



Research Article

Theoretical foundations of emergent constraints: relationships between climate sensitivity and global temperature variability in conceptual models

Mark S Williamson*, Peter M Cox and Femke J M M Nijse

Exeter Climate Systems, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Laver Building, North Park Road, Exeter EX4 4QE, UK

*Correspondence Mark S Williamson, Exeter Climate Systems, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Laver Building, North Park Road, Exeter EX4 4QE, UK; E-mail: m.s.williamson@exeter.ac.uk

Received 9 February 2018; Accepted 29 February 2018

Abstract

Background. The emergent constraint approach has received interest recently as a way of utilizing multi-model General Circulation Model (GCM) ensembles to identify relationships between observable variations of climate and future projections of climate change. These relationships, in combination with observations of the real climate system, can be used to infer an emergent constraint on the strength of that future projection in the real system. However, there is as yet no theoretical framework to guide the search for emergent constraints. As a result, there are significant risks that indiscriminate data-mining of the multidimensional outputs from GCMs could lead to spurious correlations and less than robust constraints on future changes. To mitigate against this risk, Cox et al. (Cox et al. Emergent constraint on equilibrium climate sensitivity from global temperature variability. *Nature* 2018a; 553: 319, hereafter CHW18) proposed a theory-motivated emergent constraint, using the one-box Hasselmann model to identify a linear relationship between equilibrium climate sensitivity (ECS) and a metric of global temperature variability involving both temperature standard deviation and autocorrelation (Ψ). A number of doubts have been raised about this approach, some concerning the application of the one-box model to understand relationships in complex GCMs, which are known to have more than the single characteristic timescale.

Objectives. To study whether the linear Ψ –ECS proportionality in CHW18 is an artefact of the one-box model. More precisely, we ask ‘Does the linear Ψ –ECS relationship feature in the more complex and realistic two-box and diffusion models?’.

Methods. We solve the two-box and diffusion models to find relationships between ECS and Ψ . These models are forced continually with white noise parameterizing internal variability. The resulting analytical relations are essentially fluctuation–dissipation theorems.

Results. We show that the linear Ψ -ECS proportionality in the one-box model is not generally true in the two-box and diffusion models. However, the linear proportionality is a very good approximation for parameter ranges applicable to the current state-of-the-art CMIP5 climate models. This is not obvious—due to structural differences between the conceptual models, their predictions also differ. For example, the two-box and diffusion, unlike the one-box model, can reproduce the long-term transient behaviour of the CMIP5 abrupt4xCO₂ and 1pcCO₂ simulations. Each of the conceptual models also predicts different power spectra with only the diffusion model's pink 1/f spectrum being compatible with observations and GCMs. We also show that the theoretically predicted Ψ -ECS relationship exists in the piControl as well as historical CMIP5 experiments and that the differing gradients of the proportionality are inversely related to the effective forcing in that experiment.

Conclusions. We argue that emergent constraints should ideally be derived by such theory-driven hypothesis testing, in part to protect against spurious correlations from blind data-mining but mainly to aid understanding. In this approach, an underlying model is proposed, the model is used to predict a potential emergent relationship between an observable and an unknown future projection, and the hypothesized emergent relationship is tested against an ensemble of GCMs.

Key words: Emergent constraint, equilibrium climate sensitivity, fluctuation-dissipation theorem, global temperature variability.

1. Introduction

Emergent constraints (Hall and Qu 2006; Allen and Ingram, 2002) provide a promising way to relate observations of the present day to future projections of the climate. The usual approach is to take a model ensemble (such as the multi-model CMIP5 ensemble; Taylor *et al.*, 2012) and use it to find a relationship via a scatter plot between an observable plotted on the x axis and the future projection plotted on the y axis, each point on the plot being one member of the model ensemble. The model ensemble derived relationship or *emergent relationship*, and the uncertainty in it, can then be determined from regression on the scatter plot. A measurement of the observable in the real world can be combined with the model-derived emergent relationship to produce an *emergent constraint* on the climate projection.

Hall and Qu (2006) published one of the first emergent constraints relating the strength of the snow albedo feedback in the seasonal cycle (the observable) to the strength of the snow albedo feedback in climate projections within the multi-model ensemble used in IPCC AR4 (Meehl *et al.*, 2001). Since then, many others have been published including studies on sea-ice (Boe *et al.*, 2009; Massonnet *et al.*, 2012), tropical precipitation extremes (O'Gorman, 2012), equilibrium climate sensitivity (ECS) (Annan and Hargreaves, 2006; Huber *et al.*, 2010; Fasullo and Trenberth, 2012; Brown and Caldeira, 2017; Cox *et al.*, 2018a), carbon loss from tropical land under warming (Cox *et al.*, 2013), zonal shift of Southern Hemisphere westerlies (Kidston and Gerber, 2010), cloud feedbacks (Brient and Bony, 2013; Sherwood *et al.*, 2014; Klein and Hall, 2015; Tian, 2015; Tsushima *et al.*, 2016; Lipat *et al.*, 2017), strengthening of the hydrological cycle (DeAngelis *et al.*, 2015), the climate-carbon cycle feedback (Wenzel *et al.*, 2014) and CO₂ fertilization effect (Wenzel *et al.*, 2016), future changes in ocean net primary production (Kwiatkowski *et al.*, 2017), permafrost melt (Chadburn *et al.*, 2017) and changes in natural sources and sinks of CO₂ (Hoffman *et al.*, 2014).

Some scepticism about emergent constraints is healthy, particularly when they are not founded on well understood physical processes. There are significant risks that indiscriminate data-mining of the multidimensional outputs from models could lead to spurious correlations (Caldwell *et al.*, 2014) and less than robust constraints on future changes (Bracegirdle and Stephenson, 2013). Care is also needed drawing statistical inferences from ensembles of small numbers of models. The problem is compounded if models within the ensemble share common components giving a smaller effective ensemble size (Pennell and Reichler, 2011; Masson and Knutti, 2011; Herger *et al.*, 2018). Observations used to guide model development also may lead to dependencies (Masson and Knutti, 2012).

To minimize these risks, a theoretical framework for finding and evaluating emergent constraints is needed. The approach described here involves a form of hypothesis testing, in which an underlying simple, conceptual model is proposed, the model is used to predict an emergent relationship between an observable and an unknown future projection, and the predicted emergent relationship is tested against results from an ensemble of more complex models. Emergent relationships are usually assumed to be univariate and linear, but these are not necessary simplifications. As

an example, we illustrate this theory-led approach using simple conceptual models of the global mean temperature as emergent constraints on ECS and test the theoretically predicted relations against observations and the CMIP5 models.

