PHILOSOPHICAL TRANSACTIONS A

rsta.royalsocietypublishing.org

Review

Cite this article: Knutti R, Rugenstein MAA. 2015 Feedbacks, climate sensitivity and the limits of linear models. *Phil. Trans. R. Soc. A* **373**: 20150146.
http://dx.doi.org/10.1098/rsta.2015.0146

Accepted: 14 July 2015

One contribution of 12 to a discussion meeting issue 'Feedbacks on climate in the Earth system'.

Subject Areas:
climatology, atmospheric science

Keywords:
feedbacks, climate sensitivity, energy balance model

Author for correspondence:
Reto Knutti
e-mail: reto.knutti@env.ethz.ch

Feedbacks, climate sensitivity and the limits of linear models

Reto Knutti and Maria A. A. Rugenstein

ETH Zurich, Institute for Atmospheric and Climate Science, Universitätstrasse 16, Zurich, Switzerland

The term 'feedback' is used ubiquitously in climate research, but implies varied meanings in different contexts. From a specific process that locally affects a quantity, to a formal framework that attempts to determine a global response to a forcing, researchers use this term to separate, simplify and quantify parts of the complex Earth system. We combine new model results with a historical and educational perspective to organize existing ideas around feedbacks and linear models. Our results suggest that the state- and forcing-dependency of feedbacks are probably not appreciated enough, and not considered appropriately in many studies. A non-constant feedback parameter likely explains some of the differences in estimates of equilibrium climate sensitivity from different methods and types of data. Clarifying the value and applicability of the linear forcing feedback framework and a better quantification of feedbacks on various timescales and spatial scales remains a high priority in order to better understand past and predict future changes in the climate system.

1. Introduction

Partly originating from control theory, the analysis of feedbacks is a powerful tool to study dynamical systems, in which one quantity affects another, thereby attenuating or amplifying the original signal (see [1] for a review). For example, warmer temperatures lead to melting of snow and ice, which exposes a darker surface that absorbs rather than reflects incoming solar radiation, which leads to more warming and melting than would have occurred if the snow cover area had been fixed. In simple systems with few components and interactions, such feedback frameworks can separate cause and effect, and allow for a mathematical description of a dynamical

THE ROYAL SOCIETY PUBLISHING

system. However, there are difficulties in applying it to the global climate system, which is not closed, and where the interplay of different feedbacks and forcings complicate the description. Some feedbacks may only become relevant in the future, or may no longer be relevant (e.g. if there is no snow and ice left), whereas some changes may be nonlinear, abrupt or irreversible. For instance, systems like the El Niño Southern Oscillation could potentially show regime shifts, invalidating simple linear feedback formulations and potentially making feedback analysis less relevant for both understanding the past and predicting the future. Yet, despite all these potential complexities, the construction of linear feedback frameworks has been helpful in the past, if applied carefully to parts of the whole climate system, and within certain bounds on timescales and climate states that we discuss below.

The perspective provided here, focusing on the global forcing feedback framework, emerged from an overview talk presented at the Royal Society Discussion Meeting on 'Feedbacks on climate in the Earth system'. We attempt to provide an extended context and perspective to the more detailed papers in this theme issue. As a consequence, some conceptual material presented here is not novel, though we hope to stimulate potential avenues of future research.

2. The case for forcing feedback frameworks

A specific forcing might affect the climate system response on a large range of timescales. In the usual forward thinking and modelling chain, shown in figure 1*a*, the use of fossil fuels leads to greenhouse gas emissions and an increase in their atmospheric concentrations, a change in radiative forcing, which causes a climate response. In the more detailed view in figure 1*b*, the change in the CO_2 concentration causes an instantaneous forcing, which—after being adjusted for very fast responses—becomes an effective radiative forcing, defined as the change in the top of atmosphere radiative balance before the surface temperature responds (see [3] for an overview). By warming, the surface restores the radiative balance by increasing the radiation to space, but this warming causes water vapour, lapse rate, albedo, clouds, vegetation, ice sheets, permafrost and/or atmospheric chemistry to change. Those changes—directly or indirectly—affect the Earth's radiation budget, and amplify or damp the temperature response.

Equilibrium climate sensitivity (ECS) is an attempt to combine many of these changes in a tractable manner, and is one of several key numbers that are used to characterize the temperature response of the Earth to a change in forcing or the CO_2 concentration. ECS is usually defined as the equilibrium global average surface warming in response to the radiative forcing from an atmospheric CO_2 doubling, and includes the changes in water vapour, lapse rate, surface albedo and clouds (see magenta box in figure 1). By definition, in equilibrium, the ocean heat uptake is zero, but in a transient climate, it damps the warming. The transient climate response (TCR) characterizes the warming at the time of CO_2 doubling after a 1% per year increase in the CO_2 concentration (see violet box). The transient climate response to cumulative carbon emissions (TCRE, light blue box) characterizes the warming as a function of the total emitted CO_2, and is relevant to estimate the carbon budgets, and emission reductions required for stabilizing global temperature (see [4] for an overview). In some sense, the definition of ECS is arbitrary and has survived only because of historical development, convenience in modelling and the lack of better alternatives. The early generations of climate models included only the water vapour, lapse rate, albedo and cloud feedbacks, and had no appropriate representation of land ice, vegetation, chemistry or biogeochemical cycles, nor did they include a dynamical ocean component. Doubling the atmospheric CO_2 concentration for a few decades in such a model was therefore a benchmark to characterize the overall temperature response to a well-defined forcing, and a measure of the total feedback on timescales of decades to centuries.

From figure 1, it becomes clear that ECS and TCR are rather limited characterizations of a much larger and interactive system. Other feedbacks such as vegetation, chemistry or land ice are now included in some climate models as their relevance is better understood. Some feedbacks operate on very long timescales that are determined by the internal dynamics of the system, and

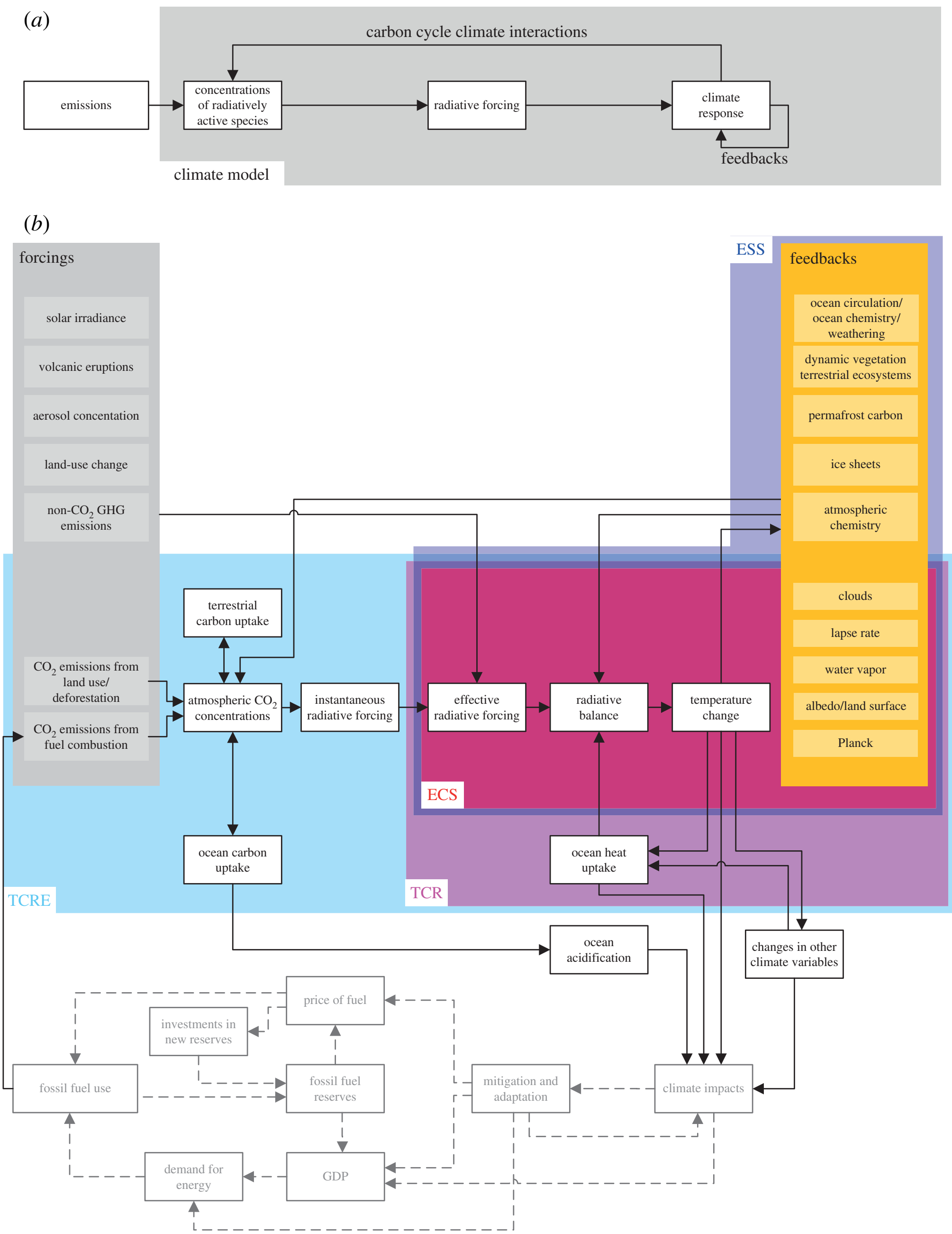

Figure 1. (*a*) Simplistic and generalized modelling chain (adapted from [2], figure 10.1) and (*b*) more refined distinction between feedbacks acting in and on the climate system. Greenhouse gas emissions perturb the radiative balance, which force the temperature to respond. Temperature change is causing various feedbacks (yellow box, interactions between the feedbacks are not marked) to act back onto the radiative balance, which again causes the temperature to adjust. The equilibrium climate sensitivity (ECS, magenta box) covers only some of the feedbacks. In a transient reference framework (transient climate response TCR, violet box), the rate of ocean heat uptake affects the radiative balance and temperature change in return. The transient climate response to cumulative carbon emissions (TCRE, light blue box) characterizes the temperature response to emissions and includes carbon emissions, uptake and release of the land biosphere and the ocean. The Earth system sensitivity (ESS, blue box) includes more feedbacks, generally but not exclusively acting on longer than century timescales. The separation of forcings (grey box) and feedbacks (yellow box) is in some sense arbitrary and has to be defined for each problem. Climate change—through temperature and other variables' change will impact socio-economic systems, which finally will feed back on emissions. See text for further discussion. The feedback loops sketched act on different timescales. (Online version in colour.)

their response is not proportional to temperature. Thus, a more recent concept is an equilibrium Earth system sensitivity (ESS, dark blue box) which encompasses all climate (but not human) feedbacks. The separation of ECS and ESS is often made along timescales, with the argument that those feedbacks included in ECS essentially scale with surface temperature, whereas the others in ESS partly have their intrinsic (and often slower) timescales. However, this does not apply to atmospheric chemistry which responds quickly. Here, the reason is a historic one, as the early climate models simply did not simulate interactive chemistry. This supports the argument that the separation of ECS and ESS is somewhat arbitrary in the real world where a lot of processes interact.

