Multi-objective Optimization of GENIE Earth System Models

By Andrew R. Price[1], Richard J. Myerscough[2], Ivan I. Voutchkov[1], Robert Marsh[2] and Simon J. Cox[1]

[1] School of Engineering Sciences, University of Southampton, Highfield, Southampton SO17 1BJ, UK
[2] School of Ocean and Earth Sciences, University of Southampton, National Oceanography Centre, European Way, Southampton SO14 3ZH, UK

The tuning of parameters in climate models is essential to provide reliable long-term forecasts of Earth system behaviour. We apply a multi-objective optimization algorithm to the problem of parameter estimation in climate models. This optimization process involves the iterative evaluation of Response Surface Models (RSMs) followed by the execution of multiple Earth system simulations. These computations require an infrastructure that provides high performance computing for building and searching the RSMs and high-throughput computing for the concurrent evaluation of a large number of models. Grid computing technology is therefore essential to make this algorithm practical for members of the GENIE project.

Keywords: Multi-Objective Optimization, High Performance Computing, High Throughput Computing, Grid Computing

1. Introduction

The GENIE project (Lenton *et al.* 2007) has developed a component framework for the construction, execution and management of Earth System Models capable of simulation over millennial timescales. Constituent models of the Earth system at varying resolutions, complexity and dimensionality can be coupled together to form a hierarchy of climate models. Grid computing technology is a key enabler within the project to aid in flexible coupling of component models, subsequent execution of the resulting ESMs and sharing and management of the data that they generate. In particular, a Grid-enabled problem solving environment is exploited for concurrent execution of multiple instances of GENIE models on computational grids.

At their heart, climate models solve a system of equations, derived from physical laws, that represent the major physical processes of the Earth system and their interactions. The numerical solution of these equations is only practical if approximations are made where the representation of the physical processes on small spatial and temporal scales are simplified or parametrized. Often, the parameters of a model do not relate to real-world physical quantities so a key stage in climate model development is the selection of values that result in the simulation of a reasonable climatology. However, the non-linear response of a model to its parameters and the often conflicting attainment goals make this a difficult problem to solve.

Parameter estimation has been the subject of much study within the GENIE project (see Price *et al.* 2006 for further details). Population based multi-objective

design search and optimization algorithms from the engineering domain have proved useful for parameter estimation in GENIE models. In particular, the Non-dominated Sorting Genetic Algorithm (NSGA-II) with surrogate modelling has been applied successfully for parameter estimation in a number of models from the GENIE framework (Price *et al.* 2006; Matsumoto et al. 2008). This optimization process involves the iterative evaluation of Response Surface Models (RSMs) followed by the execution of multiple Earth system simulations. These computations require an infrastructure that provides high performance computing for building and searching the RSMs and high-throughput computing for the concurrent evaluation of a large number of models. Grid computing technology is essential to make this algorithm practical for members of the GENIE project.

In this paper we present the application of the NSGA-II algorithm with surrogate modelling to the tuning of a new model from the GENIE framework. In section 2 the new GENIE model is presented and the parameter estimation problem formalized. The multi-objective optimization algorithm is presented in section 3 and the use of surrogate modelling is detailed. This method presents two forms of computational challenge requiring both high-performance and high-throughput computing infrastructure to provide a timely result. We present the software environment used to meet these challenges in section 4. The results of the optimization of this model are presented in section 5 and we draw conclusions in section 6.

2. Parameter Estimation Problem

The GENIE-1 Earth System Model consists of a 3D frictional-geostrophic ocean model (GOLDSTEIN) coupled to a 2D energy-moisture balance model (EMBM) and a thermodynamic sea-ice model. The model is computationally efficient, designed for studies which require long timescales of integration.

