ENVIRONMENTAL RESEARCH
LETTERS

PURPOSE-LED PUBLISHING™

Part of the
ENVIRONMENTAL RESEARCH
SERIES

LETTER • OPEN ACCESS

Emulating long-term CMIP6 projections of sterodynamic sea-level change using a three-layer energy balance model

To cite this article: Víctor Malagón-Santos *et al* 2025 *Environ. Res. Lett.* **20** 084034

ENVIRONMENTAL RESEARCH
LETTERS

OPEN ACCESS

RECEIVED
28 February 2025

REVISED
25 June 2025

ACCEPTED FOR PUBLICATION
27 June 2025

PUBLISHED
8 July 2025

LETTER

Emulating long-term CMIP6 projections of sterodynamic sea-level change using a three-layer energy balance model

Víctor Malagón-Santos[1,*], Chris Smith[2,3], Hege-Beate Fredriksen[4], Tim H J Hermans[5], Tamsin L Edwards[6] and Aimée B A Slangen[1,7]

[1] Department of Estuarine & Delta Systems, NIOZ Royal Netherlands Institute for Sea Research, PO Box 140, 4400 AC, Yerseke, The Netherlands
[2] Department of Water and Climate, Vrije Universiteit Brussel, 1050 Brussels, Belgium
[3] Energy, Climate and Environment Program, International Institute for Applied Systems Analysis, 2361 Laxenburg, Austria
[4] Norwegian Polar Institute, Tromsø, Norway
[5] Institute for Marine and Atmospheric Research Utrecht, Utrecht University, Utrecht, The Netherlands
[6] Department of Geography, Kings College, London, United Kingdom
[7] Department of Physical Geography, Faculty of Geosciences, Utrecht University, Utrecht, The Netherlands
[*] Author to whom any correspondence should be addressed.

E-mail: victor.malagonsantos@gmail.com

Keywords: emulation, energy balance model, simple climate model, thermal expansion, ocean dynamic sea level

Supplementary material for this article is available online

Abstract

Multi-century projections of sea-level change are crucial for understanding long-term climate impacts. However, projecting ocean dynamic processes affecting sea-level change faces two main challenges: (1) the ocean's thermal inertia and dynamics can lead to substantial nonlinearities both on a global (thermal expansion) and regional scale (dynamic sea level); and (2) simulating ocean dynamic sea-level change over multiple centuries with global climate models is computationally intensive. To address these challenges, we use an energy balance model (EBM) to emulate the thermal responses of models participating in the Coupled Model Intercomparison Project 6 (CMIP6) and evaluate its optimal layer configuration for reproducing both thermal expansion and dynamic sea-level change. We compare results between a two and a three-layer EBM configuration, as fitting more than three layers can lead to EBM parameter overfitting. We find that both configurations perform similarly when emulating thermal expansion, with performance being highly dependent on accurate EBM forcing. To emulate dynamic sea level, we couple the EBM to a multivariate pattern scaling approach that relates the response of layer temperatures to regional changes in dynamic sea level. The latter demonstrates clear advantages of a three-layer configuration under high-emission scenarios, leading to an 18% reduction in emulator error at 2300 while capturing nonlinearities more effectively. Coupling the regional emulator with a simple climate model to propagate climate uncertainties further highlights the advantage of using a three-layer approach, leading to more stable parameter fitting and reducing uncertainty in probabilistic projection by up to one global mean standard deviation. Our findings suggest that multi-layer EBMs can more accurately mimic long-term (up to 2300) CMIP6 projections of sterodynamic sea-level changes while limiting computational burden and reducing uncertainty in emulators due to statistical fitting.

1. Introduction

Due to climate system inertia (Levermann *et al* 2013) and potential instabilities in the Greenland and Antarctic ice sheets (Lenton *et al* 2019), global-mean sea level is projected to continue rising for centuries, even if temperatures stabilize (Fox-Kemper *et al* 2021). Multi-century projections of sea-level change (SLC) are not only essential to understanding long-term effects such as sea-level commitment, but can also help guide risk-averse coastal practitioners and policy makers (Turner *et al* 2023, Palmer and Weeks 2024). Process-based multi-century SLC projections, however, are challenging due to computational constraints and limited understanding of processes affecting SLC over those timescales. On a global scale, major challenges arise from limited understanding of ice-sheet instability mechanisms (Edwards *et al* 2019, Pattyn and Morlighem 2020) and model spread in the distribution of ocean heat content in the deep ocean (Cheng *et al* 2022). On a regional scale, those challenges not only come from translating ice loss into a regional pattern driven by gravitational, rotational, and deformational effects (Farrell and Clark 1976, Mitrovica *et al* 2001, Slangen *et al* 2012), but also from understanding non-linear processes driving vertical land motion (e.g. Oelsmann *et al* 2024) and changes in ocean water density (Church *et al* 2010, Johnson and Wijffels 2011) and circulation patterns (Couldrey *et al* 2021). The latter is known as sterodynamic SLC (Gregory *et al* 2019) and, according to simulations from General Circulation (or related) Models (GCMs), undergoes substantial non-linear changes over multi-century timescales (Yuan and Kopp 2021). The present study focuses on modeling long-term, multi-century (up to 2300) sterodynamic SLC as an essential ingredient for developing multi-century regional SLC projections.

GCMs are the primary source of data for projecting sterodynamic SLC, but the high computational demands of these models limit the number of simulations extending beyond the 21st century. The relatively few long-term sterodynamic SLC simulations available from the Coupled Model Intercomparison Project (CMIP) provide some understanding of multi-century non-linear processes in the ocean, but the robust estimation of structural uncertainty is challenging due the limited number of projections. Emulation—a technique that reproduces complex models using physics-based, statistical, or machine-learning methods—can alleviate these computational limitations, enabling efficient long-term projections across a broader range of scenarios and supporting uncertainty assessment in distinct sea-level contributions (e.g. Nauels *et al* 2017, Palmer *et al* 2018, Thomas and Lin 2018, Bulthuis *et al* 2019, Berdahl *et al* 2021, Edwards *et al* 2021, Fox-Kemper *et al* 2021, 2020, Yuan and Kopp 2021, 1). For instance, temperature uncertainties arising from key parameters controlling the climate system, such as equilibrium climate sensitivity and transient climate response, can be estimated using reduced-complexity models such as the Finite-amplitude Impulse Response (FaIR) model (Smith *et al* 2018, Leach *et al* 2021). By using FaIR probabilistic temperatures as SLC emulator input, temperature uncertainties can be incorporated in SLC projections (e.g. Fox-Kemper *et al* 2021, Yuan and Kopp 2021). Designing emulators of distinct sea-level contributors which use the same inputs facilitates consistent projections across different SLC components (e.g. ice sheet and glacier contributions as in Edwards *et al* (2021)), aligning them with common temperature pathways and enabling their integration into comprehensive SLC projections (Fox-Kemper *et al* 2021, Kopp *et al* 2023).

