SUPPLEMENTARY MATERIAL

for "The realized warming fraction: A multi-model sensitivity study"

Patrik L. Pfister[1,2] and Thomas F. Stocker[1,2]

[1] Climate and Environmental Physics, Physics Institute, University of Bern, 3012 Bern, Switzerland

[2] Oeschger Center for Climate Change Research, University of Bern, 3012 Bern, Switzerland

S1 Extended Results

In this supplementary section, we first present two tables as discussed in the main text. Table S1 shows the influence of ensemble subset selection on the relation between the equilibrium climate sensitivity (ECS) and the realized warming fraction (RWF). Table S2 lists cross-correlations between the diagnostic EBM parameters for the three analyzed model ensembles.

We also include three subsections describing complementary results, which were only briefly touched upon in the main text.

Table S1: Median and 90%-spread of realized warming fraction (RWF) and influential physical quantities in the three ensembles presented in this study (ESMs, EMICs and the B3D-LPX ensemble). For comparison, we also list the full CMIP5 (IPCC, 2013, Andrews et al., 2015) and EMIC-AR5 ensembles (Eby et al., 2013, , this study), as well as the subsets presented by Frölicher and Paynter (2015) (F2015). Note that the "full" EMIC-AR5 ensemble does not include the UMD and IAP models, due to lack of ECS extrapolation (Section S2.1). The "full" CMIP5 ensemble does not include HadCM3, for which we have not received the required model output. N, ε and γ are not listed for CMIP5 (and the Frölicher and Paynter (2015) model selection) because we are lacking N estimates for some models.

	members	RWF	ECS	ε	γ	N
ESMs	15	0.46 (0.37, 0.64)	3.86 (2.55, 5.54)	1.41 (1.09, 1.63)	0.72 (0.56, 0.90)	1.38 (1.10, 1.56)
EMICs	9	0.61 (0.50, 0.69)	3.30 (2.62, 3.97)	0.99 (0.84, 1.10)	0.69 (0.58, 1.01)	1.44 (1.17, 1.86)
CMIP5 full	26	0.52 (0.38, 0.68)	3.44 (2.31, 5.25)			
EMIC-AR5 full	13	0.59 (0.39, 0.76)	3.30 (1.97, 4.19)	0.97 (0.76, 1.26)	0.69 (0.51, 1.13)	1.33 (0.98 1.83)
ESMs F2015	12	0.52 (0.39, 0.64)	3.80 (2.42, 5.67)			
EMICs F2015	8	0.61 (0.49, 0.80)	3.44 (2.15, 3.99)	0.99 (0.74, 1.04)	0.62 (0.51, 0.97)	1.40 (0.89, 1.64)
B3D-LPX	1068	0.60 (0.42, 0.75)	2.88 (1.43, 5.90)	1.09 (0.81, 1.42)	0.66 (0.45, 0.98)	1.13 (0.62, 1.73)

S1.1 TCR variance decomposition

Analogously to the RWF spread, we also decompose the TCR spread based on the EBM equation $\mathrm{TCR} = R/(\varepsilon\gamma - \lambda)$, using partial derivatives (Equation 4 in the main text). We do this only for the ESM ensemble, to compare the results with Geoffroy et al. (2012). We find that the relative TCR

Table S2: Cross-correlations in various ensembles. Stars mark significant correlations, at a confidence level of 95% (*), 99% (**) or 99.9% (***).

(a) EMICs

	ECS	ε	γ	R	λ	$\varepsilon\gamma$
RWF	−0.46	−0.54	−0.59	0.21	−0.48	−0.67
ECS		0.35	−0.17	0.03	0.60	−0.03
ε			0.32	0.38	−0.09	0.60
γ				0.34	−0.34	0.95***
R					−0.77*	0.42
λ						−0.33

(b) ESMs

	ECS	ε	γ	R	λ	$\varepsilon\gamma$
RWF	−0.70**	−0.66**	−0.34	0.19	−0.69**	−0.55*
ECS		0.21	−0.34	−0.06	0.91***	−0.10
ε			0.51	0.35	0.03	0.85***
γ				0.07	−0.35	0.88***
R					−0.39	0.26
λ						−0.21

(c) Bern3D-LPX constrained ensemble

	ECS	ε	γ	R	λ	$\varepsilon\gamma$
RWF	−0.86***	−0.01	−0.18***	0.05	−0.85***	−0.25***
ECS		0.02	−0.09**	0.02	0.77***	−0.12***
ε			−0.66***	−0.05	−0.04	0.00
γ				0.05	−0.06	0.73***
R					−0.04	0.02
λ						−0.13***

spread contributions of λ, R, γ and ε at CO_2 doubling amount to 43%, 34%, 13% and 10%, respectively. This is roughly consistent with Geoffroy et al. (2012), who find contributions of 56%, 26%, 6% and 8%. The quantitative differences are not surprising given three methodological differences between their study and ours. Firstly, they use a two-layer EBM where ε and γ correspond to deep ocean heat uptake, as opposed to total heat uptake in our one-layer EBM. Secondly, they use an analysis of variance (ANOVA) to decompose the TCR spread contributions, while we use a simpler spread decomposition based on partial derivatives. Thirdly, their selection of ESMs is different, with an overlap of 10 out of 15 models. Our qualitative agreement is thus a confirmation of their finding that λ and R explain most of the TCR spread in the CMIP5 ensemble.

