Supplementary Material: An analytical decomposition of the Zero Emissions Commitment in terms of aggregate climate metrics

B. M. Sanderson

April 1, 2026

Contents

This supplement provides step-by-step derivations of the analytical expressions used in the main text. The presentation includes intermediate algebraic steps, physical motivation for each approximation, and worked limiting cases.

S1 Full Level 1 derivation

We distinguish two levels of analytical approximation throughout: *Level 1* retains all eleven model parameters without further reduction (though it relies on several benign linearisation assumptions; see section S1.3), while *Level 2* (Section S2) condenses these into aggregate climate metrics (e.g. TCR/ECS, an effective carbon timescale) at the cost of additional approximations.

This section derives the complete parametric expression for ZEC (eq. (S14)) from the 12-parameter coupled model described in the main text. The key idea is to split ZEC into two physically transparent pieces: a *thermal* term representing unrealised warming that continues to emerge from the ocean after emissions cease, and a *carbon* term representing cooling caused by the drawdown of atmospheric CO_2 once anthropogenic emissions stop. Together these define the "Level 1" expression, which uses all model parameters and reproduces the numerical forward model to within 0.018–0.064 K across nine ESMs.

S1.1 Thermal component: ramp-forced disequilibrium

Definition of ZEC. ZEC is defined as the temperature anomaly measured from the moment emissions stop:

$$\mathrm{ZEC}(\Delta t) \;=\; T(t_0 + \Delta t) \;-\; T(t_0), \tag{S1}$$

where t_0 is the cessation time (the end of the 100-year emission phase in the flat10 protocol) and Δt is the time elapsed since cessation. A positive ZEC means the planet continues to warm; a negative ZEC means it cools.

Equilibrium target. If atmospheric CO_2 were held fixed at its cessation-time value $C(t_0)$, the radiative forcing would remain at $F_0 = a\ln\big(1 + C(t_0)/C_0\big)$ and the climate system would eventually reach the equilibrium temperature

$$T_{\mathrm{eq}} \;=\; \frac{F_0}{\lambda_{\mathrm{eff}}}, \tag{S2}$$

where λ_{eff} ($\mathrm{W\,m^{-2}\,K^{-1}}$) is the effective radiative feedback parameter. This equilibrium is never reached in practice because CO_2 is not held fixed (it declines after cessation) but T_{eq} provides the natural reference scale for the thermal component.

Two-box climate model. The climate system is represented by two coupled heat reservoirs: a well-mixed surface layer with heat capacity C_f and a deep ocean layer with heat capacity C_s. Heat exchange between the layers proceeds at a rate proportional to their temperature difference, governed by the coupling parameter γ ($\mathrm{W\,m^{-2}\,K^{-1}}$).

This two-box system has two eigenmodes, each characterised by a timescale τ_j^{clim} and a weight q_j (with $q_1 + q_2 = 1$). The fast eigenmode ($j = 1$, timescale $\tau_1^{\mathrm{clim}} \sim 4$ yr) is dominated by the surface-layer adjustment; the slow eigenmode ($j = 2$, timescale $\tau_2^{\mathrm{clim}} \sim 200$–$500$ yr) is dominated by the gradual equilibration of the deep ocean. Note that the eigenmodes are linear combinations of both boxes, so the "fast" and "slow" labels refer to the eigenmode timescales, not to the individual boxes.

Response to a linearly increasing (ramp) forcing. During the emission phase, atmospheric CO_2 rises roughly linearly, so the forcing ramps up approximately as $F(t) \approx (F_0/t_0)\, t$. One can show (by convolving this increase with the two-exponential impulse response) that by the end of the ramp, each eigenmode j has accumulated a *disequilibrium*, a gap between the warming it has delivered and the warming it would deliver at equilibrium. The warming deficit of mode j at cessation is

$$\Delta T_j(t_0) \;=\; T_{\mathrm{eq}}\, q_j\, \frac{\tau_j^{\mathrm{clim}}}{t_0}\Big(1 - e^{-t_0/\tau_j^{\mathrm{clim}}}\Big). \tag{S3}$$

The factor $\tau_j^{\mathrm{clim}}/t_0$ is the ratio of the mode's response timescale to the ramp duration. It captures an intuitive result:

- A mode whose timescale is *short relative to the ramp* ($\tau_j^{\mathrm{clim}} \ll t_0$) adjusts continuously as the forcing increases, keeping pace with the ramp and accumulating only a small deficit $\sim \tau_j^{\mathrm{clim}}/t_0$.
- A mode whose timescale is *long relative to the ramp* ($\tau_j^{\mathrm{clim}} \gg t_0$) barely responds during the ramp and retains most of its equilibrium share as unrealised warming.

Worked limiting cases. Examining these limits symbolically shows which mode dominates the "warming in the pipeline":

- **Fast mode** ($t_0 \gg \tau_1^{\mathrm{clim}}$): when the ramp duration is much longer than the fast timescale, the exponential $e^{-t_0/\tau_1^{\mathrm{clim}}} \to 0$ and eq. (S3) reduces to $\Delta T_1(t_0) \approx T_{\mathrm{eq}}\, q_1\, \tau_1^{\mathrm{clim}}/t_0$. The fast mode has essentially kept pace with the forcing ramp, and only a fraction $\tau_1^{\mathrm{clim}}/t_0$ of its equilibrium share remains unrealised. For example, with $\tau_1^{\mathrm{clim}} \sim 4$ yr and $t_0 = 100$ yr, this fraction is $\sim 4\%$.
- **Slow mode** ($t_0 \ll \tau_2^{\mathrm{clim}}$): when the ramp is short relative to the slow timescale, we expand the exponential to second order, $1 - e^{-t_0/\tau_2^{\mathrm{clim}}} \approx t_0/\tau_2^{\mathrm{clim}} - t_0^2/(2\tau_2^{\mathrm{clim}^2})$, giving $\Delta T_2(t_0) \approx T_{\mathrm{eq}}\, q_2 \left[1 - t_0/(2\tau_2^{\mathrm{clim}})\right]$. Nearly the entire equilibrium share of the slow mode remains unrealised, reduced only by a small correction $t_0/(2\tau_2^{\mathrm{clim}})$. For $\tau_2^{\mathrm{clim}} \sim 250$ yr and $t_0 = 100$ yr, this gives $\Delta T_2 \approx 0.8\, T_{\mathrm{eq}}\, q_2$: roughly 80% of the slow mode's equilibrium share is still unrealised at cessation.

In summary, for representative numbers, **85–90% of the total unrealised warming resides in the slow eigenmode** under ramp forcing. This asymmetry is a direct consequence of the ramp protocol; a step-change in forcing would distribute the disequilibrium more evenly between modes.

Forcing projection into the eigenmode basis. To see why this asymmetry arises, it helps to decompose the forcing into the eigenmode basis. Radiative forcing enters the two-box system through the vector $\mathbf{b} = (1/C_f,\, 0)^\top$, since only the surface box is directly heated. The eigenvectors of $\mathbf{A}$ (derived in section S3) are approximately $\mathbf{v}_1 \approx (1,\, 0)^\top$ (fast mode, nearly pure surface) and $\mathbf{v}_2 \approx (1,\, 1+\lambda_f/(\varepsilon\gamma))^\top$ (slow mode, coupled surface–deep). Projecting $\mathbf{b}$ onto these eigenvectors shows that the forcing excites *both* modes: a large fast-mode amplitude (because $\mathbf{v}_1$ is nearly aligned with $\mathbf{b}$) and a smaller but non-zero slow-mode amplitude (because $\mathbf{v}_2$ has a surface component). The eigenmode weights q_j in the step-response formula (eq. (S24)) quantify this projection: typically $q_1 \approx 0.55$–0.75 and $q_2 = 1 - q_1 \approx 0.25$–$0.45$.

For a *step* in forcing, both modes are excited simultaneously and the disequilibrium of each mode is proportional to its weight q_j; the unrealised warming is therefore split roughly $q_1 : q_2$. Under *ramp* forcing, however, the fast mode keeps pace with the slowly increasing forcing (its disequilibrium shrinks as τ_1^{clim}/t_0), while the slow mode lags far behind (its disequilibrium remains $\sim q_2$). The ramp protocol thus amplifies the slow mode's share of the pipeline relative to its equilibrium weight, concentrating the post-cessation warming commitment in the mode that delivers heat most slowly.

Post-cessation relaxation. After emissions cease, suppose that forcing is held constant (this assumption is relaxed when the carbon term is introduced). Each mode independently relaxes toward its equilibrium share at rate $1/\tau_j^{\text{clim}}$. The warming delivered by mode j during the interval $[t_0,\, t_0 + \Delta t]$ is

$$\Delta T_j(t_0)\left(1 - e^{-\Delta t/\tau_j^{\text{clim}}}\right). \tag{S4}$$

Summing over both modes and substituting the disequilibrium from eq. (S3), we obtain the thermal component of ZEC:

$$\boxed{\text{ZEC}_{\text{thermal}}(\Delta t) \;=\; T_{\text{eq}} \sum_j q_j \,\frac{\tau_j^{\text{clim}}}{t_0}\Big(1 - e^{-t_0/\tau_j^{\text{clim}}}\Big)\Big(1 - e^{-\Delta t/\tau_j^{\text{clim}}}\Big).} \tag{S5}$$

This term is always positive (given $T_{\text{eq}} > 0$, i.e. positive net forcing): unrealised warming can only add heat to the surface.