In CHW18, the theoretical linear relationship between a measure of the variability of annual mean global surface air temperature, the observable Ψ and the ECS (the future projection) was used to derive an emergent constraint on ECS. Colman and Power (2018) also found a correlation between the tropical decadal temperature standard deviation and ECS in the CMIP5 models. A number of doubts have been raised about CHW18 (Brown *et al.*, 2018; Po-Chedley *et al.*, 2018; Rypdal *et al.*, 2018; Cox *et al.*, 2018b), some concerning the theory and the application of the one-box model to understand relationships in complex GCMs which are known to have more than the single characteristic timescale (MacMynowski *et al.*, 2011, Geoffroy *et al.*, 2013). In section 2, we investigate whether the relation in CHW18 derived for the one-box model still holds in more realistic yet still analytically soluble conceptual models, namely the often used two-box and diffusion models. It is known the two-box and diffusion models unlike the one-box model are able to reproduce the long-term transient behaviour of the CMIP5 GCM abrupt4xCO2 and 1pcCO2 simulations (Caldeira and Myhrvold, 2013; Geoffroy *et al.*, 2013). Although we find the one-box linear proportionality between Ψ and ECS is generally no longer true in the two-box and diffusion models, we show the linear proportionality holds to a good approximation for both when the range of their parameters are applicable to the complex CMIP5 GCMs. This gives us increased confidence in the theoretical foundation of CHW18.

It is important to note that each of these conceptual models differs structurally, predicts different temperature responses and will not be able to reproduce all of the features of the global mean temperature response. One could loosely view these conceptual models as zeroth-order approximations and GCMs as higher order approximations of the real world. The often used quote ‘all models are wrong but some are useful’ is quite apt as a guiding principle for this manuscript. The usefulness of the model will depend on the question asked of it.

In section 3, conceptual model predictions are compared with the CMIP5 ensemble and observations, particularly the power spectra and autocorrelation functions. The pink power spectrum of global mean temperature in observations and CMIP5 models can only be reproduced by the diffusion model. However, if one is interested in the shorter timescale behaviour for use as an emergent constraint on ECS, we find the simplest conceptual one-box model will serve as a good approximation.

Also in section 3, Ψ versus ECS emergent relationships for both piControl and historical CMIP5 experiments are shown. Both have the theoretically predicted linear proportionality although they have differing gradients. This difference in gradient is theoretically expected to scale inversely with the effective forcing, and this is also observed in the CMIP5 models.

The current article can be seen as a companion article to CHW18, as it examines and tests the appropriateness of the theory used to inform that study.

2. Conceptual models relating global temperature variability to ECS

Caldeira and Myhrvold (2013) (hereafter CM13) fitted three different conceptual models, namely the one-box, two-box and diffusion models to the annual global mean air temperature time series of the CMIP5 abrupt4xCO2 experiments (Taylor *et al.*, 2012). These fits were then tested against the 1pcCO2 CMIP5 experiments. CM13 showed that while the one-box model was a poor fit to either experiment on longer timescales both the two-box and diffusion models did good jobs. Here we use these conceptual models to analyse the annual global mean air temperature variability in the CMIP5 historical experiments with a view to obtaining ECS as a function of Ψ as found in CHW18 for the one-box model.

Each of the conceptual models has differing numbers of free parameters, the one-box and diffusion models have three, and the two-box model has five. None of these are assumed to be fixed. These parameters are essentially fitted to each of the CMIP5 models and the observations in the historical period via Ψ (introduced in equation (12)). The models are introduced in order of complexity and completeness, the one-box model being the simplest analytically, while the diffusion model is harder to solve but reproduces more of the observed temperature response.

The historical time series can be approximated as the sum of the responses to the forcing resulting from changes in the atmospheric composition. These include greenhouse gases, tropospheric and stratospheric aerosols from large volcanic eruptions and solar variability. There is also a response to fast, internal variability that is parameterized here as the response to random, white noise forcing. We seek to isolate the response to the latter and relate it to ECS.

A relation between the system response to random fluctuations and its sensitivity is essentially a fluctuation-dissipation theorem (Kubo, 1966; Leith, 1975).

ECS is defined as the equilibrium temperature change due to the constant forcing $Q_{2\times CO_2}$ from the doubling of CO_2 ,

$$ECS = \frac{Q_{2\times CO_2}}{\lambda}, \quad (1)$$

and λ is the climate feedback factor.

Linearity of the conceptual models allows each temperature response $T_i(t)$ to each forcing $Q_i(t)$ to be added to give the total response, i.e. if the total forcing is $Q(t) = \sum_i Q_i(t)$, then the total temperature response is just the sum of the temperature responses to each of the individual forcings $T(t) = \sum_i T_i(t)$ (principle of superposition). Linearity means that by suitable detrending the response from the trend in emissions can be removed from the total temperature response to leave the residual response, $\Delta T(t)$, to the random forcing. For this study, we assume this detrending can be carried out to a good approximation and work with just the residual temperature. For notational ease, we also refer to ΔT as T , i.e. $\Delta T = T$.

Although the theory we derive here assumes external, random forcing, we have shown that the Ψ -ECS linear proportionality will theoretically become more tightly defined in the presence of common forcing across a model ensemble (Cox *et al.*, 2018b). The gradient of the relationship does however change, being roughly inversely proportional to the amplitude of the forcing (see section 3).

The superposition principle implies the response to any forcing can be written as the convolution of the linear response function $g(t)$ (the response to delta function forcing) with the forcing, i.e.

$$T(t) = \int_0^t g(t-s)Q(s)ds. \quad (2)$$

Each model can therefore be characterized by $g(t)$. We will be interested in their response in the stationary limit, i.e. when $t \gg \tau$ where τ is the longest timescale in the model. The residual response is found when $Q(t)$ is a Gaussian random variable with zero mean and a standard deviation of σ_Q , $Q(t) = \sigma_Q dW_t$, turning equation (2) into a stochastic integral

$$T(t) = \sigma_Q \int_0^t g(t-s)dW_s \quad (3)$$

where W_s is a Wiener process.

In the following we choose to use the two observables variance σ_T^2 and autocorrelation $\alpha_T(t)$ as fitting parameters for the conceptual models, as they can be easily estimated for a given time series, be it a CMIP5 model or observations. These can be computed for the residual temperature by using equation (3) and the relevant model $g(t)$ via the autocovariance $R(t)$

$$R(t) = \lim_{P \rightarrow \infty} \frac{1}{P} \int_0^P T(s-t)T(s)ds. \quad (4)$$

$$\sigma_T^2 = R(0), \quad (5)$$

$$\alpha_T(t) = \frac{R(t)}{R(0)}. \quad (6)$$

Another useful quantity we use to compare the simple models to the CMIP5 models and observations is the power spectrum of T , $|T(\omega)|^2$, which can be found from the Fourier transform of the autocovariance

$$|T(\omega)|^2 = \int_{-\infty}^{\infty} R(t)e^{-i\omega t} dt. \quad (7)$$

2.1. Hasselmann one-box model

The simplest, one-box (or one-timescale) model for the evolution of $T(t)$ is

$$C \frac{dT}{dt} = Q(t) - \lambda T \quad (8)$$

In this model, the climate system can be thought of as a single well-mixed box with effective heat capacity C forced by $Q(t)$ and adjusting to this forcing with climate sensitivity λ proportional to the temperature anomaly. The single well-mixed box can be roughly thought of as representing the atmosphere, surface mixed ocean layer and the land.

