulation entities record the details for the specific implementation of the model including the total number of timesteps to be performed and the output data to generate. This data structure within the database captures all of the information that is required for the study and its creation amounts to ~100 lines of configurable script for the scientist to edit.

Once a study has been defined the coordinator typically circulates the unique identifier to members of the project. If a user wishes to contribute compute resource for the progression of the experiment then this is the only piece of information that they require. They simply contribute to the study by invoking a "worker" script, specifying the experiment identifier and providing details of the amount of work they would like to submit to a particular resource. A user session is shown in Figure 1.

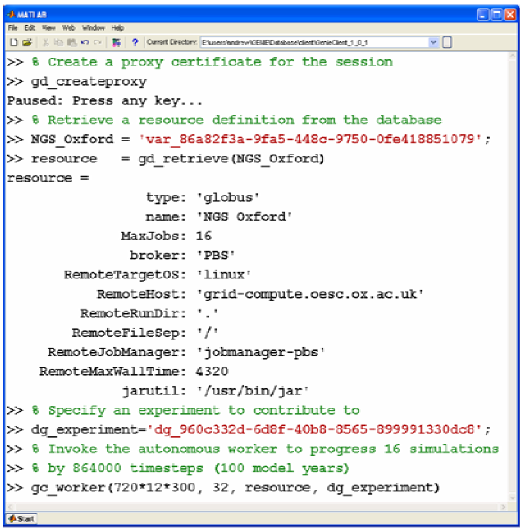

Figure 1: Typical user session contributing to an ensemble study.

The first action is to create a time-limited proxy certificate (to authenticate all actions on the Grid). The user then retrieves a resource definition from the database and invokes the autonomous "worker" script. The "worker" allows a user to specify the number of timesteps to progress the members of the ensemble by and the number of compute jobs to submit to the specified resource.

The user specifies a resource by creating a data structure in Matlab capturing the information needed by the system to exploit that resource. We provide a utility function that prompts the user for the necessary information and then upload an appropriate data structure to the database. Once the resource definition is available in the database it is more common for the user to retrieve the definition and use it to submit work to the resource it describes. The worker script progresses through four stages of execution (see Figure 2):

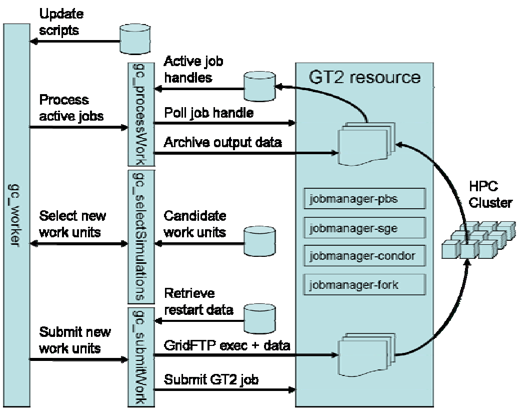

Figure 2: Scripted workflow of the autonomous "worker" interfacing to a resource managed by the Globus Toolkit (v2.4).

Stage 1: The worker interrogates the database for details of any changes to the scripts that define the experiment. Any new or updated scripts that have been added to the experiment are downloaded and installed in the user's local workspace. This provides the experiment coordinator the means to dynamically update the study and ensures that all participants are running the same study.

Stage 2: The worker invokes a post-processing script to ensure that any completed work units are archived to the database before new work units are requested. A list of active job handles is retrieved from the database and each handle is polled for its status. Any completed compute tasks are processed and the output from each is uploaded to the database. The files from which the model may be restarted are tagged as such. In the event of a model failure the post-processing will attempt to record the problem to allow future invocations of the worker to perform corrective action.

Stage 3: The worker invokes a selection script that queries the database for available compute tasks. Based on a point-in-time assessment of the progress of the study the system returns a list of work units that are available for progression. The list of jobs is prioritised so that the entire ensemble is progressed as evenly as possible; all inactive work units are returned first with those having least achieved timestep given highest priority.

Stage 4: The final action of the worker script is to submit the specified number of compute tasks to the resource that the user provided. Working through the list of compute tasks obtained in stage 3 the script retrieves the appropriate restart files from the database and transfers them, along with the model binary, to a unique directory on the remote compute resource. Once the model run is set up the script submits the job to the scheduler of the resource. The job handle returned by the scheduler is uploaded to the database. Subsequent invocations of a worker script on this experiment are then informed about all active compute tasks and can act accordingly.

