



An optimally tuned ensemble of the “eb_go_gs” configuration of GENIE: parameter sensitivity and bifurcations in the Atlantic overturning circulation

R. Marsh¹, A. Sóbester², E. E. Hart², K. I. C. Oliver¹, N. R. Edwards³, and S. J. Cox²

¹National Oceanography Centre Southampton, University of Southampton, Waterfront Campus, European Way, Southampton SO14 3ZH, UK

²Faculty of Engineering and the Environment, University of Southampton, Southampton SO17 1BJ, UK

³Department of Earth and Environmental Sciences, CEPSAR, The Open University, Milton Keynes MK7 6AA, UK

Correspondence to: R. Marsh (rm12@soton.ac.uk)

Received: 24 August 2012 – Published in Geosci. Model Dev. Discuss.: 11 February 2013

Revised: 19 July 2013 – Accepted: 11 September 2013 – Published: 21 October 2013

Abstract. The key physical parameters for the “eb_go_gs” configuration of version 2.7.4 of GENIE, an Earth system model of intermediate complexity (EMIC), are tuned using a multi-objective genetic algorithm. An ensemble of 90 parameter sets is tuned using two ocean and two atmospheric state variables as targets. These are “Pareto-optimal”, representing a range of trade-offs between the four tuning targets. For the leading five parameter sets, simulations are evaluated alongside a simulation with untuned “default” parameters, comparing selected variables and diagnostics that describe the state of the atmosphere, ocean and sea ice. Further experiments are undertaken with these selected parameter sets to compare equilibrium climate sensitivities and transient climate responses. The pattern of warming under doubled CO₂ is strongly shaped by changes in the Atlantic meridional overturning circulation (AMOC), while the pattern and rate of warming under rising CO₂ is closely linked to changing sea ice extent. One of the five tuned parameter sets is identified as marginally optimal, and the objective function (error) landscape is further analysed in the vicinity of the tuned values of this parameter set. “Cliffs” along some dimensions motivate closer inspection of corresponding variations in the AMOC. This reveals that bifurcations in the AMOC are highly sensitive to parameters that are not typically associated with MOC stability. Specifically, the state of the AMOC is sensitive to parameters governing the wind-driven circulation and atmospheric heat transport. For the GENIE configuration presented here, the marginally optimal parameter set

is recommended for single simulations, although the leading five parameter sets may be used in ensemble mode to admit a constrained degree of parametric uncertainty in climate prediction.

1 Model calibration and parameter space analysis

Earth system models of full complexity are computationally expensive, due to the resolution of physical and biogeochemical processes on short timescales and space scales. Relatively small (O(10)-member) ensembles of relatively short (centennial) simulations are commonplace. Simulations and predictions with such models are nonetheless sensitive to the parameters for key unresolved processes, such as turbulent mixing of ocean tracers and cloud physics. The uncertainty due to such parameter sensitivity may be considerable, and is the subject of much ongoing research (Slingo et al., 2009). A range of non-linear behaviour in the Earth system (Lenton et al., 2008) may be likewise sensitive to key parameters, but even less is known of such sensitivities in complex models. The comparative computational affordability of Earth system models of intermediate complexity (EMICs) facilitates quantifying their uncertainties due to mixing and transport parameter choices in particular. Additionally, through carefully designed experiments and optimization studies across the space of these parameters, the locations of sub-domains of parameter space within which model behaviour may be regarded

as plausible (i.e. neither unphysical nor in unacceptable disagreement with observations) may be identified, yielding insights into these sub-domains and into the optimal composition of ensembles that best enable the meeting of observational targets. Consequently, the dependence of emergent non-linear behaviour on key model parameters may become apparent.

One particular family of EMICs is built using the Grid ENabled Integrated Earth system modelling (GENIE) framework. At the core of many GENIE models is the most basic climate model in which atmosphere, ocean and sea ice all play an active role, configured on a 36×36 equal-area-partitioning of the Earth surface with 16 depth levels in the ocean. This climate core has been used extensively, in studies of past, present and future Earth system dynamics (Cao et al., 2009). In the present study, we report the results from an objective tuning of the most recently documented version of the basic climate model (Marsh et al., 2011). As such, this paper is the second in a series that document the development, evaluation and benchmarking of GENIE.

Edwards and Marsh (2005) reported on an early parameter sensitivity study of C-GOLDSTEIN, the predecessor to GENIE, which has almost identical ocean and climate dynamics in a simpler (although less flexible) computational implementation. They used a semi-random ensemble of 1000 simulations, with which they addressed both the inverse problem of parameter estimation and the direct problem of quantifying the uncertainty due to mixing and transport parameters. Subsequent analysis of the model by Edwards et al. (2011) included a statistical process they described as pre-calibration, referring to the infeasibility of a full Bayesian model calibration. The goal of this effort was the identification of uncontroversially implausible values of certain inputs and outputs. They encountered a region of collapsed Atlantic meridional overturning circulation, presenting itself as a cliff-edge catastrophe in the freshwater forcing dimension of the input space. They also concluded that the exact location of this implausible subspace is a function of several other model parameters.

The identification of plausible sets was also the goal of the study by Holden et al. (2010), with deterministic emulators for five different aspects of the climate state as the key tool. The plausible ensemble was achieved by building, and then using a statistical filtering process known as approximate Bayesian computation. Emulators appeared in a supporting role in the calibration study by Price et al. (2009), where four observational target sets were selected and a multi-objective, emulator-assisted optimizer was used to identify effective ensembles.

Here we use an updated version of GENIE (Marsh et al., 2011) to probe into the structure of the landscapes of four similar targets, related to ocean and atmosphere prognostic variables, in the space of 13 of the GENIE parameters. A goal is to identify a diverse range of parameter sets, each of which provides a different Pareto-optimal fit to observations

representing a different balance of processes, thus providing an effective database for testing robustness when inter-model comparisons are not readily available. We discuss a simple evolutionary heuristic for the fast identification of ensembles that simultaneously optimize these targets, also examining the plausibility of these Pareto-optimal parameter sets through a set of further diagnostics for sea ice and the ocean circulation, and additional experiments to compare equilibrium and transient climate sensitivities. One of the five lowest-error parameter sets is used to investigate Atlantic meridional overturning circulation (AMOC) sensitivity to selected parameters further, with a focus on narrow regions of parameter space that host AMOC bifurcations. Following a brief description of the the model, we outline objective targets and the tuning method. The results of model tuning and analysis of parameter sensitivity are presented sequentially in three parts. We conclude with a discussion of novelties in the method and thereafter present results.

2 The model

The EMIC at the centre of this study is *eb_go_gs* configuration of *GENIE-1*, comprising an energy and moisture balance model (EMBM) for the atmosphere, coupled with the Global Ocean Linear Drag Salt and Temperature Equation Integrator (GOLDSTEIN) 3-D ocean model and a dynamic and thermodynamic sea ice model. This set-up largely follows that of Edwards and Marsh (2005), but uses an updated model version as integrated into the GENIE framework, described by Marsh et al. (2011). This version, based on GENIE version 2.7.4, includes revised wind forcing. Here, we tune the set-up referred to as “3636s16l” by Marsh et al. (2011), a standard model resolution with horizontal resolution of equal-area grid cells of 10 degrees longitudinal extent and 16 depth levels in the ocean, which has been used for a wide range of studies. In the following sub-sections, the three components of *eb_go_gs* are outlined briefly.

2.1 Atmosphere

The EMBM represents the atmosphere as a single 2-D layer with an advective–diffusive transport scheme for heat and moisture. The prognostic variables are air temperature and specific humidity, representative of the total atmospheric air column. Planetary albedo and the annual-average wind fields for advective transports are prescribed, while transport and ancillary parameters are typically calibrated using data assimilation techniques. Physical processes represented by the model are greatly simplified, examples being the instantaneous return of continental precipitation to coastal ocean points via a runoff map and the parameterization of outgoing long-wave radiation by an empirical polynomial function. An implicit numerical scheme is used to allow long EMBM time steps.

2.2 Ocean

GOLDSTEIN comprises a reduced physics (frictional–geostrophic) 3-D ocean model (Edwards et al., 1998) featuring spatially variable drag, realistic global bathymetry, multiple islands, and wind-stress forcing from a prescribed 2-D annual-mean wind-stress forcing field (Edwards and Marsh, 2005; Marsh et al., 2011). The prognostic variables are temperature and salinity. The tracer transport scheme employs an isoneutral and eddy-induced mixing scheme and an efficient convection scheme. Unlike primitive-equation ocean models, momentum advection and acceleration terms are neglected in the equation of motion, allowing the use of time steps which are long relative to those generally used in 3-D ocean models.

2.3 Sea ice

The third component of *eb_go_gs* is a dynamic and thermodynamic sea-ice model (Edwards and Marsh, 2005; Marsh et al., 2011) (herein referred to as GS). Sea ice is transported with the surface ocean current and is subject to a diffusive process with a strength controlled by a tunable parameter. An implicit numerical scheme for sea-ice transport is available (Marsh et al., 2011) and is used in this study.

3 Objective targets and tuning method

Table 1 lists the 13 dimensions of the EMBM–GOLDSTEIN–GS parameter space with respect to which we are investigating the model sensitivities. Twelve of these parameters comprise the set used by Edwards and Marsh (2005). We have added a further atmospheric parameter, r_κ , controlling the scaling (reduction) of meridional heat diffusivity over Antarctica and the Southern Ocean south of 56° S, introduced to parameterize the partial isolation of the atmosphere in this region, reducing atmospheric temperatures in the Southern Hemisphere high latitudes (see Appendix A in the publication by Cao et al., 2009, and also Marsh et al., 2011). Further tunable parameters of *eb_go_gs* have been fixed at their default values for the present study.

Table 2 lists the four observational fields that we use to define the targets of the ensemble selection study. These data sets, denoted by $S_{T_{\text{ocn}}}$, $S_{S_{\text{ocn}}}$, $S_{T_{\text{atm}}}$ and $S_{Q_{\text{dry}}}$ in what follows, are the same as those used for model–data comparison in Marsh et al. (2011), and are similar to those used in earlier studies. A notable exception to some earlier studies (e.g. Edwards and Marsh, 2005; Price et al., 2009) includes the replacement of specific humidity with a climatology of relative humidity (for reasons discussed in Lenton et al., 2006, and Marsh et al., 2011). The observational data are aligned with the model grid points through linear interpolation. In the case of the 3-D temperature and salinity fields, some of the values for some grid points of the model ocean are filled with the value of the closest available points from the observational fields.