The linear response function for this model is

$$g(t) = \Theta(t) \frac{e^{-\frac{t}{\tau_H}}}{\lambda \tau_H} \quad (9)$$

where the timescale in the model $\tau_H = \frac{C}{\lambda}$ and $\Theta(t)$ is the Heaviside step function. When the forcing $Q(t)$ is Gaussian white noise, equation (8) is known as the Hasselmann model (Hasselmann, 1976). Variance and autocorrelation for the one-box model can be computed from equations (3) and (4) to be

$$\sigma_T^2 = \frac{\sigma_Q^2}{2\lambda^2\tau_H}, \quad (10)$$

$$\alpha_T(t) = e^{-\frac{|t|}{\tau_H}}. \quad (11)$$

These equations can be combined with equation (1) to give (Cox *et al.*, 2018a)

$$ECS = \sqrt{2} \frac{Q_{2 \times CO_2}}{\sigma_Q} \Psi, \quad (12)$$

where Ψ is defined as

$$\Psi = \frac{\sigma_T}{\sqrt{-\log \alpha_{1T}}}, \quad (13)$$

and $\alpha_{1T} = \alpha_T(1\text{year})$. It was equation (12), namely the linear proportionality between the observable Ψ , estimated from time series of T and the future projection ECS, which was used to guide the search for an emergent constraint in CHW18. The magnitude of proportionality between Ψ and ECS, the ratio of the effective forcing due to doubling CO_2 and the mean amplitude of the effective forcing in the experiment $\sigma_Q \sqrt{2} \frac{Q_{2 \times CO_2}}{\sigma_Q}$ cannot be observed but is fortunately weakly correlated with ECS ($r = -0.02$) across the CMIP5 model ensemble (Cox *et al.*, 2018a). By linearly regressing Ψ against ECS, the magnitude of proportionality is therefore determined by the model ensemble itself.

The power spectrum of the one-box model is

$$|T(\omega)|^2 = \frac{\sigma_Q^2}{\lambda^2(1 + \omega^2\tau_H^2)}. \quad (14)$$

This model predicts a red power spectrum temperature response, that is, the power scales inversely to the square of the forcing frequency ω .

2.2. Two-box model

The two-box model (Gregory, 2000; Held *et al.*, 2010; Geoffroy *et al.*, 2013) consists of two well-mixed layers, one representing the upper mixed layer of the ocean plus the lower atmosphere, with effective heat capacity C and

temperature T , and a second well-mixed box representing the deep ocean with heat capacity C_0 and temperature T_0 . Heat transport between the two boxes is proportional to the temperature difference between the two boxes with constant of proportionality γ . The equations describing the evolution of temperature are therefore

$$C \frac{dT}{dt} = Q(t) - \lambda T - \gamma(T - T_0), \quad (15)$$

$$C_0 \frac{dT_0}{dt} = \gamma(T - T_0). \quad (16)$$

This model has two timescales, a fast τ_f and slow one τ_s . The linear response function is the sum of the fast and slow modes with amplitudes $\frac{a_f}{\tau_f}$ and $\frac{a_s}{\tau_s}$,

$$g(t) = \frac{\Theta(t)}{\lambda} \left(\frac{a_f}{\tau_f} e^{-\frac{t}{\tau_f}} + \frac{a_s}{\tau_s} e^{-\frac{t}{\tau_s}} \right). \quad (17)$$

This model has been extensively used in previous climate applications and here we use the notation and expressions for the amplitudes and timescales in terms of the quantities in equation (15) as given in [Geoffroy et al. \(2013\)](#). They also fitted this model to abrupt4xCO2 CMIP5 experiments for which they found two widely separated timescales, typical values being $\tau_f \sim 4$ years and $\tau_s \sim 250$ years, while the dimensionless mode parameters, a_f and a_s , were roughly of equal size ($a_f \sim 3/5$ and $a_s \sim 2/5$).

The autocovariance function for Gaussian white noise forcing can be found by using equations (3) and (4) and in contrast to the one-box model features two modes:

$$R(t) = \frac{\sigma_Q^2}{2\lambda^2} \left\{ \frac{a_f^2}{\tau_f} e^{-\frac{|t|}{\tau_f}} + \frac{a_s^2}{\tau_s} e^{-\frac{|t|}{\tau_s}} + \frac{2a_f a_s}{\tau_f + \tau_s} \left(e^{-\frac{|t|}{\tau_f}} + e^{-\frac{|t|}{\tau_s}} \right) \right\} \quad (18)$$

giving

$$\sigma_T^2 = \frac{\sigma_Q^2}{2\lambda^2} \left(\frac{a_f^2}{\tau_f} + \frac{a_s^2}{\tau_s} + \frac{4a_f a_s}{\tau_f + \tau_s} \right), \quad (19)$$

$$\alpha_T(t) = \frac{(a_f^2 \tau_s e^{-\frac{|t|}{\tau_f}} + a_s^2 \tau_f e^{-\frac{|t|}{\tau_s}})(\tau_f + \tau_s) + 2a_f a_s \tau_f \tau_s (e^{-\frac{|t|}{\tau_f}} + e^{-\frac{|t|}{\tau_s}})}{(a_f^2 \tau_s + a_s^2 \tau_f)(\tau_f + \tau_s) + 4a_f a_s \tau_f \tau_s}. \quad (20)$$

This general result simplifies for typical fitted parameters to the CMIP5 models ([Geoffroy et al., 2013](#)) as the variance and the autocorrelation are dominated by the fast mode. These can be approximated in the limit ($\tau_s \gg \tau_f$, $a_s \sim a_f$) by:

$$\sigma_T^2 \approx \frac{\sigma_Q^2 a_f^2}{2\lambda^2 \tau_f}, \quad (21)$$

$$\alpha_T(t) \approx e^{-\frac{|t|}{\tau_f}}. \quad (22)$$

The approximate expressions are therefore very similar to the one-box model for the CMIP5 models. Combining these expressions with the equation for ECS gives

$$ECS = \sqrt{2} \frac{Q_{2 \times \text{CO}_2} \Psi}{\sigma_Q a_f} \quad (23)$$

so that the linear relationship between Ψ and ECS found in the one-box model also approximately holds for the two-box model. The constant of proportionality is however different, and it has an extra factor in the denominator a_f , which is roughly constant and is approximately $a_f \sim \lambda/(\lambda + \gamma)$ over the CMIP5 model range of parameters. Relative standard deviation in a_f is 13%. The reason for the approximate equivalence between the ECS relations in one- and two-box models is due to the wide separation in timescales between the two modes fitted to the CMIP5 models. As in the one-box case, the ‘constant’ of proportionality between ECS and Ψ , $\sqrt{2} \frac{Q_{2 \times CO_2}}{\sigma_Q a_f}$, is weakly correlated with ECS ($r = 0.03$) across the CMIP5 models and one can linearly regress Ψ against ECS for a theoretical emergent relationship.