How would we go about estimate ECS in the real world? The Earth today is not in equilibrium, and other, non-greenhouse gas forcings (aerosols, dust, land use, solar and volcanic), which are smaller than greenhouse gas forcings, are still important locally. Attempting to capture the importance of these other forcings, scenarios of future climate now prescribe emissions of many gases [5]. Climate change is also expressed in other variables—from ranges of species' habitats to hail grain size—including their variance and extremes. The state of the changed climate system causes impacts, leading to adaptation and mitigation, which, in turn, influence the economy and fossil fuel exploitation and use (grey in figure 1*b*, sketched only roughly to indicate the incompleteness of process understanding), which further influence greenhouse gas emissions. It is tempting to broaden the definition of ECS to include more feedbacks to simplify the comparison with the real world. Even impacts and the human response in terms of adaptation and mitigation could be included in a broader concept of sensitivity [6], encompassing most or all relations shown in figure 1. However, the decision to incorporate an additional process into 'sensitivity' must consider the need to reduce complexity, in order to have a tractable system that is useful for understanding. The human component is a hypercomplex interaction of nature and societies. Humans, as biological systems, may, in theory, be described by the laws of physics and chemistry, and could be parametrized similar to other ecosystems, but human decisions, ideas and inventions can (and have done in history) literally change the course of the world, and thus introduce a problem of predictability of the first kind (sensitivity of the outcome to initial conditions). If climate sensitivity is defined in such broad terms to include human behaviour, it is apt to be unpredictable and fails to provide insights into the climate system.

The idea of the feedback framework in climate science is to break down complex processes and quantify their sensitivities. For long-term warming, ECS or TCR may be useful numbers and they explain the largest fraction of uncertainty [7], but for adaptation purposes, global temperature is of very limited value. For regional change and changes other than temperature, the feedbacks and processes that matter most may be different (e.g. soil moisture, vegetation or air pollution) from the ones that are most important for TCR or ECS.

Why would we stick to an arguably narrow framework of climate sensitivity, which describes only a limited number of the feedbacks in the real world? A number of reasons partly explain why we have done so for a long time. First, many changes in climatic variables approximately scale with temperature [8,9]. As a result, global temperature is probably the best proxy for aggregated impacts, even though the relation is likely nonlinear. Global temperature is relatively easy to measure, records extend further back than measurements of most other climate variables, and temperature is more straightforward to reconstruct from palaeodata than other quantities. Together, this provides a way of comparing current and future climate with the climate that would have been without anthropogenic emissions. If we had to reduce climate change to a single aggregate number, for example to agree on a single climate target, global temperature is an obvious choice. Second, in the global forcing feedback framework, the radiative forcings and their responses are assumed to be additive, as discussed further in §3. This is key for the relevance of the radiative forcing definition, as it means that $1\,\mathrm{W\,m^{-2}}$ of a forcing can be 'traded' against $1\,\mathrm{W\,m^{-2}}$ from a different forcing when designing policies towards a climate target, and the total warming is proportional to the total forcing. This additivity is also a key assumption for detection and attribution studies, to break down the observed changes into parts caused by different forcings. Third, many earlier studies (partly based on slab ocean rather than on dynamical ocean

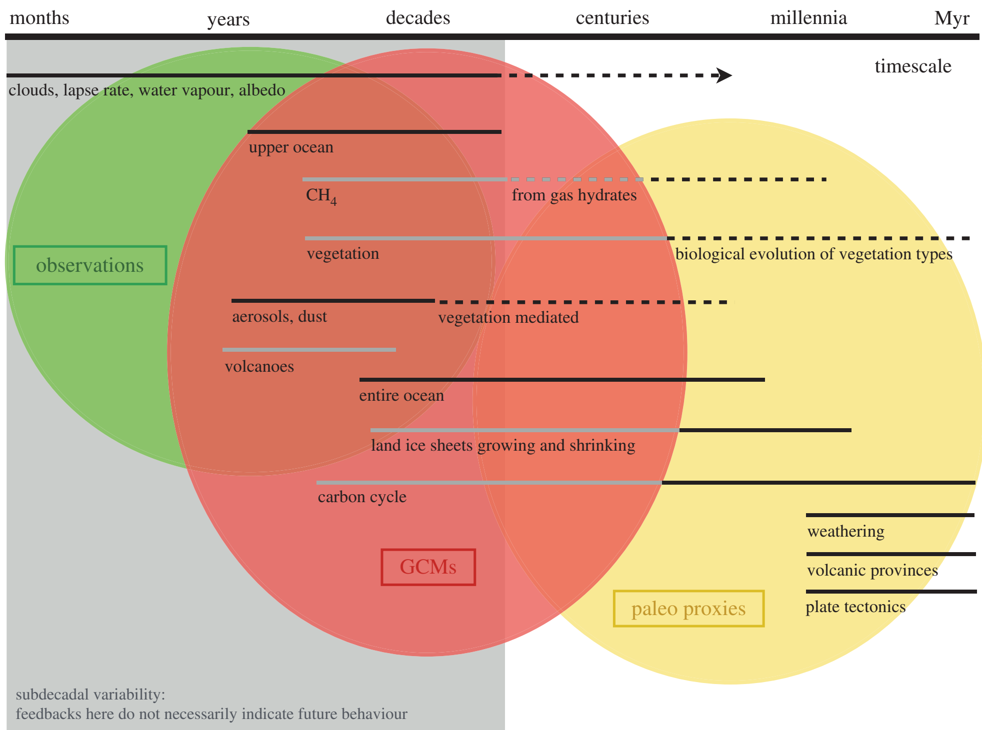

Figure 2. Timescales of climate relevant processes. Light grey bars indicate processes that act on timescales that a GCM can resolve, but are usually assumed to be (partly) inactive or non-existent. Dashed lines indicate timescales where specific feedbacks are weaker or only operate under certain circumstances. The arrow for clouds, lapse rate, water vapour and albedo indicates that those feedbacks operate on short timescales, but, because the surface warming takes centuries or more to equilibrate, these feedbacks continue to change and affect the overall response of the systems up to millennia. This can apply similarly to other feedbacks that respond quickly but continue to change over long timescales in response to other feedbacks. The coloured ellipses each cover different methods used to estimate climate sensitivity. The vertical ordering of the feedbacks is arbitrary. Models of intermediate complexity (EMICs) can bridge the gap between GCMs and palaeo proxies, for example by including carbon cycles, weathering and ice sheets. Usually, there are trade-offs between simulating very long timescales and the level of detail of short timescale processes. (Online version in colour.)

models) indicated that the global feedback parameter (the inverse of the equilibrium warming per unit forcing) is roughly constant for various forcings and climate states. This is equivalent to a description of a Taylor expansion neglecting higher-order terms, as shown in §3. To the degree that this is justified, the global feedback can be used in simple energy balance models to estimate the future warming from future emissions or forcings, or in integrated assessments models. Such models, where the forcing is seen as the cause, and warming as the effect, are known to be a simplification of the real world, but have been crucial for understanding how models of various degrees of complexity respond to perturbations, and to which degree past and future climate change can, to first order, be described as an energy balance problem [10].

Figure 1 does not indicate typical process timescales, but it is obvious that cloud droplet formation acts on different temporal and spatial scales than weathering of rocks or land-use changes. Climate sensitivity was defined with a century timescale in mind and, as such, can characterize only certain processes. Figure 2 compares the most common process timescales. The direction—warming or cooling, positive or negative feedback—is not taken into account in this representation, and some processes have different sensitivities for warming than for cooling (see discussion in §3). The coloured ellipses indicate different methods to define sensitivity in broad terms.

Climate sensitivity is not a quantity that can be measured, and it characterizes only a part of the relevant processes and feedbacks, but it is an emerging property of the system. From past climate, it can be approximated by relating equilibrium warming to radiative forcing. In global climate models (GCMs), climate sensitivity is normally not tuned, but it results from aggregating or parametrizing small-scale processes and ignoring long-term ones (red ellipse in figure 2). GCM-based estimated of TCR and ECS ignore certain processes even within the time frames they consider (grey bars within the red ellipse).

On short timescales (green ellipse), the observed surface warming, ocean heat uptake, and an estimate of radiative forcing, provide an estimate of the anthropogenic contribution to the observed warming and the global feedbacks [10], and therefore, ECS and TCR [11–24]. Those again can be used for probabilistic projections that are conditional on, i.e. constrained by, past warming [11,25–32]. Observations and simulations of the response to natural external forcings (volcanic or solar) [33–36] or unforced climate variations on short or very long timescales (green and yellow ellipse in figure 2), or the climatology and seasonal cycle may provide information on feedbacks [37–39], but the inferred numbers (in $\mathrm{W\,m^{-2}\,K^{-1}}$) may differ from those on the century timescale. Both the short-term and proxy methods are often called 'observational', but it is important to note that they rely on models and assumptions as much as GCMs. Their radiative forcing is derived from a GCM, the magnitude and timescales of internal climate variability often come from climate model control runs or statistical models, and in many cases, strong assumptions about linearity and spatial aggregation are made, as discussed in the next sections. Information from palaeoclimate combined with models [40–49] provides further support for an ECS value in the consensus range of 1.5–4.5°C, but also highlights that feedbacks for warmer or colder states and on longer timescales may differ from those today.