Given uncertainties in ocean observations at high latitudes, where property distributions play a critical role in supporting the overturning circulation, erroneous observations in these regions may represent an inappropriate tuning target. Consequently, a more (less) realistic overturning circulation may follow less (more) agreement with observed property distributions at high latitudes. While there is thus an argument for removing high-latitude property observations from the tuning metric, it is uncertain whether revising the metric in this way would yield optimal circulation states that were more or less consistent with observations, so we choose to use global observations in the two oceanic objective functions defined below.

Numerical inter-annual variability in GENIE is generally negligible. We therefore compute root-mean-square (RMS) errors using these observed fields along with corresponding output fields from *eb_go_gs* for the last year of a 5000 yr spin-up model integration ($S_{T_{\text{ocn}}}$, $S_{S_{\text{ocn}}}$, $S_{T_{\text{atm}}}$ and $S_{Q_{\text{dry}}}$) to obtain four *objective functions*:

$$f_{T_{\text{ocn}}}(\mathbf{x}) = \sqrt{\frac{1}{N_{T_{\text{ocn}}}} \sum_{i=1}^{N_{T_{\text{ocn}}}} \frac{[S_{T_{\text{ocn}},i}(\mathbf{x}) - S_{T_{\text{ocn}},i}]^2}{\widehat{\sigma}_{T_{\text{ocn}}}^2}} \quad (1)$$

$$f_{S_{\text{ocn}}}(\mathbf{x}) = \sqrt{\frac{1}{N_{S_{\text{ocn}}}} \sum_{i=1}^{N_{S_{\text{ocn}}}} \frac{[S_{S_{\text{ocn}},i}(\mathbf{x}) - S_{S_{\text{ocn}},i}]^2}{\widehat{\sigma}_{S_{\text{ocn}}}^2}} \quad (2)$$

$$f_{T_{\text{atm}}}(\mathbf{x}) = \sqrt{\frac{1}{N_{T_{\text{atm}}}} \sum_{i=1}^{N_{T_{\text{atm}}}} \frac{[S_{T_{\text{atm}},i}(\mathbf{x}) - S_{T_{\text{atm}},i}]^2}{\widehat{\sigma}_{T_{\text{atm}}}^2}} \quad (3)$$

$$f_{Q_{\text{dry}}}(\mathbf{x}) = \sqrt{\frac{1}{N_{Q_{\text{dry}}}} \sum_{i=1}^{N_{Q_{\text{dry}}}} \frac{[S_{Q_{\text{dry}},i}(\mathbf{x}) - S_{Q_{\text{dry}},i}]^2}{\widehat{\sigma}_{Q_{\text{dry}}}^2}}, \quad (4)$$

where $\mathbf{x} = \{W, \kappa_h, \kappa_v, \lambda, \kappa_t, \kappa_q, \beta_T, \beta_q, F_a, l_d, l_s, \kappa_{hi}, r_\kappa\}$, and the 13 dimensions of this vector correspond to the parameters defined in Table 1. $N_{T_{\text{ocn}}}$ and $N_{S_{\text{ocn}}}$ are the number of ocean grid cells (in all three dimensions), while $N_{T_{\text{atm}}}$ and $N_{Q_{\text{dry}}}$ are the number of atmospheric grid cells (in latitude and longitude).

The variances $\widehat{\sigma}_{T_{\text{ocn}}}^2$ in the four expressions above are designed to weight the root-mean-square error in a way that makes the values of the four functions approximately comparable. This is the same formulation as used in Price et al. (2009), so the results reported there are directly comparable with ours, except for $f_{Q_{\text{dry}}}$, as indicated earlier. With these functions we formulate a multi-objective search, the goal of which is to build ensembles comprising *non-dominated* parameter sets. Such vectors, also known as *Pareto-optimal* points of the parameter space, have the property that each outperforms all other points in the set along one of the four dimensions of the output space.

This methodology builds on earlier tuning work by Edwards and Marsh (2005), who sampled the entire parameter space of the (pre-GENIE) C-GOLDSTEIN model through a

Table 1. The GENIE parameter space sampled in the first generation ensemble.

Var. no.	Parameter	Notation	Minimum	Maximum	Units
Ocean					
1.	Wind-scale coefficient	W	1	3	–
2.	Isopycnal diffusivity	κ_h	1×10^2	1×10^4	$\text{m}^2 \text{s}^{-1}$
3.	Diapycnal diffusivity	κ_v	2×10^{-6}	2×10^{-4}	$\text{m}^2 \text{s}^{-1}$
4.	Inverse friction coefficient	λ^{-1}	0.5	5	days^{-1}
Atmosphere					
5.	Heat diffusivity	κ_t	1×10^6	1×10^7	$\text{m}^2 \text{s}^{-1}$
6.	Moisture diffusivity	κ_q	5×10^4	5×10^6	$\text{m}^2 \text{s}^{-1}$
7.	Heat advection coefficient	β_T	0	1	–
8.	Moisture advection coefficient	β_q	0	1	–
9.	Freshwater flux factor	F_a	0	1	–
10.	Heat diffusivity width	l_d	0.5	2	radians
11.	Heat diffusivity slope	l_s	0	0.25	–
13.	Meridional heat diffusivity scaling south of 56°S	r_κ	0	1	–
Sea ice					
12.	Sea ice diffusivity	κ_{hi}	1×10^2	1×10^5	$\text{m}^2 \text{s}^{-1}$

Table 2. Observational target data for the calculation of the objectives. Oceanic properties are from the World Ocean Atlas 2005 (WOA05; http://www.nodc.noaa.gov/OC5/WOA05/pr_woa05.html; Antonov et al., 2006; Locarnini et al., 2006) from the National Oceanographic Data Center (NODC; <http://www.nodc.noaa.gov/>). NCEP_Renalysis 2 data (Kanamitsu et al., 2002) provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at <http://www.esrl.noaa.gov/psd/> were used to compute the climatologies of the atmospheric properties.

	Model field	Notation	Units	Climatology	Observational reanalysis data product
Ocean	Temperature	$S_{T_{\text{ocn}}}$	$^\circ\text{C}$	Annual average	World Ocean Atlas 2005 (Antonov et al., 2006)
	Salinity	$S_{S_{\text{ocn}}}$	PSS	Annual average	World Ocean Atlas 2005 (Locarnini et al., 2006)
Atmosphere	Temperature	$S_{T_{\text{atm}}}$	K	Long-term annual average (1979–2010) at 1000 mb level	NCEP_Renalysis 2 (Kanamitsu et al., 2002)
	Relative humidity	$S_{Q_{\text{dry}}}$	%	Long-term annual average (1979–2010) at 1000 mb level	NCEP_Renalysis 2 (Kanamitsu et al., 2002)

semi-random, space-filling sampling plan of 1000 individual simulations, and used the responses of the model as the basis for parameter sensitivity analysis. The drawback of this approach is poor scalability with increased problem dimensionality: the number of runs required for a factorial design increases exponentially with the number of parameters to be tuned, although even simple Latin hypercube designs (such as used by Edwards and Marsh, 2005) are much more efficient.

Multi-objective evolutionary search methods, such as the one adopted here (similar to that used by Price et al., 2009), are more robust to this *curse of dimensionality*, and

they are also more readily scalable in terms of the number of targets/constraints. The multi-objective genetic algorithm achieves this by progressively “learning” the plausible regions of the search space through a sequence of generations, during which the selective pressure of the optimizer biases each population towards non-dominated individuals (see Deb et al., 2002). This will, ultimately, result in a fuller understanding of the tuning landscape than a process that either samples uniformly (Edwards and Marsh, 2005) or considers each target (objective function) in isolation, the latter being prone to the risk of biasing the tuning process towards solutions that excel on individual targets, if model structural error

Table 3. Parameter values for GMD11 and the five selected points. The parameter set for the recommended point 18 are emphasized in bold font. See Table 1 for parameter definitions.

Variable	GMD11	Point no.					Units
		18	37	49	56	74	
W	1.531	2.367	2.363	2.367	2.500	2.363	–
κ_h	1.49×10^3	5.54×10^3	5.45×10^3	5.54×10^3	5.45×10^3	5.54×10^3	$\text{m}^2 \text{s}^{-1}$
κ_v	4.45×10^{-5}	4.45×10^{-5}	2.54×10^{-5}	1.43×10^{-4}	5.46×10^{-5}	4.45×10^{-5}	$\text{m}^2 \text{s}^{-1}$
λ^{-1}	2.710	4.161	4.941	4.161	4.161	4.723	days^{-1}
κ_t	5.21×10^6	5.11×10^6	4.09×10^6	3.98×10^6	3.98×10^6	3.98×10^6	$\text{m}^2 \text{s}^{-1}$
κ_q	1.17×10^6	1.95×10^6	1.18×10^6	1.95×10^6	1.95×10^6	1.17×10^6	$\text{m}^2 \text{s}^{-1}$
β_T	1.04×10^{-3}	1.04×10^{-3}	1.04×10^{-3}	1.28×10^{-3}	1.04×10^{-3}	1.04×10^{-3}	–
β_q	0.165	0.844	0.430	0.842	0.841	0.415	–
F_a	0.727	0.976	0.886	0.852	0.852	0.635	–
l_d	1.410	1.143	1.174	1.214	1.120	1.115	radians
l_s	0.0	0.120	0.057	0.026	0.057	0.120	–
κ_{hi}	2.00×10^4	3.65×10^4	9.88×10^4	9.88×10^4	9.92×10^4	6.99×10^3	$\text{m}^2 \text{s}^{-1}$
r_κ	0.6	0.764	0.830	0.756	0.861	0.754	–

is not properly accounted for. Of course, the genetic multi-objective search comes with no mathematical guarantees of convergence (not even to locally non-dominated parameter sets), but experience shows that this class of heuristics is better suited to problems with high dimensionality, especially those that exhibit discontinuities (which, as we shall see, are a feature of the tuning landscape being considered here).

On a complex, multi-dimensional search space containing islands of implausibility, a standard choice for a multi-objective search heuristic is some type of evolutionary algorithm. These require few assumptions in relation to the shapes of the objective landscapes and their derivatives, and promise robust performance even in conditions such as those encountered by Edwards et al. (2011), whose study identified potential discontinuities caused by AMOC collapse.