In this study, we build on the work by Yuan and Kopp (2021), who used an energy balance model (EBM) coupled to a pattern scaling approach to emulate, respectively, global-mean thermosteric sea-level rise (GMTSLR) and ocean dynamic sea level change (hereinafter, ODSLC), combinedly known as sterodynamic SLC (2019). The EBM they used was a two-layer model (2 LM) that mimics the thermal response of GCMs and has been extensively used in literature to emulate both global surface air temperature (GSAT) (e.g. Geoffroy *et al* 2013a, 2013b, Nicholls *et al* 2020, Jackson *et al* 2022) and GMTSLR (Palmer *et al* 2018, Fox-Kemper *et al* 2021). Yuan and Kopp (2021) demonstrated that by using GSAT and deep ocean temperature output from the 2 LM with a bivariate pattern scaling regression approach, non-linearities in ODSLC are better explained than with a univariate approach with GSAT as an explanatory variable alone. Other studies, however, suggested a three-layer model (3 LM) may be better suited to emulate the thermal response of GCMs (e.g. Cummins *et al* 2020), although this has only been demonstrated for GSAT. Here, we explore whether considering three layers in the EBM leads to better emulation of both GMTSLR and ODSLC compared to a two-layer configuration, particularly on multi-century projections, where non-linearities are more pronounced. We aim to produce emulators that are traceable to the latest generation of CMIP models (CMIP6, Eyring *et al* 2016) and optimize the emulator design to enable integration into a framework for assessing structural uncertainty in SLC projections (Kopp *et al* 2023) in combination with other SLC emulation methods.

We address three key objectives: (1) fit and compare the two- and three-layer EBM configurations against multi-century CMIP6 simulations of GMTSLR; (2) evaluate these two EBM configurations to project ODSLC via regional emulation using a multivariate pattern scaling approach; and (3) propagate temperature uncertainties through the regional emulator to determine the impact of different EBM layer configurations on probabilistic ODSLC projections.

Table 1. CMIP6 models, runs, and scenario availability along with end year of simulation for GMTSLR and ODSLC simulations running beyond 2100. Only models whose EBM parameters could be robustly calibrated using the 4xCO_2 experiment are considered. All model data is available from 2015 onwards.

Model	References	Run	SSP1-2.6	SSP2-4.5	SSP5-3.4-over	SSP5-8.5
ACCESS-CM2	Bi *et al* (2020)	r1i1p1f1	2300			2300
ACCESS-ESM1-5	Ziehn *et al* (2020)	r1i1p1f1	2300			2300
CanESM5	Swart *et al* (2019)	r1i1p1f1	2300		2300	2180
GISS-E2-1-G	Kelley *et al* (2020)	r1i1p3f1	2500	2500		2500
IPSL-CM6A-LR	Boucher *et al* (2020)	r1i1p1f1	2300		2300	2300
MRI-ESM2-0	Yukimoto *et al* (2019)	r1i1p1f1	2300		2300	2300

2. Data and methods

2.1. CMIP6 climate model data

GMTSLR and ODSLC—or *zostoga* and *zos* in CMIP variable naming conventions, respectively—are typically assessed individually by GCMs. Since the focus of this study is on multi-centennial simulations of GMTSLR and ODSLC, we only select model simulations from the latest generation of CMIP models (CMIP6) (Earth System Grid Federation 2025) that extend beyond 2100, and model runs that feature an abrupt-4xCO_2 experiment to fit EBM parameters to (see section 2.2). For every model, we consider both the historical and scenario simulations, usually starting at 1850 and 2015, respectively. Shared Socioeconomic Pathways (SSPs) scenarios are a core part of CMIP6 experiments, designed to explore how global society, demographics, and economics might evolve in the future (O'Neill *et al* 2014, Meinshausen *et al* 2020). SSP scenario availability and simulation length vary between the selected models, with SSP1-2.6 and SSP5-8.5 as scenarios with the highest availability (table 1), representing a sustainability-focused and a fossil-fueled development scenario, respectively.

Both GMTSLR and ODSLC can be subject to model drift because GCMs are typically spun up for only a few centuries while the deep ocean can take millennia to reach an equilibrium (Gupta *et al* 2013). Therefore, we correct for model drift by (i) subtracting the smoothed long-term linear change estimated from the entire available simulation of the pre-industrial control run, (ii) re-gridding all simulations to a common 1° by 1° grid, and (iii) removing the area-weighted mean at each time step to ensure de-drifted fields of ODSLC time series have a global mean of zero (Slangen *et al* 2012, Hermans *et al* 2021). As regional sea level is affected by sea-level pressure anomalies, referred to as the inverted barometer effect, we add this by using the sea-level pressure from each model simulation (*psl* in CMIP convention), following Stammer and Hüttemann (2008).

We emulate only the long-term response of sterodynamic SLC to increased radiative forcing and associated temperature change, and not the variability driven by internal climatic oscillations. Hence, we focus on annual data and compute yearly means for both GMTSLR and ODSLC from the monthly data provided by the CMIP6 models. There are other forms of internal climate variability that can exert an influence on interannual (Hamlington *et al* 2020), decadal (Frederikse *et al* 2016), and interdecadal (Lyu *et al* 2016) timescales, and this is especially relevant to ODSLC (Nandini-Weiss *et al* 2024). To further reduce the effect of internal climate variability on longer-than-annual timescales, we filter the ODSLC simulations using low-frequency component analysis, applying a low-pass filter of 30 years (Wills *et al* 2020, Malagón-Santos *et al* 2023).

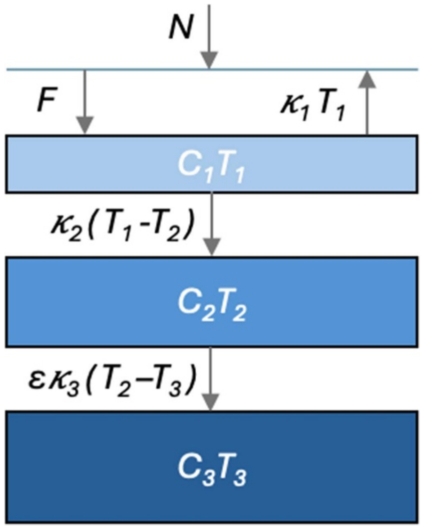

Figure 1. Layout of an EBM composed of three layers, each of those with different heat capacities (C) represented by their thickness. The upper layer represents both the atmosphere and ocean surface, while the other two layers represent the deep ocean. The blue horizontal line depicts the top of the atmosphere with no heat capacity, and arrows represent heat transfer between adjacent layers governed by heat transfer coefficients (κ).