Two pressing questions become clear from figure 2. The first is why different lines of evidence point to different ECS values. Specifically, some but not all recent studies on the twentieth-century warming find rather low ECS values (median at or less than 2°C) [17–19,21]. Climate models show a large spread in ECS, with the spread half as big as the actual value. The highest uncertainty can be attributed to the cloud feedbacks (traceable to certain cloud types and regions), and the lapse rate feedback [50–53]. But all comprehensive climate models indicate sensitivities above 2°C, and those that simulate the present-day climate best [54–57] even point to a best estimate of ECS in the range of 3–4.5°C. The second question is how to infer present-day ECS from the climate sensitivity in warmer or colder states, from shorter or longer timescales, or for a non-CO_2 perturbation ('mapping'). Both questions are partly rooted in the use of simple linear forcing feedback models with a constant feedback parameter, discussed in depth in the following sections.

3. Climate sensitivity, timescales and commitment

(a) General concepts

In equilibrium, the global radiation budget, the sum of net incoming solar shortwave and outgoing terrestrial longwave radiation, is closed ($R=0$). The degree of imbalance ($R\neq 0$) at some time following a perturbation can be ascribed to the temperature response itself (ΔT), and changes induced by the temperature response, called feedbacks ($\alpha\Delta T$), thus $R=R(\Delta T,\,\alpha(T))$ [1,58]. To study how a small change in the radiation budget ΔR is related to the temperature response, one can use the Taylor expansion of R, in T and $\alpha(T)$

$$\Delta R=\frac{\partial R}{\partial T}\Delta T+\frac{\partial \mathrm{R}}{\partial \alpha}\frac{\partial \alpha}{\partial T}\Delta T+\mathrm{O}((\Delta T)^2). \tag{3.1}$$

The perturbation of the radiation budget is the effective radiative forcing F minus the heat flux or top of the atmosphere (TOA) radiative imbalance N, which is non-zero as long as the system is not in equilibrium. The reference height of the heat flux is usually the tropopause. Over time scales longer than a year, this is the same as the heat flux into the ocean, ice and land. The first term on the right-hand side describes the strongest negative feedback, sometimes called the Planck feedback.

Increased temperatures lead to increased TOA outgoing longwave radiation. Other feedbacks would have to be stronger than the Planck feedback to lead to a runaway climate. The second term on the right-hand side describes the sum of the feedbacks, which scale with the temperature response:

$$\frac{\partial R}{\partial \alpha}\frac{\partial \alpha}{\partial T} = \sum \frac{\partial R}{\partial \alpha_i}\frac{\partial \alpha_i}{\partial T},$$

with $i =$ water vapour, lapse rate, albedo and cloud feedback. These are the common physical feedbacks analysed in CMIP5-type climate models (see the violet TCR box in figure 1). The feedbacks can be positive (e.g. water vapour) or negative (e.g. lapse rate) and sometimes difficult to determine (e.g. for the cloud feedbacks). Processes that involve several of the feedbacks can lead to correlations between them. For example, the sum of the water vapour and lapse rate feedback is better constrained than the individual parts [59]. Finally, the last term on the right-hand side of equation (3.1) is the sum of all higher-order terms of the Taylor expansion, representing the nonlinearities of individual process and the interaction between the different feedbacks.

The linear approximation generally neglects the last term, because the temperature response from interactions between the feedbacks is usually small. Focusing on the linear term helps to distinguish and quantify the single feedbacks' influence on the final response [60]. However, it is not clear what a 'small perturbation' comprises and when higher-order terms should be taken into account, such as for high emission scenarios or palaeoclimate studies with large perturbations or additional active feedbacks (figure 2). Another limitation arises, because the climate system may include thresholds and tipping points, where the linear assumptions are not justified [58]. As discussed in §1, part of why the linear approximation is so widely used is its simplicity, convenience and lack of alternatives; its validity is not in all cases examined. Studies investigating limitations of the linearization would help to strengthen trust in the findings obtained within the linear framework.

All terms in equation (3.1) are globally defined and hold for large temporal integrated scales. To analyse feedbacks on a local scale a heat-flux divergence term has to be added [61,62]. The meridional structure of feedbacks tends to compensate for local nonlinearites [63].

As shown, the climate feedbacks are treated as relative contributions to the response compared with the strongly negative Planck feedback. One can define a reference temperature (increase T_0) caused by the Planck feedback (about 1.1°C for a doubling of the atmospheric CO_2 concentration). The additional temperature response caused by the feedbacks can then be described by $\Delta T = \Delta T_0/(1-f)$ with $f = \partial T/\partial R(\partial R/\partial \alpha\, \partial \alpha/\partial T)$ the feedback factor. For an ECS value of approximately 3°C, this implies that more than half of the warming is caused by feedbacks in the climate system, and less than half is a direct Planck response to forcing.

Accepting the linear assumption and adopting the naming conventions mentioned above, one can rewrite equation (3.1) as

$$F - N = \lambda \Delta T. \tag{3.2}$$

The linearization leads to the assumption that the feedback parameter λ is constant, meaning the net feedback strength is independent of the climate state ΔT and the forcing F [64]. It is assumed that the real-world climate system has an *a priori* unknown λ and climate models can help finding the value of that λ and then project ΔT into the future. When the system settles into the new equilibrium, the net heat flux, N, at the TOA is zero, and the temperature change necessary to reach the new equilibrium $\Delta T = F/\lambda$, is—by convention and as defined in §1—the ECS, if the forcing is a doubling of the preindustrial CO_2 concentration. The less efficient the Earth is at emitting energy to space (smaller λ), the higher temperature increase ΔT is necessary to restore the balance. By incorporating the heat uptake as a measure of the TCR, the global feedback (and thus ECS) can be inferred from $\Delta T = (F-N)/\lambda$. The transient response can be approximated from the ratio between temperature change and forcing, and is smaller than ECS. As a consequence, keeping F fixed at a certain time during a warming simulation would result in further surface warming for several centuries. This is the commitment warming or 'warming in the pipeline'

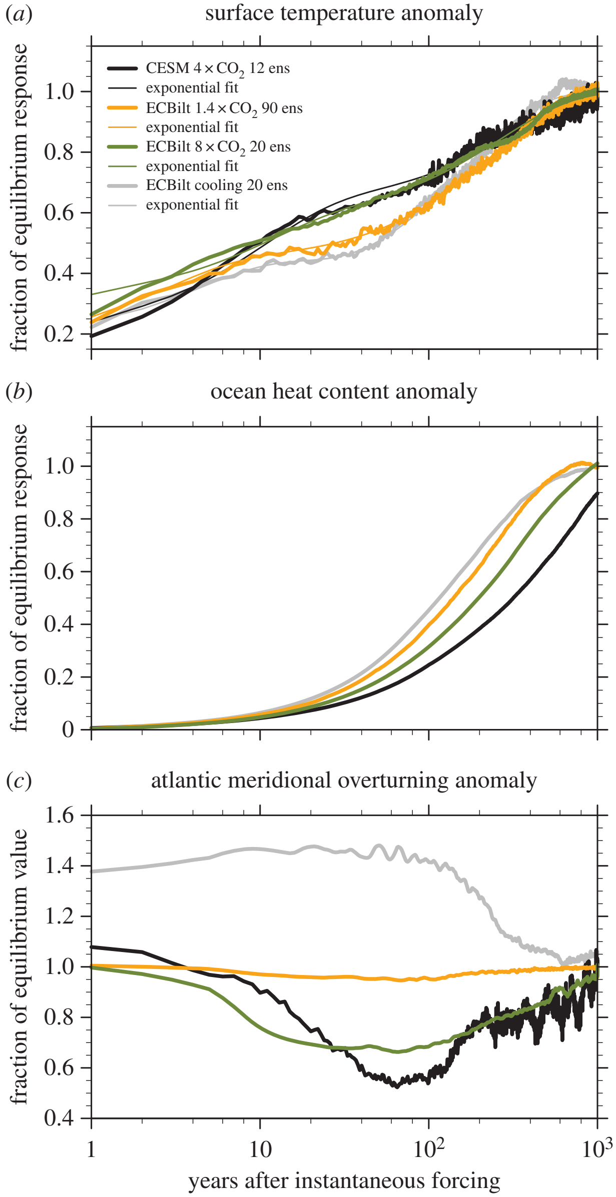

Figure 3. Response timescales, expressed as fraction of the realized equilibrium response, for the global surface temperature (*a*), the global ocean heat content (*b*) and fraction of the final equilibrium value of the maximum atlantic meridional overturning circulation at 30° N (*c*) for different models (CESM and ECBilt-CLIO) and forcing levels of 1.4 to 8 $\times$ CO_2 and cooling to 100 ppm. For CESM, only the 12 longest runs are used here. The number of ensemble members is noted in the label (ens). (Online version in colour.)

[65,66]. The magnitude of the commitment warming depends on ECS, because the response timescale is longer, and, therefore, the fraction of realized equilibrium warming (discussed later in figure 3) is smaller, for higher ECS. In other words, if ECS is high, the current temperature (expressed as a fraction) is further away from the equilibrium temperature for that forcing. As a consequence, TCR becomes less sensitive to ECS for high ECS (i.e. a high and very high ECS are difficult to separate in their short-term response as, indeed, in many other observables), which often results in probability density functions with fat tails to high values [67,68].

The description of equilibration in equation (3.2)—as N approaches 0—is a choice of a reference framework and might be more helpful for certain questions than for others. While the global energy balance has to be closed of course, the ability of equation (3.2) to physically explain different timescales is limited. There is no physical necessity that the response scales with the global mean surface temperature change, although many variables do (see discussion in §1).