In contrast to the one-box, the two-box power spectrum is

$$|T(\omega)|^2 = \frac{\sigma_Q^2}{\lambda^2} \left\{ \frac{1 + \omega^2(a_f \tau_s + a_s \tau_f)^2}{(1 + \omega^2 \tau_f^2)(1 + \omega^2 \tau_s^2)} \right\}. \quad (24)$$

which depending on the size of terms can give red and ω^{-4} scaling, although when fitted to the CMIP5 models, the spectrum is approximately red.

2.3. Diffusion equation

The diffusion equation (or heat equation) model (MacMynowski *et al.*, 2011; Caldeira and Myhrvold, 2013) consists of a continuous vertical layer, $z \geq 0$, where radiative forcing at the surface ($z = 0$) causes heating which is transported down through the water column by diffusion (parameterized by diffusivity \mathcal{D}), representing heat uptake by the deep ocean. A mixed-layer surface box has also been added in previous studies to add realism (Oeschger *et al.*, 1975; Hansen *et al.*, 1985; Fraedrich *et al.*, 2004) although here we use just the diffusion equation for simplicity. The model is described by a partial differential equation:

$$\frac{\partial T}{\partial t} = \mathcal{D} \frac{\partial^2 T}{\partial z^2} \quad (25)$$

with flux boundary conditions

$$-\rho c_p \mathcal{D} \frac{\partial T}{\partial z} \Big|_{z=0} = Q(t) - \lambda T(0, t), \quad (26)$$

$$\frac{\partial T}{\partial z} \Big|_{z=z_{\max}} = 0. \quad (27)$$

where ρ and c_p are the density and specific heat capacity of water, respectively. The maximal depth of the ocean, z_{\max} , is taken to be infinite. The temperature is now a function of both depth and time, $T(z, t)$, although our interest is only in the surface temperature $T(0, t)$.

In contrast to the one- and two-box models, the ocean is modelled as a vertical continuum rather than a finite number of well-mixed boxes. As heat is diffused down the water column with time, the effective heat capacity increases as the heat sees more ocean. This model can also be thought of as an M -box model where M is very large and each well-mixed box is very thin resulting in a continuum of (M) timescales. The diffusion model reduces to a one-box model when times of interest are larger than the time taken for heat to be well diffused throughout the water column, i.e. when $t > z_{\max}^2/\mathcal{D}$. For ocean depths of $z_{\max} = 4000$ m and typical diffusivities of $\mathcal{D} = 5 \times 10^{-5} \text{ m}^2 \text{ s}^{-1}$, this happens when $t > 10\,000$ years, and so this limiting case is not met for the application here.

The linear response function for surface temperature $T(0, t)$ can be found using Laplace transforms on equations (25) and (26) to be

$$g(0, t) = \frac{\Theta(t)}{\lambda} \left\{ \frac{1}{\sqrt{\pi \tau_D t}} - \frac{e^{-\frac{t}{\tau_D}}}{\tau_D} \operatorname{erfc} \left(\sqrt{\frac{t}{\tau_D}} \right) \right\} \quad (28)$$

A timescale, τ_D , can be identified in this model as $\tau_D = \frac{\rho^2 c_p^2 D}{\lambda^2}$. $\tau_D \sim 25$ years for the mean value of D found in CM13 (Caldeira and Myhrvold, 2013). One needs to be aware of an unphysical infinity in $g(0, t)$ at $t = 0$ because of the time dependence of the effective heat capacity. At $t = 0$, this results in zero effective heat capacity and therefore an infinite response. In reality, energy is not absorbed in an infinitely thin surface layer and thus care needs to be taken when calculating at $t = 0$.

The power spectrum at the surface can also be found using either Laplace or Fourier transforms. This is given by

$$|T(0, \omega)|^2 = \frac{\sigma_Q^2}{\lambda^2} \left\{ \frac{1}{1 + \sqrt{2}|\omega|\tau_D + |\omega|\tau_D} \right\}. \quad (29)$$

The diffusion model therefore predicts a pink spectrum, i.e. power scales inversely proportional to ω^{-1} in contrast to the red spectra predicted by the one- and two-box models.

To obtain the autocovariance function at $z = 0$, we start with the power spectrum and Fourier transform it using equation (7):

$$\begin{aligned} R(t) &= \frac{\sigma_Q^2}{2\lambda^2\tau_D} \left\{ e^{-\frac{t}{\tau_D}} \left[\operatorname{erfi}\left(\sqrt{\frac{t}{\tau_D}}\right) + \frac{1}{\pi} E_1\left(-\frac{t}{\tau_D}\right) \right] \right. \\ &\quad \left. - e^{\frac{t}{\tau_D}} \left[\operatorname{erfc}\left(\sqrt{\frac{t}{\tau_D}}\right) - \frac{1}{\pi} E_1\left(\frac{t}{\tau_D}\right) \right] \right\} \end{aligned} \quad (30)$$

this rather long exact expression can be well approximated more compactly as

$$R(t) \approx \frac{2\sigma_Q^2}{\pi\lambda^2\tau_D} E_1\left(\sqrt{\frac{\pi t}{\tau_D}}\right) \quad (31)$$

where the exponential integral $E_1(x)$ is defined as $E_1(x) = \int_x^\infty \frac{e^{-t}}{t} dt$. Unfortunately, $R(t=0) = \sigma_T^2$ is also infinite because of the unphysical zero effective heat capacity. σ_T^2 can however be approximated by taking a very small but finite time, t_0 . Starting with equation (31) and Taylor expanding the exponential integral to zeroth order around $t = 0$ results in

$$\sigma_T^2 \approx \frac{2\sigma_Q^2}{\pi\lambda^2\tau_D} \left\{ -\gamma_{EM} - \log\left(\sqrt{\frac{\pi t_0}{\tau_D}}\right) \right\} \quad (32)$$

$$\approx \frac{c_0\sigma_Q^2}{\pi\lambda^2} \tau_D^{\frac{1-c_0}{c_0}} \quad (33)$$

where $c_0 = -2\gamma_{EM} - \log(\pi t_0)$, $\gamma_{EM} \approx 0.577$ is the Euler–Mascheroni constant and in the second line, the approximation $\log x = c_0 x^{\frac{1}{c_0}} - c_0$ has been used. This approximation gets better for larger c_0 (smaller t_0). So for small t_0

$$\sigma_T^2 \rightarrow \frac{c_0\sigma_Q^2}{\pi\lambda^2\tau_D} \quad (34)$$

and the autocorrelation function is

$$\alpha_T(t) \approx \frac{2}{c_0} E_1\left(\sqrt{\frac{\pi t}{\tau_D}}\right) \quad (35)$$

which is purely a function of τ_D .