2.2. The energy balance model

We use a physics-based emulator, consisting of an EBM (Geoffroy *et al* 2013a, 2013b) with either two or three layers to represent the thermal response of both the atmosphere and the ocean (see figure 1 for the 3 LM example). The emulator layers do not mix, but adjacent layers exchange heat, regulated by their temperature difference and heat transfer coefficients (κ). Each layer (i) has a temperature (T_i) and heat capacity (C_i) that is related to its thickness and is climate-model dependent. The response of the EBM to top-of-the-atmosphere (TOA) net downward radiative flux (N) is governed by a set of differential equations, which are as follows for a three-layer EBM configuration:

$$C_1 \frac{dT_1}{dt} = F(t) - k_1 T_1 - k_2 (T_1 - T_2) \quad (1)$$

$$C_2 \frac{dT_2}{dt} = k_2 (T_1 - T_2) - \varepsilon k_3 (T_2 - T_3) \quad (2)$$

$$C_3 \frac{dT_3}{dt} = k_3 (T_2 - T_3) \quad (3)$$

where F is radiative forcing which the depends on time (t) represented in years. The parameter κ_1 is known as the climate feedback parameter and often expressed as λ (Fredriksen and Rypdal 2017, Cummins *et al* 2020). The strength of the feedback parameter varies geographically, leading to the pattern of surface temperature change caused by ocean heat uptake affecting the radiative imbalance in a transient regime (Geoffroy *et al* 2013a). This variation in effective strength of κ_1 is represented by the efficacy factor ε (Held *et al* 2010). The TOA energy imbalance (N) is given in equation (4), whose integral represent the Earth's energy uptake

$$N(t) = F(t) - k_1 T_1(t) + (1 - \varepsilon) k_3 (T_2(t) - T_3(t)). \quad (4)$$

EBM is usually designed to emulate GSAT, calibrating parameters to match this variable (e.g. Geoffroy *et al* 2013b, Cummins *et al* 2020, Smith *et al* 2024). GSAT represents temperature from the first layer, while GMTSLR depends on all layer temperatures. Hence, Emulating GMTSLR requires two additional steps, including (1) estimating heat content (OHC) from layer temperatures and heat capacity (figure S1), and (2) multiplying this OHC by an expansion efficiency coefficient (σ) following equation (5), adapted from Kuhlbrodt and Gregory (2012):

$$\mathrm{GMTSLR} = \mathrm{OHC} * \sigma. \quad (5)$$

We estimate σ for each CMIP6 model (table S1) by minimizing the error between emulated GMTSLR in equation (5) against the actual GMTSLR simulated by the respective model for all scenarios available, using the Sequential Least Squares Programming optimization algorithm from SciPy version 1.11.2 (Virtanen *et al* 2020).

The EBM is a time-invariant system characterized by its step response. Hence, the most common method for calibrating EBM parameters consists of using a single realization of the abrupt-4xCO2 experiment, since this experiment is designed to elicit the step response in CMIP models (Good *et al* 2011). Although the EBM can theoretically have any number of layers to represent the climate system (Cummins *et al* 2020), the fitting procedure involving the abrupt-4xCO2 experiment, whose simulation length typically is 150 years, makes fitting more layers impractical. Fitting more than three layers would increase the risk of overfitting and is uninformative in trying to fit a response time beyond the characteristic timescale of the deep ocean, which is a few hundred years.

EBM parameters can be fitted to match emulated GSAT (i.e. T_1) to that of the corresponding CMIP6 model by using a goodness-of-fit method (Cummins *et al* 2020). Fitted parameters are then used to make inferences for other scenarios by inputting the appropriate forcing. EBM calibrated parameters introduced in this paper, and the F_{4xCO2} parameter which represents ERF after quadrupling preindustrial atmospheric CO_2, can be found in table S1 for both 2 LM and 3 LM configurations and the models considered in our analysis.

CMIP6 models have different radiative transfer parameterizations and transport schemes, thus their effective radiative forcing (ERF) can substantially differ between models for the same climate scenario (Smith *et al* 2020, 2021). Calculating ERFs from every model run was not part of CMIP6 design, though a subset of models produced results from the Radiative Forcing Model Intercomparison Project (Pincus *et al* 2016). Here we use model- and scenario-specific ERF estimates from Fredriksen *et al* (2023), where they used an EBM with a time-dependent climate feedback parameter (Forster *et al* 2013, Fredriksen *et al* 2021) to estimate ERFs for CMIP6 models.

2.3. Regional emulation using a multivariate pattern scaling approach

We use a point-wise regression approach to emulate ODSLC simulations from CMIP6. This regional emulation approach is a multivariate pattern scaling method that relates local ODSLC to one or more variables representing a global change. To couple pattern scaling to the EBM, we match the number of explanatory variables (T_i) to the number of EBM layers, including either two or three temperatures coming from the 2 LM or the 3 LM (hereinafter referred to as two-layer or three-layer pattern scaling, respectively). Equation (6) presents the three-layer pattern scaling approach for ODSLC that links back to the 3 LM presented in figure 1:

$$\begin{aligned}\mathrm{ODSLC}(x, y, t_s) = {} & \alpha(x, y) T_1(t_s) + \beta(x, y) T_2(t_s) \\ & + \gamma(x, y) T_3(t_s) + \xi(x, y, t_s) \quad (6)\end{aligned}$$

where x and y are longitude and latitude, respectively. Time in years is represented by t, with the subscript s representing different scenarios that may be considered in the training step of the emulator. α, β, and γ are regression coefficients relating layer temperatures to ODSLC change, while ξ is model error mostly driven by internal climate variations. To produce probabilistic ensembles of ODSLC projections, we couple the two- and three-layer pattern scaling emulators to temperature output from FaIR (Smith *et al* 2018, Leach *et al* 2021).

FaIR is a simple climate model widely used to produce future temperature projections at a global scale. FaIR calculates time series of greenhouse gas concentrations and ERF from source emissions and uses an EBM to project temperature and ocean heat content change from this ERF. FaIR (including its EBM) is calibrated to the response of CMIP6 GCMs and constrained to reproduce observed historical global mean surface temperature, ocean heat content change and CO_2 concentration changes (Smith *et al* 2024), and IPCC assessments of equilibrium climate sensitivity, transient climate response and aerosol forcing (Forster *et al* 2021). Prior to coupling with pattern scaling, we adjust FaIR so its thermal core contains the same number of layers (either two or three) as used in the emulator presented here. Following Smith *et al* (2024), both 2 LM (Smith 2024a) and 3 LM (Smith 2024b) calibrations of FaIR are produced for this study.

3. Results

3.1. Evaluating global-mean thermal expansion emulation

We evaluate the performance of the 2 LM and 3 LM for emulating GMTSLR (section 2.2) against their model-derived counterparts (figure 2) across the two emissions scenarios that are most available in the set of selected models (SSP1-2.6 and SSP5-8.5, table 1). Except for GISS-E2-1-G where a different forcing variant used in the long-term projections to that used in the abrupt-4xCO_2 runs (Miller *et al* 2021) may cause a mismatch between the ERF and EBM, both approaches show close alignment with simulated GMTSLR and exhibit a similar performance across models and scenarios. This suggests that a 2 LM already provides a good estimate of GMTSLR and that adding an additional timescale to the EBM does not substantially improve its emulation. This finding, which contrasts with previous results found for GSAT (Cummins *et al* 2020, Jackson *et al* 2022), can be explained by GMTSLR being governed by OHC (equation (5)), which is controlled by global ocean temperature and is less dependent on how heat is being distributed over different layers.