S1.2 Sum of variance contributions compared to total variance

We now compare the sum of parameter variance contributions to the total variance of the RWF. For EMICs and ESMs, the sum of contributions is comparable to the total variance. Like in the ANOVA analysis of Geoffroy et al. (2012), the sum of contributions is slightly higher than the best estimate of total variance (Figure 4 in the main text). However, the sum is well within the 90%-uncertainty range of the total variance, which is 0.0027–0.0149 K^2 for EMICs and 0.0049–0.0177 K^2 for ESMs. These Bayesian uncertainty estimates (Oliphant, 2006) are so wide because the number of ensemble members is small.

This is not the case for the large constrained Bern3D-LPX ensemble, where the sum of contributions

becomes significantly larger than the total variance. The 90%-uncertainty range of the total variance (0.0108–0.0124 K^2) cannot be reconciled with the sum of variance contributions. This discrepancy is partly resolved by the fact that ε and γ are strongly anticorrelated in this ensemble ($R = -0.66$; Table S2). Due to this anticorrelation, the total $\varepsilon\gamma$ spread contribution (Figure 4b) is smaller than the sum of separate ε and γ contributions (Figure 4a). However, even the ECS (or λ) spread contribution alone is larger than the total variance. This may be due to the anticorrelation between ECS (or λ) and $\varepsilon\gamma$. Although weak ($R = -0.12$ or -0.13), this anticorrelation is highly significant due to the large number of ensemble members, in contrast to the EMICs and ESMs which show no significant correlation between ECS and $\varepsilon\gamma$ (Table S2).

This suggests that the total RWF spread in the constrained Bern3D-LPX ensemble may be reduced compared to the ECS-induced spread, due to the counteracting cross-correlation with $\varepsilon\gamma$. Although $\varepsilon\gamma$ is separately anticorrelated with the RWF, this anticorrelation is weaker than in the EMICs and ESMs (Table S2) and secondary to the ECS effect. The anticorrelation between ECS and $\varepsilon\gamma$ is also visible in the Bern3D-LPX ensemble (Figure 2b). Returning to the constrained ensemble, we finally note that the discrepancy between the total RWF spread and the ECS-induced RWF spread is due to the long tails of the ECS distribution: the discrepancy is resolved if the ensemble is reduced to an ECS range of 2–6°C (not shown).

S1.3 Influence of the ECS state-dependence

We have argued in the main text and in Pfister (2017) that the global EBM should generally not only feature a time-dependent $\varepsilon(t)$, but also a time-dependent $\lambda(t)$. As an intermediate approach, we propose to explicitly include the state-dependence of λ_{eq} (Pfister and Stocker, 2017) in the EBM:

$$\varepsilon(t)N(t) = R(t) + \lambda_{eq}(R(t))T(t) \tag{S1}$$

The state-dependence of λ_{eq} can be estimated by independent equilibrium simulations with constant forcing (Pfister and Stocker, 2017) and explains part of the time dependence of $T(t)$ that cannot be attributed to $\varepsilon(t)$. In any model where λ_{eq} is state-dependent, only Equation S1 can unambiguously define $\varepsilon(t)$.

λ_{eq} state-dependence is equivalent to ECS state-dependence, which is found for the Bern3D-LPX model and most other EMICs (Pfister and Stocker, 2017), but also for most ESMs that have been analyzed accordingly (e.g, Jonko et al., 2013, Good et al., 2015). Accounting for state-dependence when calculating the RWF and ε is necessary irrespective of the direction of the state-dependence, i.e., whether ECS decreases or increases with increasing forcing.

This is illustrated in Figure S1 by example of the Bern3D-LPX model, in which ECS decreases with increasing forcing due to the diminishing sea-ice albedo feedback (Pfister and Stocker, 2017). Consistent