S1.2 Carbon component: CO_2 drawdown cooling

The thermal term assumed that atmospheric CO_2 (and hence forcing) remains fixed after cessation. In reality, once anthropogenic emissions stop, the land and ocean carbon sinks continue to absorb CO_2 from the atmosphere, reducing the forcing and cooling the climate. This *carbon component* is always negative (cooling) in the present framework, which assumes passive sinks only, and is what makes ZEC substantially smaller than the "warming in the pipeline."

Multi-pool carbon model. Atmospheric CO_2 is drawn down by multiple uptake processes operating on different timescales. We represent this with a sum of exponential uptake pools, each characterised by an uptake fraction f_i and a timescale τ_i. A permanent airborne fraction f_∞ represents the portion of emitted CO_2 that remains in the atmosphere indefinitely (on the timescales considered here).

During the emission phase, each pool i accumulates excess airborne carbon proportional to $f_i\,\tau_i(1 - e^{-t_0/\tau_i})$. After cessation, each pool exponentially relaxes on its own timescale τ_i.

Carbon–climate feedback. As the climate warms, carbon sinks become less efficient; for example, warmer oceans hold less dissolved CO_2, and drought-stressed vegetation takes up less carbon. We parameterise this with a linear carbon–climate feedback ζ, which reduces the effective uptake fractions as $f_i^*(t) = f_i(1 - \zeta\, T(t))$. Evaluating this

at the mean temperature during the emission phase, $\bar{T} \approx T_0/2$, the post-cessation CO_2 drawdown is:

$$\Delta C(\Delta t) \;=\; -(1-\zeta\bar{T})\,\frac{E_0}{\kappa}\sum_i f_i\,\tau_i\big(1-e^{-t_0/\tau_i}\big)\big(1-e^{-\Delta t/\tau_i}\big), \tag{S6}$$

where E_0 is the emission rate ($\mathrm{PgC\,yr^{-1}}$) and κ ($= 2.12\ \mathrm{PgC\,ppm^{-1}}$) converts between carbon mass and CO_2 concentration. The factor $(1-\zeta\bar{T})$ is less than 1, reducing the drawdown relative to the zero-feedback case: the climate system "holds on" to more CO_2 in the atmosphere when it is warmer. For fitted values of $\zeta \sim 0.03$–$0.06\ \mathrm{K^{-1}}$ and warming levels $\bar{T} \lesssim 3$ K, the product $\zeta\bar{T}$ remains well below unity, so this factor stays positive and the linearisation is well-behaved for all realistic warming levels.

From drawdown to forcing. To convert the CO_2 change into a radiative forcing change, we linearise the logarithmic forcing relationship $F = a\ln(1 + C/C_0)$ around the cessation concentration $C(t_0)$. This gives a forcing sensitivity of $a/(C_0 + C(t_0))$ ($\mathrm{W\,m^{-2}\,ppm^{-1}}$), so the forcing perturbation from pool i's drawdown is:

$$\delta F_i(\Delta t) \;\approx\; -\frac{a}{C_0 + C(t_0)}\,(1-\zeta\bar{T})\,\frac{E_0}{\kappa}\,f_i\,\tau_i\big(1-e^{-t_0/\tau_i}\big)\big(1-e^{-\Delta t/\tau_i}\big). \tag{S7}$$

It is convenient to define the *ultimate forcing change* that would result from complete drawdown of pool i (i.e. as $\Delta t \to \infty$):

$$\delta F_{i,\infty} \;\approx\; -\frac{a}{C_0 + C(t_0)}\,(1-\zeta\bar{T})\,\frac{E_0}{\kappa}\,f_i\,\tau_i\big(1-e^{-t_0/\tau_i}\big). \tag{S8}$$

This remains an approximation because the linearisation holds only while $C(t)$ stays close to $C(t_0)$; for very large drawdown (i.e. $C(t_\infty) \ll C(t_0)$) the logarithmic curvature of the forcing relationship would need to be retained. With this definition the time-dependent forcing from pool i takes the simple form $\delta F_i(\Delta t) = \delta F_{i,\infty}\,(1 - e^{-\Delta t/\tau_i})$.

Temperature response to a gradually building forcing. Each carbon pool's forcing builds on its own timescale τ_i, and each climate eigenmode j responds on timescale τ_j^{clim}. The interplay between these two timescales is captured by a *mixed response function* $\Phi_{ij}(\Delta t)$, which we now derive.

The forcing from pool i grows as $\delta F_i(t') = \delta F_{i,\infty}(1 - e^{-t'/\tau_i})$ for $t' \geq 0$ (where $t' = t - t_0$ is time after cessation). Climate eigenmode j has an impulse response $G_j(t') = (q_j/\lambda_{\mathrm{eff}})\,e^{-t'/\tau_j^{\mathrm{clim}}}/\tau_j^{\mathrm{clim}}$, so the temperature change in mode j driven by pool i is the convolution

$$\delta T_{ij}(\Delta t) \;=\; \frac{q_j\,\delta F_{i,\infty}}{\lambda_{\mathrm{eff}}}\int_0^{\Delta t}\frac{1}{\tau_j^{\mathrm{clim}}}\,e^{-(\Delta t - t')/\tau_j^{\mathrm{clim}}}\,\big(1-e^{-t'/\tau_i}\big)\,\mathrm{d}t'. \tag{S9}$$

Splitting the integrand and evaluating the two resulting integrals:

$$\int_0^{\Delta t}\frac{e^{-(\Delta t - t')/\tau_j^{\mathrm{clim}}}}{\tau_j^{\mathrm{clim}}}\,\mathrm{d}t' \;=\; 1 - e^{-\Delta t/\tau_j^{\mathrm{clim}}}, \tag{S10}$$

$$\int_0^{\Delta t}\frac{e^{-(\Delta t - t')/\tau_j^{\mathrm{clim}}}}{\tau_j^{\mathrm{clim}}}\,e^{-t'/\tau_i}\,\mathrm{d}t' \;=\; \frac{\tau_i}{\tau_j^{\mathrm{clim}} - \tau_i}\big(e^{-\Delta t/\tau_j^{\mathrm{clim}}} - e^{-\Delta t/\tau_i}\big), \tag{S11}$$

where the second integral requires $\tau_j^{\text{clim}} \neq \tau_i$ (the degenerate case does not arise in practice). Subtracting eq. (S11) from eq. (S10) and collecting terms gives:

$$\Phi_{ij}(\Delta t) \;=\; 1 - \frac{\tau_j^{\text{clim}}}{\tau_j^{\text{clim}} - \tau_i}\, e^{-\Delta t/\tau_j^{\text{clim}}} + \frac{\tau_i}{\tau_j^{\text{clim}} - \tau_i}\, e^{-\Delta t/\tau_i}. \tag{S12}$$

$\Phi_{ij}(\Delta t)$ gives the fraction of the equilibrium temperature response of climate mode j that has been realised at time Δt, when driven by a forcing that itself saturates exponentially on timescale τ_i.

Three limiting cases clarify the behaviour of Φ_{ij}:

- **Long time** ($\Delta t \to \infty$): both exponentials vanish, $\Phi_{ij} \to 1$. The system fully equilibrates.
- **Fast carbon, slow climate** ($\tau_i \ll \tau_j^{\text{clim}}$): the CO_2 is drawn down quickly, so the forcing appears almost like a step change to the slow climate mode. In this limit $\Phi_{ij} \approx 1 - e^{-\Delta t/\tau_j^{\text{clim}}}$, the standard step-response of mode j.
- **Slow carbon, fast climate** ($\tau_i \gg \tau_j^{\text{clim}}$): the climate mode adjusts rapidly relative to the forcing change, and $\Phi_{ij} \approx 1 - e^{-\Delta t/\tau_i}$, the climate mode tracks the forcing quasi-statically, and the timescale of the temperature response is set by the carbon pool, not by the climate.

Co-existing regimes in practice. In any given model, multiple (i, j) pairings span these limiting cases simultaneously. The fastest carbon pool ($\tau_i \sim 3$ yr) is much faster than the slow climate mode ($\tau_2^{\text{clim}} \sim$ 200–500 yr), so that pairing operates in the "fast carbon, slow climate" regime: the forcing from this pool appears as a near-step to the slow mode, which responds only sluggishly. At the same time, the slowest carbon pool ($\tau_i \sim 200$ yr) is much slower than the fast climate mode ($\tau_1^{\text{clim}} \sim 4$ yr), placing that pairing in the "slow carbon, fast climate" regime: the fast mode tracks the slow drawdown quasi-statically. Intermediate pairings fall between these extremes. As a result, the dominant behaviour of ZEC can shift over time as faster-acting pairings saturate and slower ones take over. The ocean heat uptake efficacy ε further modulates this picture by controlling the relative weights q_1 and q_2 of the two climate modes, and hence how much of the drawdown cooling is realised quickly versus slowly.

Full carbon component. Summing over all carbon pools and both climate eigenmodes, the carbon component of ZEC is:

$$\text{ZEC}_{\text{carbon}}(\Delta t) \;=\; \frac{1}{\lambda_{\text{eff}}} \sum_i \delta F_{i,\infty} \sum_j q_j\, \Phi_{ij}(\Delta t). \tag{S13}$$

Since every $\delta F_{i,\infty} < 0$ (drawdown reduces forcing, provided $(1-\zeta\bar{T}) > 0$; see the discussion following eq. (S6)), and $\Phi_{ij} > 0$, this term is always negative: the carbon cycle cools the surface after emissions cease.