The particular heuristic adopted here is the Non-dominated Sorting Genetic Algorithm II, or NSGA-II (Deb et al., 2002). A key challenge in all evolutionary search algorithms is the effective control of population diversity; if diversity declines too rapidly, the population may converge prematurely on a local optimum, whereas convergence onto the global optimum may be delayed unduly in an excessively diverse population. In multi-objective search algorithms, the issue of diversity manifests itself in terms of convergence not only on multi-modal landscapes but also on the uniformity of the coverage of the Pareto front. If diversity is lost prematurely, the population might be drawn towards a particular segment of the front, which basically amounts to biasing the search in favour of one of the objectives – a phenomenon we sought to avoid here. NSGA-II's means of diversity control is solution density estimation via a *crowding distance* metric associated with each individual. The fitness function balances this with the *non-domination rank* of the individual

(i.e. the order number of the Pareto front it sits on) to exert selective pressure that evens out the sampling of the front.

Given a CPU time of 2 h and 10 min for a model run with a 5000 yr spin-up and the availability of approximately 90 processors at any one time on the University of Southampton Iridis 3 supercomputer, we opted for a population size of 90 individuals, giving a generation wall-clock time of 2 h 10 min. For each of the 90 individual parameter sets, the 13 parameter values are each randomly sampled in the range between minima and maxima specified in Table 1. In total, we ran the ensemble selection search over 20 generations to obtain a non-dominated ensemble of tuned parameter sets.

4 Results

We first address the initial generation of non-dominated ensembles of equilibrium solutions and the identification of Pareto fronts in 2-D target space. We then outline the isolation of five equally plausible parameter sets, based on the appraisal of atmospheric and ocean state variables, and three sensitive model diagnostics (2-D fields). These five parameter sets are subsequently used to investigate climate sensitivity, alongside the default parameter set of Marsh et al. (2011). Finally, we focus on the sensitivity of the four objective functions and one of these diagnostics to key parameters in the vicinity of one tuned parameter set, exploring features of the “landscape” associated with variations in these parameters.

4.1 Identifying non-dominated ensembles

After 20 generations of simulations, we obtain the Pareto fronts (or, more specifically, two-dimensional projections of Pareto fronts) depicted in Fig. 1, which illustrates the relationships between objective functions that measure goodness

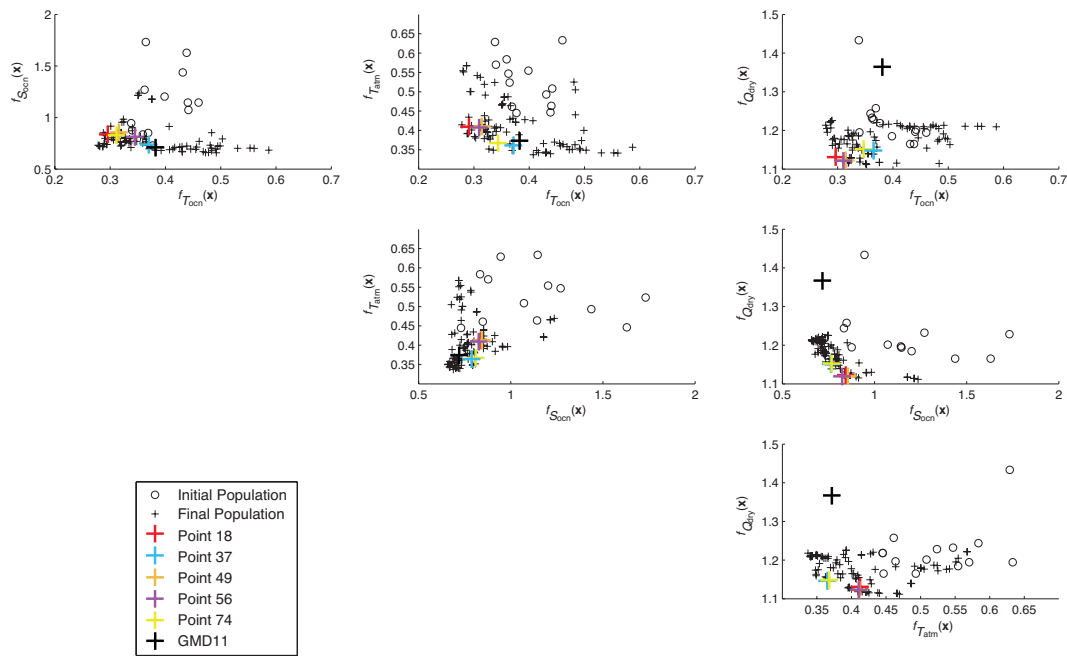


Fig. 1. Scatter plots for each pair of objective function values, as defined in Eqs. (1)–(4), for the 90 non-dominated vectors, prior to (circles) and following tuning (crosses) via 20 generations of simulations (per ensemble member). Each cross or circle therefore represents the final state of the 5000-year spin-up for each ensemble member. Five tuned ensemble members, or “points”, judged to be optimal and colour-coded, along with the objective function values for an identical spin-up of the Marsh et al. (2011) configuration (GMD11). The 2-D Pareto fronts depicted here thus illustrate relationships between four measures of model error. Note the relatively small number of successful runs in the first generation – hence the considerably fewer than 90 circles on each plot.

of fit, or model error. In particular, f_{Tocn} and f_{Socn} are reasonably well correlated in the region of the best objective values. This is consistent with obtaining water mass properties (temperature and salinity) that are closest to observations. An even clearer correlation is between f_{Tatm} and f_{Socn} , possibly due to a link between air temperature and evaporative forcing of surface salinity. The shape of the fronts related to other possible pairings indicates some level of competition between these objective functions. The apparent competition between f_{Socn} and f_{Qdry} in particular may reflect a trade-off between realism over land or oceans, or between climate zones, so more realistic ocean salinity may be obtained with less realistic land humidity, or more realistic low-latitude salinity may be obtained with less realistic high-latitude humidity (and vice versa in both cases).

Away from the optimal values, we note a tail towards high values of f_{Tocn} , with low values for the other objective function, indicating that ocean temperatures are less easily constrained, compared to the other three targets. Finally, we note that GMD11 is a clear outlier in terms of f_{Qdry} – we return to this issue below.

Mapping the Pareto front back into the 13-dimensional parameter space yielded the histograms shown in Fig. 2. Clearly some distributions are bimodal, further suggesting trade-offs between different processes/regions and objective functions.

4.2 An ensemble of five parameter sets

From the 90 Pareto-optimal parameter sets obtained through the multi-objective search, we selected five for further analysis, on the basis of their objective function values (see colour-coded crosses in Fig. 1). Specifically, we have selected points 18, 37, 49, 56, and 74, as they were the only sets to feature in the top third of the overall objective function ranges of the 90 points against all four of their objectives. The values of the input variables for these five points – along with the corresponding values for the Marsh et al. (2011) configuration (using a choice of parameter values guided by experimental experience, henceforth referred to as GMD11) – are shown in Table 3. We note here some systematic changes in the tuned parameters, compared to the GMD11 values: wind scalings, W , are increased by 54–63 %; ocean horizontal diffusivities, κ_h , are substantially larger, by a factor of 3.66–3.72; inverse frictional timescales, λ^{-1} , are substantially increased (reducing friction in the momentum balance), by a factor of 1.54–1.82; amplitudes of atmospheric heat diffusivity, κ_r , are lower by 2–24 %; moisture advection coefficients, β_q , are substantially larger, by a factor of 2.52–5.12; width scalings on atmospheric heat diffusivity, l_d , are smaller by 14–21 %; south-to-north slopes in atmospheric heat diffusivity, l_s , are slightly positive (rather than zero); sea ice diffusivities, κ_{hi} , are substantially larger, by a factor of 1.83–4.94; meridional heat

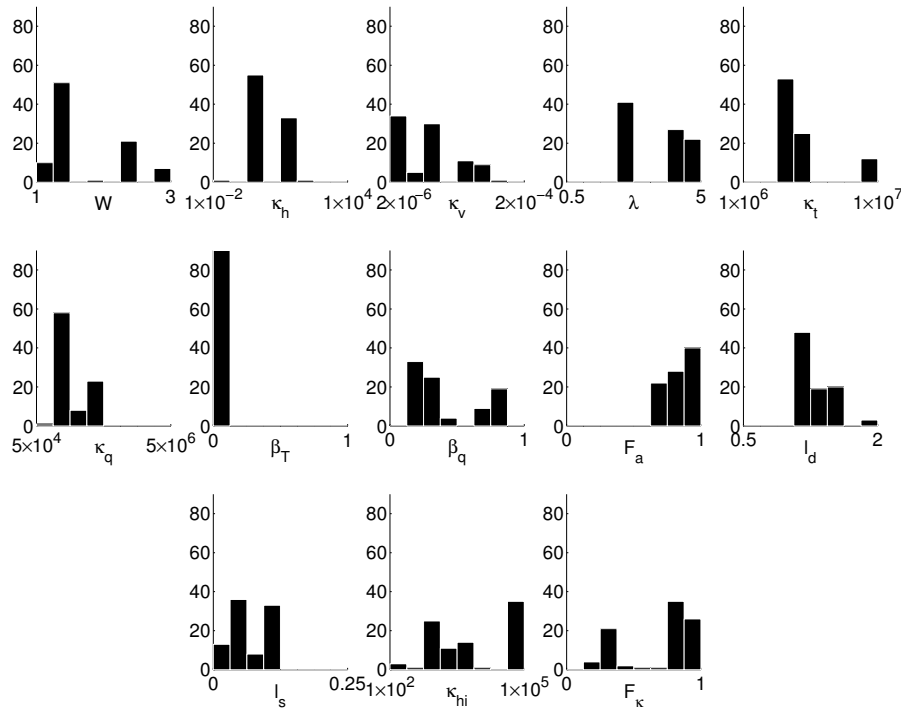


Fig. 2. Ordinary histograms of parameter values across the 90 Pareto-optimal designs, obtained by mapping the Pareto front back into the 13-dimensional parameter space and plotting number of ensemble members per parameter interval in each minimum–maximum range. See Table 1 for parameter definitions and units.

diffusivity scalings south of 56° S, r_{κ} , are increased by a factor of 1.26–1.44 (i.e. reductions of high-latitude diffusivity are more modest).