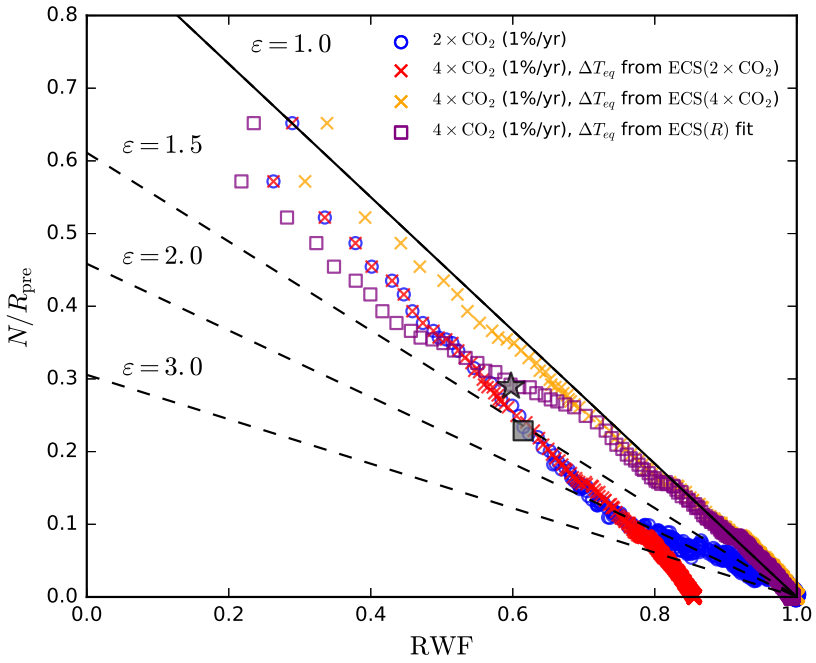

Figure S1: Relation of the RWF to the ocean heat uptake scaled with the prescribed radiative forcing N/R_{pre}, following Winton et al. (2010). Bern3D-LPX simulations with a 1%/yr CO_2 concentration increase from preindustrial concentrations are shown, stabilizing either at CO_2 doubling (blue circles) or quadrupling (red and orange crosses, purple squares). Each symbol shows a 5-year average. In the quadrupling case, the RWF was calculated with three different reference equilibrium temperatures: ECS($2\times CO_2$) (red), ECS($4\times CO_2$) (orange), and for state-dependent ECS (purple). Grey symbols show the values at year 99 as shown in Figure 2 (ECS=3.0°C), i.e., for ECS($2\times CO_2$) (square) and state-dependent ECS (star).

with Winton et al. (2010), two simulations with a gradual CO_2 forcing increase at 1%/year ($R(t)$) are compared: one stabilizing at $2\times CO_2$, the other at $4\times CO_2$.

If λ_{eq} is estimated separately for each simulation, such that the RWF reaches one in the perturbed equilibrium (as it should), $\varepsilon(t)$ calculated from Equation 1 becomes scenario-dependent. The efficacy of the 1%CO_2 experiment depends on the stabilisation level of the scenario ($2\times CO_2$ or $4\times CO_2$), even in the first 70 years where the two simulations are identical (blue circles versus orange crosses in Figure S1). Alternatively, using the same λ_{eq} in both cases leads to a misdiagnosis of the realized warming fraction (red crosses). When using λ_{eq} estimated from $\mathrm{ECS}_{2\times}$ for the $4\times CO_2$ scenario, the realized warming fraction amounts to roughly 0.85 in equilibrium where it should amount to 1.

Using a state-dependent $\lambda_{eq}(R(t))$ based on the time-dependent forcing $R(t)$ in Equation S1 amends this discrepancy and defines an $\varepsilon(t)$ that is equally valid in both simulations for the first 70 years where scenarios overlap, consistent with an RWF that equilibrates to 1 in both scenarios (purple squares). This is only shown for $4\times CO_2$ to avoid an overloaded Figure, but $2\times CO_2$ simply follows the purple squares up to year 70 and the blue circles thereafter. Note that Equation S1 is still a simplification, as λ may not only be state-dependent but also explicitely time-dependent (Gregory et al., 2015, Rose and Rayborn, 2016). Any such separate time-dependence is still included in $\varepsilon(t)$ in the state-dependent EBM (S1).

S2 Extended Methods

In this Supplementary section, we describe and discuss the analysis of the three published model ensembles in more detail than in the main text.

S2.1 EMIC-AR5 ensemble

S2.1.1 Summary

The EMIC-AR5 subset we analyze consists of 9 members, the remaining 6 members of EMIC-AR5 were excluded due to either unavailable output of ocean heat uptake (Section S2.1.2) or the requirement of temperature offset corrections (Section S2.1.3). We also provide estimates of diagnostic EBM parameters for a larger subset of 13 EMICs, including the 4 models with temperature corrections (Table S3).

S2.1.2 Extended analysis description

Our analysis of the EMIC-AR5 ensemble (Eby et al., 2013) is largely based on our earlier analysis focusing on ECS state-dependence (Pfister and Stocker, 2017). EMIC-AR5 included 15 EMICs, which are described in Eby et al. (2013) and listed in Pfister and Stocker (2017). Here we have first analyzed 13 of these models (Table S3), excluding the UMD 2.0 and IAP RAS CM models. No N output was available for the UMD 2.0 model, and the N of the IAP RAS CM model is unrealistically low: its 20-year average around CO_2 quadrupling is 0.25 Wm^{-2}, where the spread of the remaining models is 2.09 ± 0.47 Wm^{-2}. We have excluded 4 further models based on restart temperature or ECS corrections (Section S2.1.3), ending up with an ensemble of 9 uncorrected EMICs.

We use the 1%CO_2 simulation, where CO_2 is increased by 1% per year up to 4× preindustrial (experiment 4Xc), as well as the control simulation (experiment H_CTR) from EMIC-AR5. Temperature and ocean heat uptake anomalies for the 1%CO_2 simulations are obtained by subtracting the mean values of the control simulations from the 1%CO_2 time series. Twenty-year running means of these anomalies centered around year 70 (the year of CO_2 doubling) yield values of ΔT and N. These are then used to calculate RWF = $\Delta T/\mathrm{ECS}$, $\gamma = N/\Delta T$ and ε using Equation 2b.