S1.3 Total ZEC: Level 1

With both components established, adding the thermal component (eq. (S5)), which captures the continued warming from ocean heat already "in the pipeline," to the carbon

component (eq. (S13)), which captures the cooling from CO_2 drawdown, gives the complete Level 1 expression:

$$\mathrm{ZEC}(\Delta t) = \underbrace{T_{\mathrm{eq}} \sum_j q_j \, \frac{\tau_j^{\mathrm{clim}}}{t_0} (1 - e^{-t_0/\tau_j^{\mathrm{clim}}})(1 - e^{-\Delta t/\tau_j^{\mathrm{clim}}})}_{\text{thermal: unrealised warming (always} > 0)} + \underbrace{\frac{1}{\lambda_{\mathrm{eff}}} \sum_i \delta F_{i,\infty} \sum_j q_j \, \Phi_{ij}(\Delta t)}_{\text{carbon: CO}_2\text{ drawdown cooling (always} < 0)} \quad . \tag{S14}$$

Interpretation. The sign and magnitude of ZEC are determined by the race between these two terms. The structural parallel between them is instructive: expanding $\delta F_{i,\infty}$ and Φ_{ij} in the carbon term reveals an analogous product of timescale ratios:

$$\mathrm{ZEC}_{\mathrm{carbon}}(\Delta t) \;=\; \frac{T_{\mathrm{eq}}}{\lambda_{\mathrm{eff}}} \sum_i \sum_j \underbrace{(-\delta F_{i,\infty})}_{\text{forcing from pool } i} \cdot q_j \cdot \underbrace{\Phi_{ij}(\Delta t)}_{\text{climate mode } j \text{ response}} \quad , \tag{S15}$$

where each $\delta F_{i,\infty}$ contains a factor $\tau_i(1 - e^{-t_0/\tau_i})$ mirroring the thermal term's $\tau_j^{\mathrm{clim}}(1 - e^{-t_0/\tau_j^{\mathrm{clim}}})/t_0$, and Φ_{ij} plays the role of the post-cessation relaxation factor $(1 - e^{-\Delta t/\tau_j^{\mathrm{clim}}})$ but now involves *two* timescales (τ_i and τ_j^{clim}) rather than one. This makes the "race" between thermal delivery and carbon drawdown visible as a comparison of timescale ratios: the thermal term is governed by $\tau_j^{\mathrm{clim}}/t_0$ and $\Delta t/\tau_j^{\mathrm{clim}}$, while the carbon term is governed by τ_i/t_0 and $\Delta t/\tau_i$ (modified by the climate mode's response time through Φ_{ij}).

The thermal term grows as the slow ocean mode delivers stored heat to the surface; the carbon term deepens as CO_2 drawdown cools the surface. If the pipeline delivers heat faster than the carbon sinks can cool, ZEC is positive (continued warming); if the carbon sinks win the race, ZEC is negative (cooling). The main text shows that these two terms nearly cancel in the flat10 ensemble fits, with a mean ZEC_{50} of only -0.014 K despite the thermal term being $+0.216$ K and the carbon term -0.231 K.

Sources of approximation error. The Level 1 expression incurs a small positive bias (0.018–0.064 K) because it approximates the emission-phase forcing as a perfect linear ramp. In reality, the forcing trajectory is slightly concave: (i) the logarithmic CO_2–forcing relationship means each additional ppm of CO_2 contributes *less* forcing, and (ii) carbon uptake during the emission phase partially offsets the atmospheric CO_2 increase. Both effects make the actual forcing slightly below the linear ramp assumed, and the Level 1 formula correspondingly overestimates ZEC by a small amount.

S2 Reduction to aggregate metrics: Level 2

The Level 1 formula (eq. (S14)) retains all 11 parameters without further reduction, but involves several linearisation assumptions (linear ramp forcing, log-forcing linearisation, carbon–climate feedback evaluated at $\bar{T} \approx T_0/2$). This section shows how a series of physically motivated approximations reduces it to a compact expression involving only *aggregate* climate and carbon-cycle metrics, quantities that can be estimated from observations or constrained by simple experiments. This is the "Level 2" formula used in the main text.

The strategy proceeds in two steps: (i) simplify the thermal term by exploiting the timescale separation between the fast and slow modes, and (ii) collapse the multi-pool

carbon cycle to a single effective timescale and correct for the slow climate mode's partial response.

S2.1 Step 1: Simplifying the thermal term

Recall from section S1.1 that the two climate eigenmodes have very different timescales: $\tau_1^{\text{clim}} \sim 4$ yr and $\tau_2^{\text{clim}} \sim 200$–$500$ yr, while the ramp duration is $t_0 = 100$ yr and the post-cessation horizon of interest is $\Delta t = 50$ yr.

The fast mode contributes a small constant offset. Because $\tau_1^{\text{clim}} \ll t_0$ and $\tau_1^{\text{clim}} \ll \Delta t$, the fast mode's ramp-phase disequilibrium is $\sim T_{\text{eq}}\, q_1\, \tau_1^{\text{clim}}/t_0$ (about 4% of its equilibrium share), and this residual is delivered almost instantly after cessation (e-folding time $\sim$4 yr). The fast mode therefore contributes a small, rapidly delivered constant $\sim T_{\text{eq}}\, q_1\, \tau_1^{\text{clim}}/t_0$ to ZEC, which for typical parameter values amounts to a few thousandths of a Kelvin. We neglect it here, but note that in the limiting case of a fully decoupled deep ocean ($q_1 \to 1$) this term would become the dominant thermal contribution; the approximation is valid because the fitted models all have substantial ocean coupling ($q_2 \approx 0.25$–0.45).

Direct derivation via the warming pipeline. Having established that the fast mode equilibrates quickly, the total unrealised warming at cessation is well approximated by the difference between the equilibrium temperature and the warming already realised:

$$\Delta T_{\text{pipeline}} \;=\; T_{\text{eq}} - T_0 \;=\; T_0\left(\frac{1}{r} - 1\right), \tag{S16}$$

where $r \equiv T_0/T_{\text{eq}}$ is the *realised warming fraction*. This pipeline warming resides almost entirely in the slow eigenmode (since the fast mode has already delivered its share).

After cessation, the slow mode delivers this stored heat at rate $1/\tau_2^{\text{clim}}$. For $\Delta t \ll \tau_2^{\text{clim}}$ (which holds for $\Delta t = 50$–100 yr and $\tau_2^{\text{clim}} = 200$–$500$ yr), linearising the exponential gives $1 - e^{-\Delta t/\tau_2^{\text{clim}}} \approx \Delta t/\tau_2^{\text{clim}}$. The fraction of the pipeline delivered in Δt years is therefore simply $\Delta t/\tau_2^{\text{clim}}$, yielding:

$$\boxed{\text{ZEC}_{\text{thermal}}(\Delta t) \;\approx\; T_0\left(\frac{1}{r} - 1\right)\frac{\Delta t}{\tau_2^{\text{clim}}}.} \tag{S17}$$

Physical interpretation. This compact result is the product of two factors:

- $T_0(1/r - 1) = T_{\text{eq}} - T_0$: the total unrealised warming at cessation, the "warming in the pipeline."
- $\Delta t/\tau_2^{\text{clim}}$: the fraction of this pipeline delivered in Δt years. Because τ_2^{clim} is typically several hundred years, only a small fraction ($\sim$10–25%) arrives by $\Delta t = 50$ yr.

The thermal ZEC is thus the product of *how much warming is stored* and *how fast it is released*. This explains why the thermal ZEC can be modest even when the pipeline is large: the slow ocean timescale acts as a "delivery bottleneck."

S2.2 Step 2: Collapsing the carbon cycle and the q_{eff} correction

To reduce the multi-pool carbon sum to a single effective timescale, we define

$$\tau_c \;=\; \frac{\sum_i f_i\,\tau_i^2\,(1-e^{-t_0/\tau_i})}{\sum_i f_i\,\tau_i\,(1-e^{-t_0/\tau_i})}. \tag{S18}$$

This is a choice, not a derivation: τ_c is the weighted average timescale that matches the initial drawdown rate of the full multi-pool sum, with weights proportional to each pool's accumulated carbon excess at cessation. Typical values range from 30–60 years.

Similarly, the total equilibrium forcing change from complete drawdown (repeating eq. (S8) summed over all pools, for reference) is:

$$\delta F_{\text{tot},\infty} \;=\; \sum_i \delta F_{i,\infty} \;=\; -(1-\zeta\bar{T})\,\frac{a}{C_0+C(t_0)}\,\frac{E_0}{\kappa}\sum_i f_i\,\tau_i(1-e^{-t_0/\tau_i}). \tag{S19}$$

Why the fast-mode-only approximation is insufficient. As a first attempt, one might assume only the fast climate mode responds to the drawdown forcing within 50 years. Because $\tau_1^{\text{clim}} \sim 4$ yr $\ll \tau_c \sim$ 30–60 yr $\ll \tau_2^{\text{clim}} \sim$ 200–500 yr, the fast mode tracks the forcing quasi-statically while the slow mode barely responds, giving $\text{ZEC}_{\text{carbon}}^{(q_1)} \approx (q_1/\lambda_{\text{eff}})\,\delta F_{\text{tot},\infty}(1-e^{-\Delta t/\tau_c})$. But this systematically *underestimates* the cooling by ∼0.030 K because the slow climate mode does accumulate a partial response over 50 years. To correct for this, we replace the pure fast-mode weight q_1 with an *effective* climate response fraction q_{eff} that includes the slow mode's partial contribution.

The q_{eff} correction. Working with the single effective carbon timescale τ_c, the carbon ZEC can be factored as:

$$\text{ZEC}_{\text{carbon}}(\Delta t) \;=\; \frac{\delta F_{\text{tot},\infty}}{\lambda_{\text{eff}}}\,(1-e^{-\Delta t/\tau_c})\;\underbrace{\left[q_1 + q_2\,\frac{\Phi(\Delta t,\,\tau_c,\,\tau_2^{\text{clim}})}{1-e^{-\Delta t/\tau_c}}\right]}_{\equiv\, q_{\text{eff}}(\Delta t)}, \tag{S20}$$

where $\Phi(\Delta t,\,\tau_c,\,\tau_2^{\text{clim}})$ is the mixed response function (eq. (S12)) evaluated for the effective carbon timescale and the slow climate mode.