For closer examination of these points, and for comparison with GMD11, we evaluate selected variables and diagnostics that describe the state of the atmosphere, ocean and sea ice. Figures S1–12 (Supplement) show simulated, observed and difference (simulated minus observed) fields, for annual-mean surface air temperature and relative humidity (Figs. S1–6), and for annual-mean sea surface temperature and salinity (Figs. S7–12). We show corresponding Taylor diagrams for air temperature (Fig. 3a), relative humidity (Fig. 3b), sea surface temperature (Fig. 3c), sea surface salinity (Fig. 3d), and also for full-depth ocean temperature (Fig. 3e) and salinity (Fig. 3f). Figure 4a through e show annual-mean sea ice concentrations and thicknesses for this small ensemble, with Figs. 5a through e and Fig. 6a through e showing, respectively, the barotropic stream function and the Atlantic meridional overturning stream function. As a reference, Figs. 4f, 5f and 6f show the sea ice variables and ocean circulation stream functions for GMD11.

We first consider the atmosphere (Figs. S1–6). The model is cooler than observations at most locations, with particularly large errors in the Eurasian Arctic. Relative humidity is generally too homogeneous, with limited representation of low and high terrestrial values in arid subtropics and the

humid tropics, respectively, due in turn to the limited representation of the terrestrial hydrological cycle in the eb_go_gs configuration of GENIE. It is, however, notable that a greater degree of realism emerges over land for the five tuned points, compared to GMD11. We further note that relative humidities over the mid-latitude oceans are too low, while too high over the tropical/subtropical oceans, with the particular exception of anomalously low humidities over the eastern subtropical basins for points 18, 49 and 56. The surface ocean is generally characterized by a cold Atlantic sector and a warm/cold dipole in the west/east Pacific, and largest salinity errors at western boundaries and at high latitudes (Figs. S7–12).

Differences between the fits of each simulated property distribution to observations are captured in the Taylor diagrams (Fig. 3), where standard deviations of 1.0 correspond to the correct amplitude of property distribution. Parameter tuning both improves and degrades property distributions, compared to GMD11. Although differences relative to GMD11 are generally small, some are notable. The standard deviation of air temperature is somewhat improved, without compromising correlation, for points 37 and 74. In the case of relative humidity, standard deviation and correlation are improved over GMD11 at all points, most clearly for point 18. For sea surface temperature, only marginal differences arise in standard deviation and correlation for points 37 and 74.

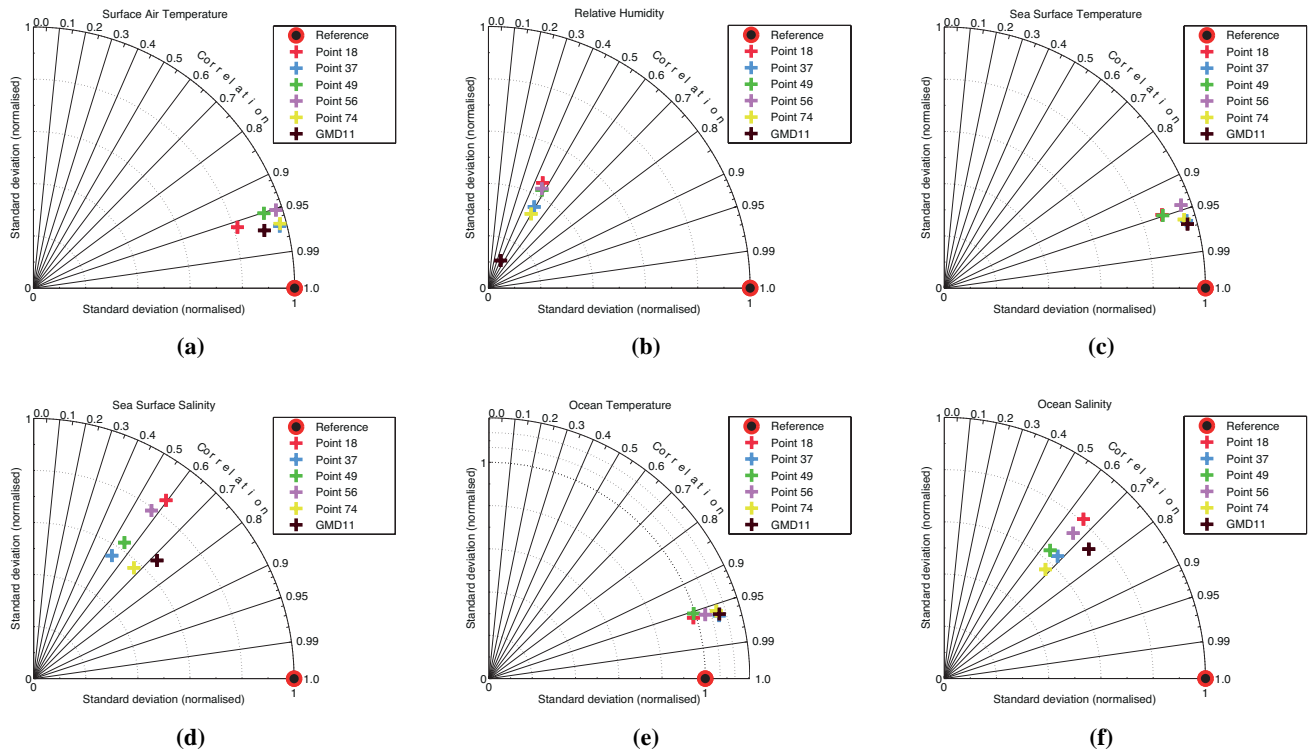


Fig. 3. Taylor diagram of (a) air temperature T_{atm} , (b) relative humidity Q_{dry} , (c) sea surface temperature, (d) sea surface salinity, (e) ocean temperature T_{ocn} , and (f) ocean salinity S_{ocn} , for the five selected points and GMD11. Standard deviations are normalized (dividing standard deviation of model data by standard deviation of observations) such that values of 1.0 correspond to the correct amplitude of property distribution.

For sea surface salinity, standard deviation is substantially improved, with reduced correlation, for points 18 and 56. As for surface temperature, the fits of full-depth temperature distributions are little altered by tuning. For full-depth salinity, standard deviation is improved for point 18, but standard deviation and correlation are both degraded for points 37, 49 and 74.

Beyond property distributions, we also evaluate aspects of the climate system that are of regional importance and likely to play key roles in the transient response to radiative forcing. We first consider sea ice distributions in the context of reanalysis data (consistent with the target data for air temperature and relative humidity) – see Fig. S13 (Supplement). Points 18 and 74 show reasonable values in the north, but too little sea ice in the south (Fig. 4a and e). Point 37 corresponds to plausible amounts of sea ice in the south, although slightly excessive in the north (Fig. 4b). There is too much sea ice in the north for point 49, yet almost none in the south (Fig. 4c), while northern sea ice is the most excessive for point 56 (Fig. 4d). In contrast to the ensemble of tuned points, sea ice in the Southern Hemisphere of GMD11 is most realistic, although excessive in the north (Fig. 4f).

The horizontal (barotropic) circulation in GENIE is unrealistically weak in general (Edwards and Marsh, 2005), although the wind-driven gyres in GMD11 are notably weaker

than for the five selected points, consistent with the large differences in W and λ^{-1} (Table 3). In quantitative terms, observations indicate a circumpolar transport of around 140 ± 6 Sv (Ganachaud and Wunsch, 2000), while the corresponding transports in GENIE are weaker by 25–50% (see Table 4). More specifically, The Antarctic circumpolar flow is strongest, and hence most realistic, for point 49 (Fig. 5c). The barotropic circulation of GMD11 is most unrealistically weak (Fig. 5f).

The Atlantic overturning circulation comprises two meridional cells: an upper cell transporting around 15 Sv and an abyssal cell transporting around 2 Sv (Ganachaud and Wunsch, 2000). For comparison with these estimates, we list maxima and minima of the AMOC stream function in Table 4. Points 18 and 56 yield the most realistic overturning stream functions (Fig. 6a and d). Points 37 and 49 (Fig. 6b and c) are characterized by overturning stream functions that are rather too intense, and the southward flow for point 37 extends considerably deeper than the 2000–3000 m depth range that is observed (see Fig. 2 in Lumpkin and Speer (2007)). The overturning corresponding to point 74 is too weak – it is close to collapse (Fig. 6e). Compared to points 18, 37, 49 and 56, the Atlantic overturning of GMD11 is unrealistically weak and shallow (Fig. 6f).

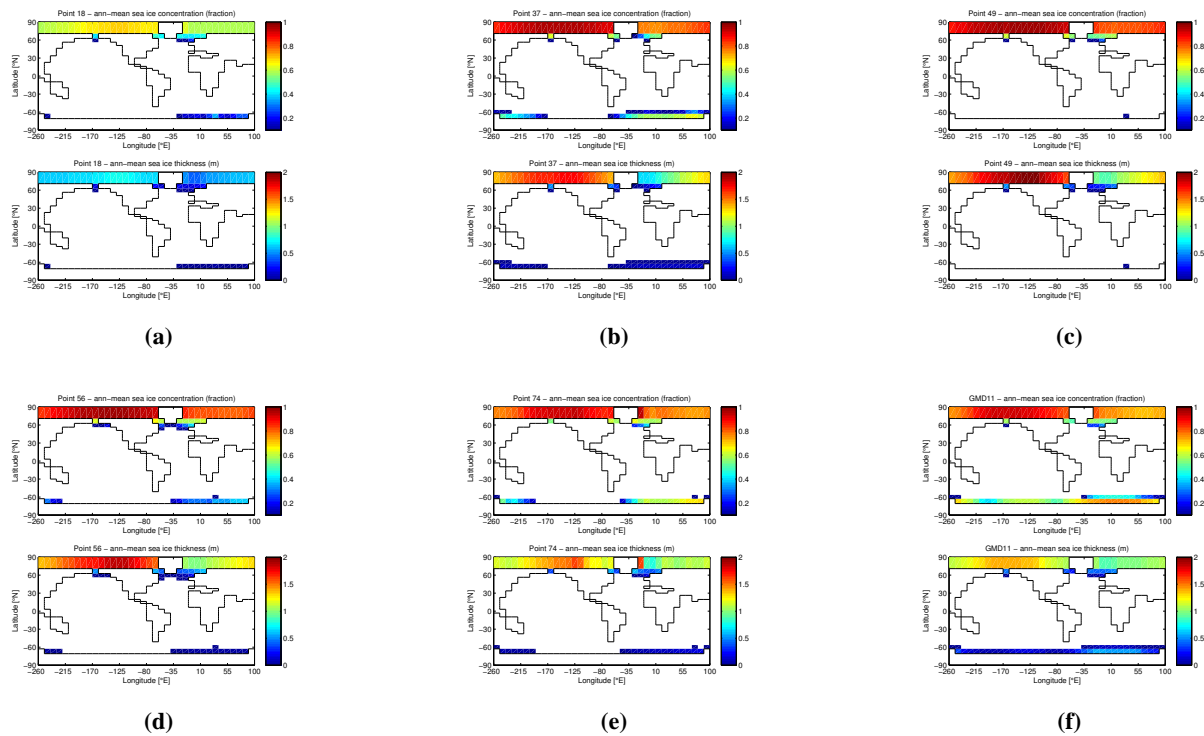


Fig. 4. Sea ice diagnostics (upper panels – concentration; lower panels – thickness) for the five tuned points and GMD11 for reference – (a) point 18, (b) point 37, (c) point 49, (d) point 56, (e) point 74, and (f) GMD11.

