Climate-carbon cycle uncertainties and the Paris Agreement

P. B. Holden[1*], N. R. Edwards[1,7], A. Ridgwell[2], R. D. Wilkinson[3], K. Fraedrich[4], F. Lunkeit[5], H. E. Pollitt[6,7], J.-F. Mercure[6,7,8], P. Salas[7], A. Lam[7], F. Knobloch[8], U. Chewpreecha[6] and J. E. Viñuales[7]

[1]Environment, Earth and Ecosystem Sciences, The Open University, Milton Keynes, MK7 6AA, UK
[2]Department of Earth Sciences, University of California, Riverside, CA 92521, USA
[3]School of Mathematics and Statistics, University of Sheffield, Sheffield S3 7RH, UK
[4]Max Planck Institute of Meteorology, KlimaCampus, Bundesstraße 53, 20146 Hamburg, Germany
[5]Meteorological Institute, University of Hamburg, Bundesstraße 55, 20146 Hamburg, Germany
[6]Cambridge Econometrics Ltd, Covent Garden, Cambridge, CB1 2HT, UK
[7]Cambridge Centre for Environment, Energy and Natural Resource Governance (C-EENRG), University of Cambridge, The David Attenborough Building, Pembroke Street, Cambridge CB2 3QZ, UK.
[8]Faculty of Science, Radboud University, PO Box 9010, 6500 GL Nijmegen, The Netherlands

*Corresponding author

P. B. Holden philip.holden@open.ac.uk
N. R. Edwards neil.edwards@open.ac.uk
A. Ridgwell andy@seao2.org
R. D. Wilkinson r.d.wilkinson@sheffield.ac.uk
K. Fraedrich klaus.fraedrich@mpimet.mpg.de
F. Lunkeit frank.lunkeit@uni-hamburg.de
H. E. Pollitt hp@camecon.com
J.-F. Mercure j.mercure@science.ru.nl
P. Salas pas80@cam.ac.uk
A. Lam al554@cam.ac.uk
F. Knobloch f.knobloch@science.ru.nl
U. Chewpreecha uc@camecon.com
J. E. Viñuales jev32@cam.ac.uk

Abstract 200/200 words

Main text 1992/2000 words, 30/30 refs, 5/5 display items

Methods 2845/3000 words, 21 refs.

The Paris Agreement(PA2105) PA aims to address the gap between existing climate policies and policies consistent with 'holding the increase in global average temperature to well below 2°C'. The feasibility of meeting the target has been questioned both in terms of the possible requirement for negative emissions(Anderson2016), and ongoing debate on the sensitivity of the climate-carbon cycle system(Friedlingstein2013). Using a sequence of ensembles of a fully dynamic three-dimensional climate-carbon cycle model, forced by emissions from an integrated assessment model of regional-level climate policy, economy, and technological transformation, we show that a reasonable interpretation of the PA is still technically achievable. Specifically, limiting peak (decadal) warming to less than 1.70°C, or end-century warming to less than 1.54°C, occurs in 50% of our simulations in a policy scenario without net negative emissions or excessive stringency in any policy domain. We evaluate two mitigation scenarios, with 200GTC and 307GTC post-2017 emissions, quantifying spatio-temporal variability of warming, precipitation, ocean acidification and marine productivity. Under rapid decarbonisation decadal variability dominates the mean response in critical regions, with profound implications for decision making, demanding impact methodologies that address non-linear spatio-temporal responses. Ignoring carbon-cycle feedback uncertainties (explaining 47% of peak warming uncertainty) becomes unreasonable under strong mitigation.

A widely-held misconception is that given ~1°C warming to-date, and considering committed warming concealed by ocean thermal inertia, the 1.5°C target of the Paris Agreement(PA2015) is already impossible. However, it is cumulative emissions that define peak warming(Allen2009). When carbon emissions cease, terrestrial and marine sinks are projected to draw down atmospheric CO_2, approximately cancelling the lagging warming. While the sign of this "zero emissions commitment" is uncertain, its contribution can be neglected for low CO_2 scenarios(Ehlert2017). Therefore, at least when considering CO_2 emissions in isolation, the 1.5°C target will remain physically achievable until the point that it has been crossed. The physical achievability of the Paris target has been demonstrated in a complex carbon cycle model with a simplified atmosphere(Steinacher2013) and updated recently using a simple carbon cycle model forced by a modified RCP2.6 scenario(Millar2017) and by policy-driven scenarios with substantial reliance on negative emissions technology(Rogelj2018). Here, we demonstrate that the target is achievable using a fully-dynamic three-dimensional climate-carbon cycle model forced with emissions from a detailed set of sectorally and regionally specific mitigation policies without net negative emissions(Pollitt2018).

We use the intermediate-complexity three-dimensional Earth system model PLASIM-GENIE (Holden2016), a model with similar ocean, atmosphere and carbon cycle dynamics to full complexity models, but with simpler parameterisations and lower spatial resolution. The model will not produce the full range of small-scale variability in high-complexity models, but it has the computational efficiency to allow a comprehensive treatment of uncertainties cognizant, for instance, of ongoing discussions on the state dependency of climate sensitivity (Geoffroy2013, Gregory2015) and ocean heat uptake efficacy (Winton2010). We evaluate climate-carbon cycle uncertainty using a 69-member history-matched ensemble

(Williamson2013) designed from 940 training simulations (see methods). The ensemble climate sensitivity is 2.6 to 4.5°C (90% confidence), which compares to 1.9 to 4.5°C in CMIP5(IPCCAR5). The transient climate response is 1.1 to 1.8°C, 1.2 to 2.4°C in CMIP5(IPCCAR5). Ensemble ocean heat uptake (1965 to 2004) is 207 to 330 ZJ, 182 to 363 ZJ (1970 to 2010) in IPCC(IPCCAR5).

We validate the history-matched ensemble in Table 1A, by comparison with the CMIP5 multi-model ensembles forced by Representative Concentration Pathway (RCP) 2.6 (mitigation scenario) and RCP8.5 ('business-as-usual' scenario) (Meinshausen2011). Under RCP8.5, the PLASIM-GENIE end-century CO_2 concentration, global warming and Atlantic Meridional Overturning Circulation (AMOC) strength(IPCCAR5,Cheng2013) are remarkably consistent with the CMIP5 ensemble, illustrating that uncertainties in transient climate sensitivity, carbon cycle sensitivity and AMOC stability capture the spread of high complexity models. Mean surface pH is also well represented, the significantly lower uncertainty in CMIP5 pH(Bopp2013) arises because these particular CMIP5 simulations were concentration forced. Overstated impacts in marine productivity are apparent relative to CMIP5(Bopp2013), but there is significant overlap in the highly uncertain distributions. Under RCP2.6 forcing, there is a less complete analysis of CMIP5 outputs. The PLASIM-GENIE ensemble understates the mean warming in RCP2.6 by 0.3°C relative to CMIP5, under-estimating the warmest ensemble members (Table 1A). We therefore apply 0.3°C to bias-correct warming estimates in the rapid decarbonisation scenarios (Table 1B).

Our future simulations are forced with emissions from policy scenarios of the simulation-based integrated assessment model E3ME-FTT-GENIE(Mercure2018a). The E3ME macroeconomic model differs fundamentally from the equilibrium models more usually used to assess climate policy by representing realistic (non-optimal) behaviour based on empirical relationships, and by relaxing the constraint of a fixed money supply. Investment in renewables therefore can in principle generate economic stimulus, for instance through increased employment(PollittMercure2017). Furthermore, the framework is suited to flexible application of a range of policy implementations that are not limited to a carbon tax, including regulations, subsidies, focussed taxation policies and public procurement. The model contains a bottom-up representation of technological diffusion in multiple-sectors (FTT) and is connected to a climate-carbon cycle model (GENIE) with a single-layer atmosphere. We consider three scenarios: 1) Current policy *CP*(Mercure2018a,b), 2) *2P0C*(Mercure2018a,b), rapid decarbonisation policies to avoid 2°C peak warming with 75% confidence (according to GENIE) and 3) *1P5C*(Pollitt2018), representing our most optimistic set of policy assumptions, avoiding 1.5°C peak warming with 50% confidence.