This defines:

$$\boxed{q_{\text{eff}}(\Delta t) \;=\; q_1 + q_2\,\frac{\Phi(\Delta t,\,\tau_c,\,\tau_2^{\text{clim}})}{1-e^{-\Delta t/\tau_c}}.} \tag{S21}$$

Physical meaning. q_{eff} represents the fraction of the equilibrium temperature response to the total drawdown forcing that has been realised by time Δt. It has two additive contributions:

1. The fast mode contributes its full weight q_1, because it equilibrates quasi-instantaneously to the forcing at every moment.

2. The slow mode contributes q_2 scaled down by the ratio $\Phi/(1-e^{-\Delta t/\tau_c})$, which measures how much of the slow mode's response has accumulated relative to the forcing change.

Typical values at $\Delta t = 50$ yr imply $q_{\text{eff}} \approx 0.58$–$0.78$, compared to $q_1 \approx 0.55$–0.75 (see section S3 for the eigenvector decomposition that determines these weights). The correction adds 0.03–0.05 to the response fraction, accounting for $\mathcal{O}(0.030)$ K of additional cooling that the uncorrected formula misses.

In the limiting case of an infinitely slow deep ocean ($\tau_2^{\text{clim}} \to \infty$): $\Phi \to 0$ and $q_{\text{eff}} \to q_1$. This is the fully decoupled limit where the eigenmodes align with the physical boxes and forcing excites only the fast (surface) mode, so $q_1 = 1$ exactly.

Assembling the Level 2 formula. Combining the thermal term (eq. (S17)) with the carbon term with the q_{eff} correction (eq. (S20)), we obtain:

$$\boxed{\text{ZEC}(\Delta t) \;\approx\; \underbrace{T_0 \,\frac{(1/r - 1)\,\Delta t}{\tau_2^{\text{clim}}}}_{\substack{\text{thermal:}\\ \text{pipeline delivery}}} \;+\; \underbrace{\frac{\delta F_{\text{tot},\infty}}{\lambda_{\text{eff}}}\, q_{\text{eff}}\, \left(1 - e^{-\Delta t/\tau_c}\right)}_{\substack{\text{carbon:}\\ \text{CO}_2\text{ drawdown cooling}}},} \tag{S22}$$

where the total equilibrium drawdown cooling $\delta F_{\text{tot},\infty}/\lambda_{\text{eff}}$ is given by eq. (S19). This quantity is always negative (drawdown reduces forcing), so the carbon term opposes the positive thermal term.

Interpretation of the Level 2 formula. The Level 2 expression captures the competition between two processes:

1. The **thermal term** is the product of the unrealised warming at cessation, $T_0(1/r - 1) = T_{\text{eq}} - T_0$, and the fraction delivered in Δt years, $\Delta t/\tau_2^{\text{clim}}$. Models with low r (large pipeline) or short τ_2^{clim} (fast delivery) have larger thermal ZEC.

2. The **carbon term** is the equilibrium drawdown cooling $\delta F_{\text{tot},\infty}/\lambda_{\text{eff}}$, scaled by the effective climate response fraction q_{eff} and the fraction of drawdown realised by time Δt. Larger removable CO_2 fraction ($1 - f_\infty^{\text{eff}}$), shorter τ_c, and larger q_{eff} all strengthen the cooling.

Both terms grow approximately linearly with Δt while Δt remains small relative to τ_2^{clim} and τ_c; the sign of ZEC is thus set by the difference of two rates and is approximately time-invariant in this regime. This separation makes it immediately apparent that ZEC uncertainty is governed by whichever rate is less well constrained. As shown in the main text, the thermal rate has roughly twice the inter-model spread of the carbon rate, explaining why thermal parameters (τ_2^{clim}, ocean coupling) dominate ZEC uncertainty.

S3 Eigenvectors and their relation to the box temperatures.

To understand why the eigenstructure matters, recall how it connects to the time-dependent solution. The homogeneous part of the two-box system (i.e. the free decay after forcing is removed) takes the matrix form $\mathrm{d}\mathbf{T}/\mathrm{d}t = \mathbf{A}\,\mathbf{T}$. For a general linear system of this kind, solutions are superpositions of modes $\mathbf{T}(t) = \sum_j c_j\,\mathbf{v}_j\,e^{\mu_j t}$, where μ_j and $\mathbf{v}_j$ are the eigenvalues and eigenvectors of $\mathbf{A}$, and the coefficients c_j are set by initial conditions. Each eigenvalue μ_j satisfies the *characteristic equation*

$$\det(\mathbf{A} - \mu\,\mathbf{I}) = 0 \qquad \Longrightarrow \qquad \mu^2 - \text{tr}(\mathbf{A})\,\mu + \det(\mathbf{A}) = 0, \tag{S23}$$

whose two roots are Eq. (8) in the main text. Each root defines a timescale $\tau_j^{\mathrm{clim}} = -1/\mu_j$ and a spatial pattern (the eigenvector $\mathbf{v}_j$) describing *how* the warming or cooling is distributed between the two boxes in that mode.

When forcing is present, the full (particular + homogeneous) solution for a step forcing F applied at $t = 0$ is

$$T_f(t) = \frac{F}{\lambda_f}\Big[1 - q_1\, e^{-t/\tau_1^{\mathrm{clim}}} - q_2\, e^{-t/\tau_2^{\mathrm{clim}}}\Big], \tag{S24}$$

where the eigenmode weights q_1 and q_2 (with $q_1 + q_2 = 1$) are determined by projecting the forcing vector $\mathbf{b} = (1/C_f,\, 0)^\top$ onto the eigenvectors. Specifically, $q_j \propto v_{j,f}$, the first (surface) component of $\mathbf{v}_j$. This is why the structure of the eigenvectors directly controls how much of the equilibrium warming is delivered quickly versus slowly.

With this motivation, we now derive the eigenvectors explicitly. The eigenvector $\mathbf{v}_j$ associated with eigenvalue μ_j satisfies $(\mathbf{A} - \mu_j \mathbf{I})\,\mathbf{v}_j = 0$. From the first row of the system matrix (eq. (S28)), the ratio of the deep-ocean to surface components of each eigenvector is

$$\frac{v_{j,s}}{v_{j,f}} = \frac{(\lambda_f + \varepsilon\,\gamma) + \mu_j\, C_f}{\varepsilon\,\gamma}. \tag{S25}$$

In the limit $C_s \gg C_f$, the approximate eigenvalues $\mu_1 \approx -(\lambda_f + \varepsilon\,\gamma)/C_f$ and $\mu_2 \approx -\lambda_f\,\gamma/[C_s(\lambda_f + \varepsilon\,\gamma)]$ yield:

$$\text{Fast mode:} \quad \frac{v_{1,s}}{v_{1,f}} \;\approx\; \mathcal{O}\big(C_f/C_s\big) \;\to\; 0, \tag{S26}$$

$$\text{Slow mode:} \quad \frac{v_{2,s}}{v_{2,f}} \;\approx\; 1 + \frac{\lambda_f}{\varepsilon\,\gamma}. \tag{S27}$$

Equation (S26) shows that the fast eigenvector is essentially $\mathbf{v}_1 \approx (1,\, 0)^\top$: this mode is an almost pure surface adjustment with negligible deep-ocean participation. The slow eigenvector, by contrast, involves *both* boxes: the deep ocean warms, and it drags the surface along. For typical fitted parameters ($\lambda_f \sim 1.2$ W m^{-2} K^{-1}, $\gamma \sim 0.7$ W m^{-2} K^{-1}, $\varepsilon \sim 1.3$), $v_{2,s}/v_{2,f} \approx 2.3$, so the deep-ocean component of the slow mode is roughly twice the surface component.

These results confirm that the numeric eigenmode indices ($j = 1$ for the fast mode, $j = 2$ for the slow mode) correspond closely to the physical box decomposition: the fast eigenmode maps almost exactly onto the surface box, while the slow eigenmode represents the coupled deep-ocean equilibration. The correspondence becomes exact in the limit $C_s/C_f \to \infty$.

Because the observable temperature is $T = T_f$ (the first component of the state vector), the eigenmode weight q_j that appears in the step-response formula is proportional to $v_{j,f}$, explaining why $q_1 > q_2$: the forcing vector $\mathbf{b} = (1/C_f,\, 0)^\top$ projects predominantly onto the fast mode.

S4 Efficacy extension

The eigenmode decomposition in the preceding sections is valid for *any* choice of two-box eigenvalues τ_j^{clim} and weights q_j; those sections did not need to specify how the eigenvalues depend on the underlying physical parameters. Here we write down the system matrix

explicitly, show that introducing an efficacy parameter $\varepsilon \neq 1$ changes the eigenvalues while leaving the decomposition structure intact, and derive the consequences for ZEC.

Physically, the "pattern effect" (the observation that the spatial pattern of warming evolves over time, changing the effective radiative feedback) can be represented in the two-box framework by an efficacy parameter ε that modifies the heat exchange between the two boxes.

The efficacy parameter ε is estimated from top-of-atmosphere (TOA) radiation data following the Gregory method. This analysis was performed for eight of the ten ESMs; MPI-ESM1-2-LR and UVic-ESCM-2-10 lack the required TOA radiation variables (rlut, rsut, rsdt) and are therefore excluded from the efficacy analysis.