Values for all quantities obtained from EMIC-AR5 are listed in Table S3. ECS and R estimates from abrupt CO_2 doubling simulations are mostly taken from Pfister and Stocker (2017), apart from corrections described below. Following Andrews et al. (2015), R estimates are based on a linear $\Delta T(t)/N(t)$ fit over the first 20 simulation years after the year of peak ocean heat uptake (Figure S2). For ECS, we mostly select the Gregory estimates extrapolated from a $\Delta T(t)/N(t)$ fit over simulation years 150–1000 ($\Delta T_{lin}^{150-1000}$ in Pfister and Stocker, 2017), because the exponential fits are problematic for some models. Only for two models, namely LOVECLIM and IGSM2, the exponential ECS fits

($\Delta T_{exp}^{150-1000}$) were used because the Gregory estimates were biased by an N offset and temperature variability under early equilibration, respectively.

In addition to those ECS choices, we note that the temperature restart offset correction described by Pfister and Stocker (2017) was consistently applied to the whole ΔT time series for the three affected models LOVECLIM, FAMOUS and SPEEDO. This correction is illustrated by the example of LOVECLIM in Section S2.1.3, Figure S3. While the offset correction only weakly affects ECS, its impact on the TCR, RWF and ε is much larger. The correction is thus crucial to obtaining reasonable estimates of these quantities for the affected models; consequently, these estimates are very sensitive to the chosen correction methods. This also applies to the additional R corrections applied to LOVECLIM and SPEEDO (Section S2.1.3). Based on this limitation, we exclude the models with corrections of any kind from our correlation analysis, but we still show their corrected results in Table S3.

S2.1.3 EMIC-AR5 diagnosis results and restart offset correction

This subsection contains one Table and two Figures illustrating the analysis of EMIC-AR5 output. Table S3 lists all diagnostic EBM quantities obtained for all EMICs, after all corrections were applied as summarized in the table footnote. Figure S2 illustrates the Gregory estimation method (Gregory et al., 2004, Andrews et al., 2015) for ECS and R. Figure S3 illustrates restart offset corrections applied to some EMICs by the example of LOVECLIM, as described further below.

Figure S2 is identical to Figure S1 in Pfister and Stocker (2017), except that the $4\times CO_2$ simulations shown here are abrupt forcing simulations. This allows an R estimate from these simulations, which is very similar to the estimate from $2\times CO_2$ for most EMICs. Note that the raw data is used for the R fits, but 50-year running means (not shown here) are used for the ECS fits to avoid misfits like in IGSM2 (Pfister and Stocker, 2017).

Figure S3 illustrates the temperature restart offset correction applied to three EMICs (Pfister and Stocker, 2017) by example of the LOVECLIM model, as well as an N restart offset correction that is only applied to LOVECLIM. The temperature offset correction (already applied to the temperatures in Figure S2) is described in the Supporting Information of Pfister and Stocker (2017) and very briefly in the following. For each EMIC, it is checked whether the first decadal warming of the 1%/year simulation is below zero or exceeds a warming corresponding to RWF = 1 (red dot in Figure S3) compared to the end of the control run. If either is true, an offset correction is applied; this is the case for LOVECLIM, FAMOUS and SPEEDO.

For LOVECLIM, we have argued that the exponential ECS estimate is probably more accurate than the Gregory estimate (Pfister and Stocker, 2017). This is based on the fact that the warming slope is still substantial after 1000 years, but the N output is close to zero, which may be related to an offset problem as suggested also by the large N spike at the beginning of the 1%/year simulation. Therefore

we apply an N offset correction (Figure S3). The magnitude of this correction is obtained by shifting N (and thereby also R) until the Gregory ECS estimate for $2\times CO_2$ matches the exponential ECS estimate. This corrected R is used in Table S3. The same shift is used for $4\times CO_2$, because the restart offset should not depend on the forcing scenario.

Table S3: Physical quantities diagnosed from the EMIC ensemble. Model label abbreviations are given in Figure S2. The RWF and ε are calculated using ECS estimates from the Gregory method (Pfister and Stocker, 2017). R_{pre} is the CO_2 forcing prescribed to the model, R is the effective radiative forcing diagnosed as shown in Figure S2. ECS and R values marked by asterisks and/or daggers were corrected as specified in the footnotes.

	B3	DC	GE	I2	LO	ME	ML	UV	C2	C3	FA	SP	MI	B3new
R [$\mathrm{Wm^{-1}}$]	3.04	3.47	3.49	3.43	3.14†	3.87	3.21	3.43	5.44	4.44	2.95	1.31††	3.31	3.39
R_{pre} [$\mathrm{Wm^{-1}}$]	3.75	3.81	n.a.	3.73	3.85	4.05	3.15	3.79	3.69	3.73	n.a.	n.a.	3.78	3.75
N [$\mathrm{Wm^{-1}}$]	1.22	1.33	1.40	0.74	1.17	1.44	1.59	1.66	1.78	1.92	1.32	1.14	1.13	1.24
γ [$\mathrm{Wm^{-1}K^{-1}}$]	0.62	0.64	0.57	0.49	1.23	0.60	1.02	0.86	0.86	0.99	0.53	1.07	0.69	0.78
RWF [1]	0.58	0.70	0.61	0.84	0.46	0.63	0.49	0.50	0.63	0.59	0.57	0.28	0.68	0.52
ECS_{corr} [K]	3.44	2.96	4.04	1.80‡	2.08*‡	3.81	3.19	3.86	3.27	3.30	4.41*	3.85*	2.39	3.08
ε [1]	1.05	0.79	0.97	0.72	1.45	0.99	1.03	1.03	1.13	0.95	0.97	0.84	0.93	1.32

† R is corrected by shifting N to match corrected ECS (Figure S3)
†† R from year 1–150 fit is used because years 1–20 are biased due to centennial variability in SPEEDO (Figure S2)
* temperatures are shifted based on restart offset corrections (Figure S3)
‡ exponential extrapolation is used for ECS estimation instead of the Gregory method

S2.2 Ensemble of 15 CMIP5 models

S2.2.1 Summary

We analyze an ESM subset of 15 models from CMIP5 (Taylor et al., 2012), based on the overlap of published results by Andrews et al. (2015) and Gregory et al. (2015). We use ECS and R estimates from Andrews et al. (2015), TCR and γ estimates from Gregory et al. (2015), and calculate RWF = TCR/ECS and ε from Equation 2a (main text).