Most models, namely ACCESS-CM2, ACCESS-ESM1-5, CanESM5, and IPSL-CM6A-LR (figures 2(a)–(f), (i) and (j)), exhibit a consistent performance across different scenarios. This suggests that performance is primarily governed by a

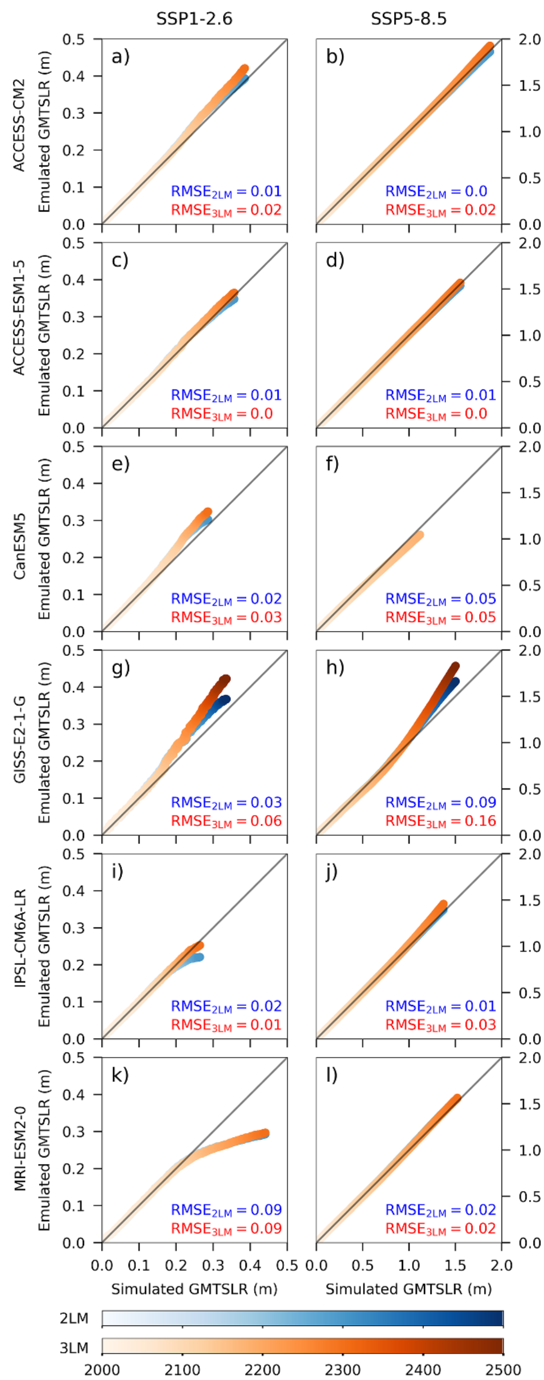

Figure 2. Simulated versus emulated GMTSLR (m) for (left column) SSP1-2.6 and (right column) SSP5-8.5. GMTSLR is referenced to the 1995–2014 average. RMSE is computed from 2100 onwards.

robust EBM parameterization, while the influence of the specific scenario is relatively limited. In contrast, Jackson *et al* (2022) found that ERF-dependent parameter calibration was necessary to achieve close GSAT emulation for both the historical and the future period. Bespoke parameter calibration for distinct ERF scenarios may be needed for close GSAT emulation as this is more sensitive to EBM configuration. In contrast, GMTSLR is less sensitive to such configurations, making it more robust for out-of-sample

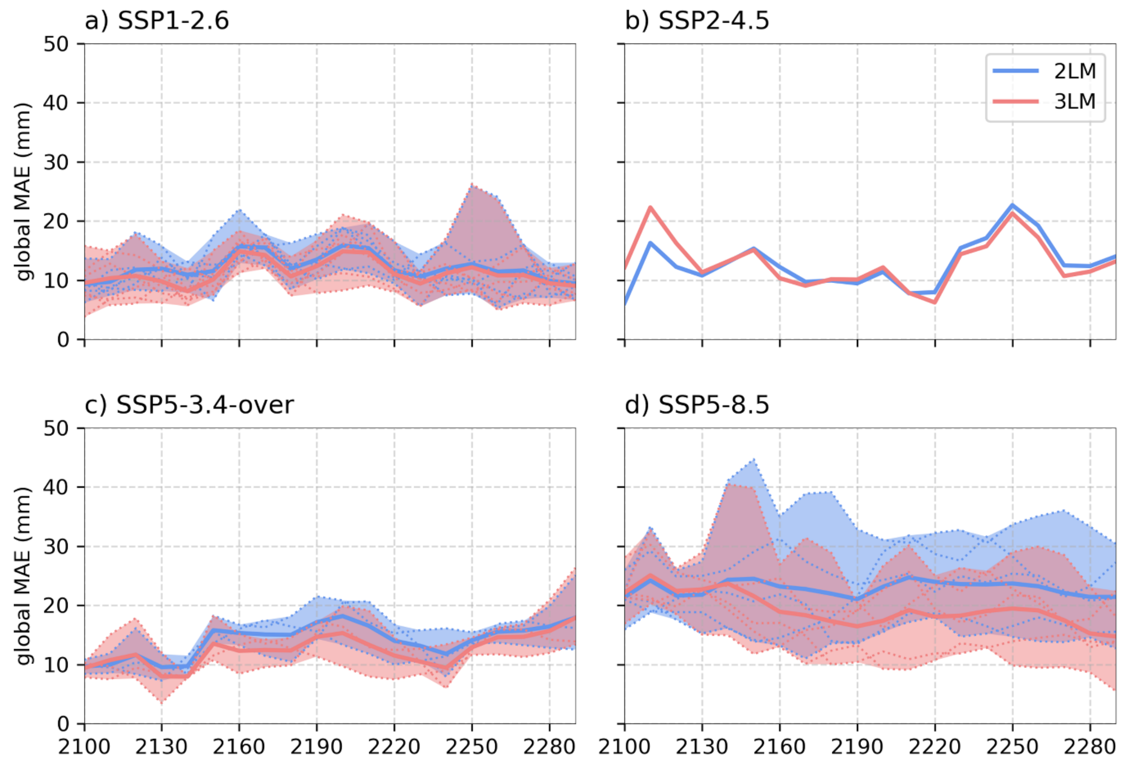

Figure 3. Global mean absolute error (mm) for 2 LM (blue) and 3 LM (red), using in-sample emulation to extend projections. Dotted lines represent individual simulations, shaded colors depict range between the lowest and highest error for a given approach across models. Only one model provides a simulation for SSP2-4.5 (table 1).

applications. However, exceptions exist (e.g. MRI-ESM2-0; figures 2(k) and (l)), illustrating that accurate out-of-sample emulation is not always guaranteed and may be influenced by scenario-dependent forcing and model-specific dynamics.