S4.1 How efficacy modifies the eigenstructure

With efficacy $\varepsilon \neq 1$, the system matrix of the two-box model becomes:

$$\mathbf{A} = \begin{pmatrix} -(\lambda_f + \varepsilon\,\gamma)/C_f & \varepsilon\,\gamma/C_f \\ \gamma/C_s & -\gamma/C_s \end{pmatrix}. \tag{S28}$$

What ε does physically. When $\varepsilon > 1$, heat is extracted from the surface layer *more efficiently* during the transient period (when the surface is warmer than the deep ocean). This represents the observation that during transient warming, the pattern of surface temperature change preferentially warms regions where climate feedbacks are least stabilising, causing the effective feedback to be stronger (more negative) than the equilibrium value.

Effect on eigenvalues. The trace and determinant of $\mathbf{A}$ are:

$$\mathrm{tr}(\mathbf{A}) = -\frac{\lambda_f + \varepsilon\,\gamma}{C_f} - \frac{\gamma}{C_s}, \tag{S29}$$

$$\det(\mathbf{A}) = \frac{\lambda_f\,\gamma}{C_f\,C_s}. \tag{S30}$$

A key observation: **the determinant is independent of ε.** Since the eigenvalues μ_1, μ_2 satisfy $\mu_1 + \mu_2 = \mathrm{tr}(\mathbf{A})$ and $\mu_1 \cdot \mu_2 = \det(\mathbf{A})$, increasing ε increases the sum of the (negative) eigenvalues while keeping their product fixed. The consequence is a widening of the timescale separation:

- The fast mode speeds up (τ_1^{clim} decreases): the surface adjusts more rapidly because the enhanced coupling extracts heat more efficiently.
- The slow mode slows down (τ_2^{clim} increases): the deep ocean must absorb more heat through the same coupling before the temperature difference equilibrates.
- The ratio $\tau_2^{\mathrm{clim}}/\tau_1^{\mathrm{clim}}$ widens.

Numerically, increasing ε from 1.0 to 1.4 typically lengthens τ_2^{clim} by $\sim$23% and shortens τ_1^{clim} by $\sim$18%.

S4.2 Consequences for TCR and the realised warming fraction

The equilibrium sensitivity ECS $= F_{2\times}/\lambda_f$ is unaffected by ε, because at equilibrium $T_f = T_s$ and the heat exchange term vanishes: the two boxes are at the same temperature and no heat flows between them.

The *transient* response, however, is suppressed by efficacy. To see how, note that in the $C_s \gg C_f$ limit the approximate eigenvalues and eigenmode weights are:

$$\tau_1^{\text{clim}}(\varepsilon) \;\approx\; \frac{C_f}{\lambda_f + \varepsilon\,\gamma}, \qquad\qquad q_1(\varepsilon) \;=\; \frac{\lambda_f}{\lambda_f + \varepsilon\,\gamma}, \tag{S31}$$

$$\tau_2^{\text{clim}}(\varepsilon) \;\approx\; \frac{C_s\,(\lambda_f + \varepsilon\,\gamma)}{\lambda_f\,\gamma}, \qquad\qquad q_2(\varepsilon) \;=\; \frac{\varepsilon\,\gamma}{\lambda_f + \varepsilon\,\gamma}. \tag{S32}$$

Increasing ε speeds up the fast mode (enhanced surface coupling extracts heat more efficiently) and slows down the slow mode (the deep ocean must absorb more heat before equilibrating). The eigenmode weights shift accordingly: q_2 grows at the expense of q_1, so more of the equilibrium response is channeled through the slow mode.

The realised warming fraction is therefore:

$$r(\varepsilon) = \frac{T_0}{T_{\text{eq}}} \approx 1 - q_2(\varepsilon)\,\frac{\tau_2^{\text{clim}}(\varepsilon)}{t_0} = 1 - \frac{\varepsilon\,C_s}{\lambda_f\,t_0}. \tag{S33}$$

The product $q_2\,\tau_2^{\text{clim}} = \varepsilon\,C_s/\lambda_f$ is linear in ε: the larger weight assigned to the slow mode and the longer slow timescale compound rather than cancel. Since τ_2^{clim} increases with ε, the realised fraction r decreases: more warming is pushed into the pipeline at cessation. This is the key distinction from simply increasing ECS: $\varepsilon > 1$ **reduces** r **without changing ECS**. It makes the transient warming fraction smaller while leaving the equilibrium target unchanged.

S4.3 Consequences for ZEC

How does this efficacy-dependent eigenstructure affect the post-cessation temperature evolution?

Thermal term. From eq. (S17), the thermal ZEC scales as:

$$\text{ZEC}_{\text{thermal}}(\Delta t) \;\propto\; T_0\,\frac{(1/r - 1)\,\Delta t}{\tau_2^{\text{clim}}}. \tag{S34}$$

Using the explicit ε-dependence from eqs. (S32) and (S33), the pipeline $(1/r - 1) \approx q_2\,\tau_2^{\text{clim}}/t_0$ and the delivery rate $1/\tau_2^{\text{clim}}$ give:

$$\frac{(1/r - 1)}{\tau_2^{\text{clim}}} \;=\; \frac{q_2(\varepsilon)}{t_0} \;=\; \frac{\varepsilon\,\gamma}{t_0\,(\lambda_f + \varepsilon\,\gamma)}. \tag{S35}$$

The τ_2^{clim} in the pipeline and the τ_2^{clim} in the delivery rate cancel exactly, leaving the thermal ZEC rate dependent on ε only through $q_2(\varepsilon)$. Since q_2 increases monotonically with ε (more of the equilibrium response is channelled through the slow mode), the thermal ZEC *increases* with ε: the growth of the pipeline outpaces the slowing of its delivery.

Carbon term. With $\varepsilon > 1$, τ_2^{clim} increases, which reduces $\Phi(\Delta t, \tau_c, \tau_2^{\text{clim}})$ and hence pushes q_{eff} toward q_1. The magnitude of the carbon cooling is therefore reduced: less of the slow mode's partial response accumulates within Δt years.

Net ZEC. Increasing ε pushes ZEC positive through two reinforcing channels: the thermal term grows (larger q_2) while the carbon cooling weakens (q_{eff} shifts toward q_1). For typical flat10 parameters, increasing ε from 1.0 to 1.4 shifts ZEC_{50} by $\mathcal{O}(0.010\text{–}0.030)$ K, a modest effect in the flat10 regime, but one that would grow for higher-emission scenarios.

S4.4 Structural asymmetry: buildup vs. drawdown

The efficacy framework reveals a subtle structural asymmetry in the climate system's response before and after emissions cease.

During buildup. While emissions are ongoing, the surface is substantially warmer than the deep ocean ($T_f > T_s$), and heat flows downward. With $\varepsilon > 1$, this heat extraction is enhanced, so the effective feedback experienced by the fast box during the transient is:

$$\lambda_{\text{eff,transient}} = \lambda_f + (\varepsilon - 1)\,\gamma\,(T_f - T_s)/T_f > \lambda_f \qquad (\text{for } \varepsilon > 1). \tag{S36}$$

This stronger effective feedback *suppresses* transient warming. As the system relaxes toward equilibrium, the temperature difference $T_f - T_s$ shrinks toward zero, and the effective feedback returns to its equilibrium value λ_f. The post-cessation warming therefore proceeds under a *weaker* effective feedback than during the buildup phase.

This asymmetry implies that the post-cessation warming rate should *exceed* what a constant-λ model fitted to the buildup phase would predict. A model fitted with $\varepsilon = 1$ absorbs the pattern effect into an inflated feedback $\lambda_f^* = \lambda_f + (\varepsilon - 1)\gamma$, correctly reproducing the transient response but then applying this same (too strong) feedback to the relaxation phase, systematically underestimating the thermal ZEC.

S4.5 Sensitivity of ZEC_{50} to ε

From eq. (S35), the thermal ZEC rate depends on ε through $q_2(\varepsilon)$. Differentiating:

$$\frac{\partial\,\text{ZEC}_{\text{thermal}}(\Delta t)}{\partial \varepsilon} \;\propto\; \frac{\partial q_2}{\partial \varepsilon} \;=\; \frac{\lambda_f\,\gamma}{(\lambda_f + \varepsilon\,\gamma)^2} \;>\; 0, \tag{S37}$$

which is always positive: increasing efficacy increases the thermal ZEC. Higher ε channels more of the equilibrium response through the slow mode (q_2 grows), enlarging the warming pipeline that is delivered after cessation. Meanwhile, the carbon cooling weakens (as q_{eff} shifts toward q_1). Both effects push ZEC positive, so in models where the thermal term dominates, increasing ε amplifies ZEC_{50}.

S4.6 Fitting considerations

Adding ε as a free parameter creates a fundamental identifiability problem: in the fast-box equation, ε and γ enter only as the product $\varepsilon\gamma$. From the temperature record alone, only this product is constrained; one cannot separately determine how much of the effective coupling is due to the "true" heat exchange rate and how much is due to the pattern effect.

Breaking this degeneracy requires additional independent constraints, such as multiple forcing scenarios, top-of-atmosphere radiative flux data, or separate ocean heat content measurements for well-mixed and deep layers.

In practice, two approaches are viable:

1. **Prescribed** ε: fix ε at a value diagnosed from abrupt-4×CO_2 simulations (which provide an equilibrium feedback estimate independent of the transient), then fit the remaining parameters.

2. **Sensitivity analysis**: fit the standard model with $\varepsilon = 1$ (so the fitted γ should be interpreted as the effective coupling $\varepsilon\gamma$), then use the analytical expressions above to explore how ZEC estimates shift across a plausible range of ε (1.0–1.8).