We use a new version of GENIE-1 which employs an ocean vertical mixing scheme of increased complexity relative to the default included in GOLDSTEIN. In brief, an ocean mixing scheme aims to represent the effect of unresolved turbulence on simulated water mass properties. Vertical (diapycnal) turbulent mixing is thought to play a key role in sustaining or "closing" the thermohaline circulation; studies suggest the diffusion of heat from the near-surface to deep ocean facilitates the upwelling of dense bottom waters against the large-scale stable ocean stratification (Rahmstorf 2003, Wunsch & Ferrari 2004). In terms of physical complexity, a wide spectrum of mixing schemes currently exists; ranging from a simple dependence on concentration gradients, through to explicit consideration of local energetics. A survey of existing literature reveals the choice of mixing scheme exerts an important control on the resulting large-scale ocean circulation, in particular the thermohaline circulation stability (Marzeion *et al.* 2007). Our implementation of an increased complexity mixing scheme thus forms the basis for a larger study currently in progress, in which we explore the implications of vertical mixing complexity for large-scale ocean circulation stability. Future publications will provide a full description and discussion of both the scientific development of the new mixing scheme and the results of our investigations.

In this paper, we focus on the multi-objective optimization of the final implementation of this scheme as incorporated into GENIE-1. Due to the fundamental modification to the ocean physics arising from use of the stratification-dependent

mixing scheme, we consider it necessary to re-evaluate the standard GOLDSTEIN parameter set. In addition to the principal transport parameters identified in Edwards & Marsh (2005), we include two new parameters related to the stratification dependent mixing scheme whose values are poorly constrained but which strongly influence the model climatology. The aim of the tuning exercise is to find the optimal values for these free parameters that produce an equilibrium model end-state with the closest fit to equivalent observational data. In order to reach equilibrium the model is integrated from a uniform initial state for at least 4,000 model years to reach a steady end-state. On a typical desktop PC (Pentium IV 2.6GHz) an integration over 4,000 model years typically requires approximately 7 hours of CPU time. The aim of the optimization process is to reduce the root mean square errors between model fields s_i and observational data S_i across four physical fields (i).

$$f_i(x) = \sqrt{\frac{(s_i(x) - S_i)^2}{\widehat{\sigma_i}^2}}, i = 1..4 \tag{2.1}$$

The four physical data sets to which the model is tuned consist of 3D ocean temperature (OCNTEMP) and salinity fields (OCNSAL) and 2D atmospheric surface air temperature (SURFTEMP) and humidity (SURFHUM) fields. These data sets are an average over approximately five decades (spanning the period 1948–2002) and obtained from the World Ocean Atlas (Levitus *et al.* 1998) and the National Centers for Environmental Prediction (NCEP Reanalysis data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.cdc.noaa.gov/). The observational data are interpolated onto the model grid and weighted by the variance of the data for each field $\widehat{\sigma_i}^2$.

3. Multi-Objective Optimization

Many optimization problems consist of competing design/attainment goals and some form of compromise must be enforced to obtain a result. In "traditional" single objective optimization such compromise is typically achieved by optimizing a weighted sum of individual design goals. The results of employing such a summation can be highly dependent upon the weightings used. Multi-objective methods avoid this problem by investigating the problem in the objective space and seeking a set of non-dominated solutions, usually referred to as the Pareto Front (PF). A Pareto set of design solutions contains members that are all in some sense optimal; in order to improve a single member in one objective it must get worse in one or more of its other objectives. A solution is said to be "non-dominated" if there is no other set member that exceeds its performance in *all* attainment goals.

(a) Non-dominated Sorting Genetic Algorithm

Population based optimization algorithms map well to multi-objective problems because they maintain a family of solutions that evolve over successive generations. A suitable algorithm can seek to generate a population of Pareto optimal solutions in the objective space of the problem. There are many multi-objective algorithms based on evolutionary programming or genetic algorithms (see Zitzler *et al.* 2000) that will perform such multi-objective optimizations. Arguably, the most popular
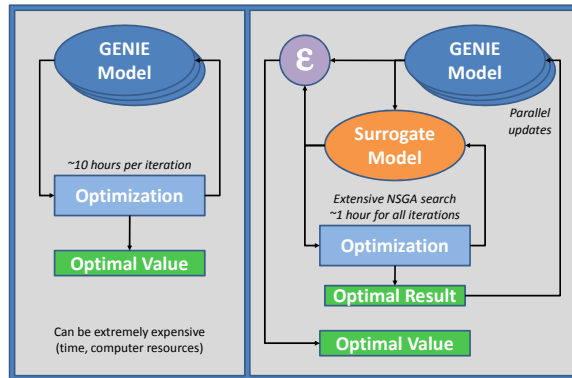
Figure 1. Multi-objective optimization improved using surrogate modelling.