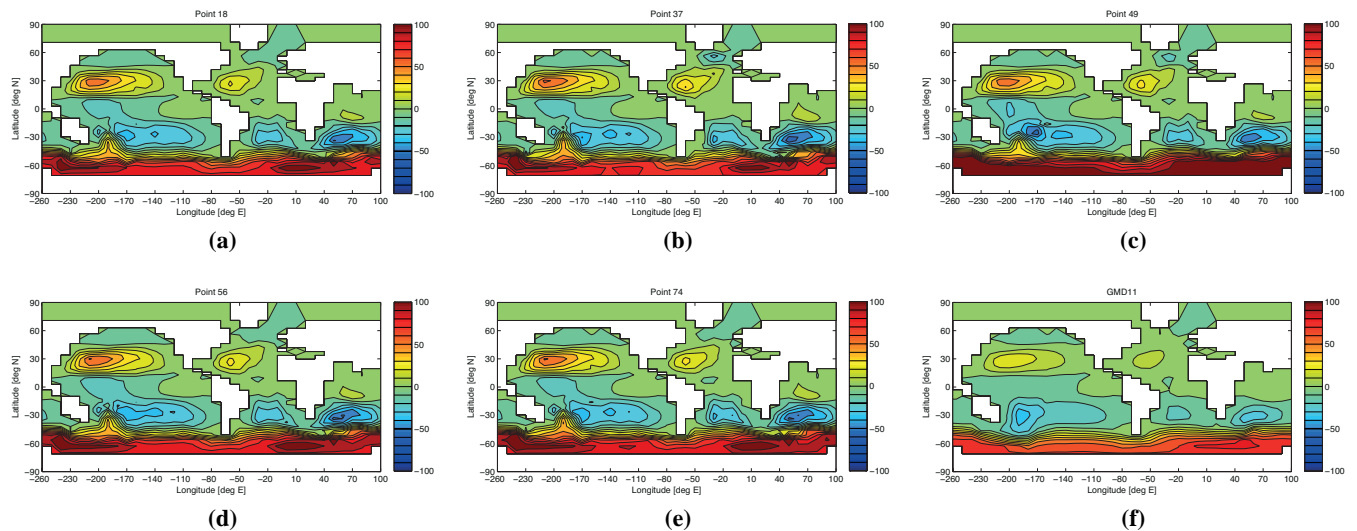


Fig. 5. Barotropic stream function – (a) point 18, (b) point 37, (c) point 49, (d) point 56, (e) point 74, and (f) GMD11.

With these states of sea ice and ocean circulation in mind, we undertake equilibrium climate sensitivity and transient climate response experiments. First, we repeat the spin-up experiments under doubled CO_2 . Figure 7 shows the change of air temperature due to CO_2 doubling, subtracting air temperature for the original $1\times\text{CO}_2$ experiments from that for the $2\times\text{CO}_2$ experiments, for the five selected points and GMD11. While patterns of temperature change are broadly similar,

regional differences are considerable, with the most uniform warming for point 37 and strongest regional variation for point 49. The most striking differences are over the North Atlantic, likely associated with varying responses of AMOC – see below. Large differences are also apparent in and around Antarctica, in proportion to the amount of sea ice in the $1\times\text{CO}_2$ simulation.

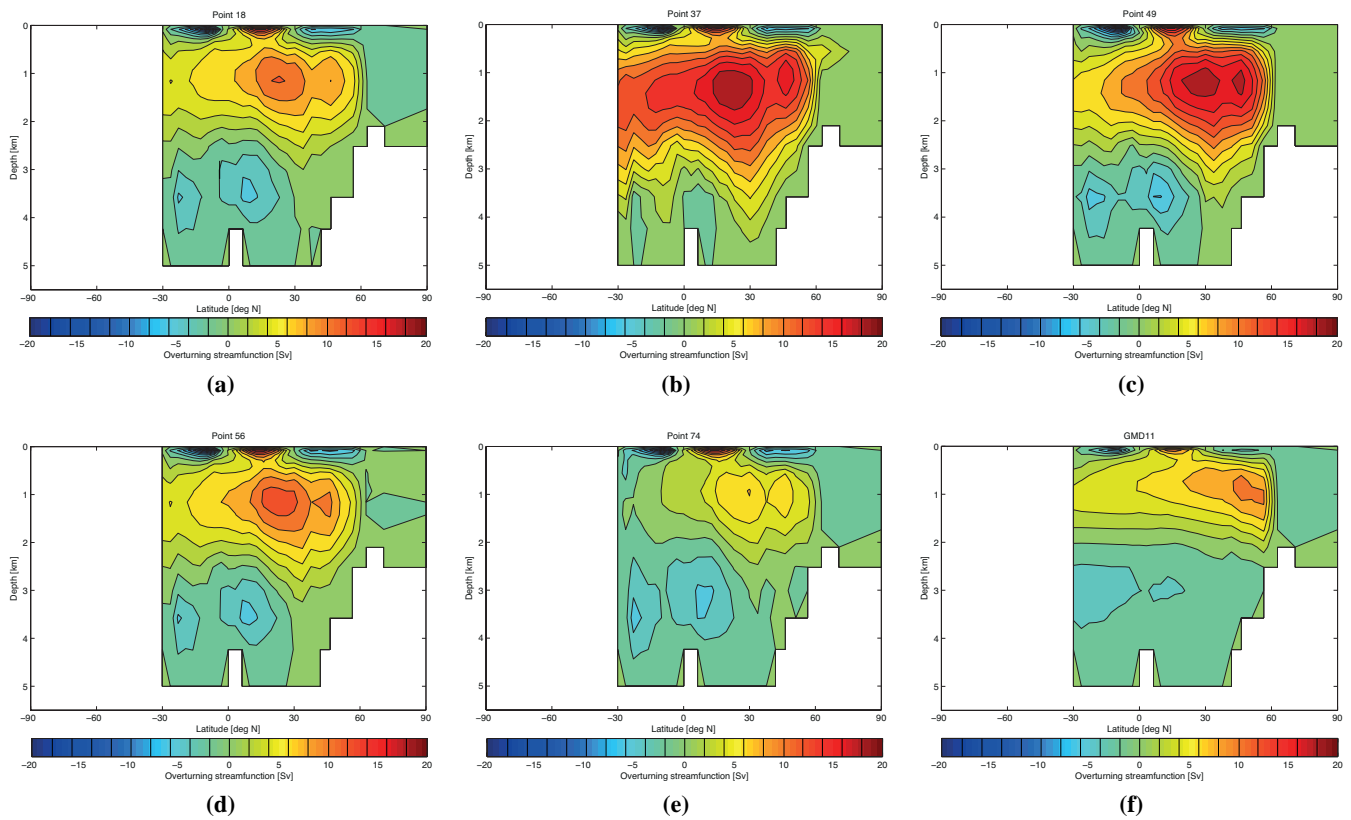


Fig. 6. Atlantic overturning stream function – (a) point 18, (b) point 37, (c) point 49, (d) point 56, (e) point 74, and (f) GMD11.

Table 4. Selected transport metrics for the five selected points and GMD11. Circumpolar transport is defined as the net eastward transport through Drake Passage. Maximum and minimum AMOC transport are defined below 700 m to exclude shallow wind-driven cells, and represent the strength of the upper and abyssal cells, respectively. Transports for the recommended point 18 are emphasized in bold font.

Transport (Sv)	GMD11	Point no.				
		18	37	49	56	74
Circumpolar	71.2	82.3	79.2	105.0	86.6	88.2
Maximum AMOC	10.9	12.3	19.5	19.0	13.1	8.6
Minimum AMOC	-3.0	-4.6	-1.6	-6.1	-4.7	-5.1

Atlantic overturning stream functions under CO₂ doubling are shown in Fig. 8. In five of the six cases, the AMOC has indeed collapsed and a reverse cell of varying intensity is established in place of clockwise overturning. Strikingly, however, the AMOC persists, only slightly reduced, in the case of point 37, and this is associated with pan-Arctic warming (see Fig. 7b). In an opposite sense, the strongest zonal contrasts evident for point 49 (Fig. 7c) are associated with the most intense upwelling of deep water in the North Atlantic (Fig. 8c).

In a second experiment to investigate transient climate responses, we increase CO₂ at 1% per annum (compounded) for 100 yr, continuing simulations from the end of each initial spin-up. Figure 9 shows the change of air temperature,

relative to initial state, averaged over the Northern Hemisphere (Fig. 9a), the Southern Hemisphere (Fig. 9b), and globally (Fig. 9c). While centennial temperature rises are in the range of more complex climate models (Hegerl et al., 2007), there is a substantial degree of spread, with GMD11 and point 49 the most and least sensitive cases, respectively. We also note more spread in the rise of air temperature averaged over the Southern Hemisphere.