We adopt the second approach in the main text. The sensitivity analysis (Figure 9 in the main text) shows how ZEC_{50} varies across this range, bounding the structural uncertainty from the pattern effect.

S5 Extension to longer timescales

The Level 2 formula (eq. (S22)) was derived for $\Delta t = 50$ yr, exploiting the fact that $\Delta t/\tau_2^{\mathrm{clim}} \ll 1$ to linearise several exponentials. The question is then the temporal range over which these approximations remain valid. This section quantifies the accuracy of each approximation at longer horizons and shows how to extend the formula when needed.

Thermal term accuracy. The key approximation is $(1 - e^{-\Delta t/\tau_2^{\mathrm{clim}}}) \approx \Delta t/\tau_2^{\mathrm{clim}}$. For a representative $\tau_2^{\mathrm{clim}} = 200$ yr (the short end of the fitted range ∼200–500 yr, chosen as a conservative case that maximises the deviation from linearity), the relative error of this linearisation is:

Δt (yr)	Exact	Linear approx	
50	0.221	0.250	(12% error)
100	0.393	0.500	(27% error)
200	0.632	1.000	(58% error)

At $\Delta t = 50$ yr the linearisation is adequate; by $\Delta t = 100$ yr it introduces substantial error; and by $\Delta t = 200$ yr it is clearly inappropriate (the linear "approximation" exceeds 1, which is physically impossible for a $1 - e^{-x}$ function bounded by unity).

Fix: Keep the exponential expression. For horizons beyond 50 years, the thermal term should retain the full exponential form rather than using the linear approximation:

$$\mathrm{ZEC}_{\mathrm{thermal}}(\Delta t) \;\approx\; T_0\left(\frac{1}{r} - 1\right)\left(1 - e^{-\Delta t/\tau_2^{\mathrm{clim}}}\right). \tag{S38}$$

This corrected form captures the saturation of thermal delivery as $\Delta t \to \infty$: eventually all of the pipeline warming has been delivered and $\mathrm{ZEC}_{\mathrm{thermal}}$ asymptotes to $T_0(1/r - 1) = T_{\mathrm{eq}} - T_0$.

Carbon term accuracy. Two separate approximations degrade with Δt:

1. **Single τ_c.** At short times, the effective single timescale τ_c is dominated by the fast and medium uptake pools that are actively drawing down CO_2. At longer times, the slow pool (with $\tau_i \sim$ 200–400 yr) becomes increasingly important, and a single τ_c cannot simultaneously capture rapid early drawdown and slow late drawdown. The resulting error grows to ∼0.030–0.050 K at ZEC_{200}.

2. q_{eff} **convergence.** As Δt increases, the slow climate mode has more time to respond to the drawdown forcing, so q_{eff} increases toward unity:

Δt (yr)	Typical q_{eff}
50	0.58–0.78
100	0.70–0.88
200	0.85–0.95

   By $\Delta t = 200$ yr, $q_{\text{eff}} \approx 1$ and the distinction between fast-mode-only and full response becomes negligible. Conversely, the correction matters most at shorter horizons where the slow mode has had less time to respond.

Practical guidance.

- **ZEC_{50}**: the Level 2 formula with all linearisations is accurate to within ∼0.03 K.
- **ZEC_{100}**: Keep the exponential in the thermal term and recompute q_{eff} at $\Delta t =$ 100 yr. This gives useful accuracy (∼0.05 K).
- **ZEC_{200} and beyond**: the Level 1 formulation (keeping track of all pools and both modes) is recommended, as the single-timescale approximation becomes increasingly unreliable.

S6 Parameter–ZEC correlation scatter plots

Figures S1 and S2 show the full 3×9 scatter grid of ensemble parameter values versus ZEC at three horizons (25, 50, and 100 years), pooled across all nine ESMs. Each small point represents one ensemble member; coloured markers denote deterministic fits for individual ESMs. Pearson correlation coefficients and linear fits are shown for each panel.

S7 Asymptotic ZEC and identifiability of f_∞

Table S1 reports the permanent airborne fraction f_∞, slow carbon timescale τ_3, and analytically implied ZEC_∞ for each ESM's ensemble. The values are computed from the fitted parameters by letting all transient carbon pools decay to zero and allowing the climate to fully equilibrate to the residual forcing from the permanent airborne fraction. The 10th–90th percentile ranges are wide for every model, reflecting the fundamental tradeoff between f_∞ and τ_3: with only 100 years of post-cessation data, a large permanent fraction with a moderate slow pool is statistically indistinguishable from a small permanent fraction with a very slow pool. The resulting ZEC_∞ is dominated by slow-process carbon

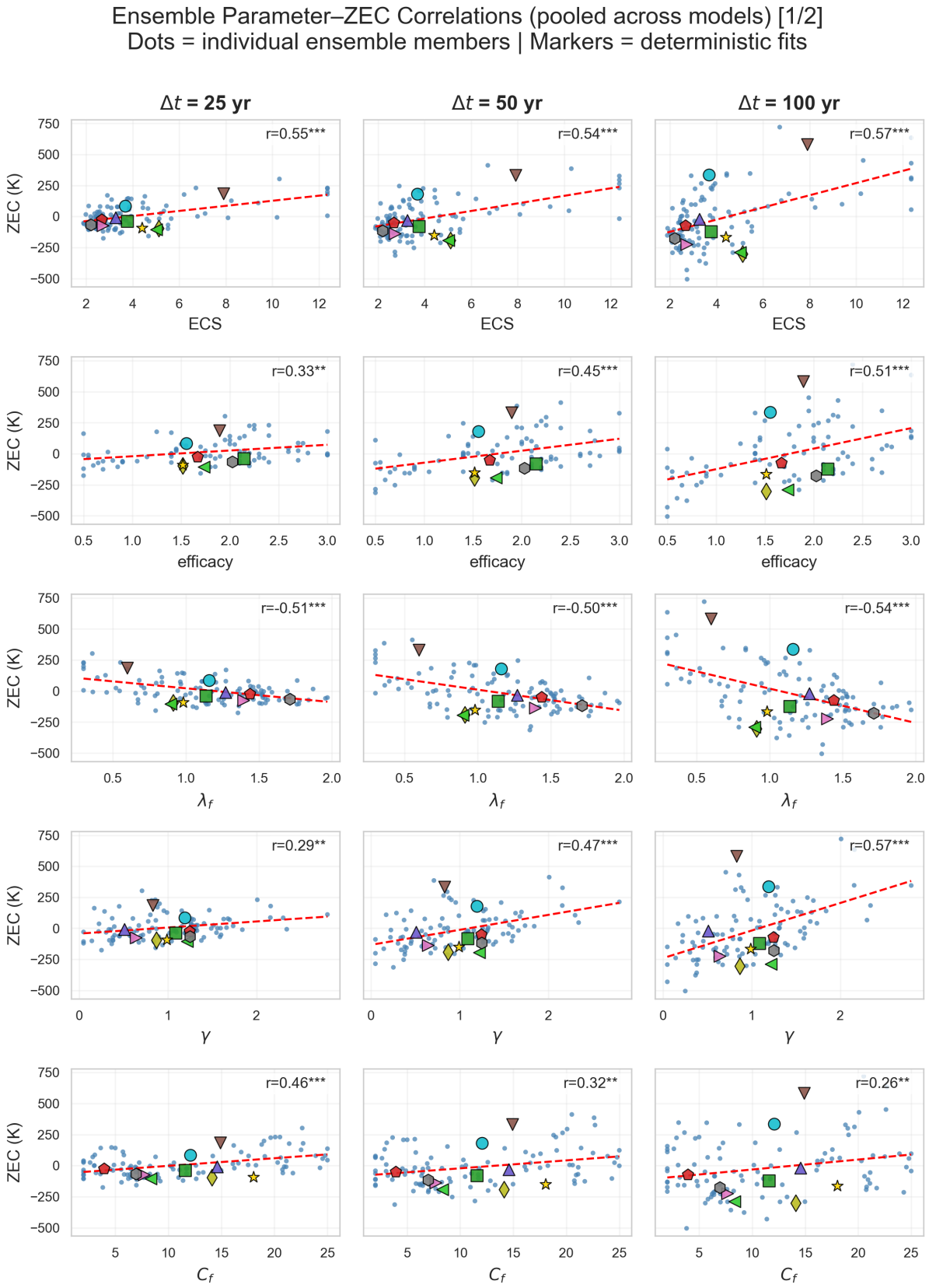

Figure S1: Ensemble parameter–ZEC correlations (3 horizons, first 5 parameters), pooled across nine ESMs. Small blue points: ensemble members. Coloured markers: deterministic fits. Red lines: linear fits with Pearson r annotated. Stars denote significance at $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***). Continued in fig. S2.