methods from the literature are NSGA-II (Deb *et al.* 2002) and SPEA2 (Zitzler *et al.* 2000). In this work we use a modified NSGA-II method (Voutchkov & Keane, 2006) which has been improved for gains in performance over the standard algorithm by utilising the GA engine in the OptionsMatlab software package. This will be referred to as OptionsNSGA2. The algorithm maintains a population of solutions which "evolve" over generations of the algorithm to generate a number of Pareto optimal model instances. At the end of the process the modeller selects a subset of the members of the Pareto front for further study. While this method yields high quality solutions to the estimation problem, it places considerable demands on computational resource. At each generation multiple evaluations of an Earth system model must be performed. For such methods to be practical these evaluations must be executed concurrently. The Grid computing software exploited within GENIE enables us to exploit appropriate Grid computing resource to perform function evaluations in a timely and reliable fashion.

(*b*) *Non-dominated Sorting Genetic Algorithm with Surrogate Modelling*

Even with concurrent execution of the GENIE models, the large number of generations $O(10^2)$ required to reach an optimal solution can make the method prohibitively expensive. For example, a standard application of the OptionsNSGA2 algorithm to a GENIE model would typically require several weeks of continuous compute time to achieve a high quality result (see Figure 1a). In order to perform the parameter estimation exercise in a timely fashion we therefore employ the OptionsNSGA2 method with surrogate modelling (see Figure 1b). The use of surrogate models with the OptionsNSGA2 algorithm can reduce, by an order of magnitude, the total number of simulation years required for a high quality result in the calibration of a GENIE model. In the surrogate modelling approach one seeks to build a representation of the individual objective functions of the problem using Response Surface Modelling techniques. This provides surrogate models of the underlying problem which can be extensively searched at significantly less cost than the true expensive functions (Earth system simulations). The main challenge with this method is to refine the surrogate models so that they provide a sufficiently accurate representation of the objective functions, at least in the optimal regions of the parameter space. This is typically achieved through an iterative process in

which the surrogate models are updated using the true response of the underlying function. By adding such evaluations to the data cache of existing points the quality of subsequent RSMs is improved until good agreement between the GENIE model and the RSMs is achieved. The resulting members of the Pareto Front are taken from the evaluations of the GENIE model and not from the RSMs thus ensuring the results relate to the true response of the model.

(c) Kriging

There are many Response Surface Models that could be used to provide a cheap surrogate model for the optimization process. Voutchkov & Keane (2006) discuss some of the merits of different RSM methods but we limit our discussion here to the Kriging method (Jones *et al.* 1998) that we have chosen to employ. Kriging is a curve-fitting technique that originated in the field of geological surveying. This method has been found to work very well for a wide range of multi-objective problems. However, there is a computational cost to building the Krig models of the underlying functions. The curvature of each Krig is controlled by a set of hyper-parameters that must themselves be tuned (optimized) to provide the best fit of the surface to the sampled data. This is achieved by maximising a concentrated likelihood function (CLF) over a set of sampled data points. The evaluations of the CLF involve the inversion of a matrix of correlation measures (an $O(N^3)$ operation) and consequently the tuning of the Krig can incur significant computational expense. For the typical scale of GENIE parameter estimation problem (in this case 12 model parameters) the Krigs are built using in excess of 120 sampled data points and the hyper-parameter tuning process is expensive. Since a separate Krig is built for each objective function there is scope for concurrency and these calculations are therefore performed in parallel on the best resource available.