It would be unreasonable to require a user to manually invoke the worker script on a regular basis and the most common mode of operation is to set up a scheduled task (Windows) or cron job (Linux) to automatically initiate the Matlab session. The user typically provides a script to run one or more workers that submit work to the resources available to the client. Through the regular invocation of the worker script the entire ensemble is progressed to completion without any additional input by the user. The progress of the study can be monitored at any point through a function call that queries the database for summary information about the experiment.

4 Results

We present the results of an end-to-end study of a small suite of GENIE-2 models. The first stage of the study performs a tuning of the IGCM atmosphere to provide a stable coupling to the GOLDSTEIN ocean component. Using an optimal parameter set we then define a number of ensembles in the database system for the IGCM coupled to the GOLDSTEIN model running on three computational meshes with varying horizontal resolution. Project members collaborate to progress the simulations through to completion on resource available to them both locally and nationally.

4.1 Multiobjective Tuning

An important aspect of Earth system model development is the tuning of free parameters so that the simulated system produces a reasonable climatology. In Price [4] the OPTIONS design search and optimisation package [6] was used to apply a Genetic Algorithm to vary 30 free parameters in the IGCM and minimise a single objective measure of the mismatch between annually averaged model fields and observational data. This measure of model-data mismatch was reduced by ~36% and the resulting parameters are now the preferred set for further study using the IGCM. However, while this tuning produced a good improvement in the sea-surface temperature (SST), surface energy fluxes and horizontal wind stresses, there was little or no improvement in the precipitation and evaporation fields which comprised part of the tuning target. Analysis of the model state at this point (GAtuned) in parameter space also showed that seasonal properties of the model were not a good match for seasonally averaged observational data.

Recent studies have therefore adopted a multiobjective tuning strategy in order to provide targeted improvements in seasonally averaged fields of the model across physically significant groupings of state variables. In general, multiobjective optimisation methods seek to find a set of solutions in the design space that are superior to other points when all objectives are considered but may be inferior in a subset of those objectives. Such points lie on the Pareto front and are those for which no other point improves all objectives [9]. We have exploited an implementation of the non-dominated sorting genetic algorithm (NSGA-II) [10]. As described in [4] the GENIE model is exposed as a function that accepts as input the tuneable parameters and returns, after simulation, the multiple objective measures of model mismatch to observational data. The NSGA-II algorithm maintains a population of solutions like a GA but uses a different selection operator to create each successive generation. The algorithm ranks each member of the parent population according to its non-domination level and selects the members used to create the next population from the Pareto-optimal solutions that have maximum separation in objective space. The method seeks to reduce all objectives while maintaining a diverse set of solutions. The result of the optimisation is a set of points on the Pareto front from which a user can then apply domain knowledge to select the best candidates.

The IGCM atmosphere component was tuned using the NSGA-II algorithm applied over 50 generations using a population size of 100. 32 free parameters in the genie_ig_sl_sl model (IGCM atmosphere, slab ocean, slab sea-ice) were varied and 2 constraints were applied. Three objective functions were defined to provide improvements in the seasonal averages of the surface energy fluxes (OBJ1), the precipitation and evaporation properties (OBJ2) and the surface wind stresses (OBJ3). The model runs were performed on a ~1400 node Condor pool at the University of Southampton with each generation taking approximately three hours to complete. The algorithm generated a Pareto optimal set of solutions comprising 117 members. The results of the tuning exercise are summarised in Table 1 comparing the 'best' result (arbitrarily selected as the point with minimum sum of the three objectives) of the Pareto study with the objectives evaluated at the default and GAtuned points in parameter space.

We first point out that the precipitation and evaporation in the model (OBJ2) saw little improvement in the original GAtuned study over the default parameters even though these fields were part of that tuning study. The GApareto result has produced significant improvements in all three objective functions as desired and provided a better match to the observational data. The scale of these improvements is also greater then the original GAtuned study, but we note that this is primarily due to the seasonal rather than annual averages that have been used to comprise the tuning target. That is to say that we are comparing the seasonal measures of fitness at a point that was obtained when optimising annual averages – we should expect a greater improvement in the seasonal study. The improvements in the precipitation and evaporation fields are not as great as the improvements made in the other two objectives and we note that there are probably fundamental limits to the model's ability to simulate precipitation and evaporation that tuning alone cannot overcome.