Time series for the PLASIM-GENIE ensembles forced with the three policy scenarios are illustrated in Fig 1, and ensemble distributions are summarised in Table 1B. Note that the time series of ensemble median values do not correspond to fixed simulations, thus the distribution of peak decadal warming (Table 1B) show slightly higher values as individual trajectories cross owing to decadal variability. Steady-state decadal variability of mean surface temperature in PLASIM-GENIE is ±0.08°C (one standard deviation).

Small differences in assumptions can make significant differences to cumulative emissions budgets under strong mitigation, noting that 0.1°C incremental warming is equivalent to ~50GTC(Allen2009). Here, we consider both maximum and end-century change, as the former is most relevant for impact assessment and most consistent with the text of the Paris Agreement, with change expressed relative to a preindustrial (1856-1885) baseline taken from ensembles of 1805-2105 AD transient simulations. RCP2.6 non-CO_2 forcing is applied for both mitigation scenarios, and RCP8.5 non-CO_2 forcing for the current-policy scenario.

Bias-corrected median peak warming estimates (Table 1B) are 1.82°C (2P0C) and 1.70°C (1P5C), and 2100 estimates are 1.71°C and 1.54°C. Correlations suggest an increasing relative contribution of carbon-cycle processes to warming under rapid decarbonisation (Table S1). The response of the maximum value of Atlantic meridional overturning circulation (AMOC) in the mitigation scenarios is notable. The simulated expected peak weakening to 84% of preindustrial (Table 1B) arises from natural variability (steady-state decadal variability is 0.9Sv); the median response through the Century is steady (Fig1). However, in one 1P5C and two 2P0C simulations the AMOC reduces to ~50% of its present-day strength. We therefore cannot rule out significant AMOC weakening under mitigation, but note the suggestion of a reduction in the probability of this unlikely event under accelerated decarbonisation.

We now consider the mean climate-change patterns for a range of impact-relevant climate stressors: decadal DJF surface air temperature (Fig 2A), decadal JJA precipitation (Fig 3A), annual surface ocean acidity (Fig 4A) and annual marine primary productivity (Fig 4D). Patterns are 1P5C ensemble averages of (2090 minus 1990) change, expressed per 1°C mean ensemble warming. The mean patterns of changes of temperature and precipitation are broadly consistent with CMIP5 projections. Changes in pH (Fig 4A) result from increased concentrations of dissolved CO_2 and the associated reduction in carbonate ion concentrations approximately uniform across the surface ocean, except in the Arctic where amplified CO_2 uptake is apparent under melting sea ice(Yamamoto2012). This pattern is robust, explaining more than 95% of the variability in the ensemble (quantified through singular vector decomposition); a similar robust pattern of acidification was found in CMIP5 (Bopp2013). Changes in primary productivity (Fig 4D) are dominated by large reductions of up to ~10% per °C of warming that are simulated in the Equatorial Pacific. Significant reductions are also simulated in mid-latitude Pacific and Indian oceans, and in the Equatorial and high-latitude Atlantic. Despite the simplified ecosystem model(Ridgwell2007), the patterns and magnitudes of productivity change are consistent with CMIP5 analysis; in RCP8.5, decreases of up to 30-50% are simulated in these regions(Bopp2013), attributed to increased ocean stratification and slowed circulation, with consequent reductions in nutrient supply(Steinacher2010). Increases in productivity are apparent in the Arctic and in parts of the Southern and Indian Oceans, here likely attributable to increased nutrient supply(Rykaczewski2010). In stark contrast to pH, the pattern of productivity change explains only 20% of ensemble variability.

The ensemble-projections are now used to quantify spatio-temporal uncertainty by evaluating the adequacy of the approximations made in "pattern scaling"(Santner1990), a widely used approach to estimating climate fields for impacts evaluation. In pattern scaling an average climate response is calculated, typically as a multi-decadal average pattern of change. The pattern, normalised per °C global mean warming, is then scaled as appropriate for scenarios of interest. The strengths and limitations of pattern scaling, including modified approaches, have recently been reviewed(Tebaldi2014). It is known to be less accurate under strong mitigation(Wu2010).

Figures 2B, 3B, 4B and 4E plot the normalised mean field difference (1P5C – CP), capturing non-linear scenario-dependent feedbacks, and examining the pattern-scaling approximation of a scenario-invariant pattern. The temperature pattern differences reveal modest changes, for instance in the northern Atlantic, where the stronger AMOC leads to relatively warmer temperatures under mitigation. The largest precipitation pattern differences are associated with the Indian and SE Asian monsoons. The magnitudes of pH change patterns are very different in the two scenarios, approximately -0.1pH unit per °C under current policy and -0.03 per °C for rapid decarbonisation. This difference reflects the different response times of pH and temperature to changing CO_2. The 2090 temperature is influenced by cumulative excess CO_2 but the surface pH in 2090 is determined by 2090 CO_2 with no significant lag; mitigation acts at the timescale of natural CO_2 sinks to reduce acidification impacts on the surface ocean. In contrast, the patterns of change of marine productivity in the two scenarios are spatially different, with amplified relative reductions in the Atlantic, Indian and Southern Oceans, and a reduced relative reduction in the Equatorial Pacific.

The most important error when using pattern scaling arises from the neglect of variability. This emerges from two distinct sources, the neglect of model uncertainty and the neglect of natural variability, both of which alter the pattern of change itself. It is well established that natural variability, which has a magnitude that differs with location, is a critical limiting factor for the accuracy of climate projections and impact evaluation(Deser2012). If we assume that the spread of climate model outputs encompasses possible reality, then model error can be estimated by applying the patterns from different climate models to test robustness of the impacts that result. However, internal variability is generally not considered, and pattern scaling impacts are derived from climate means. Under strong mitigation we argue this neglect may be inappropriate. The signal-noise ratio in strong mitigation scenarios is of order one and, for instance, decadal variability will be a significant contributor to the uncertainty in determining peak (~2050 AD) climate change.

In the final columns of Figs 2, 3 and 4, each 1P5C simulation anomaly field is normalised by its respective warming, and the RMS ensemble variability about the 1P5C scenario mean is plotted. For the climate fields (Figs 2 and 3), comparison of variability about the mean fields 30-year averages (predominantly parametric uncertainty) and 10-year averages (internal and parametric uncertainty) relative to a 30-year baseline, indicates that the two sources of variability are comparable in amplitude. For the ocean impact fields (Fig 4) the variability is derived from annual averages. In all fields, the uncertainties in the patterns (1P5C - CP) are dominated by the variability about the pattern (right panels). The uncertainties often dominate even the mean response. For instance, in parts of the Arctic, RMS uncertainty of ~3°C per °C warming compares to a mean signal of ~3°C (Fig 2, Table S2), while RMS uncertainty of precipitation is comparable to the mean signal in monsoon regions (Fig3, Table S2). Simulations forced by current-policy emissions are associated with significantly lower fractional uncertainty (Table S2), reflecting an increased signal-noise ratio, and demonstrating that the assumptions of pattern scaling are well justified under high-emission scenarios.

The implications of our findings for policy-making are important: if policy and market-based responses to climate change are sufficient to uphold the level of ambition of the Paris Agreement, climate change impacts could still be of large amplitude in sensitive regions such as the Arctic. However, in these scenarios, uncertainties from model error and internal variability can dominate expected mean patterns. Consequently, we argue that a paradigm shift in impacts evaluation is now essential to support decision making. Estimates based on mean patterns of change will be insufficient. Instead, statistical methodologies developed to address non-linear spatio-temporal feedbacks(Holden2015) will need to be extended to high-complexity models. Holding the increase in (multi-decadal) global average temperature above pre-industrial to 1.5 °C appears still to be possible, but results in a world where the superposition of climate change onto natural variability is key to understanding impacts on *inter alia* ecosystems, biodiversity, ice sheets and permafrost stability.

Author contributions

PBH, NRE and RDW designed and coordinated the Earth system modelling. HP, JFM and NRE designed and coordinated the energy-economy modelling. PBH, NRE, RDW and HP wrote the article with contributions from all. PBH performed the PLASIM-GENIE simulations. UC performed the E3ME-FTT simulations. All authors developed model components and/or provided scientific support: PBH (ESM coupling), KF and FL (atmosphere), NRE (ocean), AR (biogeochemistry), HP and JFM (energy-economic), PS and JFM (power sector), AL and JFM (transport sector), FK and JFM (household heating), JV (geopolitics), RDW (statistics).