Rearranging equation (34) and combining with the equation for ECS (equation (1)) gives

$$ECS = \sqrt{\frac{\pi}{c_0}} \frac{Q_{2 \times CO_2}}{\sigma_Q} \Psi_D \quad (36)$$

where

$$\Psi_D = \sigma_T \sqrt{\tau_D}. \quad (37)$$

Or in terms of observables

$$\Psi_D = \frac{\sigma_T \sqrt{\pi}}{E_1^{-1}\left(\frac{c_0 \alpha_{1T}}{2}\right)} \quad (38)$$

where $E_1^{-1}(x)$ is defined as the inverse of the exponential integral. The linear Ψ -ECS proportionality is not true for the diffusion model. However, comparing equation (37) with similar for one- and two-box models (equation (13)), if

$$\sqrt{\tau_D} \propto \frac{1}{\sqrt{-\log \alpha_{1T}}} \quad (39)$$

is approximately true then the linear ECS- Ψ proportionality is also approximately true for the diffusion model. By plotting one against the other in [Figure 1](#), this is the case for the range of values of τ_D applicable to CMIP5 models ($\tau_D \in [10, 60]$ years in CM13). α_{1T} is calculated from the exact formula, equation (30).

3. Comparison with CMIP5 models and observations

Theoretical autocorrelation functions and power spectra predicted by the conceptual models are shown in [Figure 2](#) for typical values found in fits to the CMIP5 models ([Caldeira and Myhrvold, 2013](#); [Geoffroy et al., 2013](#)). One- and two-box autocorrelation functions and power spectra are very similar for timescales < 100 years. Power spectra in these models have the same $|T(\omega)|^2 \propto \omega^{-2}$ red power spectra. In contrast to the box models, the diffusion model has a faster drop off in autocorrelation but a slower approach to equilibrium and a power spectrum that predicts a $|T(\omega)|^2 \propto \omega^{-1}$ pink power spectrum.

For comparison with the conceptual plots, the CMIP5 historical runs (coloured lines) and the HadCRUT4 historical observational data set ([Morice et al., 2012](#)) (thick black line) are shown in [Figure 3](#). The power spectra of the HadCRUT4 observations and CMIP5 models show approximately a $|T(\omega)|^2 \propto \omega^{-1}$ pink spectrum most closely resembled by the diffusion model. The dotted white line is shown as a guide to this proportionality. High sensitivity CMIP5 models also generally have higher autocorrelation. The HadCRUT4 autocorrelation is more representative of the low sensitivity models, consistent with the findings of CHW18.

We have used detrended CMIP5 historical simulations as a comparison to observations can also be made and an emergent constraint obtained. However, conceptual model theoretical relations have been derived assuming white noise external forcing as a parameterization of internal variability. The CMIP5 piControl experiments are the closest analogue to this simplification and one may wonder whether the same relations hold in these experiments as it is known the forced response may not always be the same as the response to internal variability ([Lucarini and Sarno, 2011](#); [Gritsun and Lucarini, 2017](#)). Power spectra and autocorrelation functions for the piControl experiments are broadly the same as [Figure 3](#) (not shown). The linear Ψ -ECS emergent relationships are also similarly strong in both piControl and historical simulations having correlations of $r = 0.68$ and $r = 0.77$, respectively. The higher correlation in the historical experiment resulting in a reduced uncertainty emergent constraint is theoretically expected when there is common forcing across the model ensemble (see [Cox et al., 2018b](#)). In this case, the common forcing in the historical experiment is provided by the increasing concentrations of greenhouse gases, aerosols and volcanic eruptions.

There are differences in the emergent relationships however. In [Figure 4a](#), the emergent relationships for the piControl and historical have different gradients. This is due to increased effective forcing in the historical simulations from

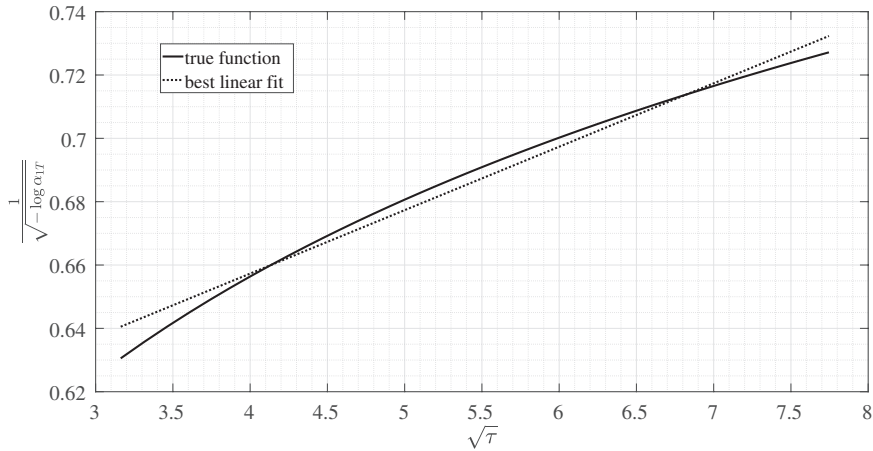


Figure 1. $\sqrt{\tau_D}$ versus $\frac{1}{\sqrt{-\log \alpha_{1T}}}$ for the diffusion model. If these two functions are linearly proportional, then Ψ is also linearly proportional to ECS for the diffusion model. Although it is slightly nonlinear, for this range of values (solid line) linearity seems to be a good approximation (dotted line). α_{1T} is calculated from the exact formula, equation (30) with $t_0 = 10^{-6}$ years (~ 1 min) and $\tau_D \in [10, 60]$ years. This spans the range of values of τ_D found in fits to CMIP5 models (Caldeira and Myhrvold, 2013).