To investigate the character and likely causes of differences in Fig. 9 further, we plot in Fig. 10 the spatial patterns of warming after 100 yr of CO₂ rise. These patterns are rather different from those in Fig. 7, with generally stronger north–south contrasts (stronger Northern Hemisphere warming) and less marked amelioration of warming in the North

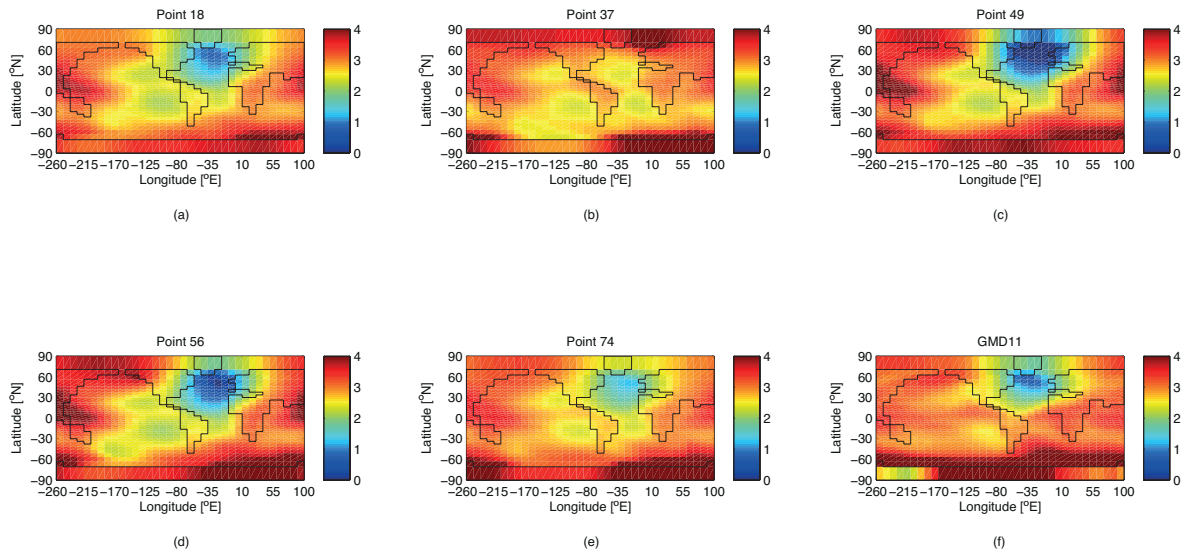


Fig. 7. Change of air temperature due to CO₂ doubling – (a) point 18, (b) point 37, (c) point 49, (d) point 56, (e) point 74, and (f) GMD11.

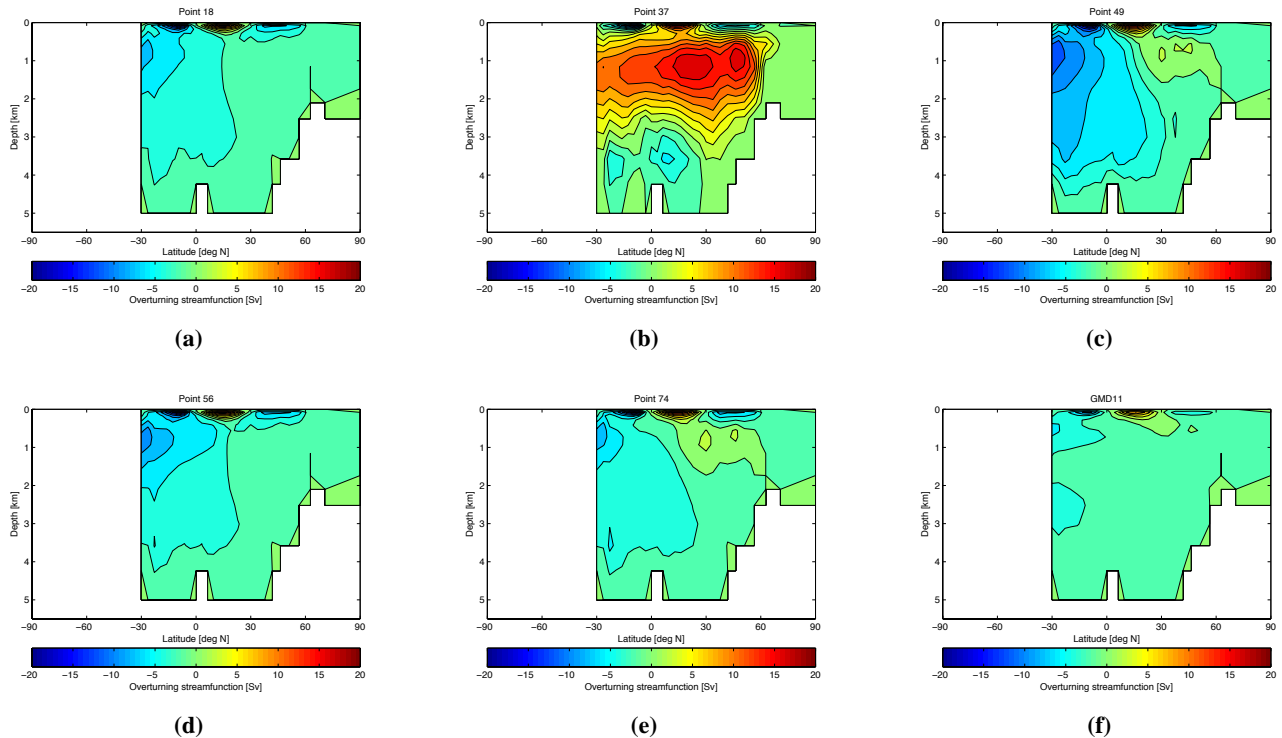


Fig. 8. Atlantic overturning stream function under CO₂ doubling – (a) point 18, (b) point 37, (c) point 49, (d) point 56, (e) point 74, and (f) GMD11.

Atlantic sector. Again, there are large differences between the six cases, with modest and strong high-latitude warming of the Southern Hemisphere for point 49 and GMD11, respectively, the latter clearly related to retreat of substantial sea ice in the initial state of GMD11 (not shown).

Finally, we consider the extent of AMOC influence on North Atlantic warming by plotting in Fig. 11 time series of

maxima in the Atlantic meridional overturning stream function. In relative terms, the responses amount to centennial reductions in the range 2.0–3.5 Sv. In the case of point 56, the AMOC actually strengthens slightly over the first 30 yr, before declining. In summary here, we conclude that AMOC influences on patterns of warming are broadly similar at centennial timescales.

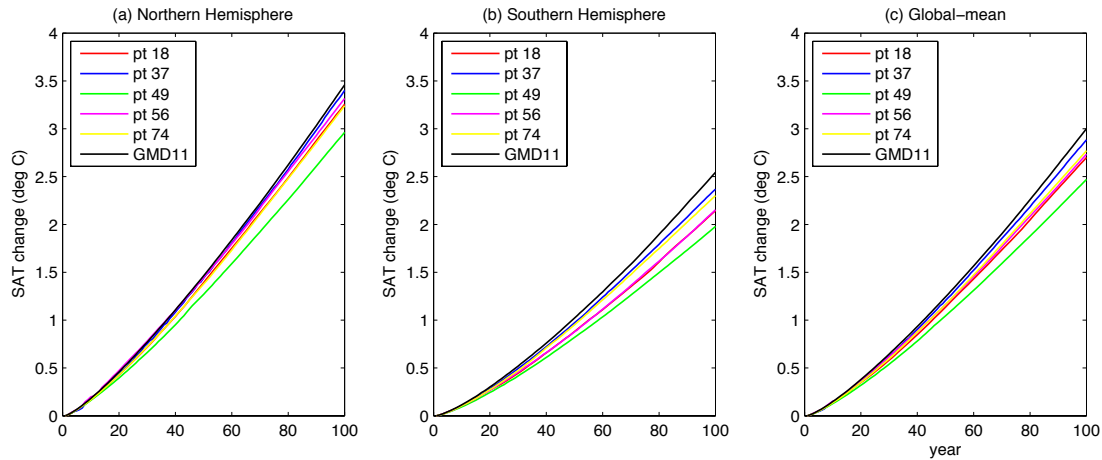


Fig. 9. Change of air temperature relative to initial state, under CO₂ rise at 1% per annum (compounded), for the five selected points and GMD11, averaged over: **(a)** the Northern Hemisphere; **(b)** the Southern Hemisphere; **(c)** globally.

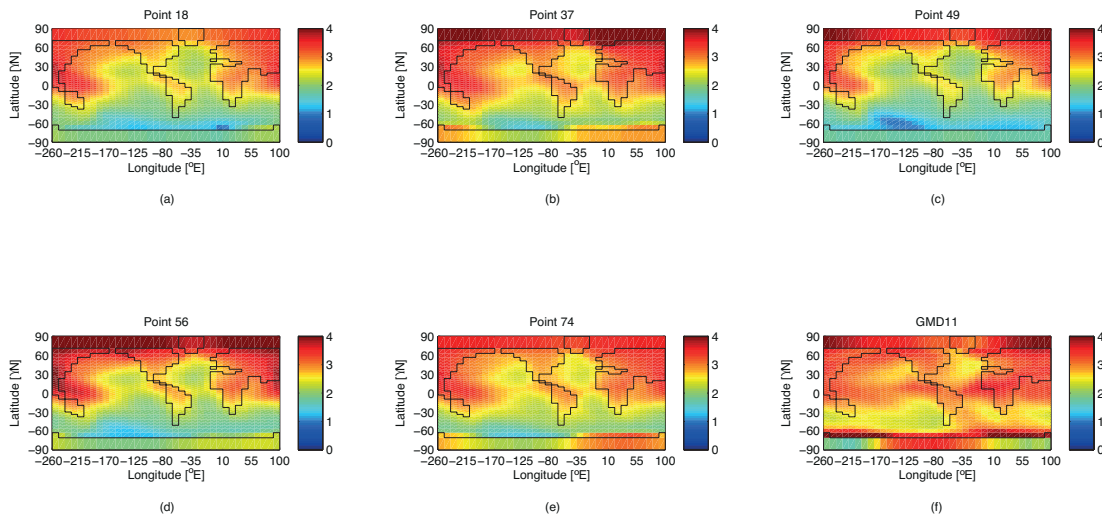


Fig. 10. Change of air temperature after 100 years of CO₂ rise at 1% per annum (compounded) – **(a)** point 18, **(b)** point 37, **(c)** point 49, **(d)** point 56, **(e)** point 74, and **(f)** GMD11.

Table 5 lists the corresponding equilibrium climate sensitivity and transient climate response, for each selected point and GMD11. Equilibrium climate sensitivity ranges from 2.742 °C (point 18) to 3.074 °C (GMD11), close to the canonical value of 3 °C (Knutti and Hegerl, 2008). Transient climate response ranges from 1.621 °C (point 49) to 1.937 °C (GMD11), towards the lower end of the the 5–95 percentile range of 1.5–2.8 °C featured in the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (Hegerl et al., 2007). In summary, the simulated pre-industrial climate of GMD11 is most sensitive to CO₂ forcing, consistent with the extensive sea ice and relatively weak AMOC in that state.