To study the validity of the assumptions discussed above, and to analyse different processes and timescales, step forcing experiments are useful. The forcing, F, does not vary in time (as it does in reality), but is prescribed as an instantaneous increase or decrease and then held constant, to let the system approach a new equilibrium. A fully equilibrated state is never reached in the real world, because boundary conditions (e.g. orbital forcing, tectonics) always change, and some feedbacks have very long response timescales. Nevertheless, these experiments are the cleanest method of studying the timescales of different processes involved in the radiative restoration or equilibration. Climate model intercomparisons reveal a large spread in timescales for a certain responses [51,69]. This indicates a large uncertainty when analysing climate change impacts and risks. Step experiments can be further used to predict the response to a more realistic time varying CO_2 forcings [70–73].

(b) Coupled model results

We use two models to illustrate some of above concepts, and to highlight the limitations of the linear forcing feedback framework. First, the Community Earth System Model (CESM v. 1.0.4), a comprehensive ocean–atmosphere–land–sea ice model, is used with fixed vegetation [74–76]. A set of 120 ensemble members branched off from different control run years—thus, different in their initial oceanic, atmospheric, and sea ice state—are forced with an instantaneous quadrupling of the CO_2 concentration from the preindustrial value. All members are run for 2 years, 12 for 100 years, six for 250 years and one member for 1300 years. Its final state is regarded as being equilibrated to calculate the fraction of equilibration shown in figure 3*a*, although the deep Southern Ocean is still adjusting. The novel result here is that the forced response (shown here as the anomaly to the control run) and, therefore, the changes in the global feedback can be estimated on all timescales owing to the many ensemble members. Most GCM studies using the energy balance equation (equation (3.2)) are done with a 150 year time series and one or a few simulations for each model [51,62,77].

The second model is ECBilt-CLIO, a model of intermediate complexity, with a three level quasi-geostrophic atmosphere with simple parametrizations for the diabatic processes and a free-surface ocean general circulation model coupled to a thermodynamic-dynamic sea–ice model [78,79]. We conducted five step forcing experiments composed of instantaneously increasing the CO_2 concentration 1.4, two, four, eight and 16 times above the preindustrial concentration, and one step experiment with reduced forcing. In this cooling scenario, the CO_2 concentration is instantaneously set to 100 ppm, thus 0.35 times the preindustrial value of 280 ppm. For each of the six ECBilt-CLIO experiments we simulate—depending on the signal to noise ratio—10 to 90 realizations of the same forcing from different initial conditions, all of which are run for 1000 years. One member per experiment is run for 10 000 years until equilibrium. These simulations provide insights into how the global feedback changes with different forcing levels, and from transient to equilibrium.

Figure 3*a* shows the realized temperature response at a certain time (relative to equilibrium) for the ensemble average of four different experiments: the $4 \times CO_2$ CESM (black), $1.4 \times CO_2$ ECBilt-CLIO (orange), $8 \times CO_2$ ECBilt-CLIO (green) and the cooling ECBilt-CLIO (grey) all in thick lines. The assumption that λ is independent of forcing level, and climate state or temperature implies that at all times the fraction of equilibration is the same in all experiments, which is not the case. There are roughly three timescales that all experiments have in common: a short timescale lasting up to a few years, a decadal timescale and a century timescale, consistent with processes operating on different timescales as shown in figure 2. Despite the instantaneous forcing, the realized warming is only 30–50% after a decade, and 60–80% after a century, confirming the commitment warming idea discussed above and in §3c.

A minimized-least-squares fit of a sum of three exponentials to the dimensionless temperature response function $\theta(t)$ of the form

$$\theta(t) = 1 - \left(\theta_0 e^{-t/\tau_0} + \theta_1 e^{-t/\tau_1} + \theta_2 e^{-t/\tau_2}\right),$$

is shown as thin lines in figure 3*a*. The choice of the exponential function is arbitrary—a sum of two exponentials, a fit to a heat diffusion equation or a transfer function might also be valuable for certain purposes [69,80]. The timescales (τ) reveal the differences: τ_0 ranges from 0.4 to 4.7 years, τ_1 from 2.6 to 50 years and τ_2 from 194 to 310 years. The models differ most on decadal timescales, with the weakest forcing case (orange) having a small warming initially (relative to equilibrium) and an increased rate of warming after 100 years, whereas the stronger forcings (green and black) lead to initially stronger warming and a slower increase on the century timescale. The amount of realized warming at a given time differs up to 15% between the experiments. The rate of temperature change involved when approaching a cooler state is initially smaller, but after some decades, it is larger than in the warming situation.

Figure 3*b* shows the oceanic timescales, which are of course much longer, leading to a smaller fraction of realized warming. Models with initially large atmospheric warming have a delayed oceanic response. The spread of realized warming or cooling is up to 30% around year 400. One reason for the differences is that stronger warming leads to a higher ocean stratification, which reduces diffusive heat uptake [81,82]. Finally, figure 3*c* shows one of several reasons why the oceanic heat uptake efficiency changes over time. The Atlantic Meridional Overturning Circulation (AMOC) decreases owing to the freshwater and heat-flux forcing, but reaches its control run strength after around 1000 years. It responds within decades and a decreased AMOC on decadal timescales leads to an increased heat uptake (figure 3*b*) [83] and reduced surface warming. The magnitude of the AMOC reduction depends on the magnitude of warming. In the cooling case, after strengthening for a decade, the AMOC reduces by a few Sverdrups and stays at its new state without restrengthening as it does in the warming case.

(c) The limits of linear models

So far, we have shown that not only different models show different timescales of equilibration, but also that within one model the response timescales depend on the forcing magnitude and sign. To analyse the constancy of λ, figure 4*a* shows TOA radiative imbalance (N in equation (3.2)) versus the surface temperature anomaly (ΔT) for all experiments. The slope of the regression line through the points of one experiment corresponds to λ, and it should be a straight line [64]. The annual averages of each ensemble member are depicted by small dots, whereas the large dots are initial condition ensemble averages. Annual averages are shown until year 150, after which decadal averages reduce the large internal variability, which dominates over the small forced signal close to the equilibrium. The standard way to estimate the climate feedback parameter λ, effective climate sensitivity (the intersect of the regression line with the horizontal axis), is to linearly regress annual averages of year 1–150 of one realization of a $2 \times CO_2$ step forcing simulation per model [50,51].

There are several known issues with this regression method and the linear assumptions described in §3a. It is unclear how much of the signal in the first year is impacted by the initial conditions and by the tropospheric adjustment to the application of the forcing [84–87]. Figure 4*a* shows a very large spread of responses for the first years. For example, the 120 ensemble members CESM (black) differ by more than $2\,W\,m^{-2}$ and by 1 K for the same forcing in the first year. A deviation from a constant λ has been found in earlier studies not only for the annual timescale, but also the first two decades [62,88,89] and is treated so far inconsistently, by cutting off a few years before regressing N against ΔT, leading to an ambiguous definition of the effective radiative forcing and effective climate sensitivity. Efficacy factors are used for different forcing

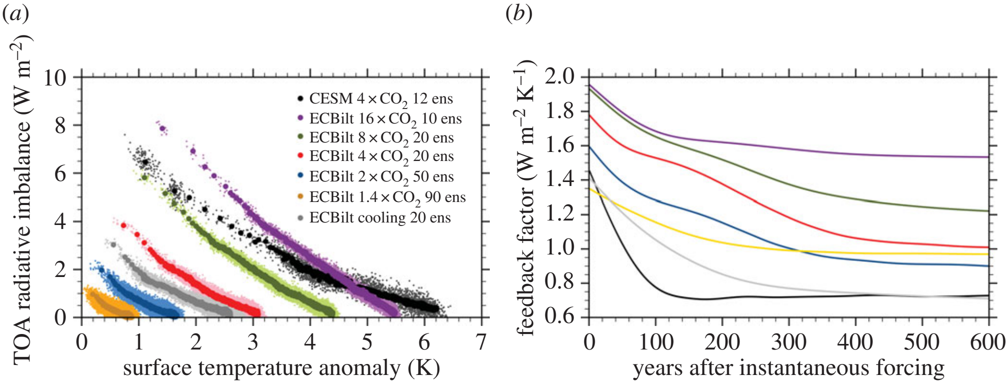

Figure 4. (*a*) Surface temperature equilibration (imbalance N versus temperature change ΔT) for different forcing levels (colours) and models (CESM and ECBilt-CLIO) where each ensemble member annual averages is a small dot and the ensemble mean a large dot (annual until year 150, decadal averages until equilibration). The number of ensemble members is noted in the label (ens). For the first 2 years, 120 CESM ensemble members are used, afterwards only 12. (*b*) Time evolution of the climate feedback parameter λ, according to the 'moving bin regression', thus $\Delta N/\Delta T$. For the cooling case (grey), absolute values are shown. (Online version in colour.)

agents (to account for the different spatial forcing distribution, shortwave versus longwave and top of atmosphere versus surface forcings) and the ability of the ocean to cool the atmosphere by taking up heat [90,91]. A dependency of λ on century timescales has been studied in just a few models [92,93] and can be ascribed to the cloud, albedo and water vapour feedback depending nonlinearly on temperature. Closely related is the dependency on the forcing level, i.e. the temperature dependency not only over equilibration time—thus, temperature—but also as climate base state (e.g. that surface albedo feedbacks will be weaker in a much warmer world without snow and ice) [77,94–96]. State dependency also applies to palaeo studies [47,97,98]. Finally, fully coupled GCMs, with a deep ocean, can amplify feedback magnitudes of lapse rate and short wave cloud feedbacks compared with their slab ocean version [95,99,100]. Recently, it has been suggested that the non-constancy in the global λ is caused by the evolving spatial surface temperature pattern, which (through ΔT) enhances certain local feedbacks at different times [62]. Further, it has been shown that the evolving sea surface temperature pattern alone could explain the time or state dependency of λ [50,101].