S2.2.2 Extended analysis description

In contrast to our EMIC-AR5 analysis, ECS and R estimates for the ESMs were obtained from abrupt CO_2 quadrupling simulations (Andrews et al., 2015). This may introduce an ECS bias if those ESMs are state-dependent, which at least some of them are (Good et al., 2015, Pfister and Stocker, 2017). Due to shorter simulations, ECS was extrapolated from years 20–150, opposed to years 150–1000 in the EMICs. This may also bias ECS, but these biases may partly cancel: The state-dependence likely leads to an overestimation of ECS($2\times CO_2$) (Good et al., 2015), and the short time scale to an underestimation (Knutti and Rugenstein, 2015, Proistosescu and Huybers, 2017). The R time scale is consistent with our EMIC estimates.

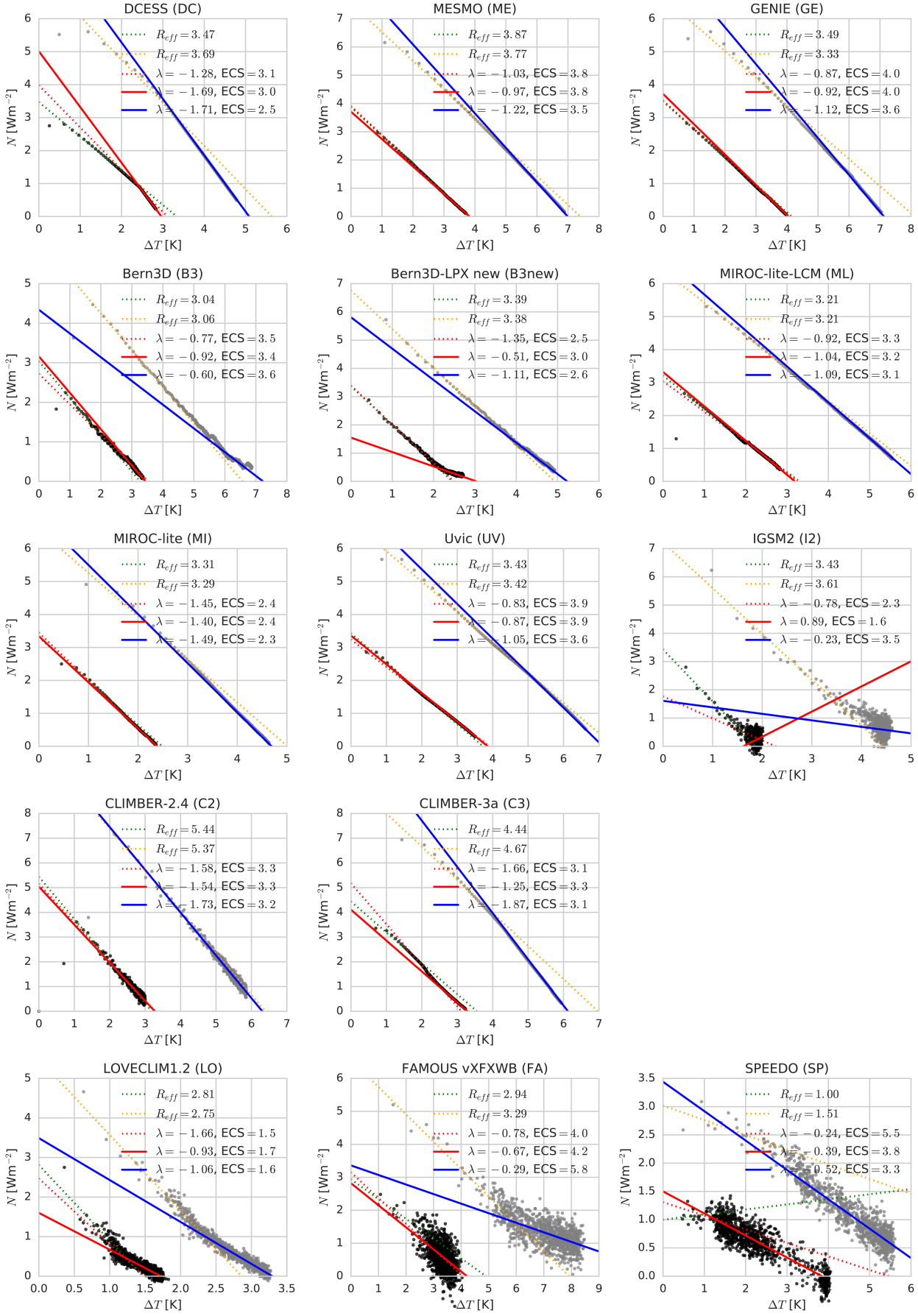

Figure S2: Estimates of ECS and effective radiative forcing R of the EMIC ensemble using the Gregory method (Gregory et al., 2004, Andrews et al., 2015). Raw model output is indicated by black ($2\times CO_2$) and grey ($4\times CO_2$) dots. All linear fits shown here are performed on this raw output (solid lines: years 150–1000, red dotted: years 20–150, green dotted: years 1–20). R is estimated from the first 20 years after N peaks. R values in the legend are in Wm^{-2}, λ values are in Wm^{-2}K^{-1}, ECS values in K.