Figure 2 presents the workflow of the optimization process and highlights the type of resource typically targeted for each stage of the calculation. For the benefit of the reader we briefly summarise the algorithm, which is embedded in the OptionsNSGA2_RSM procedure of the OptionsNSGA2 software package:

1. An initial sampling of the parameter space is performed using formal space filling Design of Experiments methods. A LP_τ (Statnikov & Matuzov 1995) sequence is typically employed and the expensive simulations are performed using Condor high-throughput computing.

2. The hyper parameters of the Krigs are then tuned concurrently to provide the best representation of the response of the objective measures of model fitness to data across the parameter space. These calculations are targeted at high-performance computing clusters.

3. The OptionsNSGA2 algorithm is applied over the cheap surrogate models to perform extensive multi-objective searches over the design space.

4. The results of these searches are used to select update points to be evaluated on the underlying simulation code. These points are selected so that the Krigs in the following iteration will be built from higher quality data and consequently yield an improvement in the Pareto optimal points. The updates are selected from five sources:
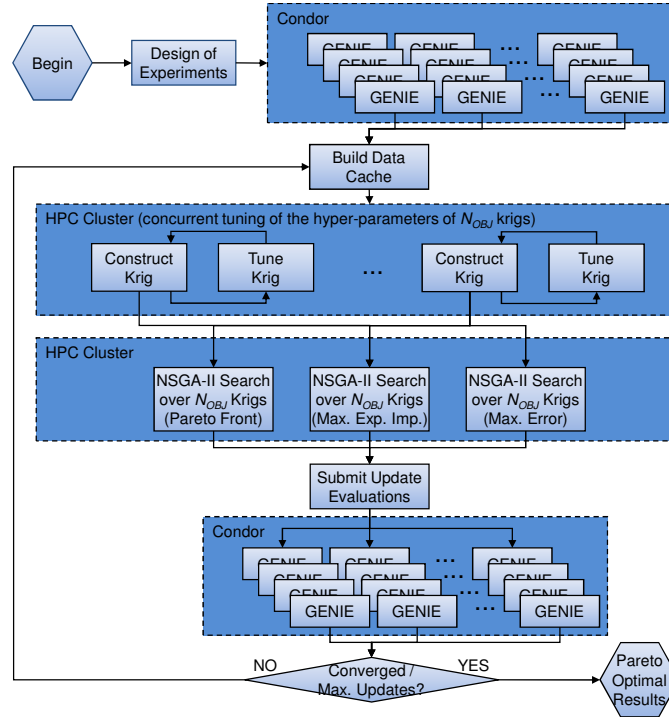
Figure 2. Workflow schematic for the OptionsNSGA2 with surrogate modelling process.

(a) Pareto optimal points in the Kriging representation of the objective space

(b) Points of maximum RMS error in the krigs

(c) Points of maximum Expected Improvement in the krigs

(d) Random points to help escape from local minima

(e) Points from a secondary OptionsNSGA2 process performed directly on the underlying expensive function

5. The update points are evaluated concurrently by running the expensive simulations, again using a Condor pool.

6. The algorithm iterates by returning to point 2 until convergence.

4. Grid and Distributed Computing Infrastructure

The process is made practical using Condor high throughput computing (Thain *et al.* 2005) which enables the $O(10^2)$ update evaluations of the expensive simulations to be executed concurrently across workstations that would otherwise sit idle. This work was carried out on the University of Southampton condor pool which is composed of $\sim 1,100$ laboratory work stations and office PCs running various versions of Windows operating system. The algorithm operates by improving the training data for the Krig hyper-parameters and increasing the accuracy of the stochastic process models over which the searches are performed. The tuning of the hyper-parameters of each Krig is performed concurrently on a high-performance

computing cluster. At the end of the process the pareto-optimal members of the global data cache are returned for further analysis.