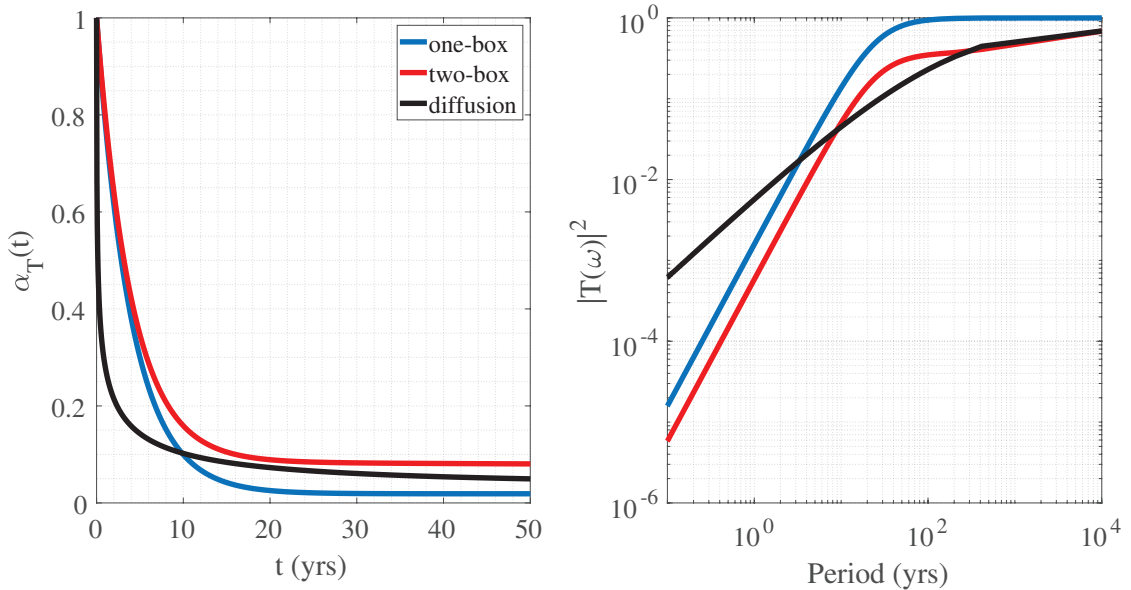


Figure 2. Autocorrelation, $\alpha_T(t)$, (left) and power spectrum, $|T(\omega)|^2$, (right) for the three conceptual models. $\lambda = \sigma_Q = 1$ are the same in all curves, while $\tau = \tau_f = 4$ years for the one- and two-box models. For the two-box model $\tau_s = 250$ years and $a_f = 3/5$, $a_s = 2/5$ (these are the mean values found by Geoffroy et al. (2013) in fits to the CMIP5 models). For the diffusion model, $\tau_D = 25$ years. Power spectra and autocorrelation functions are roughly the same for one- and two-box models at timescales less than a decade. For short periods, the diffusion model has a $|T(\omega)|^2 \propto \omega^{-1}$ (power proportional to period) pink spectrum, whereas the one- and two-box models show a $|T(\omega)|^2 \propto \omega^{-2}$ (power proportional to the square of the period) red spectra.

residual volcanic, aerosol and greenhouse gas forcing remaining after the detrending procedure. From equation (12), an inverse relationship with the magnitude of the effective forcing σ_Q is expected. When Ψ is divided by the estimated forcing, $\sigma_N \sim \sigma_Q$, in Figure 4b, gradients are very similar. Forcing has been inferred in the CMIP5 models from the net top-of-atmosphere radiative flux N where $N \approx Q - \lambda T$ with standard deviation σ_N .

4. Discussion and conclusions

All three conceptual models have both physical similarities and deficiencies relative to the CMIP5 models and the real Earth system. The one-box model only really has any physical justification when the timescales of interest are dominated by the well-mixed atmosphere and ocean surface layer. This has led some to question the use of the one-box model by CHW18 to motivate their emergent constraint between ECS and Ψ , a statistic dominated by the fast timescale processes, e.g. Rypdal *et al.* (2018). However, in this article, we have shown that a near-linear relationship is to be expected between ECS and Ψ for more realistic conceptual models. For the one- and two-box models, we were able to find analytical relations to show this. Semi-analytical relations for the diffusion model also show a similar near-linear relationship.

Even though a linear proportionality between Ψ and ECS is expected in the conceptual models for regions of parameter space applicable to CMIP5 models, each of the conceptual models predicts different temperature responses. The one-box model cannot reproduce the long timescale behaviour of the two-box or diffusion model, and neither the one- or two-box models can mimic the observed and CMIP5 power spectra. Of the three, the diffusion model reproduces the power spectra of the CMIP5 models and the observations most closely although it is more difficult to work with and has some deficiencies as an analogue to the real climate system. Combining a well-mixed atmosphere-surface ocean box with a diffusive continuous deep ocean (Oeschger *et al.*, 1975; Hansen *et al.*, 1985; Fraedrich *et al.*, 2004), although adding another layer of complexity and making the model less analytically amenable, would add physical realism. We suspect this would produce a similar linear relation to the one- and two-box models as well as mimicking the CMIP5 and HadCRUT4 power spectra due to the timescale separation between surface mixed and deep layers.

In conceptual models, we therefore expect to find emergent relationships between ECS and short-term variability (e.g. as measured by Ψ). However, the underlying models considered here remain deliberately very simple compared with the GCMs we are using to define emergent constraints. It is therefore vital that we continue to check that our conceptual models provide useful insights into the spread of projections from GCMs. We see this as a form of hypothesis testing, in which a conceptual model is proposed, an emergent relationship between variability and sensitivity is predicted based-on that conceptual model, and then that predicted emergent relationship is checked against the

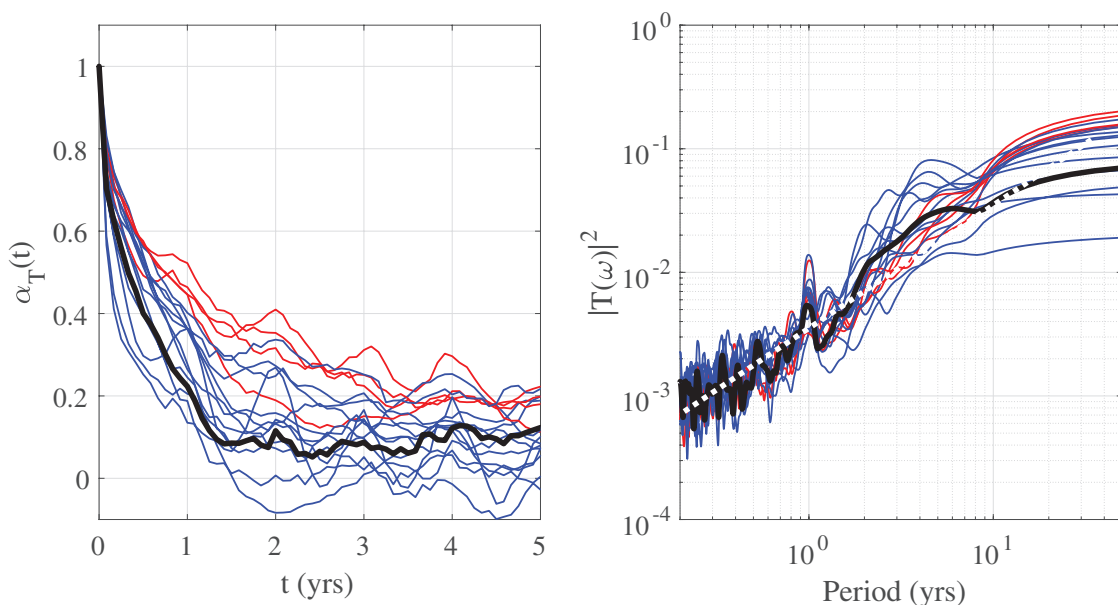


Figure 3. Autocorrelation, $\alpha_T(t)$, (left) and power spectrum, $|T(\omega)|^2$, (right) for the CMIP5 model historical runs. The CMIP5 models used are the same as in CHW18 (Cox *et al.*, 2018a). Red lines are higher sensitivity models ($\lambda < 1 \text{ W m}^{-2} \text{ K}^{-1}$), while blue lines are lower sensitivity models ($\lambda \geq 1 \text{ W m}^{-2} \text{ K}^{-1}$). The black line is the (historical) HadCRUT4 observational dataset. The white dotted line in the right-hand power spectrum plot is a guide to show the pink spectrum predicted by the diffusion model. The power spectra have been smoothed with a 25 point moving average window.