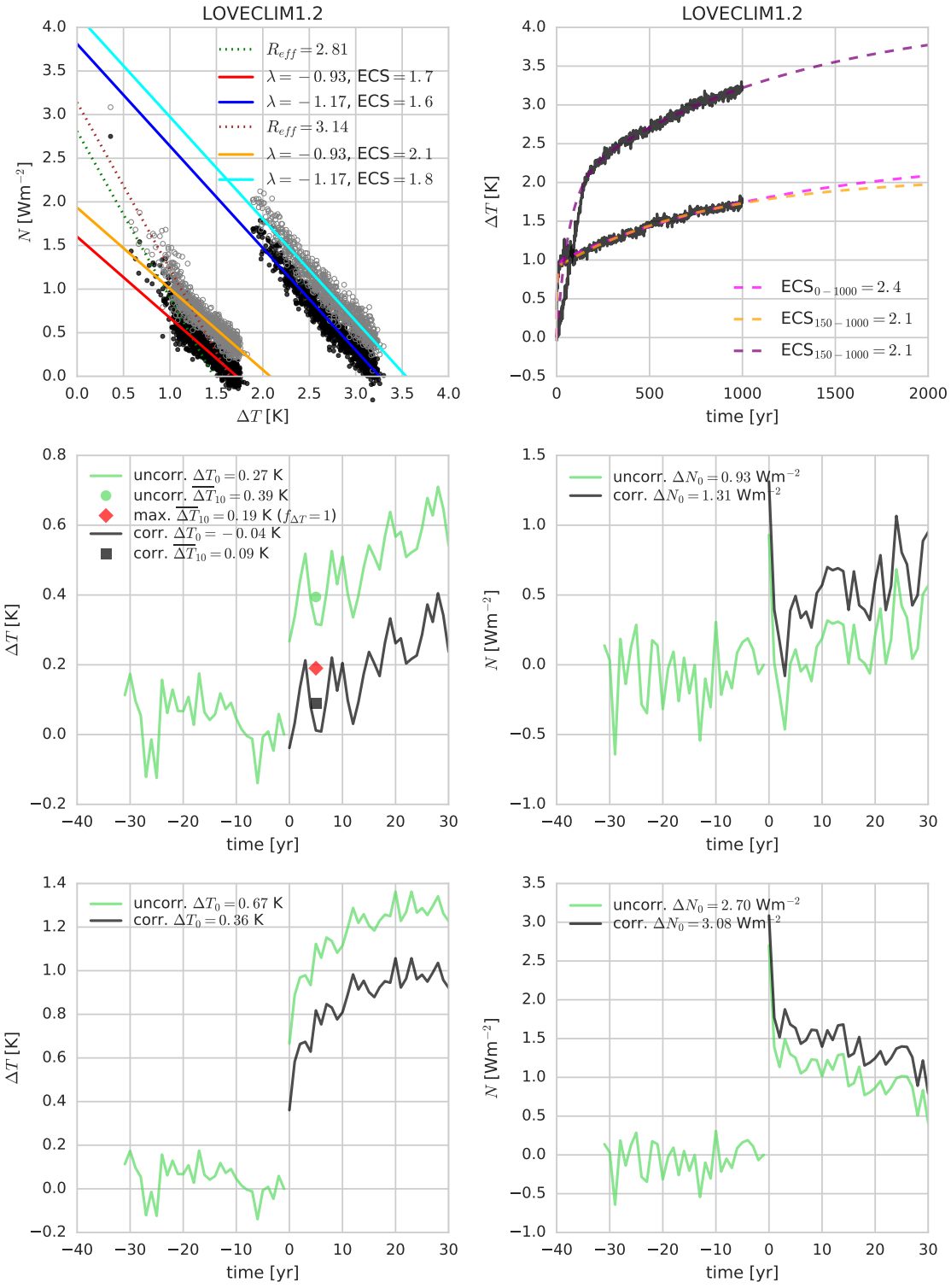

Figure S3: Temperature (ΔT) and ocean heat uptake (N) restart anomaly corrections applied to LOVECLIM model output. The top row shows the Gregory fits for ECS and R, and the exponential fits for ECS (Pfister and Stocker, 2017). Black filled dots in the first panel are only temperature-corrected, grey circles are additionally heat uptake-corrected (see text). Remaining rows show the temperature and ocean heat uptake time series of the end of the control simulation (up to year 0) and the start of the 1%/year 4×CO_2 simulation (middle row) or the start of the abrupt 2×CO_2 simulation (bottom row); uncorrected data are shown in green, corrected data in black. The green dot and black square in the ΔT panel denote averages of the first 10 years, the red diamond is the maximum allowed offset (corresponding to RWF = 1) below which anomalies would not be corrected (LOVECLIM is well above). Note that the difference between red dot and black dot is the difference between the control mean and the last control decade.