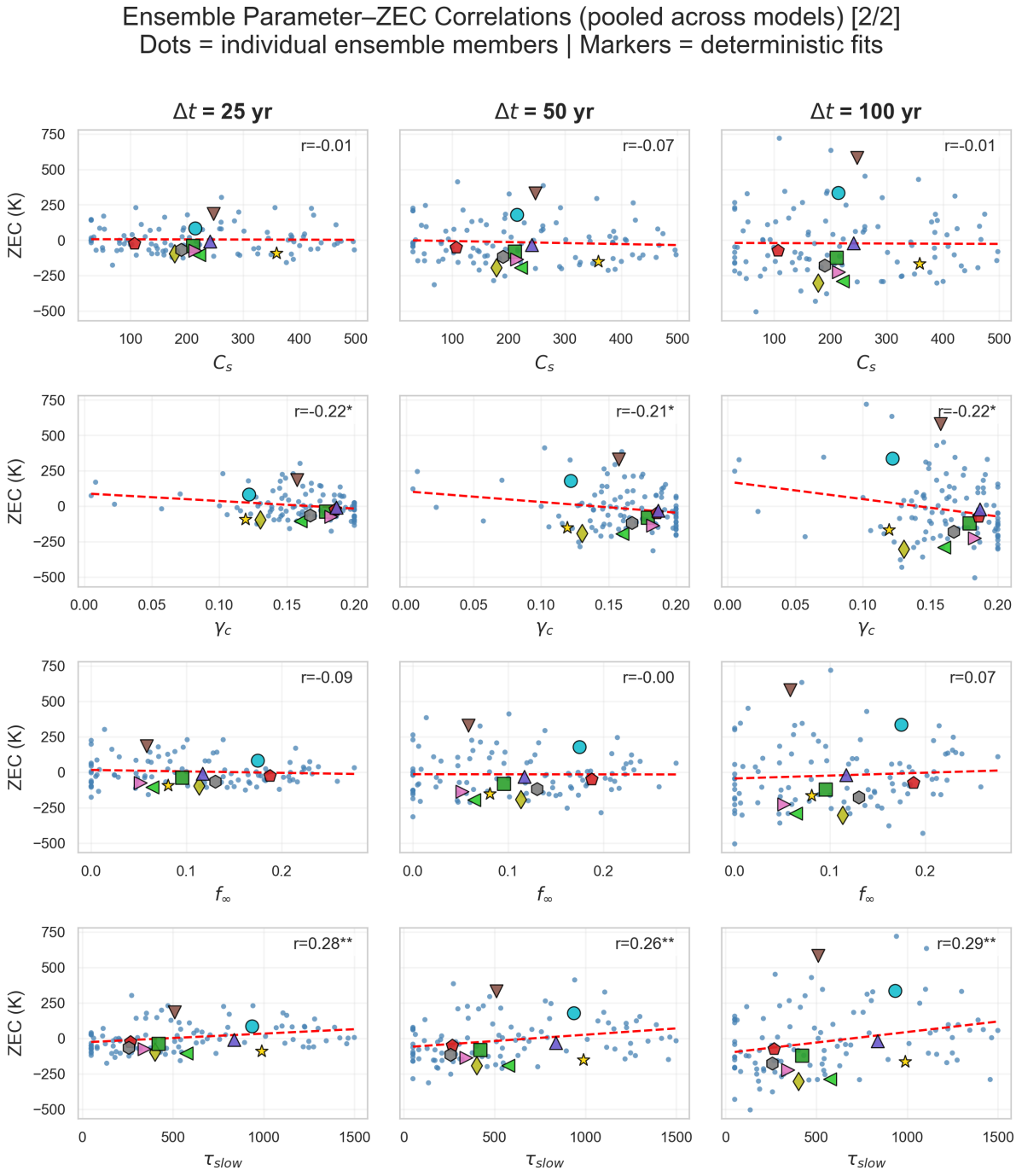

Figure S2: As fig. S1, for remaining 4 parameters.

uptake (deep ocean overturning and slow soil equilibration) and is uniformly negative; as such, temperatures would eventually decline as the slow pool absorbs its share of excess CO_2 over centuries to millennia.

Table S1: Permanent airborne fraction (f_∞), slow carbon timescale (τ_3), and implied asymptotic ZEC (ZEC_∞) for each ESM ensemble. Values show median [10th–90th percentile] across 10 ensemble members.

Model	f_∞	τ_3 (yr)	ZEC_∞ (K)
ACCESS-ESM1-5	0.21 [0.08–0.27]	848 [207–1252]	−0.86 [−2.75 – −0.44]
CESM2	0.03 [0.00–0.16]	228 [161–379]	−2.15 [−2.89 – −1.27]
CNRM-ESM2-1	0.09 [0.00–0.12]	685 [148–1093]	−2.20 [−3.81 – −1.86]
GFDL-ESM4	0.10 [0.03–0.14]	622 [180–1144]	−1.23 [−1.72 – −0.88]
GISS	0.11 [0.07–0.22]	388 [132–1440]	−1.11 [−1.18 – −0.77]
HadCM3LC-Bris	0.09 [0.00–0.19]	1249 [610–1500]	−1.99 [−5.46 – −0.35]
MPI-ESM1-2-LR	0.10 [0.01–0.15]	383 [50–1146]	−1.19 [−1.87 – −0.81]
NorESM2-LM	0.20 [0.09–0.22]	50 [50–309]	−0.64 [−0.85 – −0.48]
UKESM	0.13 [0.06–0.20]	514 [194–652]	−1.64 [−2.45 – −0.89]
UVic-ESCM-2-10	0.07 [0.00–0.21]	640 [328–1221]	−1.02 [−2.99 – −0.63]

S8 Fitted parameter values

Tables S2 and S3 report the deterministic fitted parameter values for each ESM, together with the 5th–95th percentile range from the perturbed-start ensemble.

Table S2: Deterministic fitted parameter values for each ESM, with ensemble 5th–95th percentile range in brackets. The deterministic fit can lie outside the ensemble range because the ensemble is generated by perturbed-start optimisation with cost-function filtering, and need not bracket the global optimum.

Carbon cycle parameters Parameter	ACCESS	CESM2	CNRM	GFDL	GISS
f_∞	0.334† [0.041–0.247]	0.000† [0.021–0.192]	0.000 [0.000–0.171]	0.121 [0.000–0.204]	0.230 [0.080–0.260]
f_1	0.403 [0.212–0.623]	0.349 [0.296–0.564]	0.357 [0.256–0.683]	0.479 [0.192–0.558]	0.397 [0.384–0.437]
τ_1 (yr)	6.1 [3.1–20.1]	2.0 [2.0–9.9]	2.0 [2.0–14.0]	2.0 [2.0–12.5]	2.0 [2.0–4.0]
f_2	0.252 [0.070–0.443]	0.294 [0.194–0.397]	0.326 [0.084–0.485]	0.382 [0.178–0.529]	0.358† [0.202–0.356]
τ_2 (yr)	25.1 [24.7–93.8]	27.9 [25.1–78.5]	27.7 [21.1–98.7]	65.0 [20.8–65.8]	70.2 [66.4–97.0]
τ_3 (yr)	300 [205–1380]	253 [65–1032]	250 [152–1430]	300 [108–745]	303 [50–854]
ζ (K^{-1})	0.200† [0.007–0.189]	0.137 [0.111–0.145]	0.155 [0.137–0.192]	0.200 [0.141–0.200]	0.187 [0.161–0.200]
Climate parameters Parameter	ACCESS	CESM2	CNRM	GFDL	GISS
C_f ($\mathrm{W\,yr\,m^{-2}\,K^{-1}}$)	5.8 [3.0–20.1]	7.5 [6.4–22.6]	7.7 [4.7–14.9]	7.8 [2.5–24.8]	5.8 [2.0–6.3]
C_s ($\mathrm{W\,yr\,m^{-2}\,K^{-1}}$)	79 [65–407]	99 [34–316]	98 [47–447]	100 [30–408]	80 [37–167]
λ_f ($\mathrm{W\,m^{-2}\,K^{-1}}$)	1.247 [0.559–1.515]	1.098 [0.368–1.301]	1.196 [0.460–1.438]	1.312 [0.614–1.639]	1.514 [1.082–1.822]
γ ($\mathrm{W\,m^{-2}\,K^{-1}}$)	1.361 [0.387–2.014]	0.455 [0.193–2.092]	0.535 [0.409–2.288]	0.762 [0.617–1.528]	0.689 [0.478–2.029]
ECS (K)	2.98 [2.45–6.90]	3.38 [2.86–10.51]	3.10 [2.59–9.35]	2.83 [2.27–6.13]	2.45 [2.04–3.43]
ε	1.400 [0.590–2.865]	0.700 [0.545–2.730]	0.800 [0.658–2.620]	1.150† [1.423–2.865]	1.050 [0.763–2.683]

† Deterministic value lies outside the ensemble 5th–95th percentile range.

Table S3: Deterministic fitted parameter values for each ESM, with ensemble 5th–95th percentile range in brackets. The deterministic fit can lie outside the ensemble range because the ensemble is generated by perturbed-start optimisation with cost-function filtering, and need not bracket the global optimum.

Carbon cycle parameters Parameter	HadCM3	MPI	NorESM	UKESM	UVic
f_∞	0.248† [0.024–0.169]	0.000 [0.000–0.118]	0.140 [0.004–0.208]	0.170† [0.000–0.136]	0.225† [0.028–0.216]
f_1	0.428 [0.320–0.747]	0.485 [0.416–0.561]	0.408 [0.351–0.489]	0.455 [0.375–0.690]	0.255 [0.157–0.534]
τ_1 (yr)	5.9 [2.7–20.3]	3.0 [2.0–7.1]	2.0 [2.0–6.4]	2.0† [5.4–20.7]	2.0 [2.0–17.1]
f_2	0.312 [0.009–0.454]	0.356 [0.080–0.405]	0.435† [0.092–0.311]	0.354 [0.000–0.380]	0.361 [0.088–0.453]
τ_2 (yr)	24.9 [13.3–74.4]	65.0 [45.3–92.9]	59.6 [22.4–81.6]	54.9 [5.0–87.7]	20.4 [10.8–29.2]
τ_3 (yr)	300† [359–1441]	298 [50–652]	290 [50–854]	248† [284–971]	300 [262–1292]
ζ (K^{-1})	0.112 [0.038–0.194]	0.200 [0.149–0.200]	0.200 [0.138–0.200]	0.167 [0.142–0.175]	0.190 [0.175–0.200]
Climate parameters Parameter	HadCM3	MPI	NorESM	UKESM	UVic
C_f ($W\,yr\,m^{-2}\,K^{-1}$)	12.1† [12.4–22.9]	6.2 [5.1–11.9]	6.0 [2.0–19.4]	5.4 [4.1–22.1]	10.1 [6.3–24.9]
C_s ($W\,yr\,m^{-2}\,K^{-1}$)	201 [122–486]	80 [65–361]	80† [94–335]	78 [30–447]	150 [94–386]
λ_f ($W\,m^{-2}\,K^{-1}$)	1.292 [0.525–1.464]	1.467 [1.113–1.670]	1.800 [1.439–1.963]	0.628 [0.300–0.982]	1.263 [0.663–1.523]
γ ($W\,m^{-2}\,K^{-1}$)	0.401 [0.329–2.043]	0.724 [0.181–1.079]	0.900 [0.706–1.675]	1.150 [0.298–1.328]	0.556 [0.092–1.287]
ECS (K)	2.87 [2.56–7.92]	2.53 [2.23–3.34]	2.06 [1.89–2.59]	5.91 [3.78–12.37]	2.94 [2.44–5.91]
ε	0.650 [0.550–2.595]	–	2.300 [1.260–2.542]	1.450 [0.838–2.333]	–

† Deterministic value lies outside the ensemble 5th–95th percentile range.