3.2. Emulating ocean dynamic sea level change

We compare the two- and three-layer pattern scaling approach to emulate ODSLC, using as predictors temperatures from either a 2 LM or a 3 LM. We consider two distinct emulator applications: in-sample (i.e. intra-scenario), used to extend ODSLC projections beyond their simulations period; and out-of-sample (inter-scenario) emulation which employs available SSP scenarios from models to emulate unavailable or new scenarios.

We evaluate in-sample emulation performance by taking, for a specific SSP, varying ODSLC simulation lengths for training and taking the 30 subsequent years from the simulations for validation. The global-weighted mean absolute error (MAE) shown in figure 3 shows varying degrees of performance between the two emulators. For instance, for SSP1-2.6, SSP2-4.5, and for SSP5-3.4-over (figures 3(a)–(c), respectively), both the 2 LM and 3 LM approaches show consistent MAE and minimal divergence between them. In contrast, for SSP5-8.5 (figure 3(d)), the MAE of the 3 LM is consistently lower than that of the 2 LM beyond 2150, amounting to relative difference of 18% at 2300. These findings suggest that a 3 LM pattern scaling method is more robust in capturing non-linearities associated with intensified ocean dynamic processes, while the difference between both approaches is negligible in low-forcing scenarios.

Due to varying sterodynamic conditions across the world's oceans, the magnitude of non-linearities in ODSLC as represented by GCMs varies spatially. In addition to thermosteric changes, ODSLC is affected by changes in salinity changes (halosteric effects) and shifts in ocean circulation patterns. Adding an additional temperature term in the emulator generally improves model performance, (figures S2(a) vs (b)), but both approaches reveal spatially varying errors which could be partly related to salinity and ocean circulation changes not well represented by the emulator (figure S2).

We compare different emulator performance (figure 4(a)) over long-term timescales (2270–2300) by computing their regional MAE differences (3 LM minus 2 LM). We focus on the SSP5-8.5 scenario for such comparison, as differences between both approaches are usually more significant compared to lower emission scenarios (figures 3 and S3). The 3 LM reduces MAE across most regions (blue-green shading), particularly in areas where the 2 LM shows the largest errors, such as the North Pacific (figure 4(b)) and North Atlantic Ocean (figure 4(c)) for most

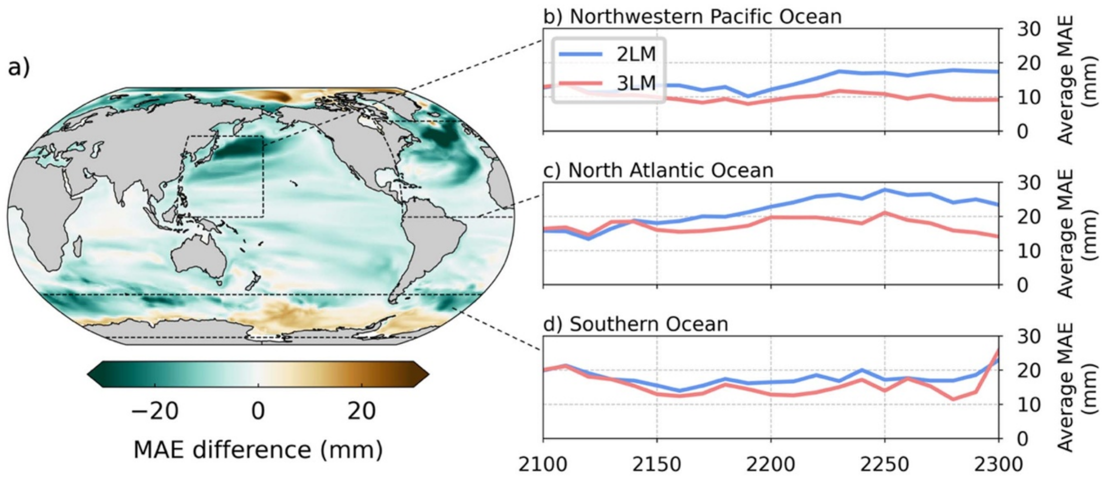

Figure 4. (a) Difference in the multi-model mean absolute error (MAE) between CMIP6 simulated and emulated ODSLC in 2270–2300 relative to 2015 when using a two-layer and a three-layer pattern scaling approach (3 LM-MAEminus 2 LM-MAE) for SSP5-8.5. (b)–(d) Time series of the average MAE for Northwestern Pacific, North Atlantic, and Southern Ocean, respectively.

of the 2100–2300 projection period. Regions with already small errors in the 2 LM, such as much of the equatorial and mid-latitude oceans, the improvements provided by the 3 LM are less pronounced. However, some areas are characterized by strong non-linear dynamics driven by ocean circulation shifts and changes in deep water formation, such as the Arctic and Antarctic, leading to abrupt changes over long-time scales that both approaches struggle to capture (e.g. figure 4(d) at 2300). This indicates that the 3 LM is generally better suited to emulate ODSLC patterns, but results over certain areas subject to significant dynamical shifts must be interpreted carefully.

We perform out-of-sample emulation by using the lowest and highest emissions scenarios available (SSP1-2.6 and SSP5-8.5), for a given model in the training step of the emulator, to replicate an intermediate scenario. Intermediate scenarios are available for CanESM5, GISS-E2-1-G, and IPSL-CM6A-LR, providing SSP5-4.3-over, SSP2-4.5, and SSP5-3.4-over, respectively (table 1). We find that the error difference between the two emulation approaches is minimal for out-of-sample ODSLC reproduction regardless of scenario. For instance, the global mean RMSE when emulating CanESM5 is 8 cm under SSP5-3.4-over, and 13 cm for IPSL-CM6A-LR, over the entire projection period from 2015 to 2300 for both the 2 LM and the 3 LM. Similarly, emulating SSP2-4.5 for GISS-E2-1-G leads to small differences, with both configurations leading to an RMSE of 6 cm over the same period. Out-of-sample emulation using a SSP5-8.5 to predict SSP1-2.6, and vice versa, leads to consistently higher RMSE across models but minimal difference between both approaches. These findings suggest that a 2 LM is sufficient to reach reasonably good out-of-sample emulation results.

3.3. Probabilistic ODSLC projections using FaIR temperatures

The pattern scaling parameters can be combined with probability distributions (e.g. figure S1) from simple climate models, such as FaIR (section 2.3), to generate probabilistic projections of ODSLC. This enables the propagation of climate uncertainties to ODSLC and allows for the combination of other SLC components based on a consistent temperature evolution.