TCR and γ estimates are taken from Gregory et al. (2015), where they are averaged over 20 years like our EMIC estimates. Note that the ocean heat uptake efficiency is named κ in Gregory et al. (2015). The remaining quantities N, ε and RWF can be calculated from these available estimates as described for the EMICs. Our model selection of 15 ESMs thus consists of the overlap between Gregory et al. (2015) and Andrews et al. (2015). For the remaining models analyzed by Andrews et al. (2015), consistent γ estimates are missing such that we cannot estimate N and ε. We do not use the γ estimates from Kuhlbrodt and Gregory (2012) to fill the model gaps, because these are obtained using linear regression and underestimate γ compared to the year 61-80 averages (not shown). In contrast to N, we can calculate the RWF for the full ESM ensemble of Andrews et al. (2015) (Table S1), using TCR estimates from IPCC (2013) that are identical to Gregory et al. (2015) apart from rounding.

S2.3 Constrained Bern3D-LPX ensemble

S2.3.1 Summary

Steinacher et al. (2013) generated 5000 Bern3D-LPX parameter sets based on estimated prior distributions of 19 model parameters. The 1068 parameter sets that best matched a suite of present-day and historic observations were selected, forming the constrained ensemble that was used to simulate future projections. The idealized projections we analyze here were first presented in Steinacher and Joos (2016).

S2.3.2 Extended analysis description

Idealized $2\times CO_2$ simulations were performed by Steinacher and Joos (2016) using the constrained Bern3D-LPX ensemble. The scenario underlying these simulations is a CO_2 increase at a rate of 1 percent per year up a concentration of $2\times CO_2$, which is reached in year 70, and constant concentration thereafter. Unfortunately, this is not identical to any of the idealized scenarios of the EMIC-AR5 intercomparison.

Nevertheless, we employ these $2\times CO_2$ simulations from the Bern3D-LPX ensemble to estimate the RWF and related quantities at CO_2 doubling. These quantities are estimated from 10-year means around year 70 to average out some of the variability imposed on the LPX model (Stocker et al., 2013, Pfister, 2017). The 10-year window length is selected over the usual 31-year window length (Pfister, 2017) to minimize the underestimation of the RWF in the Bern3D-LPX ensemble compared to EMIC-AR5 and CMIP5. Such underestimation occurs because the concentration stabilizes in year 70 in the Bern3D-LPX simulations while it continues to increase in the $4\times CO_2$ scenarios of EMIC-AR5 and CMIP5. This should not affect our semi-quantitative considerations, as the underestimation using the year 65–75 averaging window (where scenarios only differ over over years 70–75) should be small.

S2.3.3 Forcing estimation

More importantly, we cannot estimate R for the Bern3D-LPX ensemble members consistently with the EMICs and ESMs. Simulations with abrupt CO_2 increases are missing, and a linear fit for R cannot be obtained from a gradual forcing increase. However, we know that the main forcing spread in the Bern3D-LPX ensemble is caused by a prior scaling factor applied to the prescribed CO_2 forcing. To account for this, we simply scale an estimated mean R with this scaling factor that is different for each ensemble member. The estimated mean $R = 3.04$ Wm^{-2} was obtained from the abrupt $2\times CO_2$ simulation of the Bern3D-LPX model in EMIC-AR5 (Figure S2). This is the same model version as used by Steinacher and Joos (2016), run with standard parameter values. As most posteriori parameter distributions of the constrained Bern3D-LPX ensemble are centered roughly around these standard values (Steinacher et al., 2013), $R = 3.04$ should roughly correspond to the mean R of the Bern3D-LPX ensemble.

While we thus account for the directly prescribed R spread, we do not account for any additional R spread that may be caused by changes in other prior parameters. While R is only minorly affected by modifications in single parameters such as the global mean λ (squares in Figure 3e) and ocean diffusivities (not shown), the simultaneous variation of multiple parameters may result in a more substantial R spread. By not accounting for this additional spread contribution in R, it is instead implicitly included in the ε spread as calculated from Equation 1.

S2.3.4 ECS estimation using different methods

We have also obtained new ECS estimates for the Bern3D-LPX ensemble by using the Gregory method (Gregory et al., 2004), linearly extrapolating years 150 to 1000 to equilibrium. This method and time scale is consistent with our ECS diagnosis for the EMIC ensemble. It is different from the ECS estimation by Steinacher and Joos (2016) that used an exponential extrapolation; see Pfister and Stocker (2017) for a comparison of the two methods. However, the resulting ECS estimates are very similar apart from few outlier members where the exponential fit fails. Therefore, the resulting skill-weighted ECS likely ranges are in very close agreement between the two methods (1.94–4.21°C and 1.95–4.22°C), with an identical median (2.88°C). These results indicate that the ECS median and range specified by Steinacher and Joos (2016) are robust with respect to the two different ECS estimation methods we have tested.

Bibliography

Andrews, T., Gregory, J. M. and Webb, M. J. (2015). The Dependence of Radiative Forcing and Feedback on Evolving Patterns of Surface Temperature Change in Climate Models, *J. Clim.* **28**(4): 1630–1648.