To quantify the dependency of λ on the forcing level and the temperature or integration time, we calculate the local derivative ($\Delta N/\Delta T$) in each point. The radiative imbalance is regressed against temperature for all ensemble members of each model in a certain temperature bin—a few Kelvins wide—which is moved in small steps throughout the temperature range. Different bin widths are used for each simulation, according to the level of forcing and density of points in the $N-\Delta T$-space. This 'moving bin regression' circumvents the common problem of either putting more weight to later years when using annual averages, or not addressing the first years when averaging over decades before regressing N against ΔT. The evolution of λ over the temperature range obtained by this moving bin regression is then transferred back to the time domain, shown in figure 4*b*. The apparent time dependence is a temperature dependence. Time, in our case, is characterized by how close a state is to the equilibrium state. After year 600, all model simulations show a near-constant λ (cut off in figure 4*b*). The feedback parameter decreases especially strongly within the first hundred years. For CESM, λ reduces from 1.5 to 0.7 $\mathrm{W\,m^{-2}\,K^{-1}}$. The CMIP5 model mean value obtained with the standard regression method is $1 \pm 0.5\,\mathrm{Wm^{-2}\,K^{-1}}$. Accordingly, the effective climate sensitivity increases in CESM from 4.2 to 6.8 K for $4\times CO_2$. In the runs with a strong CO_2 forcing, the time it takes to reach a roughly constant λ level is several hundred years shorter, and the absolute value is higher, than for the lower CO_2 forcing levels. Even after several hundred years, λ has a small trend. Using 5 year instead of annual averages lead to the same result on timescales longer than 10 years.

4. Are the current concepts of feedbacks and climate sensitivity still useful?

(a) What have we learned from simple models?

Describing a complex system like the climate with a very simple model inevitably means that many factors are ignored, or assumed to be constant. The results above show that the global temperature response to different forcing magnitudes and timescales cannot be fully described with the assumption of a constant feedback parameter λ even in models that ignore long-term Earth system feedbacks (ice sheets, dynamic vegetation, permafrost), non-CO_2 forcings, chemistry and land-use change. In our models, the feedback parameter varies by about 50% or more between different forcing magnitudes and over time as the system approaches equilibrium. The concept of a universal constant climate sensitivity as a fundamental climate system property is very likely wrong, even when ignoring many feedbacks and forcings. This could be an explanation—next to model biases in feedback strength—for the questions outlined in §1 (figure 2). The inconsistency of ECS estimates based on the observed warming and those based on GCMs with freely evolving SST evolution could be partly caused by the assumption of a constant λ. The estimates based on the observed warming, which use an effective radiative forcing estimated from GCMs together with the assumption of a constant λ, would be biased low, if λ would, in fact, not be constant but time or temperature dependent, as shown in figure 4*b*. In the same way, a state and temperature dependency of λ makes the mapping of GCM, palaeo-proxy and short-term observational estimated sensitivities a lot more difficult.

Does this imply the zero-order linear energy balance model is useless? A model is always wrong with regard to reality in a strict sense, but the constant feedback parameter model may still be an adequate approximation for some purposes. As an example, in our case, running the CESM model for 200 years and ignoring the first 150 years for the regression of N against ΔT, would allow us to predict the further evolution of the model. We argue that the quote 'modelling for insight, not numbers' makes an essential point here [102]. We have to conclude that the global linear forcing feedback model may be of limited value to estimate quantities like the ECS of the real world, or at least we have to be more careful in understanding and quantifying in which range of forcings, timescales and climate states a simple model with a constant feedback parameter can be adequately used. But irrespective of whether the numbers tell us much about the real world, such simple models are, and have been, valuable tools to understand fundamental properties of the system [103].

For example, the fact that the transient response simulated in models (or observed, e.g. as the twentieth-century warming), particularly on short timescales, becomes less sensitive to ECS at high sensitivities, and that it is, therefore, harder to constrain the upper bound on ECS [67,68,104] has, in fact, been noted decades ago with simple energy balance models. Wigley & Raper [105] pointed out 'that the response of the climate system to high-frequency forcings such as volcanic eruptions and the seasonal insolation cycle must be virtually independent of the sensitivity. High-frequency information is therefore of little value in trying to estimate, empirically, the climate sensitivity. This is an obvious, but little appreciated result'. Wigley & Schlesinger [106] wrote that 'the observed global warming over the past 100 years can be shown to be compatible with a wide range of CO_2-doubling temperature changes (ECS)', and as a consequence, 'it may be very difficult to determine ΔT_{2x} (ECS) from observational data'.

Recent evidence from observations and models that the climate system will continue to warm for a constant forcing, the commitment warming [65,66], can be traced back to Siegenthaler & Oeschger [107], and Wigley & Schlesinger [106], who noted that 'at any given time, the climate system may be quite far removed from its equilibrium with the prevailing CO_2 level', and Schlesinger [108], who wrote that 'sequestering of heat into the ocean's interior is responsible for the concomitant warming being only about half that which would have occurred in the absence of the ocean. These studies also indicate that the climate system will continue to warm towards its yet unrealized equilibrium temperature change, even if there is no further increase in the CO_2 concentration'. These same authors also demonstrated the causes, shown in figure 3, namely that the characteristic timescales to reach equilibrium range from decades to centuries. These response timescales, and as a consequence the commitment warming, depend on the feedback strength and sensitivity of the model. Hansen *et al*. [109] noted that 'the response times are particularly sensitive to (i) the amount that the climate response is amplified by feedbacks and (ii) the representation of ocean mixing. If ECS is 3°C or greater for a doubling of the carbon dioxide concentration, then most of the expected warming attributable to trace gases added to the atmosphere by man probably has not yet occurred'.

The basic ideas of additive feedbacks enhancing the Planck response also go back to work by Hansen *et al*. [110], and earlier pioneering work, both conceptual and based on climate models [111–114]. All of those old insights are qualitatively still correct, and helpful as thought experiments. More recent work has helped to clarify some of the concepts and point to their limitations (see discussion in §§2 and 3). As GCMs become more complex and include more feedback processes, simple models are necessary to aggregate, approximate and understand the complex models [101,103,115].

(b) Have we made progress?

ECS was initially used as a model benchmark that was simple to calculate and well defined, an overall measure of the response to increased atmospheric CO_2. It is neither a characterization of all aspects of climate change, nor the most relevant number for policy for all questions. The anchoring on ECS as the holy grail of climate science, because the early report by Charney [116] is not helpful. Some feedbacks like clouds were challenging back then [117] and still are [54,55,118–120], and as a result, the uncertainty in climate projections has not decreased much [121]. But observations and models have greatly improved, palaeoclimate has given us a substantially improved view of what has and could happen, we know how to model many processes more realistically, and we have a better understanding of the robust results and key uncertainties. Charney based his conclusions on essentially two GCMs, citing five sources, of which a single one was actually published [122], the other sources were in press, submitted or labelled as personal communication. The published model by Manabe dates back to 1969 and has a limited computational domain with equal areas of land and ocean, an idealized topography, no heat transport by ocean currents and fixed cloudiness. Thus, the fact that the range for climate sensitivity today is similar as was guessed by Charney over three decades ago based on sketchy evidence should not be interpreted as a lack of progress, and using the range of ECS as a measure of success for climate research fails to characterize the state of research.

(c) Possible ways forward

There are 'top-down methods', in which a global feedback is inferred from a global energy balance equation, and 'bottom-up methods', in which the total feedback is an emerging property of the myriad processes that we try to model quantitatively based on insight into each process and data obtained to constrain or parametrize it. There are, of course, methods in between that combine elements. All methods have in common that they are a fusion of models and observations, and there is no pure observational constraint on ECS. Either we define a simple conceptual model like an energy balance model, aggregate the inputs and constrain ECS, such as relating forcing to cooling in the last glacial maximum. We then use complex models to argue that the simple model is correct and consequently use simple models to predict future warming. Alternatively, we use a complex model directly and relate whatever observations we have straight to model quantities, and use a constrained set of models for prediction [43,45,123,124]. In this case, the mapping of a palaeoclimate sensitivity to a modern ECS is not prescribed, but is implicit in the GCM by the fact that feedback changes spatially and as a function of the climate state in the GCM. In all of those questions, the treatment of uncertainties is key. In an energy balance approach, the uncertainties for different time periods are dominated by either uncertainties in radiative forcing, feedback, ocean heat uptake or natural variability. For palaeoclimate, the perturbations are large and the response is close to equilibrium, but forcing and response are uncertain. The strength of the feedbacks may differ, and additional feedbacks may become relevant, as discussed with the difficulties in defining ECS versus ESS in §2. For short timescales and forcings other than CO_2, the feedbacks are different, and variability is large. For climatological constraints, the problem is that climate models have common biases pointing to common problems in representing key feedbacks, because many relevant processes are not resolved but parametrized. Therefore, all methods have uncertainties in the climate models, the observations, the forcings, in structural and statistical assumptions (e.g. priors in Bayesian methods, or assumptions about constant feedback parameters), and in how the estimated sensitivity relates to the present-day ECS in which we are interested.

All methods, but in particular the 'bottom-up' which attempts to simulate each individual process accurately, of course require a detailed process understanding to ensure that no important feedbacks are overlooked. This again requires high-quality long-term and spatially resolved observations, and larger computing capacity to improve (and at some point eliminate where possible) parametrizations of key processes in climate models. New approaches in data assimilation and bridging the gap between numerical weather prediction and climate modelling could be important steps in that direction [125,126].