Acknowledgements

The authors acknowledge C-EERNG and Cambridge Econometrics for support, and funding from EPSRC (JFM, fellowship no. EP/ K007254/1); the Newton Fund (JFM, PS, JV, EPSRC grant no EP/N002504/1 and ESRC grant no ES/N013174/1), NERC (NRE, PH, HP, grant no NE/P015093/1), CONICYT (PS), the Philomathia Foundation (JV) and Horizon 2020 (HP, JFM; Sim4Nexus project).

Author information

Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to PBH (philip.holden@open.ac.uk).

References

Adoption of the Paris Agreement FCCC/CP/2015/L.9/Rev. 1 (UNFCCC, 2015); http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf

Allen, M.R. et al. Warming caused by cumulative carbon emissions towards the trillionth tonne *Nature* **458** 1163-1166 (2009)

Anderson, K. and Peters, G. The trouble with negative emissions *Science* **6309** 182-183 (2016)

Bopp, L. et al. Multiple stressors of ocean ecosystems in the 21st century: projections with CMIP5 models. *Biogeosci.* **10**, 6225-6245 (2013)

Cheng, W., Chiang, J.C.H. & Zhang, D. Atlantic Meridional Overturning Circulation (AMOC) in CMIP5 Models: RCP and Historical Simulations. *J. Clim.* **26**, 7187-7197 (2013)

Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA (2013)

Deser, C., Knutti, R., Solomon, S. and Phillips, A.S.: Communication of the role of natural variability in future North American climate, *Nature Climate Change*, 2, 775 (2012)

Ehlert, D. and Zickfeld, K. What determines the warming commitment after cessation of CO_2 emissions? *Env. Res. Lett.* **12** 0015002 (2017)

Friedlingstein, P., et al. Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks *Journal of Climate* **27** 511-526 (2103)

Geoffroy, O., Saint-Martin, D., Olivié, D. J. L., Voldoire, A., Bellon, G., and Tytéca, S.: Transient Climate Response in a Two-Layer Energy-Balance Model. Part I: Analytical Solution and Parameter Calibration Using CMIP5 AOGCM Experiments, Journal of Climate, 26, 1841-1857, 10.1175/jcli-d-12-00195.1, 2012.

Gregory, J. M., Andrews, T., and Good, P.: The inconstancy of the transient climate response parameter under increasing CO_2, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 373, 2015.

Holden, P.B, Edwards, N.R., Garthwaite, P.H. and Wilkinson, R.D. Emulation and interpretation of high-dimensional climate outputs, *J. App. Stats.* **42**, 2038-2055 (2015)

Holden, P.B. et al. PLASIM–GENIE v1.0: a new intermediate complexity AOGCM *Geosci. Mod. Dev.* **9** 3347-3361 (2016)

Mercure, J.-F. et al. Environmental impact assessment for climate change policy with the simulation-based integrated assessment model E3ME-FTT-GENIE *Energy Strategy Reviews* **20** 195-208 (2018a)

Mercure , J.-F. et al, Macroeconomic impact of stranded fossil-fuel assets, in review *Nature Climate Change* (2018b)

Meinshausen, M. et al. The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change* **109**, 213-241 (2011)

Millar, R.J. et al. Emission budgets and pathways consistent with limiting warming to 1.5°C, Nature Geoscience, doi 10.1038/NGE03031 (2017)

Pollitt, H. & Mercure J.-F. The role of money and the financial sector in energy-economy models used for assessing climate and energy policy. *Climate Policy* http://dx.doi.org/10.1080/14693062.2016.1277685 (2017)

Pollitt, H. Policies for limiting climate change to well below 2°C, in review *Nature Climate Change* (2017)

Ridgwell, A. et al. Marine geochemical data assimilation in an efficient Earth System Model of global biogeochemical cycling *Biogeosciences* **4**, 87–104 (2007)

Rogelj, J. et al Scenarios towards limiting global mean temperature increase below 1.5°C Nature Climate Change doi:10.1038/s41558-018-0091-3 (2018)

Rykaczewski, R. R. and Dunne, J. P.: Enhanced nutrient supply to the California Current Ecosystem with global warming and in- creased stratification in an earth system model, Geophys. Res. Lett., 37, L21606, doi:10.1029/2010GL045019, 2010.

Santner B.D., Wigley T.M.L., Schlesinger M.E., Mitchell J.F.B. *Developing climate scenarios from equilibrium GCM results*, Hamburg, Germany (1990)

Steinacher, M. et al. Projected 21st century decrease in marine productivity: a multi-model analysis. *Biogeosciences* **7**, 979–1005 (2010).

Steinacher, M., Joos, F., and Stocker, T. F.: Allowable carbon emissions lowered by multiple climate targets, *Nature*, **499**, 197-201 (2013)

Tebaldi, C. & Arblaster, J.M. Pattern scaling: Its strengths and limitations, and an update on the latest model simulations. *Climatic Change* **122**, 459-471 (2014)

Williamson, D. et al. History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble. *Climate dynamics* **41**, 1703–1729 (2013)

Winton, M., Takahashi, K., and Held, I. M.: Importance of Ocean Heat Uptake Efficacy to Transient Climate Change, Journal of Climate, 23, 2333-2344, 10.1175/2009jcli3139.1, 2010.

Wu, P., Wood, R., Ridley, J. & Lowe, J. Temporary acceleration of the hydrological cycle in response to a CO_2 rampdown. *Geophys Res Lett* **37**, L12705 (2010)

Yamamoto, A., Kawamiya, M., Ishida, A., Yamanaka, Y., and Watanabe, S. Impact of rapid sea-ice reduction in the Arctic Ocean on the rate of ocean acidification, *Biogeosciences* **9**, 2365–2375 (2012).

Tables

A	RCP2.6		RCP8.5	
	CMIP5	PLASIM-GENIE	CMIP5	PLASIM-GENIE
Warming (°C)	1.0 ± 0.4 (0.3, 1.7)	0.7 ± 0.2 (0.4, 1.0)	3.7 ± 0.7 (2.6, 4.8)	3.6 ± 0.6 (2.6, 4.4)
CO_2 (ppm)		402 ± 19 (373, 429)	985 ± 97 (794, 1142)	1010 ± 110 (829, 1185)
AMOC (% change)		-6 ± 10 (-17, 4)	(-60, -15)	-32 ± 12 (-54, -16)
Surface pH (pH)	-0.07 ± 0.001	-0.04 ± 0.01 (-0.069, -0.028)	-0.33 ± 0.003	-0.33 ± 0.04 (-0.41, -0.27)
Productivity (%)	-2.0 ± 4.1	-2.7 ± 1.2 (-4.8, -1.2)	-8.6 ± 7.9	-15.1 ± 4.1 (-21.7, -7.43)

B	Current policies	2P0C policies	1P5C policies
Peak decadal warming (°C)	(2.54, 3.12, **4.18**, 5.17, 5.47)	(1.09, 1.19, **1.52**, 1.95, 2.02)	(1.04, 1.11, **1.40**, 1.74, 1.85)
Peak annual CO_2 (ppm)	(649, 703, **863**, 996, 1048)	(394, 405, **446**, 485, 493)	(381, 391, **429**, 458, 468)
Min decadal AMOC (%)	(33, 44, **68**, 80, 87)	(43, 76, **83**, 90, 95)	(51, 74, **84**, 90, 94)
Max annual surf acidification (pH)	(-0.50, -0.47, **-0.39**, -0.31, -0.27)	(-0.22, -0.19, **-0.15**, -0.12, -0.10)	(-0.19, -0.17, **-0.14**, -0.10, -0.09)
2100 decadal warming (°C)	(2.54, 3.12, **4.18**, 5.17, 5.47)	(0.73, 1.10, **1.41**, 1.81, 1.87)	(0.63, 0.97, **1.24**, 1.61, 1.67)
2105 annual CO_2 (ppm)	(649, 703, **863**, 996, 1048)	(371, 382, **415**, 445, 453)	(357, 367, **394**, 416, 427)
2100 decadal AMOC (%)	(33, 45, **69**, 83, 91)	(43, 79, **90**, 102, 104)	(52, 82, **92**, 101, 107)
2105 annual surf acidification (pH)	(-0.50, -0.47, **-0.39**, -0.31, -0.27)	(-0.19, -0.17, **-0.13**, -0.10, -0.09)	(-0.16, -0.15, **-0.11**, -0.09, -0.08)
2105 annual productivity (%)	(-33.7, -24.3, **-13.8**, -4.6, -3.5)	(-9.5, -5.0, **-3.0**, -1.1, -0.8)	(-5.7, -4.1, **-2.2**, -0.7, -0.1)
Bias corrected peak warming (°C)		(1.39, 1.49, **1.82**, 2.25, 2.32)	(1.34, 1.41, **1.70**, 2.04, 2.15)
Bias corrected 2100 warming (°C)		((1.03, 1.40, **1.71**, 2.11, 2.17)	(0.93, 1.27, **1.54**, 1.91, 1.97)