Eby, M., Weaver, A. J., Alexander, K., Zickfeld, K., Abe-Ouchi, A., Cimatoribus, A. A., Crespin, E., Drijfhout, S. S., Edwards, N. R., Eliseev, A. V., Feulner, G., Fichefet, T., Forest, C. E., Goosse, H., Holden, P. B., Joos, F., Kawamiya, M., Kicklighter, D., Kienert, H., Matsumoto, K., Mokhov, I. I., Monier, E., Olsen, S. M., Pedersen, J. O. P., Perrette, M., Philippon-Berthier, G., Ridgwell, A., Schlosser, A., Schneider von Deimling, T., Shaffer, G., Smith, R. S., Spahni, R., Sokolov, A. P., Steinacher, M., Tachiiri, K., Tokos, K., Yoshimori, M., Zeng, N. and Zhao, F. (2013). Historical and idealized climate model experiments: an intercomparison of Earth system models of intermediate complexity, *Clim. Past* **9**(3): 1111–1140.

Frölicher, T. L. and Paynter, D. J. (2015). Extending the relationship between global warming and cumulative carbon emissions to multi-millennial timescales, *Environ. Res. Lett.* **10**(7): 075002.

Geoffroy, O., Saint-Martin, D. and Ribes, A. (2012). Quantifying the sources of spread in climate change experiments, *Geophys. Res. Lett.* **39**(24). L24703.

Good, P., Lowe, J. A., Andrews, T., Wiltshire, A., Chadwick, R., Ridley, J. K., Menary, M. B., Bouttes, N., Dufresne, J. L., Gregory, J. M., Schaller, N. and Shiogama, H. (2015). Nonlinear regional warming with increasing CO_2 concentrations, *Nat. Clim. Change* **5**(2): 138–142.

Gregory, J. M., Andrews, T. and Good, P. (2015). The inconstancy of the transient climate response parameter under increasing CO2, *Phil. Trans. R. Soc. A* **373**(2054).

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., Lowe, J. A., Johns, T. C. and Williams, K. D. (2004). A new method for diagnosing radiative forcing and climate sensitivity, *Geophys. Res. Lett.* **31**(3).

IPCC (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*, Stocker, T. F., D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P. M.

257 Midgley (eds.), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
258 1535 pp.

259 Jonko, A. K., Shell, K. M., Sanderson, B. M. and Danabasoglu, G. (2013). Climate Feedbacks in
260 CCSM3 under Changing CO2 Forcing. Part II: Variation of Climate Feedbacks and Sensitivity with
261 Forcing, *J. Clim.* **26**(9): 2784–2795.

262 Knutti, R. and Rugenstein, M. A. A. (2015). Feedbacks, climate sensitivity and the limits of linear
263 models, *Phil. Trans. R. Soc. A* **373**(2054).

264 Kuhlbrodt, T. and Gregory, J. M. (2012). Ocean heat uptake and its consequences for the magnitude of
265 sea level rise and climate change, *Geophys. Res. Lett.* **39**(18). L18608.

266 Oliphant, T. E. (2006). A Bayesian perspective on estimating mean, variance, and standard-deviation
267 from data. https://scholarsarchive.byu.edu/facpub/278.

268 Pfister, P. L. (2017). *Energy Balance and Ocean Interactions in Idealized Future Projections based on*
269 *Reduced-Complexity Climate Models*, PhD thesis, Climate and Environmental Physics, University of
270 Bern. 243 pp.

271 Pfister, P. L. and Stocker, T. F. (2017). State-Dependence of the Climate Sensitivity in Earth System
272 Models of Intermediate Complexity, *Geophys. Res. Lett.* **44**(20): 10,643–10,653.

273 Proistosescu, C. and Huybers, P. J. (2017). Slow climate mode reconciles historical and model-based
274 estimates of climate sensitivity, *Sci. Adv.* **3**(7).

275 Rose, B. E. J. and Rayborn, L. (2016). The Effects of Ocean Heat Uptake on Transient Climate
276 Sensitivity, *Curr. Clim. Change Rep.* **2**(4): 190–201.

277 Steinacher, M. and Joos, F. (2016). Transient Earth system responses to cumulative carbon dioxide
278 emissions: linearities, uncertainties, and probabilities in an observation-constrained model ensemble,
279 *Biogeosciences* **13**(4): 1071–1103.

280 Steinacher, M., Joos, F. and Stocker, T. F. (2013). Allowable carbon emissions lowered by multiple
281 climate targets, *Nature* **499**(7457): 197+.

282 Stocker, B. D., Roth, R., Joos, F., Spahni, R., Steinacher, M., Zaehle, S., Bouwman, L., Xu-Ri and
283 Prentice, I. C. (2013). Multiple greenhouse-gas feedbacks from the land biosphere under future
284 climate change scenarios, *Nat. Clim. Change* **3**(7): 666–672.

285 Taylor, K. E., Stouffer, R. J. and Meehl, G. A. (2012). An overview of CMIP5 and the experiment
286 design, *Bulletin of the American Meteorological Society* **93**(4): 485–498.

287 Winton, M., Takahashi, K. and Held, I. M. (2010). Importance of Ocean Heat Uptake Efficacy to
288 Transient Climate Change, *J. Clim.* **23**(9): 2333–2344.