The understanding of single feedbacks has increased dramatically in the past few years. The focus has moved to understanding the effect of the temperature pattern ΔT (lat, lon, time) that acts on local feedback processes and their aggregation to the global $\lambda\Delta T$ term. Analysing local scales complicates feedback analyses, because the skill of GCMs in simulating regional and local processes is reduced, and model comparisons are more difficult. Trying to understand local feedbacks also includes the evolution of the pattern of ocean heat uptake, heat convergence and TOA imbalance, and research on this subject has barely begun. Understanding regional changes though is more relevant for impact and risk assessments and might bridge the gap between the understanding of global energy budget constraints and localized impact studies. The structural problem of separating individual feedbacks in models—e.g. by keeping parts of the model fixed, or by regression, radiative kernel, or partial radiative perturbation—and comparing them to observations—in which partial derivatives are impossible—persists [52,60]. Next to the evaluation of the full-blown feedback processes in the models, a key challenge is to study the limits of using the linear framework discussed in this paper. How far can one push a GCM into being very sensitive or very insensitive to explore the range of plausible magnitudes of feedbacks and their rate of change? Do cloud, convection and aerosol parametrizations bias GCMs to be too sensitive, or not sensitive enough? For which purposes can we safely use the effective radiative forcing estimates of the linear regression methods? Over which time frames is the assumption of a constant λ justified? Can GCMs serve as a perfect model test bed for simple frameworks, as shown in figure 4? For which climatic base states, feedbacks and their interaction would it be wise to include nonlinear descriptions? For which temperatures, forcing scenarios, and locations does the rate of change of the feedback term matter? When is using a certain fit to estimate the global or regional temperature response justified? How does the coupling of ocean, atmosphere and sea ice determine the evolution of surface temperature patterns enhancing different feedback processes? How can we understand uncertainty propagation in nonlinear systems, with correlated uncertainties, and using computationally expensive climate models? In the light of all these questions, we argue to further explore various uses of feedback frameworks rather than squeezing them into a one-fits-all-concept, and to carefully explore the applicability and predictive capacity of each concept for a range of purposes.

Data accessibility. The data from the climate model simulations are available on request.

Authors' contributions. R.K. and M.A.A.R. jointly designed and wrote the paper. M.A.A.R. analysed the climate model simulations.

Competing interests. The authors declare no competing interests.

Funding. We received no funding for this study.

Acknowledgements. We thank Oliver Stebler for drawing figure 1 and three reviewers for their valuable comments.

References

1. Roe G. 2009 Feedbacks, timescales, and seeing red. *Annu. Rev. Earth Planet. Sci.* **37**, 93–115. (doi:10.1146/annurev.earth.061008.134734)
2. IPCC. 2007 *Climate change 2007: the physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge, UK: Cambridge University Press.
3. Knutti R, Hegerl GC. 2008 The equilibrium sensitivity of the Earth's temperature to radiation changes. *Nat. Geosci.* **1**, 735–743. (doi:10.1038/ngeo337)
4. Knutti R, Rogelj J. In press. The legacy of our CO_2 emissions: a clash of scientific facts, politics and ethics. *Clim. Change*. (doi:10.1007/s10584-015-1340-3)
5. Moss RH *et al.* 2010 The next generation of scenarios for climate change research and assessment. *Nature* **463**, 747–756. (doi:10.1038/nature08823)
6. Previdi M *et al.* 2013 Climate sensitivity in the Anthropocene. *Q. J. Royal Meteorol. Soc.* **139**, 1121–1131. (doi:10.1002/qj.2165)
7. Knutti R *et al.* 2008 A review of uncertainties in global temperature projections over the twenty-first century. *J. Clim.* **21**, 2651–2663. (doi:10.1175/2007jcli2119.1)
8. Tebaldi C, Arblaster JM. 2014 Pattern scaling: Its strengths and limitations, and an update on the latest model simulations. *Clim. Change* **122**, 459–471. (doi:10.1007/s10584-013-1032-9)
9. Herger N, Sanderson BM, Knutti R. 2015 Improved pattern scaling approaches for the use in climate impact studies. *Geophys. Res. Lett.* **42**, 3486–3494. (doi:10.1002/2015GL063569)
10. Huber M, Knutti R. 2012 Anthropogenic and natural warming inferred from changes in Earth's energy balance. *Nat. Geosci.* **5**, 31–36. (doi:10.1038/ngeo1327)
11. Knutti R, Stocker TF, Joos F, Plattner G-K. 2002 Constraints on radiative forcing and future climate change from observations and climate model ensembles. *Nature* **416**, 719–723. (doi:10.1038/416719a)
12. Forest CE, Stone PH, Sokolov AP, Allen MR, Webster MD. 2002 Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science* **295**, 113–117. (doi:10.1126/science.1064419)
13. Andronova NG, Schlesinger ME. 2001 Objective estimation of the probability density function for climate sensitivity. *J. Geophys. Res.* **106**, 22 605–22 612. (doi:10.1029/2000JD000259)
14. Frame D, Booth B, Kettleborough J, Stainforth D, Gregory J, Collins M, Allen M. 2005 Constraining climate forecasts: the role of prior assumptions. *Geophys. Res. Lett.* **32**, L09702. (doi:10.1029/2004GL022241)
15. Forster PMD, Gregory JM. 2006 The climate sensitivity and its components diagnosed from Earth radiation budget data. *J. Clim.* **19**, 39–52. (doi:10.1175/JCLI3611.1)
16. Gregory JM, Stouffer RJ, Raper SCB, Stott PA, Rayner NA. 2002 An observationally based estimate of the climate sensitivity. *J. Clim.* **15**, 3117–3121. (doi:10.1175/1520-0442(2002)015<3117:AOBEOT>2.0.CO;2)
17. Aldrin M, Holden M, Guttorp P, Skeie RB, Myhre G, Berntsen TK. 2012 Bayesian estimation of climate sensitivity based on a simple climate model fitted to observations of hemispheric temperatures and global ocean heat content. *Environmetrics* **23**, 253–271. (doi:10.1002/env.2140)
18. Otto A *et al*. 2013 Energy budget constraints on climate response. *Nat. Geosci.* **6**, 415–416. (doi:10.1038/ngeo1836)
19. Lewis N. 2013 An objective Bayesian improved approach for applying optimal fingerprint techniques to estimate climate sensitivity. *J. Clim.* **26**, 7414–7429. (doi:10.1175/JCLI-D-12-00473.1)
20. Skeie RB, Berntsen T, Aldrin M, Holden M, Myhre G. 2014 A lower and more constrained estimate of climate sensitivity using updated observations and detailed radiative forcing time series. *Earth Syst. Dyn.* **5**, 139–175. (doi:10.5194/esd-5-139-2014)
21. Lewis N, Curry JA. 2015 The implications for climate sensitivity of AR5 forcing and heat uptake estimates. *Clim. Dyn.* **45**, 1009–1023. (doi:10.1007/s00382-014-2342-y)
22. Johansson DJA, O'Neill BC, Tebaldi C, Haggstrom O. 2015 Equilibrium climate sensitivity in light of observations over the warming hiatus. *Nat. Clim. Change* **5**, 449–453. (doi:10.1038/nclimate2573)

23. Knutti R, Tomassini L. 2008 Constraints on the transient climate response from observed global temperature and ocean heat uptake. *Geophys. Res. Lett.* **35**, L09701. (doi:10.1029/2007GL032904)
24. Tomassini L, Reichert P, Knutti R, Stocker TF, Borsuk ME. 2007 Robust Bayesian uncertainty analysis of climate system properties using Markov chain Monte Carlo methods. *J. Clim.* **20**, 1239–1254. (doi:10.1175/jCLI4064.1)
25. Allen MR, Frame DJ, Huntingford C, Jones CD, Lowe JA, Meinshausen M, Meinshausen N. 2009 Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature* **458**, 1163–1166. (doi:10.1038/nature08019)
26. Knutti R, Stocker TF, Joos F, Plattner G-K. 2003 Probabilistic climate change projections using neural networks. *Clim. Dyn.* **21**, 257–272. (doi:10.1007/s00382-003-0345-1)
27. Rogelj J, Meinshausen M, Knutti R. 2012 Global warming under old and new scenarios using IPCC climate sensitivity range estimates. *Nat. Clim. Change* **2**, 248–253. (doi:10.1038/NCLIMATE1385)
28. Sokolov AP *et al.* 2009 Probabilistic forecast for twenty-first-century climate based on uncertainties in emissions (without policy) and climate parameters. *J. Clim.* **23**, 2230–2231. (doi:10.1175/2009JCLI3566.1)
29. Tomassini L, Knutti R, Plattner GK, van Vuuren DP, Stocker TF, Howarth RB, Borsuk ME. 2010 Uncertainty and risk in climate projections for the 21st century: comparing mitigation to non-intervention scenarios. *Clim. Change* **103**, 399–422. (doi:10.1007/s10584-009-9763-3)
30. Rogelj J, Meinshausen M, Sedlacek J, Knutti R. 2014 Implications of potentially lower climate sensitivity on climate projections and policy. *Environ. Res. Lett.* **9**, 031003. (doi:10.1088/1748-9326/9/3/031003)
31. Meinshausen M, Meinshausen N, Hare W, Raper S, Frieler K, Knutti R, Frame D, Allen M. 2009 Greenhouse-gas emission targets for limiting global warming to 2°C. *Nature* **458**, 1158–1162. (doi:10.1038/nature08017)
32. Stott PA, Kettleborough JA. 2002 Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. *Nature* **416**, 723–726. (doi:10.1038/416723a)
33. Boer GJ, Stowasser M, Hamilton K. 2007 Inferring climate sensitivity from volcanic events. *Clim. Dyn.* **28**, 481–502. (doi:10.1007/s00382-006-0193-x)
34. Wigley TML, Ammann CM, Santer BD, Raper SCB. 2005 Effect of climate sensitivity on the response to volcanic forcing. *J. Geophys. Res.* **110**, D09107. (doi:10.1029/2004JD005557)
35. Yokohata T, Emori S, Nozawa T, Tsushima Y, Ogura T, Kimoto M. 2005 Climate response to volcanic forcing: validation of climate sensitivity of a coupled atmosphere-ocean general circulation model. *Geophys. Res. Lett.* **32**, L21710. (doi:10.1029/2005GL023542)
36. Soden BJ, Wetherald RT, Stenchikov GL, Robock A. 2002 Global cooling after the eruption of Mount Pinatubo: a test of climate feedback by water vapor. *Science* **296**, 727–730. (doi:10.1126/science.296.5568.727)
37. Dessler AE. 2010 A determination of the cloud feedback from climate variations over the past decade. *Science* **330**, 1523–1527. (doi:10.1126/science.1192546)
38. Hegerl G, Crowley T, Hyde W, Frame D. 2006 Climate sensitivity constrained by temperature reconstructions over the past seven centuries. *Nature* **440**, 1029–1032. (doi:10.1038/nature04679)
39. Knutti R, Meehl GA, Allen MR, Stainforth DA. 2006 Constraining climate sensitivity from the seasonal cycle in surface temperature. *J. Clim.* **19**, 4224–4233. (doi:10.1175/JCLI3865.1)
40. Masson-Delmotte V *et al.* 2006 Past and future polar amplification of climate change: climate model intercomparisons and ice-core constraints. *Clim. Dyn.* **26**, 513–529. (doi:10.1007/s00382-005-0081-9)
41. Schneider von Deimling T, Held H, Ganopolski A, Rahmstorf S. 2006 Climate sensitivity estimated from ensemble simulations of glacial climate. *Clim. Dyn.* **27**, 149–163. (doi:10.1007/s00382-006-0126-8)
42. Hargreaves JC, Abe-Ouchi A, Annan JD. 2007 Linking glacial and future climates through an ensemble of GCM simulations. *Clim. Past* **3**, 77–87. (doi:10.5194/cp-3-77-2007)
43. Hargreaves JC, Annan JD, Yoshimori M, Abe-Ouchi A. 2012 Can the last glacial maximum constrain climate sensitivity? *Geophys. Res. Lett.* **39**, L24702. (doi:10.1029/2012gl053872)
44. Braconnot P, Harrison SP, Kageyama M, Bartlein PJ, Masson-Delmotte V, Abe-Ouchi A, Otto-Bliesner B, Zhao Y. 2012 Evaluation of climate models using palaeoclimatic data. *Nat. Clim. Change* **2**, 417–424. (doi:10.1038/Nclimate1456)