(a) *Geodise Compute Toolbox*

GENIE has adopted software from the Geodise project (Eres *et al.* 2005) to build a distributed collaborative problem solving environment for the study of Earth system models. The Geodise toolboxes integrate compute and data Grid functionality into the Matlab and Jython environments familiar to scientists and engineers.

The OptionsNSGA2 algorithm has been programmed in the Matlab Problem Solving Environment because this is a productive and familiar environment for engineers. The OptionsNSGA2 software package enables users to provide their optimization problem as a simple tunable function. For GENIE models we expose a GENIE model invocation on the Condor pool as a tunable function pair; a submission function that accepts as input a parameter set chosen by the optimizer and returns a job handle, and a retrieval function that processes the model output upon completion of the simulation and returns the objective function values to the optimizer. The Krig tuning process is similarly managed by enabling the user to write a function pair that submit the tuning binary to a resource, monitor the job handle and retrieve the optimal hyper-parameters upon completion. For this work we have exploited a licensed code for the evaluation of the hyper-parameter tuning which is deployed on the Spitfire Microsoft Compute Cluster Server at the University of Southampton. The Globus interface of the UK National Grid Service to this system enables GENIE users from other institutions to exploit this service for the OptionsNSGA2 optimization process.

(b) *Condor DAGMan*

The GENIE model used for this study needs approximately 5 to 7 hours of continuous CPU time on the range of resource available in the Condor pool targeted for this work. Furthermore, this particular Condor pool is comprised exclusively of machines running the Windows operating system and consequently Condor's native check-pointing mechanisms (which are only available for Linux platforms) could not be used. In practice the execution of a large number of compute jobs each requiring this duration of runtime presents a challenge to such a pool. The nature of the resource is that jobs of this length are at higher risk of interruption due to user activity on campus. It is found that throughput degrades considerably as the time requirements of compute jobs increases. With GENIE models, we are however able to use check-point and restart capabilities of the models to overcome these problems.

Condor DAGMan is a meta-scheduler for Condor that will manage the submission of a Directed Acyclic Graph (DAG) of compute tasks. A DAG specifies a workflow that is built from activities that depend upon the prior execution of other activities. For Condor DAGMan a simple text file is written to define the required jobs and express the explicit dependencies between those jobs. In the case of GENIE the task graph simply expresses a linear sequence of jobs in which the output (simulation checkpoint files) from one task is passed as input to the next task. The Condor DAGMan tool then manages the submission of those tasks to the Condor pool, providing some fault tolerance through job recovery mechanisms.
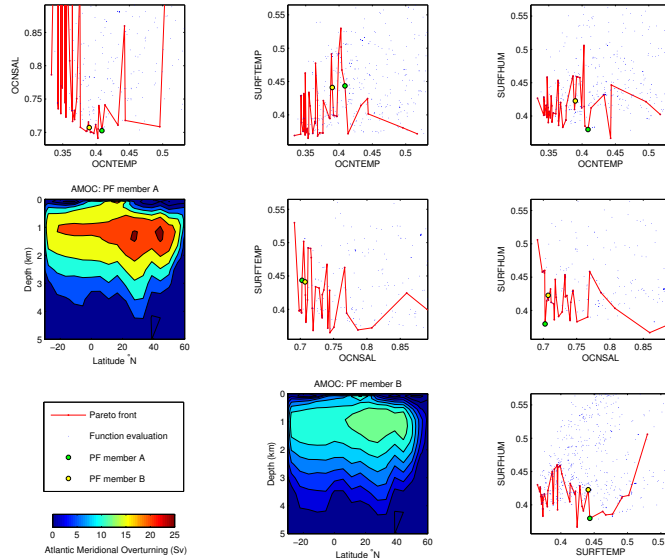
Figure 3. Objective function evaluations projected onto the 2D objective space for each pair of objectives. The final members of the Pareto Front are labelled in red. The Atlantic Meridional Circulation (AMOC) is plotted for two selected members of the PF which exhibit different strengths of overturning but are relatively close in objective space.

The simulations are divided into ten tasks and with the individual tasks then requiring less than one hour of CPU time each, the execution of a large number of long simulations is reliably completed.