Table 1: A) PLASIM-GENIE validation against multi-model ensembles of Representative Concentration Pathways. Data are expressed as 2090-1990 decadal anomalies except for CO_2 which is 2100 concentration and PLASIM-GENIE productivity which is 2105-2005 anomaly. The 1990 PLASIM-GENIE baselines are 30-year averages (1976-2005) except for ocean pH and productivity (where annual averages are used for all analysis). Ensembles are summarised as mean ± 1 standard deviation (5th and 95th percentiles), except for CMIP5 CO_2 and AMOC where the bracketed ranges represent 11-member and 10-member ensemble spreads respectively. **B) PLASIM-GENIE summary confidence intervals of the E3ME-FTT-GENIE-1 scenarios**. Minima, 5th percentile, median, 95th percentile and maxima of the 69-member ensembles. Warming, AMOC and acidification are expressed relative to a 30-year average baseline centred on 1870. Productivity is 2105-2005 anomaly. The 0.3°C bias correction under strong mitigations is implied by the RCP2.6 CMIP5 comparison (Table 1A).

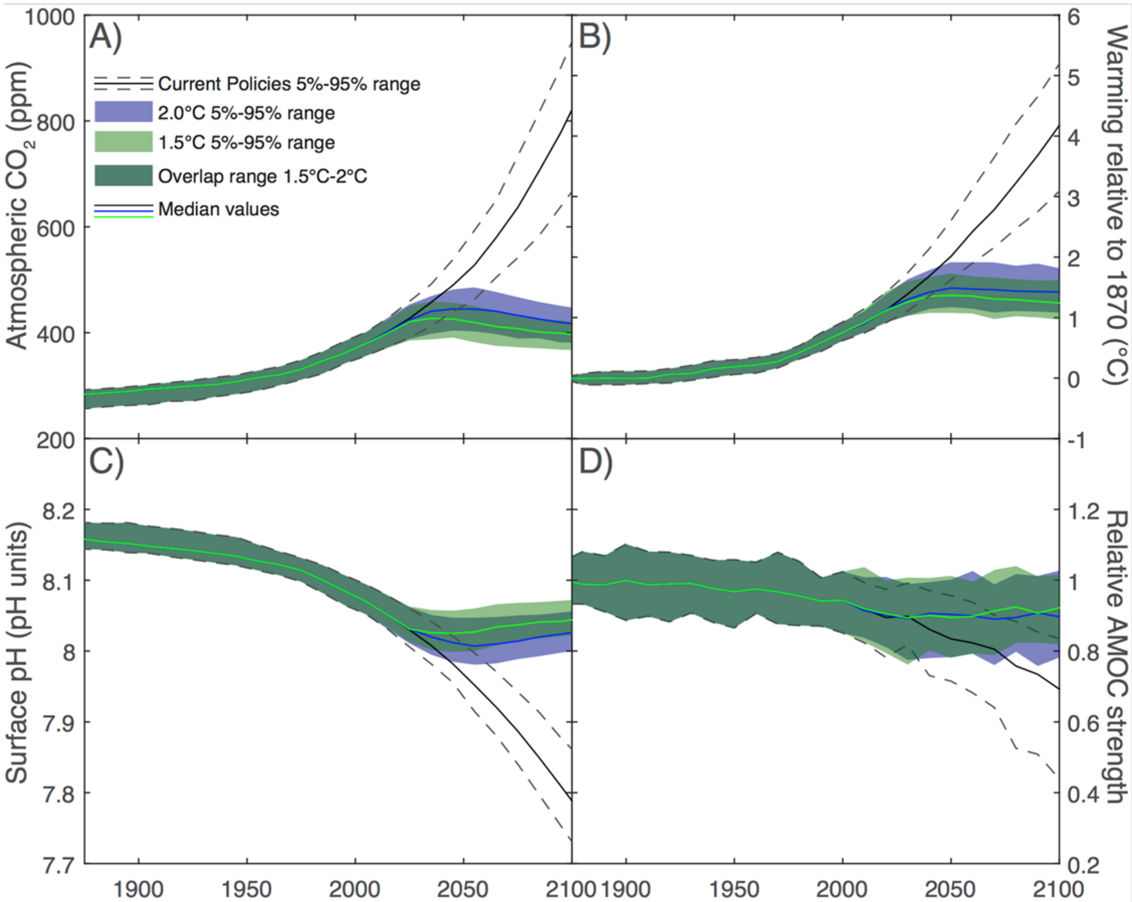

Figure 1: Summary time series of the 69-member Current-Policy, 2P0C and 1P5C E3ME-FTT-GENIE emissions-forced PLASIM-GENIE ensembles.

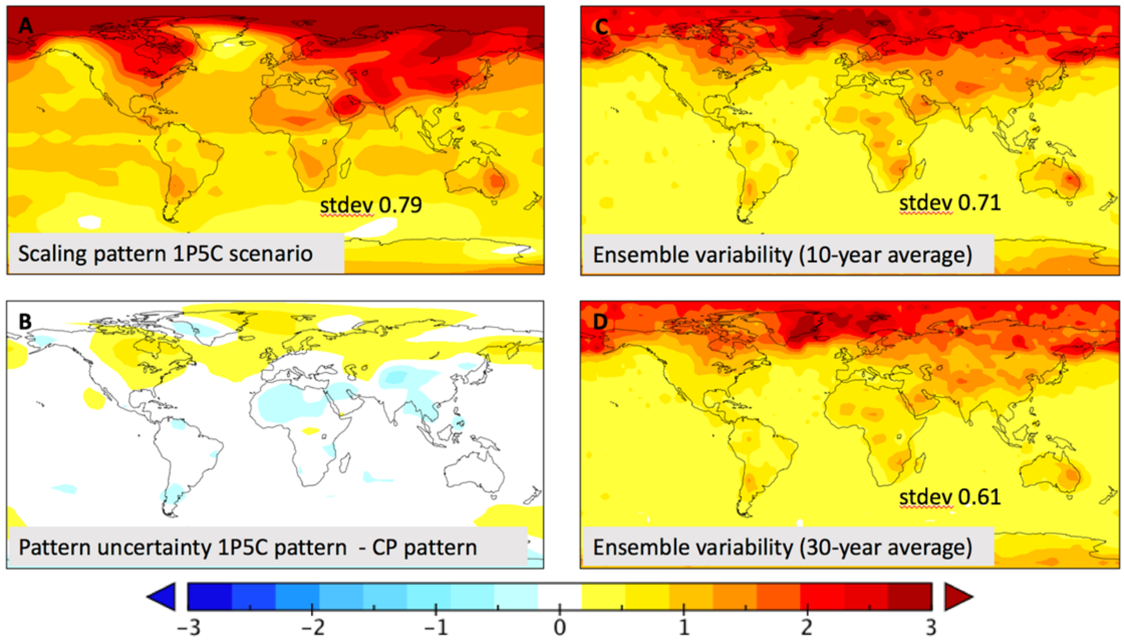

Figure 2: December-January-February surface air temperature scaling patterns and uncertainty. Scaling patterns are 1P5C and CP ensemble means (2086-2095 minus 1976-2005, °C) normalised per 1°C warming. Ensemble variability is calculated by normalising each ensemble member per 1°C warming and calculating the RMS difference with respect to the mean pattern (A). Variability is derived for both (C)10-year (2086-2095) and (D) 30-year (2076-2105) patterns to help isolate the contributions of decadal variability and parametric uncertainty.