The probability distribution of ODSLC is also influenced by the sensitivity of pattern scaling parameters to temperature changes, with more sensitive parameters leading to increased uncertainty in ODSLC probabilistic projections. The use of a 3 LM to produce probabilistic projections leads to generally narrower distributions of ODSLC along projections and across scenarios (figure 5) compared to a 2 LM, as introducing an additional term leads to more stable parameter fitting in the training step of the emulator. This demonstrates the added value of an additional layer in reducing uncertainty and improving the robustness of probabilistic ODSLC projections.

Parameter sensitivity can play a distinct role depending on scenario and timescale. For instance, the 2 LM pattern scaling approach (figure 5, blue shading) exhibits larger variability and more pronounced fluctuations in projected ODSLC. In SSP1-2.6, the 2 LM shows an initial peak in uncertainty around the mid-21st century before stabilizing, but it still maintains higher variability compared to the 3 LM approach (figure 5, red shading), which remains more stable throughout the projection period. Under SSP5-8.5, the divergence between 2 LM and 3 LM

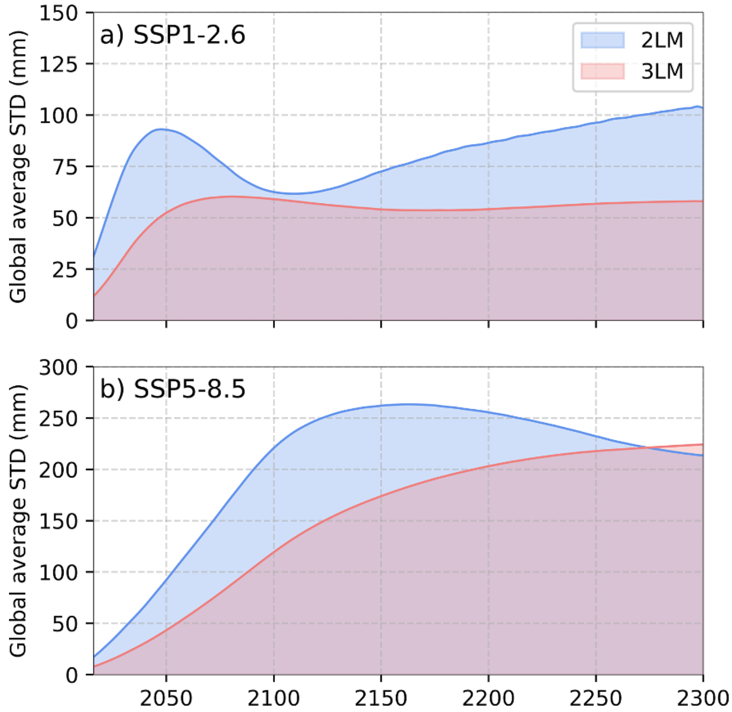

Figure 5. Global weighted-average standard deviation (STD) of local ODSLC probability distributions (mm), derived using FaIR temperature projections up to 2300 using 2 LM (blue) and a 3 LM (red) pattern scaling, for SSP1-2.6 (a) and SSP5-8.5 (b).

is even more evident, with 2 LM uncertainty growing from the late 21st to the mid-22nd century, then peaking around 2150 before gradually declining. In contrast, the 3 LM maintains a smoother and more stable evolution. Nonetheless, SSP5-8.5 shows, that over long timescales (e.g. at 2300), probability distributions from the 3 LM tend to get larger, even slightly exceeding those from the 2 LM. This global STD amplification is mainly driving by increased parameter sensitivity in polar regions (figure S4), as also noted in section 3.2.

4. Discussion and conclusions

Previous emulation approaches using an EBM to reproduce CMIP simulations of GMTSLR plus ODSLC—together known as sterodynamic sea level—have typically relied on a two-layer structure to represent the climate system (Palmer *et al* 2018, 2020, Fox-Kemper *et al* 2021, Yuan and Kopp 2021). In this study, we demonstrate that introducing an additional deep ocean layer to the emulator does not necessarily improve the emulation of GMTSRL but helps to capture non-linearities in ODSLC, especially on multi-century timescales. Differences in emulator performance stem from the forcing scenarios and models being emulated, depending not only on the system's response to increasing radiative forcing but also on model-specific dynamics.

For GMTSLR, we evaluated emulator performance for SSP1-2.6 and SSP5-8.5, as these are the scenarios most broadly available across models. We found that both the 2 LM and 3 LM EBM configurations closely approximate simulated GMTSLR for most models, showing minimal discrepancies between both approaches. Moreover, most models exhibit consistent inter-scenario similarities in performance. These findings, however, contradict recent literature on GSAT emulation (Cummins *et al* 2020, Jackson *et al* 2022). We attribute these discrepancies to the larger role of total ocean heat content driving GMTSLR, whereas GSAT is more sensitive to heat vertical exchange processes, benefiting from an additional EBM timescale and being more prone to inter-scenario parameter sensitivity.

A key advantage of estimating global ocean temperatures at multiple depths is their potential use as inputs for regional emulators to explain local oceanographic variables which are partly affected by temperature, such as ODSLC. Following the two different EBM structures compared here for GMTSLR, we also assessed the 2 LM and 3 LM for their ability to emulate ODSLC, including layer temperatures outputted from the EBM as global predictors. The contrasts between scenarios highlight that adding an additional linear term in the emulator is especially beneficial for high forcing scenarios driving pronounced non-linear ODSLC behaviors. When the emulators are coupled to a simple climate model for climatic uncertainty quantification, this difference in performance translates to lower uncertainties in probabilistic ODSLC projections when using three layers, resulting in up to one global-mean standard deviation difference. Therefore, using a three-layer EBM instead of a two-layer one would also reduce the uncertainty of total sea-level projections based on an emulation framework (Garner *et al* 2023), especially for low and intermediate emissions scenarios. However, over long timescales (at 2300), the 3 LM shows increasing uncertainty driven by higher parameter sensitivity in polar regions. We argue that halosteric and circulation processes not well represent in the emulator can lead to emulator overfitting in areas subject to strong dynamical shifts. Including these complex effects affecting ODSLC in our pattern scaling approach would require substantial emulator reformulation, which may compromise the simplicity and applicability of the pattern scaling approach and is beyond the scope of this study.

Although results demonstrate that both EBM configurations capture the thermal dynamics affecting GMTSLR on multi-centennial timescales, the analysis has limitations that could be further investigated in future work. While we used all CMIP6 models with long-term projections available, the subset is relatively small and features irregular scenario availability and simulation length, complicating efforts to draw CMIP-wide conclusions about the optimal EBM structure for emulation.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://esgf-metagrid.cloud.dkrz.de/search/cmip6-dkrz/.

Code availability statement

The code to produce the data for this manuscript is available at:

https://github.com/victor-malagon/emusdyn.