Name	**Default**	**GAtuned**	**GApareto**
OBJ1	4.47	3.68	3.23
OBJ2	3.87	3.84	3.41
OBJ3	3.11	2.32	2.08

Table 1: Values of the three objective functions evaluated at the default, GAtuned and GApareto points in parameter space.

4.2 Study of THC bi-stability

We have performed a number of ensemble studies of the GENIE-2 ig-go-sl model consisting of the 3D IGCM atmosphere code coupled to the 3D GOLDSTEIN ocean model with a slab sea-ice component. This model typically requires approximately 5 days of continuous run time to perform a 1000 year simulation of the Earth system, which is long enough for the global THC to approach equilibrium. In the Atlantic Ocean, the THC is characterised by a “meridional overturning”, which is diagnosed in the model as a streamfunction in the vertical-meridional plane. The units of this meridional overturning streamfunction are 10^6 m^3s^{-1}, or Sverdrups (Sv). The sign convention is positive for a clockwise circulation viewed from the east, corresponding to surface northward flow, northern sinking and southward deep flow. This circulation pattern corresponds to the Conveyor Belt mode of the THC, the strength of which is given by the maximum of the overturning streamfunction. In order to obtain a realistic Conveyor Belt in GENIE-2 (as in other similar models), it is necessary to include a degree of surface freshwater flux correction to ensure a sufficient Atlantic-Pacific surface salinity contrast. We have studied the Atlantic maximum overturning rate under a range of this freshwater flux correction in a three-level hierarchy of GENIE-2, across which we vary the resolution of the horizontal mesh in the ocean component.

Twelve ensemble studies were executed by three members of the project using database client installations at their local institutions. The unique identifier for each study was circulated amongst members of the project willing to contribute resource and regular invocations of the client “worker” script were scheduled at each site to submit model runs to a number of computational resources around the UK. These resources included five nodes of the UK National Grid Service (http://www.ngs.ac.uk), three Beowulf clusters available at the Universities of East Anglia and Southampton and a large Condor pool (>1,400 nodes) at the University of Southampton.

The ensemble studies of the GENIE-2 model presented here were carried out during a 14 week period from December 2005 until March 2006. In total five client installations were used to perform:

- 319 GENIE-2 simulations
- 3,736 compute tasks
- 46,992 CPU hrs (some timings estimated)
- 362,000 GENIE-2 model years

The daily breakdown of resource usage is plotted in Figure 3. The rate of progression of the studies was largely determined by the amount of available work in the system at any given time. The overheads in performing queries on the database and the upload and retrieval of model data files did not adversely impact the performance of the system.

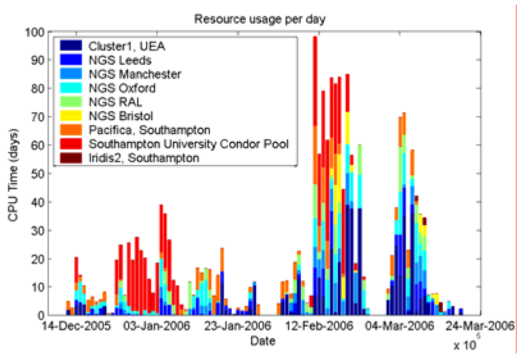

Figure 3: Resource usage for the twelve ensemble studies.

The breakdown of client contributions is presented in Figure 4a. The distribution of submitted jobs roughly reflects the amount of resource available to the client in each case. E.g. the client responsible for ~50% of the work was the only submission node on the large condor pool and was also used to submit jobs to the National Grid Service. The distribution of jobs across the computational resources is presented in Figure 4b and illustrates that work was distributed evenly to the available resources. By mediating the study through a shared central repository the coordinating scientist has had the work performed over a collected pool of resource, much of which is behind institutional firewalls and probably not directly accessible to him/her. The system also enables us to introduce new resource as a study is performed and target the most appropriate platforms.

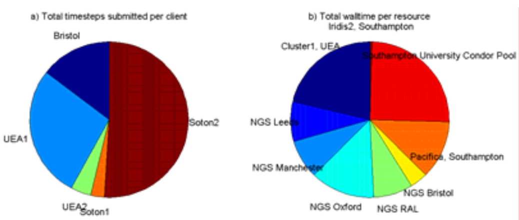

Figure 4: Distributions of client submissions and resource usage.