5. Results

The OptionsNSGA2 method with surrogate modelling has been applied to the parameter estimation problem for the new ocean mixing scheme. A total of 12 model parameters were tuned using a fixed budget of 8 update iterations of the algorithm. The final Pareto Front consists of forty-eight non-dominated solutions. This highlights a key advantage of the multi-objective method; the domain expert is able to select from a range of solutions, depending on the requirements of the study. To illustrate this point from a climate modelling perspective, we show (figure 3) streamfunctions in the meridional-vertical plane, corresponding to the Atlantic Meridional Overturning Circulation (AMOC), an important proxy for the state of the ocean circulation. The streamfunctions are obtained as top-to-bottom cumulative integrals of the zonally-averaged meridional volume flux at each depth level in the ocean. The two panels in figure 3 are for two different members of the PF; member A and member B. The maximum of the streamfunction, a measure of AMOC strength, differs significantly between the two members: 26 Sv (A) vs. 14 Sv (B). For a study requiring a single solution for the simulation of the AMOC in accordance with current scientific consensus (Lumpkin *et al.* 2008), the rejection of member A for exhibiting an excessively vigorous circulation might be appropriate. Alternatively, a statistically orientated study wishing to examine multiple optimal initial AMOC states might include members A and B, and others from the PF.
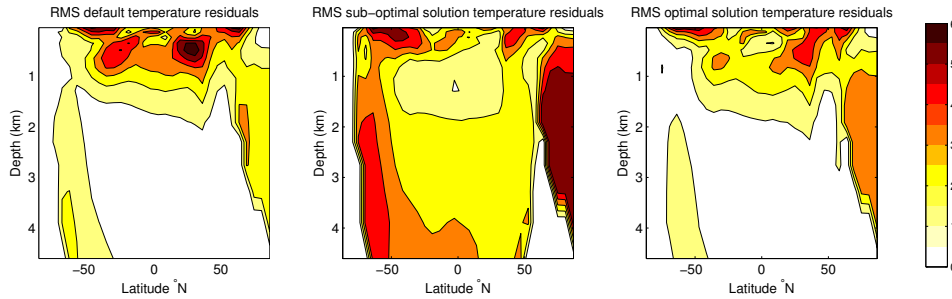
Figure 4. Differences between the Ocean temperature field and the equivalent observational data generated by a) the default parameter set, b) the parameter set of the PF member with the worst fit to the temperature data and c) the parameter set of the preferred optimal member of the PF.

Selection of a parameter set from the PF is also possible by considering a weighted measure of the four single objectives. To demonstrate the improvement in objectives compared to the default parameter set, we weight each of the single objectives by the number of residuals which constitute that objective, and choose the PF member which provides the lowest combined weighting. The panels in figure 4 present a convincing reduction in the ocean temperature residuals when compared to the default point in parameter space. For comparison, the member of the PF with the worst fit to the temperature data is also shown. These results are a small part of an extensive study of our stratification dependent mixing scheme; a complete scientific analysis will be the subject of future publications.

6. Conclusions

Parameter estimation is an important step in climate model development, especially in highly parametrised models like GENIE. The multi-objective OptionsNSGA2 method is found to be successful for this class of problem but is computationally demanding when the underlying objective function evaluations (GENIE simulations) are expensive. For timely results to be generated the OptionsNSGA2_RSM method is employed. Appropriate computing platforms are required to execute the iterative process of building surrogate models and updating those models using the true model response. This algorithm benefits from both high-performance computing to build the surrogate models and high-throughput computing to evaluate the true behaviour of the model. Within GENIE we target appropriate resource using the Geodise interfaces to Globus 2.4 and Condor. The RSMs are built using the OPTIONS optimisation package on nodes of the UK National Grid Service to which (freely available academic) licence codes have been assigned. Condor DAGMan is used to progress the large number of simulations to completion using check-points and restarts. The OptionsNSGA2_RSM software only requires the GENIE scientist to wrap their model as a pair of Matlab functions accepting as input the tunable parameters and returning, after remote execution of the simulation on a specified resource, the multiple measures of fitness to data. The time taken from a "decision to study" to an end result is in practice no more than a week. In this regard the method is very competitive with alternative approaches that often require expert

knowledge to execute. We have presented the results of this method applied to a new model in which a more complex ocean mixing scheme is introduced.