Following this analysis and experimentation, we find no clear advantage in any of the tuned points. However, as a most acceptable compromise between realism in property distributions, sea ice, horizontal transport and overturning, we judge point 18 to be marginally most plausible. If one parameter set alone is to be selected, then this is recommended for future use, although 5-member ensembles are further recommended, given the range of climate sensitivity and the associated uncertainty in patterns of temperature change. The parameter sets for each of the five selected points and GMD11 – encapsulated in the files 3636s16l_spinup_pt18.xml – are included in the Supplement. In the concluding results section, we investigate objective landscapes in more detail in the neighbourhood of point 18.

Table 5. Equilibrium climate sensitivity and transient climate response (°C) for the five selected points and GMD11. Equilibrium climate sensitivity is defined as the difference between global-mean air temperature in 1xCO₂ and 2xCO₂ simulations. Transient climate response is defined as the rise in global-mean air temperature when CO₂ doubles after rising at a rate of 1 % per annum (compounded), 71 yr into the 100 yr simulations shown in Fig. 9. Transports for the recommended point 18 are emphasized in bold font.

Climate sensitivity (°C)	GMD11	Point no.				
		18	37	49	56	74
Equilibrium	3.074	2.742	2.946	2.792	2.829	2.916
Transient	1.937	1.764	1.885	1.621	1.799	1.824

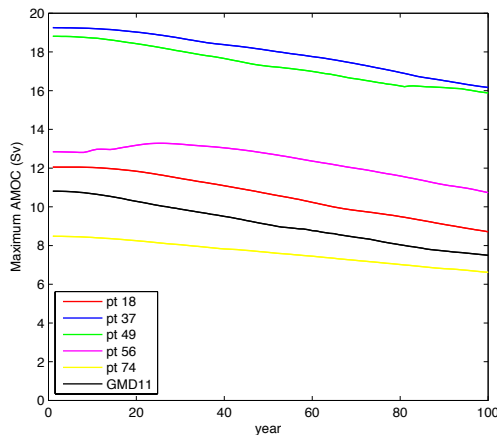


Fig. 11. Maximum of the Atlantic meridional overturning stream function over 100 yr of CO₂ rise at 1 % per annum (compounded), for the five selected points and GMD11.

4.3 Features of the landscape in an optimal ensemble member

In order to gain an understanding of the key features of the four objective landscapes (“residuals” f_{Tatm} , f_{Qdry} , f_{Tocn} and f_{Socn}), we have performed a series of 1-parameter sweeps around point 18. The resolution of these sweeps was 180 points per dimension (parameter), and the resulting 1-D sections are shown in Fig. 12.

The most striking feature of these sections through the landscape is the presence of discontinuities in some of the objective function responses. In particular, the variations of f_{Socn} with the wind-scale coefficient (W), the friction coefficient (λ), the moisture diffusivity (κ_q), the heat advection coefficient (β_T) and the freshwater flux factor (F_a) exhibit steep cliffs. Such discontinuities are most likely associated with different AMOC states. In order to understand the causes behind this phenomenon better (also observed by Marsh et al., 2004, and Edwards et al., 2011), we therefore investigate AMOC state in the neighbourhoods of these discontinuities.

In particular, we examine two metrics of the AMOC: (i) maximum (positive) intensity of the upper cell, representing the extent of northern sinking (the outflow of dense water formed in the North Atlantic) and (ii) minimum (negative) intensity of the lower cell, representing the extent of southern sinking (the inflow of dense water formed around Antarctica). These metrics are investigated on either side of each cliff, as a function of W , λ , κ_q , β_T and F_a (see Fig. 13). These “one-factor-at-a-time” studies, performed at a resolution of 180 points, cover the immediate neighbourhood of the cliffs, revealing “noise” in the transitions across each cliff.

The existence of such AMOC-transport cliffs in parameter space, associated with large-scale ocean freshwater transport, has long been recognized both in GENIE (Edwards and Marsh, 2005; Marsh et al., 2004; Lenton et al., 2006) and in other models. To our knowledge, the characterization of this transition by fine-resolution sampling of parameter space is novel. We find that the so-called cliff is instead a series of peaks (representing AMOC-on states) and troughs (representing AMOC-off states) separating the flatter regions on either side of the transition. Within the transition region, there is a strong non-monotonic dependence of steady-state AMOC on parameters expected to control the AMOC (moisture diffusivity and the freshwater flux factor), as well as three parameters that we did not expect to influence AMOC stability (wind-scale coefficient, friction coefficient and atmospheric heat advection coefficient). In this region, even the sign of the gradient $\partial\psi/\partial p_n$, where ψ is the steady-state AMOC transport and p_n is the value of parameter n , is therefore not predictable from a general understanding of controls on the AMOC.

In interpreting these results, we emphasize that these were obtained despite the use of identical initial conditions for all simulations, and despite the absence of a dynamical atmosphere or of mesoscale ocean processes recognized to give rise to chaos in climate models. It has previously been shown that $\partial\psi/\partial p_n$ undergoes frequent changes of sign within GENIE (Stephenson, 2010), likely associated with localized non-linearities in the model such as the onset or otherwise of convection within a given region of the ocean. However, the associated vacillations in AMOC transport were of sufficiently small amplitude to be negligible if large amplitude

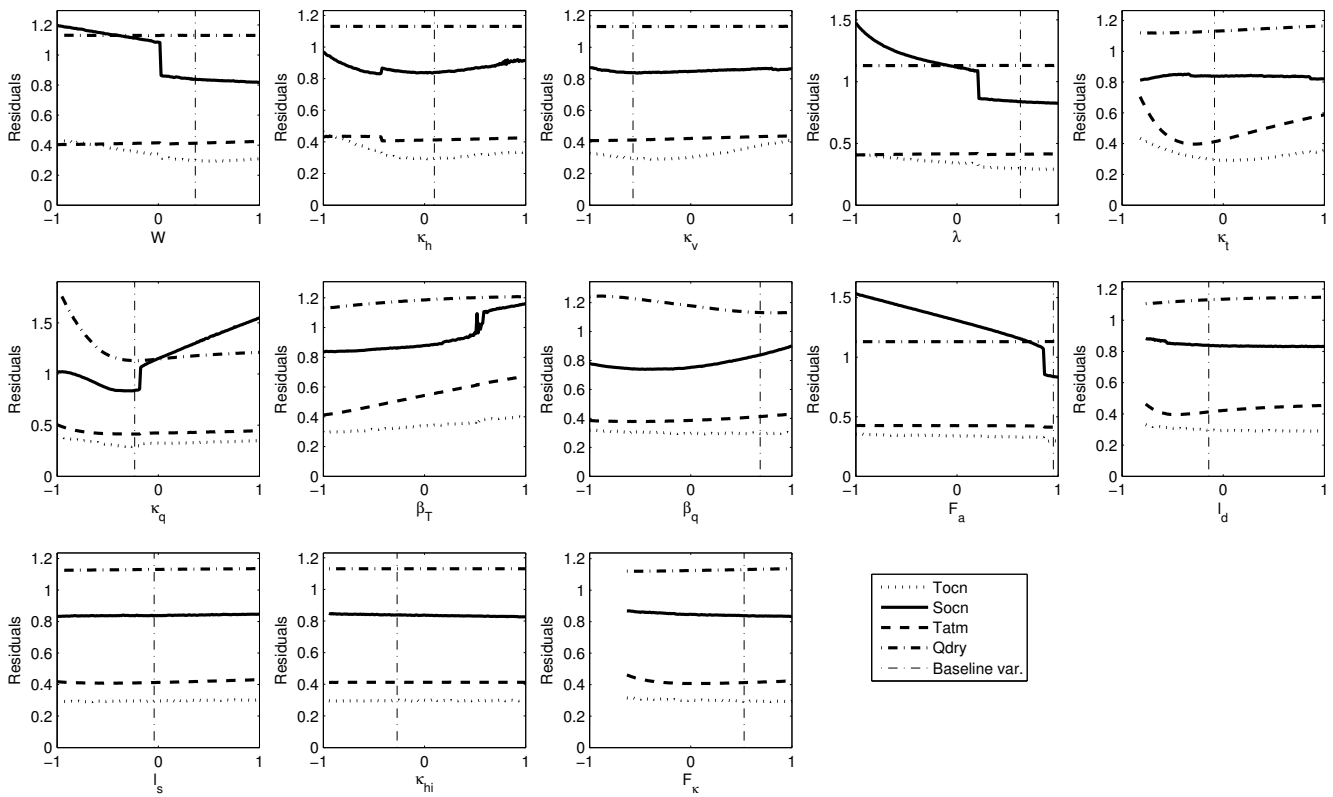


Fig. 12. 1-D slices through the objective function landscapes around point 18. By “residuals”, we refer to any one of the four objective functions, indicated by line style (see legend). The parameter axes are non-dimensionalized, -1 and 1 representing the minima and maxima of the input parameter values (Table 1), respectively. The “baseline variable” indicated by the vertical thin dot-dash line in each panel indicates the tuned parameter value for point 18 (note that this value is almost zero in the case of β_T).

changes in p_n were considered. Similar small vacillations can be observed in Fig. 13, but these vacillations attain leading order importance, in some cases exceeding 10 Sv , within the transition region. We postulate that this is due to amplification of the existing vacillations in the presence of a larger scale bifurcation with the parameter set for point 18.

5 Conclusions

A multi-objective genetic algorithm has been used to tune the key physical parameters of the climate core for an Earth system model of intermediate complexity (GENIE, version 2.7.4), previously described in Marsh et al. (2011). An ensemble of 90 parameter sets was tuned using two ocean and two atmospheric state variables as targets, defining four objective error functions. Alongside the corresponding Marsh et al. (2011) configuration (using untuned parameters), a sub-ensemble of five Pareto-optimal parameter sets was identified for more subjective evaluation of key variables and diagnostics (surface atmosphere/ocean properties, sea ice distributions and ocean circulation stream functions). While statistical analysis of surface property patterns suggests only marginal improvements over the untuned configuration in

objective terms, it was evident that sea ice and ocean circulation – key determinants of the transient climate response under radiative forcing – are more realistic after tuning, at selected points. Specifically, we emphasize stronger north–south asymmetry in sea ice (Fig. 4), stronger horizontal ocean circulation (Fig. 5) and generally a more pronounced 2-cell structure of the AMOC (Fig. 6), in the new 5-member tuned ensemble.