The data from these runs is being processed and a more detailed analysis of the results of these studies will be the subject of a separate paper [11]. We briefly present the initial findings of this experiment.

The effect of varying a fresh water flux correction parameter in the atmosphere component (IGCM) of a suite of GENIE-2 models with different ocean grid resolutions has been studied. Six of the ensembles are plotted in

Figure 5 showing the strength of the maximum Atlantic overturning circulation for ensemble members across three meshes (36x36 equal area, 72x72 equal area, 64x32 lat-lon) initialised from states with default present day circulation (r1) and collapsed circulation (r0). Maximum overturning rates are obtained as averages over the last 100 years of the 1000-year simulations.

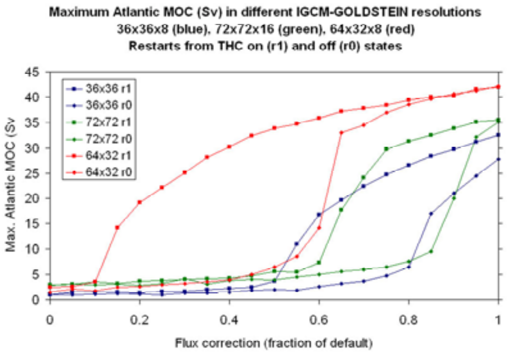

Figure 5: Maximum Atlantic overturning circulation (MOC) across six ensembles, for three different horizontal meshes of the ocean component.

Bi-stability in the Atlantic MOC is clearly evident in all three model variants. Under the same atmospheric conditions there is a range of fresh water corrections for which the ocean THC is stable in both the “on” and “off” states depending on initial conditions. The locations and widths of the hysteresis loops as a function of ocean mesh resolution are also an important feature of these results. The increase of grid resolution from 36x36 to 72x72 (effectively multiplying by volume the number of grid cells by 8) shows a slight narrowing of the hysteresis loop and steeper transitions. The use of an ocean grid at the same resolution as the atmosphere (64x32) exhibits a much wider loop and is positioned at a different point in the parameter space. The differences in these results are likely attributable to two factors; a) the interpolation employed in coupling the ocean and atmosphere models and b) the different resolutions of the ocean grids at the mid-latitudes which mean important processes are better resolved at 64x32. In the 64x32 model, there are more grid points in the east-west direction giving better resolution of zonal pressure gradients in the ocean, better THC structure, and improved northward transport of high salinity water by the THC itself, reducing the need for surface freshwater flux correction. These results tentatively support the notion that more complex models can exhibit bi-stability in the states of the THC.

5 Discussion and Future Work

Programmatic access to a shared central database provides the means to mediate the collaborative study of Earth System models in the GENIE project. In contrast to other grid-enabled task farming systems, such as Nimrod/G and GridSolve, we exploit client side scripted database workflows in place of a central agent to mediate the execution of our studies. Our system allows resource to be dynamically introduced to studies, but we achieve this at the expense of execution control. Systems with a central agent can build execution plans, target jobs at the most appropriate platform and provide guarantees for completion times of experiments. However, administrators of the agent system must ensure that this central server can communicate with all resource to be targeted. If available resource has not been exposed to the grid then there is no way of using it. By providing a rich client within the institutional boundary we maximise our ability to exploit resource. However, our system is passive and relies upon the users to contribute their resource and establish regular invocations of their client system. While this provides a scalable and robust solution it cannot provide any guarantees about completion time. A publish / subscribe approach would overcome this shortcoming but would require the development of an active message queue in the database system. We will investigate this possibility. The overheads in moving data through a client are avoided in the task farming systems but the dynamic allocation of compute tasks through multiple clients offsets this issue.

The definition and upload of a model study involves some simple edits to a template script to specify the unique experiment definition in the database. Once created the experiment coordinator circulates the unique identifier for the study and members of the project can then choose whether to contribute their resource. The effort involved in contributing to a study is minimal because the “worker” script functions autonomously. A user typically schedules a cron job or windows scheduled task to execute several times a day and then has no further interaction with the system. The scheduled task initiates the client and invokes the “worker” using the experiment identifier. As long as a valid proxy certificate exists, the system attempts to contribute work until the study is complete. A project member may stop their scheduled task at any time and withdraw their resource, introduce new resource, and drop in and out of the experiment as they see fit.

As a project we are able to exploit resources across institutional boundaries without any modification to firewalls. As long as a client can talk to the database and the GridFTP server on standard ports then we can exploit a user's local rights to resource at their institution. Our system pulls the work to a client within the institutional boundary rather than needing to push it through.