45. Schmidt GA *et al.* 2014 Using palaeo-climate comparisons to constrain future projections in CMIP5. *Clim. Past* **10**, 221–250. (doi:10.5194/cp-10-221-2014)
46. Rohling E, Members PP. 2012 Making sense of palaeoclimate sensitivity. *Nature* **491**, 683–691. (doi:10.1038/nature11574)
47. Caballero R, Huber M. 2013 State-dependent climate sensitivity in past warm climates and its implications for future climate projections. *Proc. Natl Acad. Sci. USA* **110**, 14 162–14 167. (doi:10.1073/pnas.1303365110)
48. Kohler P, Bintanja R, Fischer H, Joos F, Knutti R, Lohmann G, Masson-Delmotte V. 2010 What caused Earth's temperature variations during the last 800,000 years? Data-based evidence on radiative forcing and constraints on climate sensitivity. *Quat. Sci. Rev.* **29**, 129–145. (doi:10.1016/j.quascirev.2009.09.026)
49. Martinez-Boti MA, Foster GL, Chalk TB, Rohling EJ, Sexton PF, Lunt DJ, Pancost RD, Badger MPS, Schmidt DN. 2015 Plio-Pleistocene climate sensitivity evaluated using high-resolution CO_2 records. *Nature* **518**, 49–54. (doi:10.1038/nature14145)
50. Andrews T, Gregory JM, Webb MJ. 2015 The dependence of radiative forcing and feedback on evolving patterns of surface temperature change in climate models. *J. Clim.* **28**, 1630–1648. (doi:10.1175/JCLI-D-14-00545.1)
51. Andrews T, Gregory JM, Webb MJ, Taylor KE. 2012 Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere–ocean climate models. *Geophys. Res. Lett.* **39**. (doi:10.1029/2012gl051607)
52. Bony S *et al.* 2006 How well do we understand and evaluate climate change feedback processes? *J. Clim.* **19**, 3445–3482. (doi:10.1175/JCLI3819.1)
53. Vial J, Dufresne J-L, Bony S. 2013 On the interpretation of inter-model spread in CMIP5 climate sensitivity estimates. *Clim. Dyn.* **41**, 3339–3362. (doi:10.1007/s00382-013-1725-9)
54. Sherwood SC, Bony S, Dufresne JL. 2014 Spread in model climate sensitivity traced to atmospheric convective mixing. *Nature* **505**, 37–43. (doi:10.1038/nature12829)
55. Fasullo JT, Trenberth KE. 2012 A less cloudy future: the role of subtropical subsidence in climate sensitivity. *Science* **338**, 792–794. (doi:10.1126/science.1227465)
56. Huber M, Mahlstein I, Wild M, Fasullo J, Knutti R. 2011 Constraints on climate sensitivity from radiation patterns in climate models. *J. Clim.* **24**, 1034–1052. (doi:10.1175/2010JCLI3403.1)
57. Sanderson BM, Knutti R, Caldwell PM. 2015 Addressing interdependency in a multi-model ensemble by interpolation of model properties. *J. Clim.* **28**, 5150–5170. (doi:10.1175/JCLI-D-14-00361.1)
58. Zaliapin I, Ghil M. 2010 Another look at climate sensitivity. *Nonlinear Process. Geophys.* **17**, 113–122. (doi:10.5194/npg-17-113-2010)
59. Soden BJ, Held IM. 2006 An assessment of climate feedbacks in coupled ocean-atmosphere models. *J. Clim.* **19**, 3354–3360. (doi:10.1175/JCLI3799.1)
60. Klocke D, Quaas J, Stevens B. 2013 Assessment of different metrics for physical climate feedbacks. *Clim. Dyn.* **41**, 1173–1185. (doi:10.1007/s00382-013-1757-1)
61. Roe GH, Feldl N, Armour KC, Hwang YT, Frierson DMW. 2015 The remote impacts of climate feedbacks on regional climate predictability. *Nat. Geosci.* **8**, 135–139. (doi:10.1038/ngeo2346)
62. Armour KC, Bitz CM, Roe GH. 2013 Time-varying climate sensitivity from regional feedbacks. *J. Clim.* **26**, 4518–4534. (doi:10.1175/JCLI-D-12-00544.1)
63. Feldl N, Roe GH. 2013 The nonlinear and nonlocal nature of climate feedbacks. *J. Clim.* **26**, 8289–8304. (doi:10.1175/JCLI-D-12-00631.1)
64. Gregory JM, Ingram WJ, Palmer MA, Jones GS, Stott PA, Thorpe RB, Lowe JA, Johns TC, Williams KD. 2004 A new method for diagnosing radiative forcing and climate sensitivity. *Geophys. Res. Lett.* **31**, L03205. (doi:10.1029/2003GL018747)
65. Meehl GA, Washington WM, Collins WD, Arblaster JM, Hu A, Buja LE, Strand WG, Teng H. 2005 How much more global warming and sea level rise? *Science* **307**, 1769–1772. (doi:10.1126/science.1106663)
66. Hansen J *et al*. 2005 Earth's energy imbalance: confirmation and implications. *Science* **308**, 1431–1435. (doi:10.1126/science.1110252)
67. Roe GH, Baker MB. 2007 Why is climate sensitivity so unpredictable? *Science* **318**, 629–632. (doi:10.1126/science.1144735)

68. Knutti R, Joos F, Müller S, Plattner G, Stocker T. 2005 Probabilistic climate change projections for CO_2 stabilization profiles. *Geophys. Res. Lett.* **32**, L20707. (doi:10.1029/2005GL023294)
69. Caldeira K, Myhrvold NP. 2013 Projections of the pace of warming following an abrupt increase in atmospheric carbon dioxide concentration. *Environ. Res. Lett.* **8**, 034039. (doi:10.1088/1748-9326/8/3/034039)
70. Good P, Gregory JM, Lowe JA. 2011 A step-response simple climate model to reconstruct and interpret AOGCM projections. *Geophys. Res. Lett.* **38**, FPAGEL01703. (doi:10.1029/2010gl045208)
71. Good P, Gregory JM, Lowe JA, Andrews T. 2013 Abrupt CO_2 experiments as tools for predicting and understanding CMIP5 representative concentration pathway projections. *Clim. Dyn.* **40**, 1041–1053. (doi:10.1007/s00382-012-1410-4)
72. Geoffroy O, Saint-Martin D, Olivie DJL, Voldoire A, Bellon G, Tyteca S. 2013 Transient climate response in a two-layer energy-balance model. Part I: analytical solution and parameter calibration using CMIP5 AOGCM experiments. *J. Clim.* **26**, 1841–1857. (doi:10.1175/JCLI-D-12-00195.1)
73. Geoffroy O, Saint-Martin D, Bellon G, Voldoire A, Olivie DJL, Tyteca S. 2013 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**, 1859–1876. (doi:10.1175/JCLI-D-12-00196.1)
74. Gent PR *et al*. 2011 The community climate system model version 4. *J. Clim.* **24**, 4973–4991. (doi:10.1175/2011JCLI4083.1)
75. Bitz CM, Shell KM, Gent PR, Bailey DA, Danabasoglu G, Armour KC, Holland MM, Kiehl JT. 2012 Climate sensitivity of the community climate system model, version 4. *J. Clim.* **25**, 3053–3070. (doi:10.1175/Jcli-D-11-00290.1)
76. Danabasoglu G, Bates SC, Briegleb BP, Jayne SR, Jochum M, Large WG, Peacock S, Yeager SG. 2012 The CCSM4 ocean component. *J. Clim.* **25**, 1361–1389. (doi:10.1175/Jcli-D-11-00091.1)
77. Good P *et al*. 2015 Nonlinear regional warming with increasing CO_2 concentrations. *Nat. Clim. Change* **5**, 138–142. (doi:10.1038/nclimate2498)
78. Opsteegh JD, Haarsma RJ, Selten FM, Kattenberg A. 1998 ECBILT: a dynamic alternative to mixed boundary conditions in ocean models. *Tellus* **50A**, 348–367. (doi:10.1034/j.1600-0870.1998.t01-1-00007.x)
79. Goosse H, Fichefet T. 1999 Importance of ice-ocean interactions for the global ocean circulation: a model study. *J. Geophys. Res.* **104**, 23 337–23 355. (doi:10.1029/1999JC900215)
80. MacKay RM, Ko MKW. 1997 Normal modes and the transient response of the climate system. *Geophys. Res. Lett.* **24**, 559–562. (doi:10.1029/97gl00286)
81. Stouffer R. 2004 Time scales of climate response. *J. Clim.* **17**, 209–217. (doi:10.1175/1520-0442(2004)017<0209:TSOCR>2.0.CO;2)
82. Stouffer RJ, Russell J, Spelman MJ. 2006 Importance of oceanic heat uptake in transient climate change. *Geophys. Res. Lett.* **33**, L17704. (doi:10.1029/2006gl027242)
83. Knutti R, Stocker TF. 2000 Influence of the thermohaline circulation on projected sea level rise. *J. Clim.* **13**, 1997–2001. (doi:10.1175/1520-0442(2000)013<1997:IOTTCO>2.0.CO;2)
84. Lambert FH, Faull NE. 2007 Tropospheric adjustment: the response of two general circulation models to a change in insolation. *Geophys. Res. Lett.* **34**, L03701. (doi:10.1029/2006gl028124)
85. Bala G, Caldeira K, Nemani R. 2010 Fast versus slow response in climate change: Implications for the global hydrological cycle. *Clim. Dyn.* **35**, 423–434. (doi:10.1007/s00382-009-0583-y)
86. Andrews T, Forster PM. 2008 CO_2 forcing induces semi-direct effects with consequences for climate feedback interpretations. *Geophys. Res. Lett.* **35**, L04802. (doi:10.1029/2007GL032273)
87. Sherwood SC, Bony S, Boucher O, Bretherton C, Forster PM, Gregory JM, Stevens B. 2014 Adjustments in the forcing-feedback framework for understanding climate change. *Bull. Am. Meteorol. Soc.* **96**, 217–228. (doi:10.1175/BAMS-D-13-00167.1)
88. Williams KD, Ingram WJ, Gregory JM. 2008 Time variation of effective climate sensitivity in GCMs. *J. Clim.* **21**, 5076–5090. (doi:10.1175/2008jcli2371.1)
89. Winton M, Anderson WG, Delworth TL, Griffies SM, Hurlin WJ, Rosati A. 2014 Has coarse ocean resolution biased simulations of transient climate sensitivity? *Geophys. Res. Lett.* **41**, 8522–8529. (doi:10.1002/2014gl061523)
90. Hansen J *et al*. 2005 Efficacy of climate forcings. *J. Geophys. Res.* **110**, D18104. (doi:10.1029/2005JD005776)