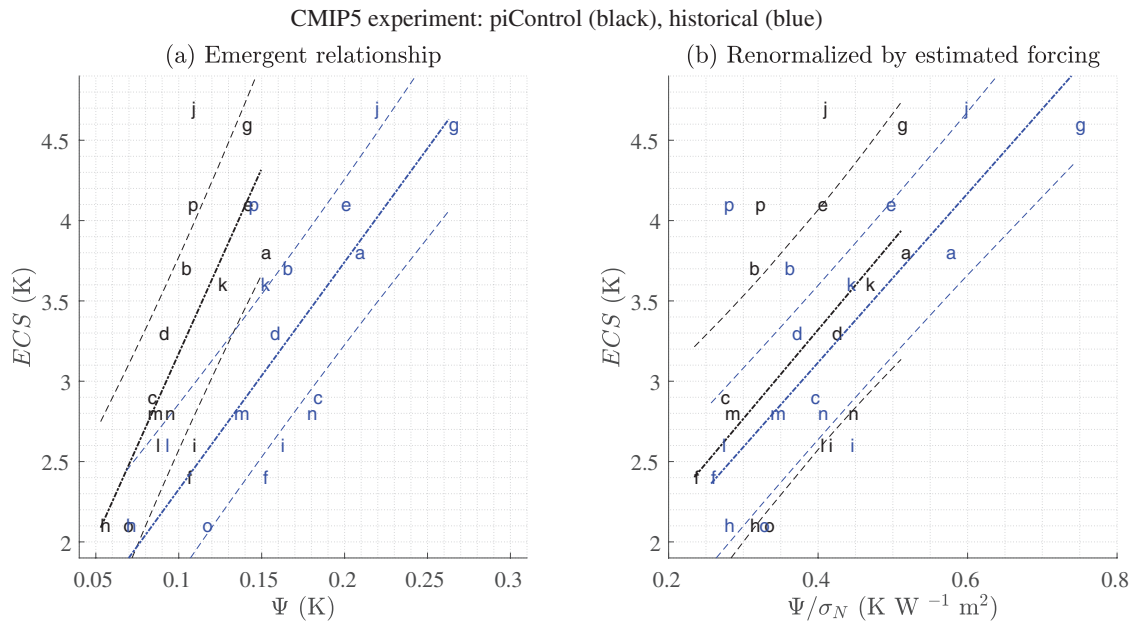


Figure 4. (a) ECS versus Ψ emergent relationships in the historical (blue) and piControl (black) CMIP5 experiments. Each letter plotted is a CMIP5 model and corresponds to the same used in CHW18 (model-letter correspondence is given in Table 1). The emergent relationship is calculated between 1880 and 2015 for the historical experiment and for the first 135 years of piControl using the same methodology of CHW18. The gradient of the emergent relationship (dashed line) is theoretically predicted to be smaller with increased forcing in the experiment. This is why the historical run with its volcanic, greenhouse gas, aerosol and internal forcing has a shallower gradient than the control run. (b) is the same plot with Ψ renormalized by the estimated forcing from σ_N , the standard deviation of the top-of-atmosphere radiative forcing. Emergent relationships then have a very similar gradient, illustrating the inverse proportionality of the gradient to the forcing in (a).

Table 1. CMIP5 models used in Figure 4 (letters in the first column identify the models) and in CHW18

	Model	ECS (K)	λ (W m ⁻² K ⁻¹)	$Q_{2\times CO_2}$ (W m ⁻²)
a	ACCESS1-0	3.8	0.8	3.0
b	CanESM2	3.7	1.0	3.8
c	CCSM4	2.9	1.2	3.6
d	CNRM-CM5	3.3	1.1	3.7
e	CSIRO-MK3-6-0	4.1	0.6	2.6
f	GFDL-ESM2M	2.4	1.4	3.4
g	HadGEM2-ES	4.6	0.6	2.9
h	inmcm4	2.1	1.4	3.0
i	IPSL-CM5B-LR	2.6	1.0	2.7
j	MIROC-ESM	4.7	0.9	4.3
k	MPI-ESM-LR	3.6	1.1	4.1
l	MRI-CGCM3	2.6	1.2	3.2
m	NorESM1-M	2.8	1.1	3.1
n	bcc-csm1-1	2.8	1.1	3.2
o	GISS-E2-R	2.1	1.8	3.8
p	BNU-ESM	4.1	1.0	3.9

Values are taken from IPCC AR5 table 9.5 (Collins *et al.*, 2013).

ensemble of full-form GCMs. This approach requires that the search for emergent constraints becomes more theory led than it has been to date, but would also guard against spurious relationships that could easily arise from blind data-mining of the many diagnostics available from modern GCMs. Most importantly, in our view, such theory-led hypothesis testing is much more likely to improve understanding of the climate system than purely statistically derived emergent constraints.

Declaration

Funding: EU Horizon 2020 Research Programme CRESCENDO project (641816 to P.M.C. and M.S.W.); the EPSRC-funded ReCoVER project (M.S.W.); the European Research Council (ERC) ECCLES project (742472 to P.M.C. and F.J.M.M.N.).

Acknowledgements

We also acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups for producing and making available their model output.

Ethical approval: none.

Conflict of interest: none.