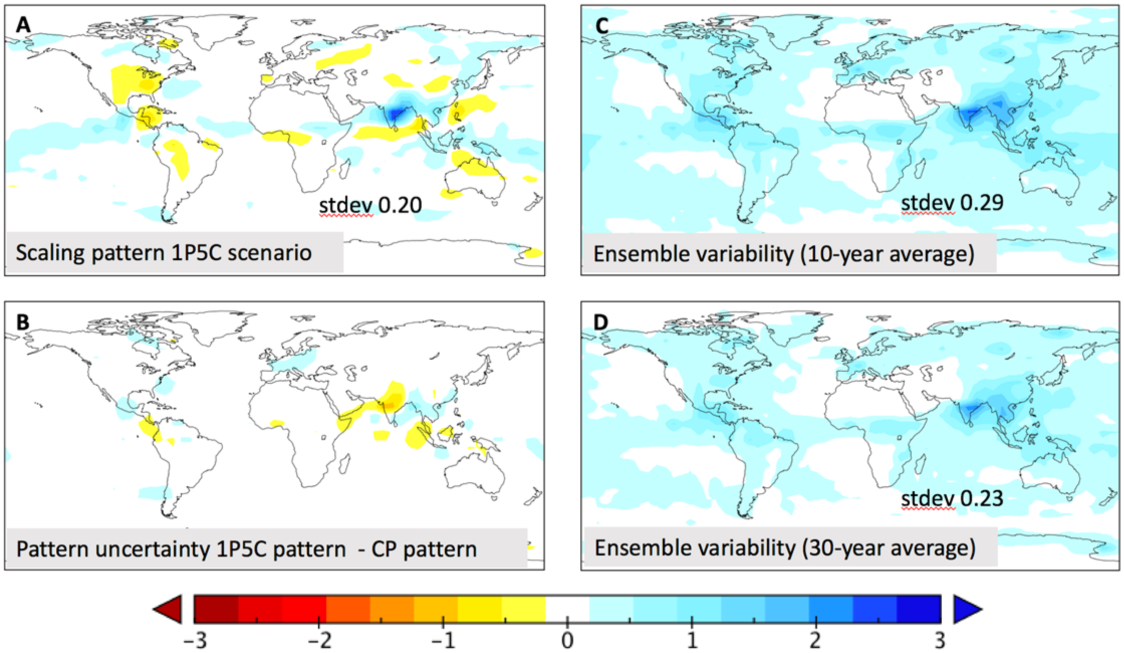

Figure 3: June-July-August precipitation scaling patterns and uncertainty. Scaling patterns are 1P5C and CP ensemble means (2086-2095 minus 1976-2005, mm/day) normalised per 1°C warming. Ensemble variability is calculated by normalising each ensemble member per 1°C warming and calculating the RMS difference with respect to the mean pattern (A). Variability is derived for both (C)10-year (2086-2095) and (D) 30-year (2076-2105) patterns to help isolate the contributions of decadal variability and parametric uncertainty.

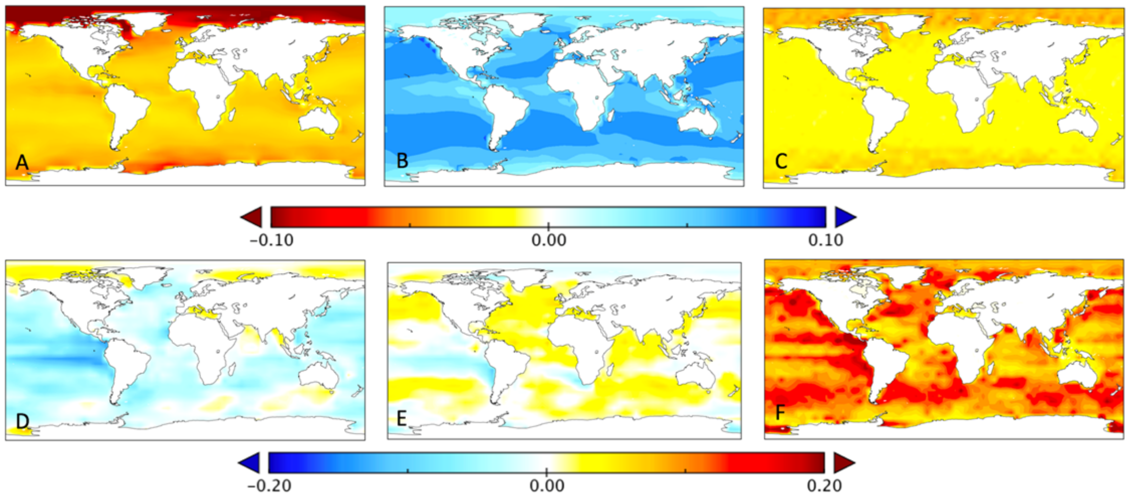

Figure 4: Ocean stressor scaling patterns and uncertainty. Top: surface pH, pH units per °C warming. Bottom: marine productivity, fractional change per °C warming. Scaling patterns (left) are 1P5C ensemble means (2105-2005), and 1P5C - CP scaling pattern difference (centre). Ensemble variability is calculated by normalising each ensemble member per 1°C warming and calculating the RMS difference with respect to the appropriate mean pattern. All data are annually averaged.

Methods

PLASIM-GENIE is a coupling of the intermediate-complexity spectral atmosphere model PLASIM (Fraedrich2012) to the Grid-Enabled Integrated Earth system model GENIE (Lenton2006). The coupling and climatology are described in detail in (Holden2016). PLASIM-GENIE is not flux corrected; the moisture flux correction required in the original tuning (Holden2016) was removed during the history-matching calibration (see below). We here apply PLASIM-GENIE with carbon-coupled biosphere modules BIOGEM and ENTS, described in (Lenton2006) for the energy-moisture balance atmosphere configuration. We apply BIOGEM with the default Michaelis-Menton phosphate-limited productivity scheme (Ridgwell2007). The carbon-cycle model has been extensively validated through model inter-comparisons (Zickfeld2013, Joos2013).

Important neglects of the PLASIM-GENIE carbon cycle are anthropogenic land-use change, peat and permafrost. These omissions tend to overstate the terrestrial carbon sink (by overstating natural forest) and they neglect potentially significant terrestrial sources (from peat and permafrost). We note that the history-matching calibration is designed to subsume such structural deficiencies (here, for instance, into CO_2 fertilization and soil respiration).

PLASIM-GENIE is freely available. Please contact the authors for information.

Atmosphere-ocean gearing. PLASIM-GENIE simulates approximately 2.5 years per CPU hour, so that 2,000-year spin-ups take one month of computing. In order to enable the exploration of parameter space, the implementation of an atmosphere-ocean gearing approach was required. The spin-up simulation time is determined by the ocean timescale, but the simulation speed of the model is determined by the atmosphere, which uses approximately

99% of the CPU demands of the physical model. In gearing mode, applied only to equilibrium spin-ups, the model alternates between a conventionally coupled mode (for 1 year) and a fixed-atmosphere mode (for 9 years), reducing spin-up CPU time by an order of magnitude. During the conventional coupling mode, atmosphere-ocean coupling variables are accumulated and saved as daily averages. These variables comprise energy fluxes, moisture fluxes and wind stresses. During the fixed atmosphere phase, the atmospheric variables are kept constant and these daily averaged fluxes are applied to the ocean. Latent heat, sensible heat and longwave radiation ocean heat loss are recalculated at every atmosphere time step during the fixed atmosphere phase, when energy conservation is therefore not imposed. This is necessary for numerical convergence because these fluxes depend upon ocean temperature, which evolves during the fixed atmosphere phase. Evaporation is not recalculated during the fixed atmosphere phase in order to ensure moisture conservation. AO-geared spin-up states are consistent with the standard model, as demonstrated by smooth spun-on historical transient simulations in all ensemble members, though we note that rapid (sub-decadal) and modest (a few Sv) AMOC adjustments are seen in some simulations, arising from different inter-annual variability.

Experimental design. Each model configuration was spun-up with a 2,000-year AO-geared quasi-equilibrium preindustrial simulation, with atmospheric CO_2 relaxed to 278ppm. Simulations were continued as emissions-forced historical transient simulations (AO-gearing off, CO_2 freely evolving). Historical forcing (1805 to 2005) comprised anthropogenic CO_2 emissions and non-CO_2 radiative forcing. Fossil fuel, cement and gas flaring emissions were prescribed from CMIP5 (https://cmip.llnl.gov/cmip5/forcing.html) and were combined with ISAM C-N land-use change emissions (Jani2013) from the HYDE land-use dataset (Ramankutty2007). Non-CO_2 forcing data was taken from Meinshausen et al (2011) implemented in PLASIM-GENIE as effective CO_2. Future (2005-2105) emissions were taken from the E3ME-FTT-GENIE scenarios, scaled by 9.82/8.62, to match estimated 2015 total emissions (Jackson2015), accounting for sources not represented in E3ME. Future land use change emissions and non-CO_2 radiative forcing were taken from RCP2.6 (1P5C and 2P0C scenarios) and RCP8.5 (CP scenario).