Acknowledgements

References

Deb, K., Pratap, A., Agarwal, S. & Meyarivan, T. 2002 A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE T. Evolut. Comput.* **6**, 182–197.

Edwards, N. R. & Marsh, R. 2005 Uncertainties due to transport-parameter sensitivity in an efficient 3-D ocean-climate model. *Clim. Dynam.* **24**, 415–433.

Eres, M. H., Pound, G. E., Jiao, Z., Wason, J. L., Xu, F., Keane, A. J. & Cox, S. J. 2005 Implementation and utilisation of a Grid-enabled problem solving environment in Matlab. *Future Gener. Comp. Sy.* **21**, 920–929. (doi:10.1016/j.future.2003.12.016)

Jones, D. R., Schonlau, M. & Welch, W. J. 1998 Efficient global optimization of expensive black-box functions. *J. Global Optim.* **13**, 455–492. (doi:10.1023/A:1008306431147)

Lenton, T. M., *et al.* 2007 Effects of atmospheric dynamics and ocean resolution on bistability of the thermohaline circulation examined using the Grid ENabled Integrated Earth system modelling (GENIE) framework. *Clim. Dynam.* **29**, 591–613.

Levitus, S., Boyer, T. P., Conkright, M. E., O'Brien, T., Antonov, J., Stephens, C., Stathoplos, L., Johnson, D. & Gelfeld, R. 1998 *NOAA Atlas NESDIS 18*. Washington, D.C.: U.S. Government Printing Office.

Lumpkin, R., Speer, K. G. & Koltermann, K. P. 2008 Transport across $48°N$ in the Atlantic Ocean. *J Phys. Oceanogr.*, **38**, 733–752. (doi:10.1175/2007JPO3636.1)

Matsumoto, K., Tokos, K. S., Price, A. R. & Cox, S. J. 2008 First description of the Minnesota Earth system model for ocean biogeochemistry. *Geosci. Model Dev.* **1**, 1–15.

Marzeion, B., Levermann, A. & Mignon, J. 2007 The role of stratification-dependent mixing for the stability of the Atlantic overturning in a global climate model, *J. Phys. Oceanogr.*, **37**, 2672–2681. (doi:10.1175/2007JPO3641.1)

Price, A. R., Voutchkov, I. I., Pound, G. E., Edwards, N. R., Lenton, T. M. & Cox, S. J. 2006 Multiobjective tuning of Grid-enabled Earth system models using a non-dominated sorting genetic algorithm (NSGA-II). In *Proc. 2nd IEEE Int. Conf. e-Science and Grid Computing*, 117. (doi:10.1109/E-SCIENCE.2006.261050)

Rahmstorf, S. 2003 Thermohaline circulation: The current climate, *Nature*, **421**, 699.

Thain, D., Tannenbaum, T. & Livny, M. 2005 Distributed computing in practice: the Condor experience. *Concurr. Comp. - Pract. E.* **17**, 323–356. (doi:10.1002/cpe.938)

Voutchkov, I. I. & Keane, A. J. 2006 Multiobjective optimization using surrogates. *Proc. 7th Int. Conf. Adaptive Computing in Design and Manufacture*, 167–175.

Wunsch, C. & Ferrari, R. 2004 Vertical mixing, energy, and the general circulation of the oceans, *Annu. Rev. Fluid Mech.*, **36**, 281–314. (doi:10.1146/annurev.fluid.36.050802.122121)

Zitzler, E., Deb, K. & Thiele, L. 2000 Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolut. Comput.*, **8**, 173–195. (doi:10.1162/106365600568202)