S9 ZEC sensitivity to emission duration and rate (all models)

Figures S3 and S4 show the full ten-model version of the ZEC sensitivity analysis presented in the main text (Fig. 9 of the main paper shows three representative models).

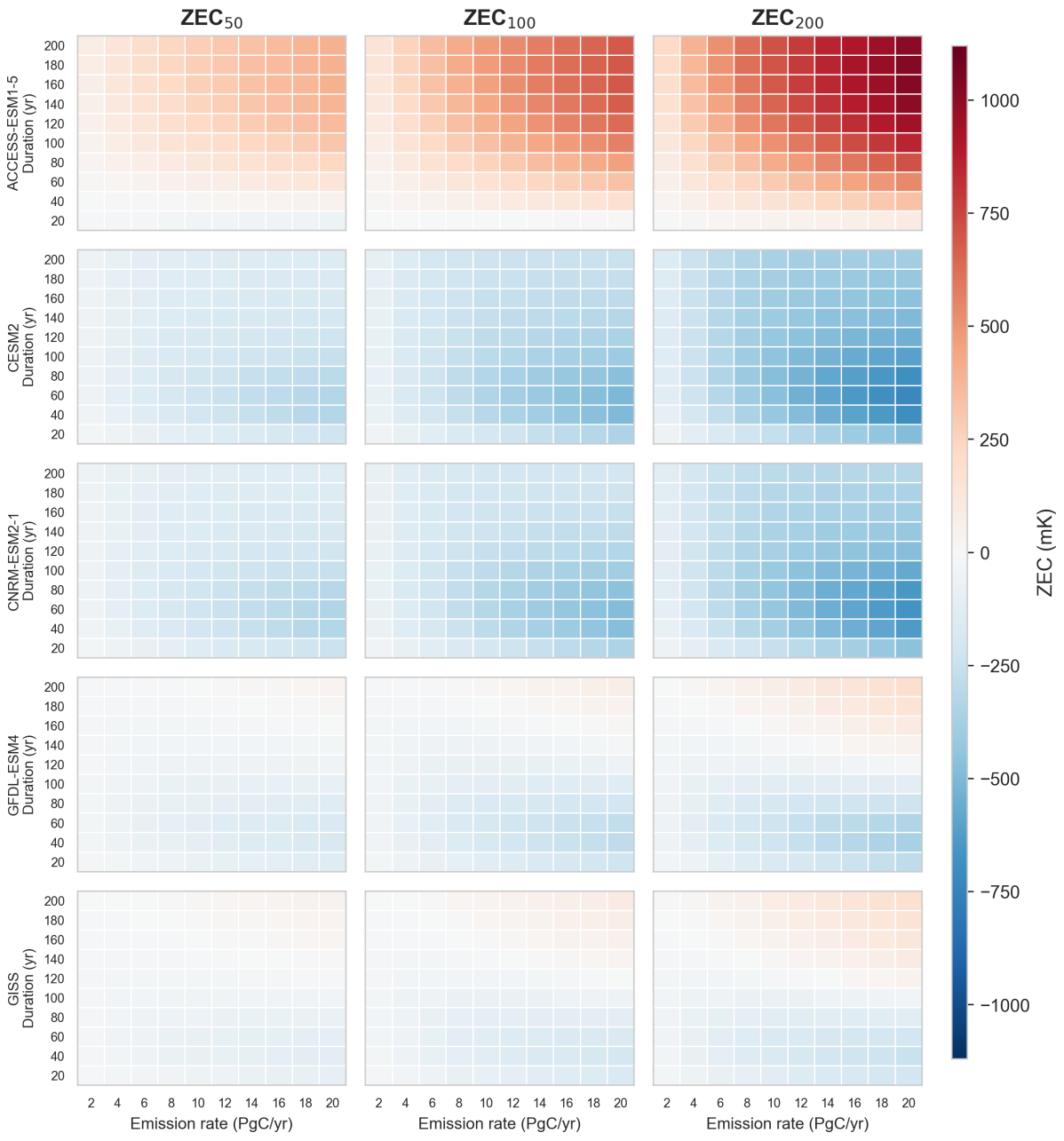

Figure S3: Sensitivity of ZEC to emission duration and rate for ACCESS-ESM1-5 through GISS. Each row shows one model; columns show ZEC at 50, 100, and 200 years after cessation (mK). The colour scale is shared across all panels (including fig. S4).

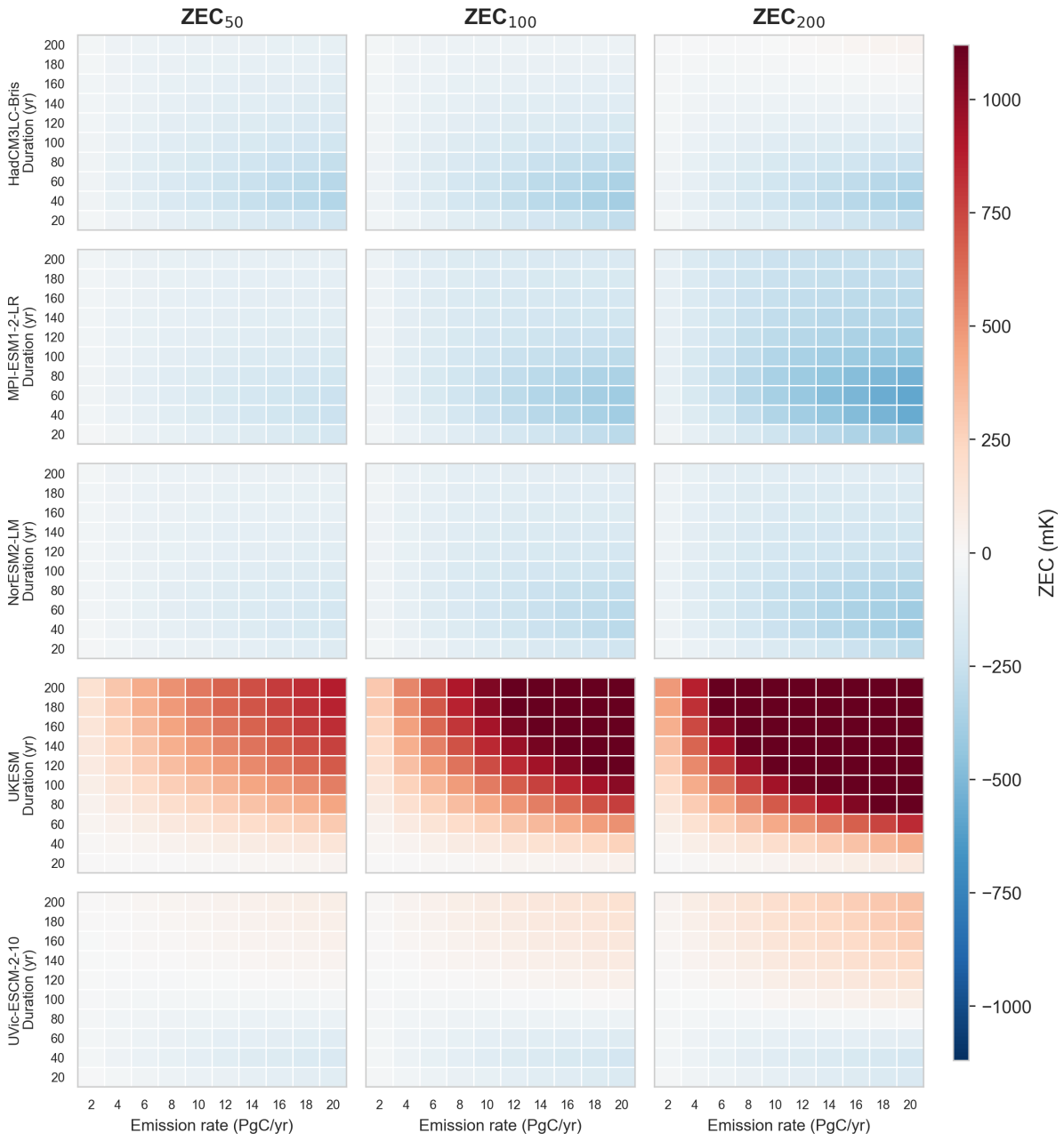

Figure S4: As fig. S3, for HadCM3LC-Bris through UVic-ESCM-2-10.