History-matched ensemble

Carefully designed ensembles of simulations are central to our approach to quantifying Earth system uncertainties. We applied a 'history matching' calibration strategy (Craig1997, Williamson2013), sampling throughout high-dimensional model input space to identify model configurations that are capable of producing reasonable simulations in the PLASIM-GENIE Earth system model, and then running the plausible configurations forward to characterise uncertainty about the future. Each configuration is required only to provide a 'plausible' simulation (Edwards2011), thereby avoiding the introduction of bias through over-fitting (Williamson2017). A configuration is ruled out only if it is inconsistent with an observation, allowing for the imperfections of both model and data. Thus, the history matching philosophy generates simulations that encompass the full range of realistic dynamical feedbacks (Holden2010).

In PLASIM-GENIE, identifying large numbers of history-matched configurations would be prohibitively demanding computationally. We render the problem tractable by using emulators (Sacks1989) to search throughout model input space. The emulators are trained on a sequence of preliminary ensembles amounting to 1.9 million years of climate simulation in total (940 completed simulations). The process produced 69 model variants, each validated by simulation, having considered hundreds of millions of randomly sampled parameter configurations in the emulator. The final models all adequately simulate ten key global-scale observational targets including surface air temperature, vegetation and soil carbon, Atlantic, Pacific and Southern Ocean circulation measures, dissolved O_2 and calcium carbonate flux, and transient temperature and CO_2 changes (Table S4).

For the purposes of the history matching, the simulator (here applied to the preindustrial spin-up state) can be considered as a function that maps from 32 input parameters (Table S3) to the eight different outputs (Table S4). Our aim is to infer the input values that lead to outputs within the plausible climate ranges as defined in Table S4. It is not possible to naively explore the simulator output over the full input parameter ranges by repeatedly evaluating the simulator, as for example, just doing one evaluation in each corner of the input space would require $2^{32} \approx 10^9$ model evaluations. Instead, we build emulators (O'Hagan2006, Sacks1989) that mimic the simulator response surface, and allow us to predict its value for any input. An initial large exploratory analysis was performed, motivated by the iterated waves approach (Williamson2017). Starting from a 100-member maximin latin hypercube ensemble, sequential series of 100-member ensembles were performed, probing regions of likely plausible space by using stepwise-selected linear regression models that were continually refitted as simulations completed. This produced 940 completed simulations that we used to train the final history match. Part of the motivation for the exploratory ensemble was to develop a general understanding of the range of model responses. Most notably it enabled us to identify regions of parameter space that satisfied the plausibility constraints without flux correction so that the associated parameter (APM, Table S3) could be fixed at zero for the final history match.

For the final history match, a variety of emulation approaches were considered, including stepwise regression, the LASSO (Tibshirani1996) which is a regularized version of linear regression, and Gaussian process regression with a combination of different mean and covariance functions (Rasmussen2004). To determine the optimal approach for each of the eight outputs, we split the data into test and training datasets and evaluated the emulators' predictive performance (RMSE, statistical coverage), repeating the process 10 times to get an average performance. The optimised emulators were used to find input values that are expected to give plausible simulations (i.e. within tabulated ranges for all emulator-filtered metrics, Table S4), to generate a sample of design points which encapsulate the uncertainty about future climate. We used an approximate Bayesian computation type approach (Marin2012), using rejection sampling to sample parameters from the prior distribution and evaluating the probability of these values leading to plausible outputs, to generate a large number of plausible future climates, considering hundreds of millions of emulator evaluations. A final 200-member candidate ensemble for the future transient simulations was then chosen using a ‘greedy’ design, adding points to maximize a criterion that combined the probability the simulation would be plausible (according to the emulator), and the distance of candidate points to the other points already in the design, so as to ensure design points fully span the 32-dimensional plausible input space.

The 200 history-matched parameter sets were applied to PLASIM-GENIE, and 183 were accepted as giving plausible preindustrial climates in the simulator. These were spun on through the industrial period (1805 to 2005) with emissions and non-CO_2 radiative forcing. Sixty-nine simulations were selected as also having plausible climate sensitivity (2005 -1870 warming between 0.6 and 1.0K) and carbon cycle (2005 CO_2 in the range 355 to 403ppm). These 69 model configurations were applied in the future transient ensembles.

It is instructive to compare history matching with the Bayesian approach to probabilistic calibration. In an ideal world, where we knew the appropriate likelihood (weighting) function, had a perfect simulator, sensible priors, and unlimited computational resource, then Bayesian inference is often the most appropriate approach for parameter estimation. History matching has been developed(Craig1997) as a philosophical (but closely related) alternative to Bayes that overcomes some of the difficulties that arise when doing inference with complex models, e.g., when we are not fully confident in our choice of likelihood, prior distributions, or lack a detailed (and informative) description of the model discrepancy. In history matching we do not weight simulations, instead we reject parameter values that lead to clearly implausible simulations, where implausibility is judged by relatively simple metrics relating the simulator output to the data, whilst taking into account the sources of error. Despite these simplifying assumptions, history-matched posteriors are not necessarily less reliable than Bayesian posteriors, because the subjective choices (particularly in the likelihood) are greatly simplified allowing us to think more carefully about each component, and as a consequence, the approach is also more transparent and easier to understand. In addition to using a history matching, we also use emulation to make the exploration of the input space more efficient. We do not have the computational resource (given the expense of the simulator) to adequately explore input space by direct sampling of the simulator, and so we use emulation to rule out regions where the model fails badly (i.e., its predictions are implausible). The emulator allows us to interpolate between parameter sets, enabling us extract more value from our simulated ensemble.

These problems (computational cost, unknown likelihood) make fully Bayesian approaches difficult for parameter estimation for climate models. However, it is possible to use an approach that approximates a Bayesian calibration. For example, Steinacher2013 took an approach that approximates a Bayesian probabilistic approach by generating an ensemble of 5,000 simulations with a 19-parameter latin hypercube design. The 5,000 parameter sets were then probability-weighted on the basis of these simulations using 26 observational constraints, and this weighting was subsequently applied to future emissions scenarios to get a distribution over future climate. They constructed a pseudo-likelihood relating simulator output to data, which averages across and weights the importance of each different data source, many of which were time series or spatial fields, using a nested structure. Whilst this has the potential to extract more information from the data than history matching can, creating an ad hoc likelihood in this way is potentially prone to error and makes it hard to keep track of the multitude of assumptions needed to form the pseudo-likelihood function. Likelihood based inference is notoriously sensitive to mis-specification, and so small changes in this complex likelihood could potentially lead to large changes in the conclusions. Moreover, it is difficult to understand the consequences of any given pseudo-likelihood, making it hard to judge scientifically any single choice. It should also be noted that even in our conservative history matching approach (which only fit the models to 8 data summaries rather than to multiple spatial fields), randomly selected parameter sets rarely satisfied the history matching constraints. In our calibration, the constraints ruled out more than 99.99% of the input space as implausible. This suggests that the Steinacher2013 approach of randomly sampling 5,000 points across the input space would have been insufficient to find the best regions of parameter space, though we note that weighting would serve to favour the best amongst these.

In total, 1140 spin-up simulations (2000 years each) were performed with the geared model and 345 transient simulations (300 years each) with the standard model, representing approximately 15 CPU years of computing.

Decarbonisation policies to meet 1.5°C and 2°C

The E3ME-FTT-GENIE modelling framework and the particular policy scenarios used here have been described in detail in elsewhere(Mercure2018a,Mercure2018b,Pollitt2018), below we give a summary of the policy choices taken as inputs to the modelling framework in deriving the emissions scenarios used here as input to PLASIM-GENIE. Three scenarios are used: a current-policy baseline, a scenario in which there is an 75% chance of limiting peak warming to 2°C and a scenario in which there is a 50% chance of limiting peak warming to 1.5°C.