Further experiments are undertaken with the five tuned parameter sets, alongside the Marsh et al. (2011) parameter set, to compare equilibrium climate sensitivities and transient climate responses. The pattern of warming under doubled CO_2 is strongly shaped by changes in AMOC, while the pattern and rate of warming under rising CO_2 is closely linked to changing sea ice extent. We conclude that the leading five tuned parameter sets should ideally be used in ensemble mode to admit a constrained degree of parametric uncertainty in climate prediction with GENIE. Nevertheless, one of the Pareto-optimal parameter sets, point 18 in the ensemble, was evaluated as marginally optimal and recommended for future use in single simulations. This parameter set was subjected to further analysis of the objective function landscape in the vicinity of its tuned values.

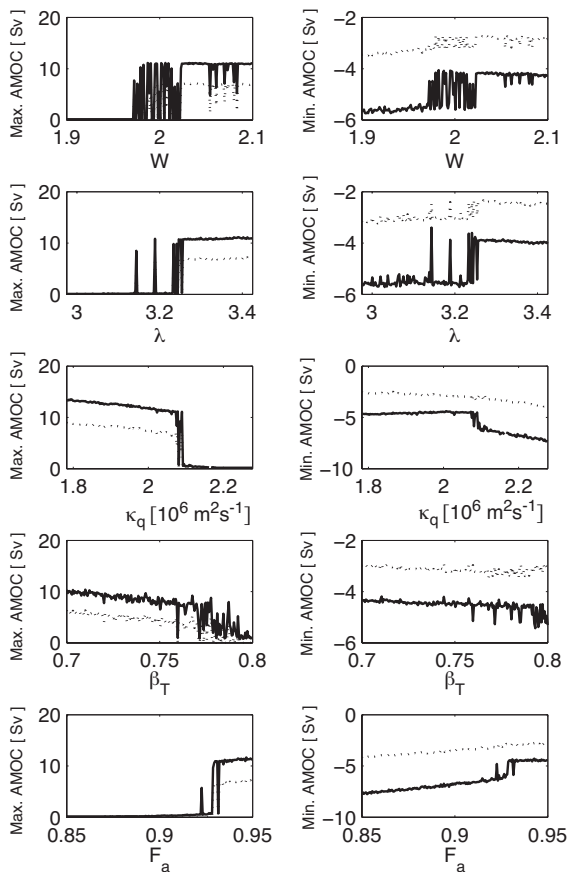


Fig. 13. 1-D slices showing the maximum (left panels) and minimum (right panels) of the overturning stream function in the Atlantic, in the vicinity of point 18. The minimum AMOC is equivalent to the maximum in the Antarctic cell in the Atlantic. The dotted lines refer to the AMOC maxima and minima at the Equator in the Atlantic.

Cliffs in the landscape are attributed to variation of the AMOC. The model AMOC is found to be highly sensitive to parameters in the proximity of a bifurcation point, manifested as vacillation between “on” and “off” AMOC states. The absence of strong AMOC variability for corresponding parameter values suggests that here the AMOC is in a bi-stable regime. This finding is complementary to previous studies that specifically addressed AMOC bistability through more extensive but less efficient parameter sweeps (Marsh et al., 2004; Lenton et al., 2006). While such studies have shown that AMOC transitions are abrupt, our fine sampling of parameter space has revealed that they are not monotonic, but are instead characterized by a region within which the dependence of AMOC transport on model parameters is highly unpredictable. Our results suggest that the limited predictability of the large-scale oscillation close to bifurcation (Knutti and Stocker, 2002) is a consequence of the location of the bifurcation being poorly defined, rather than simply uncertain or inadequately resolved by the model physics.

In summary, we have presented a tuning and evaluation of the climate core of GENIE, providing a range of plausible tuned parameter sets, and demonstrated that AMOC stability in GENIE is highly sensitive to a surprising range of model parameters.

Supplementary material related to this article is available online at <http://www.geosci-model-dev.net/6/1729/2013/gmd-6-1729-2013-supplement.zip>.

Acknowledgements. Genie-LAMP was funded by the Natural Environment Research Council. Robert Marsh acknowledges the support of the Leverhulme Trust during the completion of this manuscript. András Sóbester’s work was funded by the Royal Academy of Engineering and the Engineering and Physical Sciences Research Council under their Research Fellowship scheme. The “OptionsRSM” technology and the “Options NSGA-II” toolbox have been supplied by dezineforce and Rolls-Royce (under the VIVACE project), respectively. The simulations were run on the University of Southampton Iridis3 supercomputer. We would also like to thank Ivan I. Voutchkov for helpful suggestions regarding the implementation of the multi-objective search algorithm. Neil Edwards acknowledges financial support from EU FP7 grant no. 265170 (ERMITAGE). The authors also acknowledge Simon Mueller’s help in performing the experiments described in this paper.

Edited by: A. Ridgwell

References

- Antonov, J. I., Locarnini, R. A., Boyer, T. P., Mishonov, A. V., and Garcia, H. E.: World Ocean Atlas 2005, Volume 2: Salinity, edited by: Levitus, S., NOAA Atlas NESDIS 62, US Government Printing Office, Washington, DC, 182 pp., 2006.
- Cao, L., Eby, M., Ridgwell, A., Caldeira, K., Archer, D., Ishida, A., Joos, F., Matsumoto, K., Mikolajewicz, U., Mouchet, A., Orr, J. C., Plattner, G.-K., Schlitzer, R., Tokos, K., Totterdell, I., Tschumi, T., Yamanaka, Y., and Yool, A.: The role of ocean transport in the uptake of anthropogenic CO₂, *Biogeosciences*, 6, 375–390, doi:10.5194/bg-6-375-2009, 2009.
- Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE T. Evolut. Comput.*, 6, 182–197, doi:10.1109/4235.996017, 2002.
- Edwards, N. R. and Marsh, R.: Uncertainties due to transport-parameter sensitivity in an efficient 3-D ocean-climate model, *Clim. Dynam.*, 24, 415–433, doi:10.1007/s00382-004-0508-8, 2005.
- Edwards, N. R., Willmott, A. J., and Killworth, P. D.: On the role of topography and wind stress on the stability of the thermohaline circulation, *J. Phys. Oceanogr.*, 28, 756–778, 1998.
- Edwards, N. R., Cameron, D., and Rougier, J.: Pre-calibrating an intermediate complexity climate model, *Clim. Dynam.*, 37, 1469–1482, doi:10.1007/s00382-010-0921-0, 2011.

- Ganachaud, A. and Wunsch, C.: Improved estimates of global ocean circulation, heat transport and mixing from hydrographic data, *Nature*, 408, 453–457, 2000.
- Hegerl, G. C., Zwiers, F. W., Braconnot, P., Gillett, N. P., Luo, Y., Marengo Orsini, J. A., Nicholls, N., Penner, J. E., and Stott, P. A.: Understanding and Attributing Climate Change, in: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.
- Holden, P., Edwards, N., Oliver, K., Lenton, T., and Wilkinson, R.: A probabilistic calibration of climate sensitivity and terrestrial carbon change in GENIE-1, *Clim. Dynam.*, 35, 785–806, doi:10.1007/s00382-009-0630-8, 2010.
- Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., Hnilo, J. J., Fiorino, M., and Potter, G. L.: NCEP-DOE AMIP-II Reanalysis (R-2), *Bull. Am. Meteorol. Soc.*, 83, 1631–1643, 2002.
- Knutti, R. and Hegerl, G. C.: The equilibrium sensitivity of the Earth's temperature to radiation changes, *Nat. Geosci.*, 1, 735–743, 2008.
- Knutti, R. and Stocker, T. F.: Limited Predictability of the Future Thermohaline Circulation Close to an Instability Threshold, *J. Climate*, 15, 179–186, doi:10.1175/1520-0442(2002)015<0179:LPOTFT>2.0.CO;2, 2002.
- Lenton, T. M., Williamson, M., Edwards, N., Marsh, R., Price, A., Ridgwell, A., Shepherd, J., Cox, S., and the GENIE team: Millennial timescale carbon cycle and climate change in an efficient Earth system model, *Clim. Dynam.*, 26, 687–711, doi:10.1007/s00382-006-0109-9, 2006.
- Lenton, T. M., Held, H., Kriegler, E., Hall, J. W., Lucht, W., Rahmstorf, S., and Schellnhuber, H. J.: Tipping elements in the Earth's climate system, *P. Natl. Acad. Sci. USA*, 105, 1786–1793, doi:10.1073/pnas.0705414105, 2008.
- Locarnini, R. A., Mishonov, A. V., Antonov, J. I., Boyer, T. P., and Garcia, H. E.: *World Ocean Atlas 2005, Volume 1: Temperature*, edited by: Levitus, S., NOAA Atlas NESDIS 61, US Government Printing Office, Washington, DC, 182 pp., 2006.
- Lumpkin, R. and Speer, K.: Global ocean meridional overturning, *J. Phys. Oceanogr.*, 37, 2550–2562, doi:10.1175/JPO3130.1, 2007.
- Marsh, R., Yool, A., Lenton, T. M., Gulamali, M. Y., Edwards, N. R., Shepherd, J. G., Krznaric, M., Newhouse, S., and Cox, S. J.: Bistability of the thermohaline circulation identified through comprehensive 2-parameter sweeps of an efficient climate model, *Clim. Dynam.*, 23, 761–777, doi:10.1007/s00382-004-0474-1, 2004.
- Marsh, R., Müller, S. A., Yool, A., and Edwards, N. R.: Incorporation of the C-GOLDSTEIN efficient climate model into the GENIE framework: “eb_go_gs” configurations of GENIE, *Geosci. Model Dev.*, 4, 957–992, doi:10.5194/gmd-4-957-2011, 2011.
- Price, A. R., Myerscough, R. J., Voutchkov, I. I., Marsh, R., and Cox, S. J.: Multi-objective optimization of GENIE Earth system models, *Philos. T. R. Soc. A*, 367, 2623–2633, 2009.
- Slingo, J., Bates, K., Nikiforakis, N., Piggott, M., Roberts, M., Shaffrey, L., Stevens, I., Vidale, P., and Weller, H.: Developing the next-generation climate system models: challenges and achievements, *Philos. T. R. Soc. A*, 367, 815–831, doi:10.1098/rsta.2008.0207, 2009.
- Stephenson, G.: *Using Derivative Information in the Statistical Analysis of Computer Models*, Ph.D. thesis, University of Southampton, 197 pp., 2010.