Acknowledgment

This publication was supported by PROTECT. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869304, PROTECT contribution number 162. This publication has also received funding from the Horizon Europe project WorldTrans (Grant No. 101081661). We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups for producing and making available their model output. For CMIP the US Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

References

Berdahl M, Leguy G, Lipscomb W H and Urban N M 2021 Statistical emulation of a perturbed basal melt ensemble of an ice sheet model to better quantify Antarctic sea level rise uncertainties *Cryosphere* **15** 2683–99

Bi D *et al* 2020 Configuration and spin-up of ACCESS-CM2, the new generation Australian Community Climate and Earth System Simulator Coupled Model *J. South. Hemisph. Earth Syst. Sci.* **70** 225–51

Boucher O *et al* 2020 Presentation and evaluation of the IPSL-CM6A-LR climate model *J. Adv. Model. Earth Syst.* **12** e2019MS002010

Bulthuis K, Arnst M, Sun S and Pattyn F 2019 Uncertainty quantification of the multi-centennial response of the Antarctic ice sheet to climate change *Cryosphere* **13** 1349–80

Cheng L *et al* 2022 Past and future ocean warming *Nat. Rev. Earth Environ.* **3** 776–94

Church J A, Roemmich D, Domingues C M, Willis J K, White N J, Gilson J E, Stammer D, Köhl A, Chambers D P and Landerer F W 2010 Ocean temperature and salinity contributions to global and regional sea-level change *Understanding Sea-level Rise and Variability* pp 143–76

Couldrey M P *et al* 2021 What causes the spread of model projections of ocean dynamic sea-level change in response to greenhouse gas forcing? *Clim. Dyn.* **56** 155–87

Cummins D P, Stephenson D B and Stott P A 2020 Optimal estimation of stochastic energy balance model parameters *J. Clim.* **33** 7909–26

Earth System Grid Federation (ESGF) 2025 *CMIP6 archive at DKRZ [Online]* (available at: https://esgf-metagrid.cloud.dkrz.de/search/cmip6-dkrz/) (Accessed 15 June 2025)

Edwards T L *et al* 2021 Projected land ice contributions to twenty-first-century sea level rise *Nature* **593** 74–82

Edwards T L, Brandon M A, Durand G, Edwards N R, Golledge N R, Holden P B, Nias I J, Payne A J, Ritz C and Wernecke A 2019 Revisiting Antarctic ice loss due to marine ice-cliff instability *Nature* **566** 58–64

Eyring V, Bony S, Meehl G A, Senior C A, Stevens B, Stouffer R J and Taylor K E 2016 Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization *Geosci. Model Dev.* **9** 1937–58

Farrell W E and Clark J A 1976 On postglacial sea level *Geophys. J. Int.* **46** 647–67

Forster P M, Andrews T, Good P, Gregory J M, Jackson L S and Zelinka M 2013 Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models *J. Geophys. Res.* **118** 1139–50

Forster P *et al* 2021 The Earth's energy budget, climate feedbacks, and climate sensitivity *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* ed V Masson-Delmotte *et al* (Cambridge University Press) pp 923–1054

Fox-Kemper B 2021 Ocean, cryosphere and sea level change *AGU Fall Meeting Abstracts* vol 2021 pp U13B–09

Fox-Kemper B *et al* 2021 Ocean, cryosphere, and sea level change *Climate Change 2021: the Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* ed V Masson-Delmotte *et al* (Cambridge University Press) pp 1211–362

Frederikse T, Riva R, Slobbe C, Broerse T and Verlaan M 2016 Estimating decadal variability in sea level from tide gauge records: an application to the North Sea *J. Geophys. Res.* **121** 1529–45