The model baseline is consistent with the IEA's 'Current Policies' scenario (IEA, 2015). The baseline can broadly be considered as a continuation of current trends; existing policy remains in place and has some lagged effects that continue into the projection period, but there is no additional policy stimulus. Most policy instruments in the baseline are implicitly accounted for through the data itself (e.g. diffusion trends).

The 1.5°C and 2°C scenarios are designed as sets of policies that are added to the baseline case. In almost all countries, these policies encapsulate the measures put forward in the INDCs that were submitted to the Paris COP and complement them with other measures in order to scale up the level of ambition of decarbonisation. The scenarios are designed from a 'bottom-up' perspective. Essentially, policies are added across the full range of economic sectors sequentially until the targets are met. The 1.5°C scenario includes all the measures in the 2°C scenario, plus additional ones, as described below.

Many of the policies are specific to particular sectors, but two economy-wide policies are implemented:

- The first measure is an economy-wide programme of energy efficiency. Our 2°C scenario assumes that the programmes are in line with the IEA's analysis (IEA, 2014c) for a 450ppm scenario (excluding houses, which are treated separately, see below). They are further scaled up 25% for the 1.5°C scenario.
- The second measure is a carbon tax that is applied equally across the world. The carbon tax rates rise to \$310.2/tCO2 and \$96.4/tCO2 by 2030 in the 1.5°C and 2°C scenarios respectively, and \$886.3/tCO2 and \$274.8/tCO2 by 2050. The carbon taxes are applied to all industrial sectors, but not to road transport nor households, where separate rates are levied (since these sectors are likely to, or already have, their own specific carbon or energy tax rates).

Building on Mercure et al (2016), the following power sector policies were added to both scenarios:

- Feed-in-Tariffs - 100% of the difference between the levelised cost for wind and solar and a fixed value of \$80/MWh is paid by the grid to promote renewable uptake.
- Direct renewables subsidies – in most cases 50-60%, to provide an incentive to increase uptake, across a range of technologies (this is in addition to feed-in-tariffs). The subsidies gradually decrease over time and are phased out by 2050.
- In several countries there are immediate mandates to prevent the construction of new coal capacity.

In addition, it is assumed that electricity storage technologies advance up to 2050 such that the requirement for back-up flexible generation capacity (e.g. oil and gas peaking plants) is limited.

Combinations of policies are used to incentivise the adoption of vehicles with lower emissions (Mercure et al 2018b) in both scenarios. The list includes:

- fuel efficiency regulations of new liquid fuel vehicles
- a phase out of older models with lower efficiency
- kick-start programmes for electric vehicles where they are not available (by public authorities or private institutions, e.g. municipality vehicles and taxis)
- a tax of \$150/gCO2/km (2015 prices), to incentivise vehicle choice
- a fuel tax (increasing from \$0.10 in 2018 to \$1.00 per litre of fuel in 2050, 2015 prices) to curb the total amount of driving
- increasing/introducing biofuel mandates between current values to between 10% and 30% (40% in Brazil) in 2050, different for every country, extrapolating IEA projections (IEA 2014b) for the 2°C scenario, and to 97% in the 1.5°C scenario

Aviation is assumed to switch to biofuels gradually over the period 2020-2050 (faster in the 1.5°C scenario), but total bioenergy consumption remains within 150 EJ/yr.

The following policies were applied to homes in both scenarios:

- taxes on the residential use of fossil fuels, applied in Annex I and OPEC countries: starting at an equivalent of \$110/tCO2 (2015 values) and linearly increasing to \$240/tCO2 in 2030, constant at 2030 levels afterwards

- direct capital subsidies on renewable heating systems, applied globally: -40% on the purchase and installation of heat pumps, solar thermal systems and modern biomass boilers, phased out between 2030 and 2050
- kick-start programmes for renewable heating systems where they are not available, for a limited time period of five years (e.g. installations in publicly owned housing stock)

In some industrial sectors in East and South East Asia, a further mandate was added to electrify sectors that are currently dependent on coal (only in the 1.5°C scenario). Emissions from industrial processes are modelled as fixed in relation to real production levels from the relevant sector. In the baseline scenario, no efficiency improvements are assumed. In the 2°C and 1.5°C scenarios it is assumed that the production efficiency of process emissions improves by 3% a year over the projection period. Land-use change emissions are calculated in GENIE, with LUC assumed to follow RCP2.6 in the mitigation scenarios and RCP8.5 in the current policy baseline.

Methods References

Craig, P.S., Goldstein, M., Seheult A.H. & Smith, J.A. Pressure matching for hydrocarbon reservoirs: a case study in the use of Bayes linear strategies for large computer experiments in *Case Studies in Bayesian Statistics. Lecture Notes in Statistics* **121** Springer, New York, NY (1997)

Edwards, N.R., Cameron, D. & Rougier, J. Precalibrating an intermediate complexity climate model. *Climate dynamics* **37**, 1469–1482 (2011)

IEA (2014b) ‘World Energy Outlook’, 2014 edition, OECD/IEA, Paris.

IEA (2014c) ‘World Energy Investment Outlook’, OECD/IEA, Paris.

IEA (2015) ‘World Energy Outlook’, 2015 edition, OECD/IEA, Paris.

Holden, P.B., Edwards, N.R., Oliver, K.I.C., T. Lenton, T. M. & Wilkinson, R.D. A probabilistic calibration of climate sensitivity and terrestrial carbon change in GENIE-1. *Climate Dynamics* **35**, 785–806 (2010)

Jackson, R.B. et al. Reaching peak emissions *Nature Climate Change* **6**, 7-10 (2016)

Jain, A. K., Meiyappan, P., Song, Y., & House, J. I. CO_2 Emissions from land-use change affected more by nitrogen cycle, than by the choice of land cover Data *Global Change Biology* **19**, 2893-2906 (2013)

Joos, F. et al Carbon dioxide and climate impulse response functions for the computation of greenhouse gas metrics: a multi-model analysis *Atmos. Chem. Phys.* **13**, 2793-2825 (2013)

Fraedrich, K. A suite of user-friendly climate models: Hysteresis experiments *Eur. Phys. J. Plus* **127**, 53 (2012)

Lenton, T. M. et al. Millennial timescale carbon cycle and climate change in an efficient Earth system model *Climate Dynamics* **26**, 687–711 (2006)

Marin, J.-M., Pudlo, P., Robert, C.P. & Ryder, R.J. Approximate Bayesian Computational Methods. *Statistics and Computing* **22**, 1167–1180 (2012)

Mercure, J. F., Pollitt, H., Bassi, A. M., Viñuales, J. E. & Edwards, N. R. Modelling complex systems of heterogeneous agents to better design sustainability transitions policy. *Global Environmental Change* **37,** 102-115 (2016).

Mercure, J.F., Lam, A., Billington, S. & Pollitt, H. Integrated assessment modelling as a positive science: modelling policy impacts in road transport to meet a climate target well below 2°C, pre-print at arXiv:1702.04133. (2017)

O'Hagan, A. Bayesian Analysis of Computer Code Outputs: A Tutorial *Reliability Engineering & System Safety* **91**, 1290–1300 (2006)

Ramankutty, M. et al. Challenges to estimating carbon emissions from tropical deforestation *Global Change Biology* **13**, 51-66 (2007)

Rasmussen, C.E. *Gaussian processes in machine learning*, in *Advanced Lectures on Machine Learning*, O. Bousquet, U. von Luxburg, and G. Rätsch (Eds.) Springer, Berlin, Heidelberg, New York (2004)

Sacks, J., Welch, W.J., Mitchell, T.J. and Wynn, H.P. Design and analysis of computer experiments *Statistical Science* **4**, 409–23 (1989)

Tibshirani, R. Regression Shrinkage and Selection via the Lasso *Journal of the Royal Statistical Society. Series B (Methodological)*. **58**, 267–88 (1996)

Williamson, D. B., Blaker, A.T. & Sinha, B. Tuning without over-tuning: parametric uncertainty quantification for the NEMO ocean model *Geoscientific Model Development* **10**, 1789-1816 (2017)

Zickfeld, K. et al. Long-term climate change commitment and reversibility: An EMIC intercomparison *J. Climate* **26**, 5782– 5809 (2013)

Supplementary information

	CO_2	AMOC	Ocean Dissolved inorganic carbon	Vegetation and soil carbon	Land surface albedo	Ocean heat below 39m
1P5C scenario						
Warming	47%	0%	21%	24%	4%	59%
CO_2		1%	61%	67%	0%	34%
CP scenario						
Warming	32%	17%	12%	15%	7%	69%
CO_2		8%	61%	80%	5%	36%

Table S1: R2 Coefficient of determination between selected ensemble output metrics, all expressed as peak future change relative to a 2006-2015 baseline.