References

- Allen MR, Ingram WJ. Constraints on future changes in climate and the hydrologic cycle. *Nature* 2002; 419: 224–32.
- Annan JD, Hargreaves JC. Using multiple observationally-based constraints to estimate climate sensitivity. *Geophys Res Lett* 2006; 33. doi:10.1029/2005gl025127.
- Boe J, Hall A, Qu X. September sea-ice cover in the arctic ocean projected to vanish by 2100. *Nat Geosci* 2009; 2: 341–43.
- Bracegirdle TJ, Stephenson DB. On the robustness of emergent constraints used in multimodel climate change projections of arctic warming. *J Clim* 2013; 26: 669–78.
- Brient F, Bony S. Interpretation of the positive low-cloud feedback predicted by a climate model under global warming. *Clim Dyn* 2013; 40: 2415–31.
- Brown PT, Caldeira K. Greater future global warming inferred from earth's recent energy budget. *Nature* 2017; 552: 45.
- Brown PT, Stolpe MB, Caldeira K. Assumptions for emergent constraints. *Nature* 2018; 563: E1–E3.
- Caldeira K, Myhrvold NP. Projections of the pace of warming following an abrupt increase in atmospheric carbon dioxide concentration. *Environ Res Lett* 2013; 8: 034039.
- Caldwell PM, Bretherton CS, Zelinka MD et al. Statistical significance of climate sensitivity predictors obtained by data mining. *Geophys Res Lett* 2014; 41: 1803–08.
- Chadburn SE, Burke EJ, Cox PM et al. An observation-based constraint on permafrost loss as a function of global warming. *Nat Clim Change* 2017; 7: 340–44.
- Collins M, Knutti R, Arblaster J et al. Climate change 2013: the physical science basis. In: Stocker T F, Qin D, Plattner G K et al. (eds). *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK: Cambridge University Press, 2013.
- Colman R, Power SB. What can decadal variability tell us about climate feedbacks and sensitivity?. *Clim Dyn* 2018; 51: 3815.
- Cox PM, Huntingford C, Williamson MS. Emergent constraint on equilibrium climate sensitivity from global temperature variability. *Nature* 2018a; 553: 319.
- Cox PM, Williamson MS, Nijse FJMM, Huntingford C. Cox et al. reply. *Nature* 2018b; 563: E10–E15.
- Cox PM, Pearson D, Booth BB et al. Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability. *Nature* 2013; 494: 341.
- Meehl GA, Stocker T, Collins W et al. "Global climate projections." in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by Solomon S, Qin D, Manning M et al. UK: Cambridge University Press 2007.
- DeAngelis AM, Qu X, Zelinka MD, Hall A. An observational radiative constraint on hydrologic cycle intensification. *Nature* 2015; 528: 249–53.
- Fasullo JT, Trenberth KE. A less cloudy future: the role of subtropical subsidence in climate sensitivity. *Science* 2012; 338: 792–4.
- Fraedrich K, Luksch U, Blender R. $1/f$ model for long-time memory of the ocean surface temperature. *Phys Rev E Stat Nonlin Soft Matter Phys* 2004; 70: 037301.
- Geoffroy O, Saint-Martin D, Oliv   A et al. Transient climate response in a two-layer energy balance model. part I: analytical solution and parameter calibration using CMIP5 AOGCM experiments. *J Clim* 2013; 26: 1841–57.

- Gregory JM. Vertical heat transports in the ocean and their effect on time-dependent climate change. *Clim Dyn* 2000; 16: 501–15.
- Gritsun A, Lucarini V. *Phys D Nonlin Phenom* 2017; 349: 62.
- Hall A, Qu X. Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophys Res Lett* 2006; 33. doi:10.1029/2005gl025127.
- Hansen J, Russell G, Lacs A et al. Climate response times: dependence on climate sensitivity and ocean mixing. *Science* 1985; 229: 857–9. <http://science.sciencemag.org/content/sci/229/4716/857.full.pdf>.
- Hasselmann K. Stochastic climate models. part I. theory. *Tellus* 1976; 28: 473–84.
- Held IM, Winton M, Takahashi K et al. Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *J Clim* 2010; 23: 2418–27.
- Herger N, Abramowitz G, Knutti R et al. Selecting a climate model subset to optimise key ensemble properties. *Earth Syst Dyn Discuss* 2018; 9: 135–51.
- Hoffman FM, Randerson JT, Arora VK et al. Causes and implications of persistent atmospheric carbon dioxide biases in Earth system models. *J Geophys Res Biogeosci* 2014; 119: 141–62.
- Huber M, Mahlstein I, Wild M, Fasullo J, Knutti R. Constraints on climate sensitivity from radiation patterns in climate models. *J Clim* 2010; 24: 1034–52.
- Kidston J, Gerber EP. Intermodel variability of the poleward shift of the austral jet stream in the CMIP3 integrations linked to biases in 20th century climatology. *Geophys Res Lett* 2010; 37. doi:10.1029/2010GL042873.
- Klein SA, Hall A. Emergent constraints for cloud feedbacks. *Curr Clim Change Rep* 2015; 1: 276–87.
- Kubo R. The fluctuation-dissipation theorem. *Rep Progr Phys* 1966; 29: 255.
- Kwiatkowski L, Bopp L, Aumont O et al. Emergent constraints on projections of declining primary production in the tropical oceans. *Nat Clim Change* 2017; 7: 355.
- Leith CE. Climate response and fluctuation dissipation. *J Atmos Sci* 1975; 32: 2022–26.
- Lipat BR, Tselioudis G, Grise KM, Polvani LM. CMIP5 models' shortwave cloud radiative response and climate sensitivity linked to the climatological Hadley cell extent. *Geophys Res Lett* 2017; 44: 5739–48.
- Lucarini V, Sarno S. A statistical mechanical approach for the computation of the climatic response to general forcings. *Nonlin Proc Geophys* 2011; 18: 7–28.
- MacMynowski DG, Shin HJ, Caldeira K. The frequency response of temperature and precipitation in a climate model. *Geophys Res Lett* 2011; 38: L16711.
- Masson D, Knutti R. Climate model genealogy. *Geophys Res Lett* 2011; 38 doi:10.1029/2011GL046864.
- Masson D, Knutti R. Predictor screening, calibration, and observational constraints in climate model ensembles: an illustration using climate sensitivity. *J Clim* 2012; 26: 887–98.
- Massonnet F, Fichefet T, Gosse H et al. Constraining projections of summer arctic sea ice. *Cryosphere* 2012; 6: 1383–94.
- Morice CP, Kennedy JJ, Rayner NA, Jones PD. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 data set. *J Geophys Res Atmos* 2012; 117. doi:10.1029/2011JD017187.
- O'Gorman PA. Sensitivity of tropical precipitation extremes to climate change. *Nature Geosci* 2012; 5: 697–700.
- Oeschger H, Siegenthaler U, Schotterer U, Gugelmann A. *Tellus* 1975; 27: 168–92.
- Pennell C, Reichler T. On the effective number of climate models. *J Clim* 2011; 24: 2358–367.
- Po-Chedley S, Proistosescu C, Armour KC, Santer BD. Climate constraint reflects forced signal. *Nature* 2018; 563: E6–E9.
- Rypdal M, Fredriksen H-B, Rypdal K, Steene RJ. Emergent constraints on climate sensitivity. *Nature* 2018; 563: E4–E5.
- Sherwood SC, Bony S, Dufresne JL. Spread in model climate sensitivity traced to atmospheric convective mixing. *Nature* 2014; 505: 37–42.
- Taylor KE, Stouffer RJ, Meehl GA. An overview of CMIP5 and the experiment design. *Bul Am Meteorol Soc* 2012; 93: 485–98.
- Tian B. Spread of model climate sensitivity linked to double-intertropical convergence zone bias. *Geophys Res Lett* 2015; 42: 4133–41.
- Tsushima Y, Ringer MA, Koshiro T et al. Robustness, uncertainties, and emergent constraints in the radiative responses of stratocumulus cloud regimes to future warming. *Clim Dyn* 2016; 46: 3025–39.
- Wenzel S, Cox PM, Eyring V, Friedlingstein P. Emergent constraints on climate-carbon cycle feedbacks in the CMIP5 Earth system models. *J Geophys Res Biogeosci* 2014; 119: 2013JG002591.
- Wenzel S, Cox PM, Eyring V, Friedlingstein P. Projected land photosynthesis constrained by changes in the seasonal cycle of atmospheric CO₂. *Nature* 2016; 538: 499–501.