Fredriksen H-B, Rugenstein M and Graversen R 2021 Estimating radiative forcing with a nonconstant feedback parameter and linear response *J. Geophys. Res.* **126** e2020JD034145
Fredriksen H-B and Rypdal M 2017 Long-range persistence in global surface temperatures explained by linear multibox energy balance models *J. Clim.* **30** 7157–68
Fredriksen H-B, Smith C J, Modak A and Rugenstein M 2023 21st century scenario forcing increases more for CMIP6 than CMIP5 models *Geophys. Res. Lett.* **50** e2023GL102916
Garner G G *et al* 2023 IPCC AR6 WGI Sea Level Projections (World Data Center for Climate (WDCC) at DKRZ) (https://doi.org/10.26050/WDCC/AR6.IPCC-DDC_AR6_Sup_SLPr)
Geoffroy O, Saint-Martin D, Bellon G, Voldoire A, Olivié D J L and Tytéca S 2013a Transient climate response in a two-layer energy-balance model. Part II: representation of the efficacy of deep-ocean heat uptake and validation for CMIP5 AOGCMs *J. Clim.* **26** 1859–76
Geoffroy O, Saint-Martin D, Olivié D J, Voldoire A, Bellon G and Tytéca S 2013b Transient climate response in a two-layer energy-balance model. Part I: analytical solution and parameter calibration using CMIP5 AOGCM experiments *J. Clim.* **26** 1841–57
Good P, Gregory J M and Lowe J A 2011 A step-response simple climate model to reconstruct and interpret AOGCM projections *Geophys. Res. Lett.* **38**
Gregory J M *et al* 2019 Concepts and terminology for sea level: mean, variability and change, both local and global *Surv. Geophys.* **40** 1251–89
Gupta A S, Jourdain N C, Brown J N and Monselesan D 2013 Climate drift in the CMIP5 models *J. Clim.* **26** 8597–615
Hamlington B D, Piecuch C G, Reager J T, Chandanpurkar H, Frederikse T, Nerem R S, Fasullo J T and Cheon S-H 2020 Origin of interannual variability in global mean sea level *Proc. Natl Acad. Sci.* **117** 13983–90
Held I M, Winton M, Takahashi K, Delworth T, Zeng F and Vallis G K 2010 Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing *J. Clim.* **23** 2418–27
Hermans T H, Gregory J M, Palmer M D, Ringer M A, Katsman C A and Slangen A B 2021 Projecting global mean sea-level change using CMIP6 models *Geophys. Res. Lett.* **48** e2020GL092064
Jackson L S, Maycock A C, Andrews T, Fredriksen H-B, Smith C J and Forster P M 2022 Errors in simple climate model emulations of past and future global temperature change *Geophys. Res. Lett.* **49** e2022GL098808
Johnson G C and Wijffels S E 2011 Ocean density change contributions to sea level rise *Oceanography* **24** 112–21
Kelley M *et al* 2020 GISS-E2.1: configurations and climatology *J. Adv. Model. Earth Syst.* **12** e2019MS002025
Kopp R E *et al* 2023 The framework for assessing changes to sea-level (FACTS) v1.0-rc: a platform for characterizing parametric and structural uncertainty in future global, relative, and extreme sea-level change (climate and earth system modeling) (available at: https://egusphere.copernicus.org/preprints/2023/egusphere-2023-14/)
Kuhlbrodt T and Gregory J M 2012 Ocean heat uptake and its consequences for the magnitude of sea level rise and climate change *Geophys. Res. Lett.* **39**
Leach N, Jenkins S, Nicholls Z, Smith C J, Lynch J, Cain M, Walsh T, Wu B, Tsutsui J and Allen M R 2021 FaIRv2.0.0: a generalized impulse response model for climate uncertainty and future scenario exploration *Geosci. Model Dev.* **14** 3007–36
Lenton T M, Rockström J, Gaffney O, Rahmstorf S, Richardson K, Steffen W and Schellnhuber H J 2019 Climate tipping points—too risky to bet against *Nature* **575** 592–5
Levermann A, Clark P U, Marzeion B, Milne G A, Pollard D, Radic V and Robinson A 2013 The multimillennial sea-level commitment of global warming *Proc. Natl Acad. Sci.* **110** 13745–50
Lyu K, Zhang X, Church J A and Hu J 2016 Evaluation of the interdecadal variability of sea surface temperature and sea level in the Pacific in CMIP3 and CMIP5 models *Int. J. Climatol.* **36** 3723–40
Malagón-Santos V, Slangen A B A, Hermans T H J, Dangendorf S, Marcos M and Maher N 2023 Improving statistical projections of ocean dynamic sea-level change using pattern recognition techniques *Ocean Sci.* **19** 499–515
Meinshausen M *et al* 2020 The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500 *Geosci. Model Dev.* **13** 3571–605
Miller R L *et al* 2021 CMIP6 historical simulations (1850–2014) with GISS-E2.1 *J. Adv. Model. Earth Syst.* **13** e2019MS002034
Mitrovica J X, Tamisiea M E, Davis J L and Milne G A 2001 Recent mass balance of polar ice sheets inferred from patterns of global sea-level change *Nature* **409** 1026–9
Nandini-Weiss S D, Ojha S, Köhl A, Jungclaus J H and Stammer D 2024 Long-term climate change impacts on regional sterodynamic sea level statistics analyzed from the MPI-ESM large ensemble simulation *Clim. Dyn.* **62** 1311–28
Nauels A, Meinshausen M, Mengel M, Lorbacher K and Wigley T M L 2017 Synthesizing long-term sea level rise projections—the MAGICC sea level model v2.0 *Geosci. Model Dev.* **10** 2495–524
Nicholls Z R J *et al* 2020 Reduced complexity model intercomparison project phase 1: introduction and evaluation of global-mean temperature response *Geosci. Model Dev.* **13** 5175–90
O'Neill B C, Kriegler E, Riahi K, Ebi K L, Hallegatte S, Carter T R, Mathur R and van Vuuren D P 2014 A new scenario framework for climate change research: the concept of shared socioeconomic pathways *Clim. Change* **122** 387–400
Oelsmann J, Marcos M, Passaro M, Sanchez L, Dettmering D, Dangendorf S and Seitz F 2024 Regional variations in relative sea-level changes influenced by nonlinear vertical land motion *Nat. Geosci.* **17** 137–44
Palmer M D, Gregory J M, Bagge M, Calvert D, Hagedoorn J M, Howard T, Klemann V, Lowe J A, Roberts C D and Slangen A B A 2020 Exploring the drivers of global and local sea-level change over the 21st century and beyond *Earth's Future* **8** e2019EF001413
Palmer M D, Harris G R and Gregory J M 2018 Extending CMIP5 projections of global mean temperature change and sea level rise due to thermal expansion using a physically-based emulator *Environ. Res. Lett.* **13** 084003
Palmer M D and Weeks J H 2024 The need for multi-century projections of sea level rise *Earth's Future* **12** e2023EF004403
Pattyn F and Morlighem M 2020 The uncertain future of the Antarctic ice sheet *Science* **367** 1331–5
Pincus R, Forster P M and Stevens B 2016 The radiative forcing model intercomparison project (RFMIP): experimental protocol for CMIP6 *Geosci. Model Dev.* **9** 3447–60
Slangen A B A, Katsman C A, van de Wal R S W, Vermeersen L L A and Riva R E M 2012 Towards regional projections of twenty-first century sea-level change based on IPCC SRES scenarios *Clim. Dyn.* **38** 1191–209
Smith C J *et al* 2020 Effective radiative forcing and adjustments in CMIP6 models *Atmos. Chem. Phys.* **20** 9591–618
Smith C J *et al* 2021 Energy budget constraints on the time history of aerosol forcing and climate sensitivity *J. Geophys. Res.* **126** e2020JD033622
Smith C J, Forster P M, Allen M, Leach N, Millar R J, Passerello G A and Regayre L A 2018 FAIR v1.3: a simple emissions-based impulse response and carbon cycle model *Geosci. Model Dev.* **11** 2273–97
Smith C 2024a FaIR calibration data (available at: https://zenodo.org/records/13951079)
Smith C 2024b FaIR calibration data (available at: https://zenodo.org/records/13142999)
Smith C, Cummins D P, Fredriksen H-B, Nicholls Z, Meinshausen M, Allen M, Jenkins S, Leach N, Mathison C

and Partanen A-I 2024 Fair-calibrate v1.4.1: calibration, constraining, and validation of the FaIR simple climate model for reliable future climate projections *Geosci. Model Dev.* **17** 8569–92
Stammer D and Hüttemann S 2008 Response of regional sea level to atmospheric pressure loading in a climate change scenario *J. Clim.* **21** 2093–101
Swart N C *et al* 2019 The Canadian Earth System Model version 5 (CanESM5.0.3) *Geosci. Model Dev.* **12** 4823–73
Thomas M A and Lin T 2018 A dual model for emulation of thermosteric and dynamic sea-level change *Clim. Change* **148** 311–24
Turner F E, Malagon Santos V, Edwards T L, Slangen A B A, Nicholls R J, Le Cozannet G, O'Neill J and Adhikari M 2023 Illustrative multi-centennial projections of global mean sea-level rise and their application *Earth's Future* **11** e2023EF003550
Virtanen P *et al* 2020 SciPy 1.0: fundamental algorithms for scientific computing in Python *Nat. Methods* **17** 261–72
Wills R C J, Battisti D S, Armour K C, Schneider T and Deser C 2020 Pattern recognition methods to separate forced responses from internal variability in climate model ensembles and observations *J. Clim.* **33** 8693–719
Yuan J and Kopp R E 2021 Emulating ocean dynamic sea level by two-layer pattern scaling *J. Adv. Model. Earth Syst.* **13** e2020MS002323
Yukimoto S *et al* 2019 The Meteorological Research Institute Earth System Model Version 2.0, MRI-ESM2.0: description and basic evaluation of the physical component *J. Meteorol. Soc. Jpn.* **97** 931–65
Ziehn T, Chamberlain M A, Law R M, Lenton A, Bodman R W, Dix M, Stevens L, Wang Y-P and Srbinovsky J 2020 The Australian earth system model: ACCESS-ESM1.5 *J. South. Hemisph. Earth Syst. Sci.* **70** 193–214