The database provides a very robust and fault tolerant environment to perform our work. The entire study is held in persistent storage and progress is only achieved through the successful update of the database contents. Problems encountered on the grid, external to the database, do not have any direct impact on a study. Failed work units are automatically re-submitted by later invocations of the "worker" script. The system also provides a means to maximize our responsible use of available resource. Since the database maintains a record of the number of tasks running on each server the client can respect specified usage limits. Once each resource reaches its limit the client will move on to the next and the system therefore keeps available compute power busy.

6 Conclusions

Members of the GENIE project collaborated to perform large-scale ensemble calculations on hierarchies of Earth System Model from the GENIE framework. Using a common Grid-enabled client interfacing to the shared GENIE database, users pooled their resource to contribute to ensemble studies defined in the database. Armed with the unique system identifier for a particular study a user simply invoked an autonomous "worker" script specifying the amount of work they would like to contribute on a given resource. Through regular invocations of the client by multiple members of the project the ensembles were progressed through to completion. The system provides a robust and fault tolerant means to carry out large ensembles of lengthy model runs as progress is only achieved by the successful upload of data to the database. Previous studies of THC stability in models have been restricted to inter-comparison of disparate models of generally lower complexity than GENIE-2 [12]. Here e-Science technology has allowed us to perform the first systematic investigation of THC stability within a single model hierarchy.

Acknowledgements

The GENIE and GENIE*fy* projects are funded by the Natural Environment Research Council (NER/T/S/2002/00217 & NE/C515904) through the e-Science programme. The authors would like to acknowledge the use of the UK National Grid Service in carrying out this work.

References

[1] The GENIE project. http://www.genie.ac.uk

[2] Marsh R., A. Yool, T. M. Lenton, M. Y. Gulamali, N. R. Edwards, J. G. Shepherd, M. Krznaric, S. Newhouse and S. J. Cox, 2004. Bistability of the thermohaline circulation identified through comprehensive 2-parameter sweeps of an efficient climate model. Climate Dynamics, **23**(7–8), 761-777.

[3] Wang, H. and G. E. Birchfield, 1992. An energy-salinity balance climate model: Water vapor transport as a cause of changes in the global thermohaline circulation, J. Geophys. Res., **97**, 2335-2346.

[4] Price, A. R., Xue, G., Yool, A., Lunt, D. J., Valdes, P. J., Lenton, T. M., Wason, J. L., Pound, G. E., Cox, S. J. and the GENIE team, 2006. Optimisation of integrated Earth System Model components using Grid-enabled data management and computation. Concurrency and Computation: Practice and Experience, DOI: 10.1002/cpe.1046.

[5] The Geodise project. http://www.geodise.org

[6] Keane, A. J., 2003. The OPTIONS design exploration system: reference manual and user guide. http://www.soton.ac.uk/~ajk/options.ps

[7] Abramson, D., Buyya, R. and Giddy, J., 2002. A computational economy for grid computing and its implementation in the Nimrod-G resource broker, Future Generation Computer Systems, **18**(8), 1061-1074.

[8] YarKhan, A., Seymour, K., Sagi, K., Shi, Z. and Dongarra, J., 2006. Recent Developments in Gridsolve, International Journal of High Performance Computing Applications, **20**(1), 131-141. DOI: 10.1177/1094342006061893

[9] Srinivas, N. and Deb, K., 1995. Multiobjective function optimization using nondominated sorting genetic algorithms, Evol. Comp. J., **2**(3), 221-248.

[10]Deb, K., Pratap, A., Agarwal, S. and Meyarivan, T., 2002. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation, **6**(2), 182 – 197.

[11]Lenton, T. M., *et al*., 2006. A modular, scalable, Grid ENabled Integrated Earth system modelling framework: Effect of dynamical atmosphere model, surface grid, and ocean resolution on the stability of the thermohaline circulation. Clim. Dyn. submitted.

[12]Rahmstorf, S., Crucifix, M., Ganopolski, A., Goosse, H., Kamenkovich, I., Knutti, R., Lohmann, G., Marsh, R., Mysak, L. A., Wang, Z., and A. Weaver, 2005. Thermohaline circulation hysteresis: a model intercomparison. *Geophys. Res. Lett.*, **32**, L23605 10.1029/2005GL023655.