	1P5C scenario		CP scenario	
	Maximum expected change	Maximum variability	Maximum expected change	Maximum variability
DJF SAT (°C)	3.8	3.7	3.7	1.2
JJA SAT (°C)	2.8	2.9	3.0	1.1
DJF pptn (mm/day)	0.8	2.1	1.3	0.8
JJA pptn (mm/day)	2.5	2.4	3.5	1.3
Surface pH (pH units)	-0.12	0.06	-0.15	0.04
Marine productivity (%)	-14	41	-14	12

Table S2: Maximum change per 1°C warming (c.f. Figs 2A, 3A, 4A, 4D) and maximum variability per 1°C warming (c.f. Figs 2C, 3C, 4C, 4F) for 1P5K and CP scenarios.

Module	Parameter	Description	Units	Min	Max	Prior
PLASIM	TDISSD	Horizontal diffusivity of divergence	days	0.01	10	LOG
	TDISSZ	Horizontal diffusivity of vorticity	days	0.01	10	LOG
	TDISST	Horizontal diffusivity of temperature	days	0.01	10	LOG
	TDISSQ	Horizontal diffusivity of moisture	days	0.01	10	LOG
	VDIFF	Vertical diffusivity	m	10	1000	LOG
	TWSR1	Short wave clouds (visible)		0.01	0.5	LOG
	TWSR2	Short wave clouds (infrared)		0.01	0.5	LOG
	ACLLWR	Long wave clouds	$m^{-2}g^{-1}$	0.01	5	LOG
	TH2OC	Long wave water vapour		0.01	0.1	LOG
	RCRITMIN	Minimum relative critical humidity		0.7	1.0	LIN
	GAMMA	Evaporation of precipitation		0.001	0.05	LOG
	ALBSM	Equator-pole ocean albedo difference		0.2	0.6	LIN
	ALBIS	Ice sheet albedo		0.8	0.9	LIN[1]
	APM	Atlantic-Pacific moisture flux adjustment	Sv	0.0	0.32	LIN[2]
GOLDSTEIN	OHD	Isopycnal diffusivity	m^2s^{-1}	500	5000	LOG
	OVD	Reference diapycnal diffusivity	m^2s^{-1}	2e-5	2e-4	LOG
	ODC	Inverse ocean drag	days	1	3	LIN
	SCF	Wind stress scaling		2	4	LIN
	OP1	Power law for diapycnal diffusivity profile		0.5	1.5	LIN
BIOGEM	PMX	Maximum PO_4 uptake	mol kg^{-1} yr^{-1}	5e-7	5e-5	LOG
	PHS	PO_4 half-saturation concentration	mol kg^{-1}	5e-8	5e-6	LOG
	PRP	Initial proportion POC export as recalcitrant fraction		0.01	0.1	LIN
	PRD	e-folding remineralisation depth of non-recalcitrant POC	m	100	1000	LIN
	PRC	Initial proportion $CaCO_3$ export as recalcitrant fraction		0.1	1.0	LIN
	CRD	e-folding remineralisation depth of non-recalcitrant $CaCO_3$	m	300	3000	LIN
	RRS	Rain ratio scalar		0.01	0.1	LIN
	TCP	Thermodynamic calcification rate power		0.2	2.0	LIN
	ASG	Air-sea gas exchange parameter		0.3	0.5	LIN
ENTS	VFC	Fractional vegetation dependence on carbon density	m^2 kgC^{-1}	0.1	1.0	LIN
	VBP	Base rate of photosynthesis	kgC m^{-2} yr^{-1}	3.0	7.0	LIN
	LLR	Leaf litter rate	yr^{-1}	0.075	0.26	LIN
	SRT	Soil respiration temperature dependence	K	197	241	LIN
	VPC	CO2 fertilization Michaelis-Menton half-saturation	ppm	29	725	LOG[3]

Table S3: Prior distributions for PLASIM-GENIE varied parameters (uniform between ranges in log/linear space as stated). Notes. 1) ALBIS ice sheet albedo was fixed at 0.8 in the final ensemble. 2) APM was fixed at zero in the final ensemble (no flux correction). 3) VPC was not constrained by the emulator filtering as this parameter has no effect in the preindustrial spin up state. The final calibration step, selecting 69 simulations that satisfy present-day plausibility after the historical transient was primarily an exercise to calibrate the VPC parameter. Prior distributions are discussed and derived from Holden et al (2010, 2013a, 2013b, 2014 and 2016).

	Observations	Acceptance ranges
Emulated history match filters		
Global average surface air temperature (°C)	~14 Jones et al (1990)	11 to 17
Global vegetation carbon (GtC)	450 to 650 Bondeau et al (2007)	300 to 800
Global soil carbon (GtC)	850 to 2400 Bondeau et al (2007)	750 to 2500
Maximum Atlantic Overturning (Sv)	~19 Kanzow et al (2010)	10 to 30
Maximum Pacific Overturning (Sv)		<15 (see note)
Atlantic Circumpolar Current (Sv)	140 ± 6 Ganachaud and Wunsch (2000)	>50 (see note)
Global ocean averaged dissolved O_2 (μmol kg^{-1})	~170 Konkright et al (2002)	130 to 210
Global deep ocean $CaCO_3$ flux (GT $CaCO_3$-C yr^{-1})	~0.4 (Feely et at (2004)	0.2 to 0.8
Transient simulation history match filters		
(1866-1875) to (1996-2005) warming (°C)	~0.78 IPCC 2013 SPM	0.6 to 1.0
Atmospheric CO_2 in 2005 (ppm)	378 Keeling et al (2005)	353 to 403

Table S4: History-matching (Approximate Bayesian Computation) acceptance ranges. Acceptable simulation ranges are broadened relative to observational error, thereby acknowledging model error and avoiding over-tuning. Note: tests to minimise PMOC and maximise ACC were applied to the emulator filtering in order to favour strong ACC and minimal north Pacific intermediate water formation.

Supplementary References

Bondeau, A., et al. Modelling the role of agriculture for the 20th century global terrestrial carbon balance. *Glob. Change Biol.* **13**, 679–706 (2007).

Feely, R.A. et al. Impact of anthropogenic CO_2 on the $CaCO_3$ system in the oceans *Science* **305**, 362-366 (2005)

Ganachaud, A. & Wunsch C. Improved estimates of global ocean circulation, heat transport and mixing from hydrographic data *Nature* **408**, 453–457 (2000).

Holden, P.B, Edwards, N.R., Gerten, D. and Schaphoff S. A model-based constraint on CO_2 fertilisation, Biogeosci., 10, 339-355, (2013a)

Holden, P.B., Edwards, N.R., Mueller, S.A., Oliver, K.I.C., Death, R.M. and Ridgwell, A. Controls on the spatial distribution of ocean $d^{13}C_{DIC}$, Biogeosciences, 10, 1815-1833 (2013b).

Holden, P.B., 2014, et al, PLASIM-ENTSem v1.0: a spatio-temporal emulator of future climate change for impacts assessment, *Geosci. Mod. Dev.*, 7, 433-451, (2014).

IPCC, 2013: Summary for Policymakers in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Jones, P. D., New, M., Parker, D. E., Martin, S., & Rigor, I. G. Surface air temperature and its changes over the past 150 years. *Rev. Geophys.* **37**, 173–199 (1999)

Kanzow, T. et al. Seasonal Variability of the Atlantic Meridional Overturning Circulation at 26.58◦N, *J. Clim.* **23**, 5678–5698 (2010)

Keeling, C.D. et al. Atmospheric CO_2 and $^{13}CO_2$ exchange with the terrestrial biosphere and oceans from 1978 to 2000: observations and carbon cycle implications in *A History of Atmospheric CO_2 and its effects on Plants, Animals, and Ecosystems* editors, Ehleringer, J.R., T. E. Cerling, M. D. Dearing, Springer Verlag, New York (2005)

Konkright, M. E et al. World Ocean Atlas 2001: Objective Analysis, Data, Statistics and Figures, CD- ROM Documentation, National Oceanographic Data Center, Silver Spring, MD, (2002)