91. Winton M, Takahashi K, Held IM. 2010 Importance of ocean heat uptake efficacy to transient climate change. *J. Clim.* **23**, 2333–2344. (doi:10.1175/2009jcli3139.1)
92. Senior CA, Mitchell JFB. 2000 The time-dependence of climate sensitivity. *Geophys. Res. Lett.* **27**, 2685–2688. (doi:10.1029/2000GL011373)
93. Meraner K, Mauritsen T, Voigt A. 2013 Robust increase in equilibrium climate sensitivity under global warming. *Geophys. Res. Lett.* **40**, 5944–5948. (doi:10.1002/2013gl058118)
94. Colman R, McAvaney B. 2009 Climate feedbacks under a very broad range of forcing. *Geophys. Res. Lett.* **36**, L01702. (doi:10.1029/2008GL036268)
95. Jonko AK, Shell KM, Sanderson BM, Danabasoglu G. 2012 Climate feedbacks in CCSM3 under changing CO_2 forcing. Part I: Adapting the linear radiative kernel technique to feedback calculations for a broad range of forcings. *J. Clim.* **25**, 5260–5272. (doi:10.1175/Jcli-D-11-00524.1)
96. Jonko AK, Shell KM, Sanderson BM, Danabasoglu G. 2013 Climate feedbacks in CCSM3 under changing CO_2 forcing. Part II: variation of climate feedbacks and sensitivity with forcing. *J. Clim.* **26**, 2784–2795. (doi:10.1175/Jcli-D-12-00479.1)
97. Zeebe RE. 2013 Time-dependent climate sensitivity and the legacy of anthropogenic greenhouse gas emissions. *Proc. Natl Acad. Sci. USA* **110**, 13 739–13 744. (doi:10.1073/pnas.1222843110)
98. von der Heydt AS, Köhler P, van de Wal RSW, Dijkstra HA. 2014 On the state dependency of fast feedback processes in (paleo) climate sensitivity. *Geophys. Res. Lett.* **41**, 6484–6492. (doi:10.1002/2014GL061121)
99. Danabasoglu G, Gent P. 2009 Equilibrium climate sensitivity: Is it accurate to use a slab ocean model? *J. Clim.* **22**, 2494–2499. (doi:10.1175/2008JCLI2596.1)
100. Boer GJ, Yu B. 2003 Dynamical aspects of climate sensitivity. *Geophys. Res. Lett.* **30**, 1135. (doi:10.1029/2002GL016549)
101. Gregory JM, Andrews T, Good P. 2015 The inconstancy of the transient climate response parameter under increasing CO_2. *Phil. Trans. R. Soc. A* **373**, 20140417. (doi:10.1098/rsta.2014.0417)
102. Bloomfield SD, Updegrove DA. 1981 Modeling for insight, not numbers. *New Dir. Higher Educ.* 93–104. (doi:10.1002/he.36919813509)
103. Held IM. 2005 The gap between simulation and understanding in climate modeling. *Bull. Am. Meteorol. Soc.* **80**, 1609–1614. (doi:10.1175/BAMS-86-11-1609)
104. Allen MR *et al.* 2006 Observational constraints on climate sensitivity. In *Avoiding dangerous climate change* (eds HJ Schellnhuber, W Cramer, N Nakicenovic, T Wigley, G Yohe), pp. 281–289. Cambridge, UK: Cambridge University Press.
105. Wigley TML, Raper SCB. 1990 Natural variability of the climate system and detection of the greenhouse effect. *Nature* **344**, 324–327. (doi:10.1038/344324a0)
106. Wigley TML, Schlesinger ME. 1985 Analytical solution for the effect of increasing CO_2 on global mean temperature. *Nature* **315**, 649–652. (doi:10.1038/315649a0)
107. Siegenthaler U, Oeschger H. 1984 Transient temperature changes due to increasing CO_2 using simple models. *Ann. Glaciol.* **5**, 153–159.
108. Schlesinger M. 1986 Equilibrium and transient climatic warming induced by increased atmospheric CO_2. *Clim. Dyn.* **1**, 35–51. (doi:10.1007/BF01277045)
109. Hansen J, Russell G, Lacis A, Fung I, Rind D, Stone P. 1985 Climate response-times: dependence on climate sensitivity and ocean mixing. *Science* **229**, 857–859. (doi:10.1126/science.229.4716.857)
110. Hansen J, Lacis A, Rind D, Russell GL, Stone P, Fung I, Ruedy R, Lerner J 1984 Climate sensitivity: analysis of feedback mechanisms. In *Climate processes and climate sensitivity* (eds J Hansen, T Takahashi), pp. 130–163. Washington, DC: American Geophysical Union.
111. Manabe S, Wetherald RT. 1967 Thermal equilibrium of the atmosphere with a given distribution of relative humidity. *J. Atmos. Sci.* **50**, 241–259. (doi:10.1175/1520-0469(1967)024<0241:TEOTAW>2.0.CO;2)
112. Budyko MI. 1969 The effect of solar radiation variations on the climate of the earth. *Tellus* **21**, 611–619. (doi:10.1111/j.2153-3490.1969.tb00466.x)
113. Sellers WD. 1969 A global climatic model based on the energy balance of the earth-atmosphere system. *J. Appl. Meteorol.* **8**, 392–400. (doi:10.1175/1520-0450(1969)008<0392:AGCMBO>2.0.CO;2)

114. North GR, Cahalan RF, Coakley JA. 1981 Energy balance climate models. *Rev. Geophys. Space Phys.* **19**, 91–121. (doi:10.1029/RG019i001p00091)
115. Bouttes N, Gregory JM, Lowe JA. 2013 The reversibility of sea level rise. *J. Clim.* **26**, 2502–2513. (doi:10.1175/JCLI-D-12-00285.1).
116. Charney JG 1979 *Carbon dioxide and climate: a scientific assessment*, 22 p. Washington, DC: National Academy of Science.
117. Cess RD *et al*. 1989 Interpretation of cloud-climate feedback as produced by 14 atmospheric general-circulation models. *Science* **245**, 513–516. (doi:10.1126/science.245.4917.513)
118. Bony S, Dufresne JL. 2005 Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophys. Res. Lett.* **32**, L20806. (doi:10.1029/2005gl023851)
119. Bony S *et al*. 2015 Clouds, circulation and climate sensitivity. *Nat. Geosci.* **8**, 261–268. (doi:10.1038/Ngeo2398)
120. Stevens B, Bony S. 2013 What are climate models missing? *Science* **340**, 1053–1054. (doi:10.1126/science.1237554)
121. Knutti R, Sedláček J. 2013 Robustness and uncertainties in the new CMIP5 climate model projections. *Nat. Clim. Change* **3**, 369–373. (doi:10.1038/nclimate1716)
122. Manabe S, Wetherald RT. 1975 The effects of doubling the CO_2 concentration on the climate of a general circulation model. *J. Atmos. Sci.* **32**, 3–15. (doi:10.1175/1520-0469(1975)032<0003:TEODTC>2.0.CO;2)
123. Rowlands DJ *et al*. 2012 Broad range of 2050 warming from an observationally constrained large climate model ensemble. *Nat. Geosci.* **5**, 256–260. (doi:10.1038/ngeo1430)
124. Annan JD, Hargreaves JC. 2015 A perspective on model-data surface temperature comparison at the Last Glacial Maximum. *Quat. Sci. Rev.* **107**, 1–10. (doi:10.1016/j.quascirev.2014.09.019)
125. Palmer TN, Doblas-Reyes FJ, Weisheimer A, Rodwell MJ. 2008 Toward seamless predicion: calibration of climate change projections using seasonal forecasts. *Bull. Am. Meteorol. Soc.***89**, 459–470. (doi:10.1175/BAMS-89-4-459)
126. Rodwell MJ, Palmer TN. 2007 Using numerical weather prediction to assess climate models. *Q. J. Royal Meteorol. Soc.* **133**, 129–146. (doi:10.1002/qj.23)