Emergent constraints on climate sensitivities

Mark S. Williamson◉

College of Engineering, Mathematics and Physical Sciences,
University of Exeter, Exeter EX4 4QF, United Kingdom
and Global Systems Institute, University of Exeter, Exeter EX4 4QE, United Kingdom

Chad W. Thackeray

Department of Atmospheric and Oceanic Sciences, University of California,
Los Angeles, California 90095, USA

Peter M. Cox

College of Engineering, Mathematics and Physical Sciences,
University of Exeter, Exeter EX4 4QF, United Kingdom

Alex Hall

Department of Atmospheric and Oceanic Sciences, University of California,
Los Angeles, California 90095, USA

Chris Huntingford

Centre for Ecology and Hydrology, Wallingford OX10 8BB, United Kingdom

Femke J. M. M. Nijsse

College of Engineering, Mathematics and Physical Sciences,
University of Exeter, Exeter EX4 4QF, United Kingdom

(published 11 May 2021)

Despite major advances in climate science over the last 30 years, persistent uncertainties in projections of future climate change remain. Climate projections are produced with increasingly complex models that attempt to represent key processes in the Earth system, including atmospheric and oceanic circulations, convection, clouds, snow, sea ice, vegetation, and interactions with the carbon cycle. Uncertainties in the representation of these processes feed through into a range of projections from the many state-of-the-art climate models now being developed and used worldwide. For example, despite major improvements in climate models, the range of equilibrium global warming due to doubling carbon dioxide still spans a range of more than 3. Here a promising way to make use of the ensemble of climate models to reduce the uncertainties in the sensitivities of the real climate system is reviewed. The emergent constraint approach uses the model ensemble to identify a relationship between an uncertain aspect of the future climate and an observable variation or trend in the contemporary climate. This review summarizes previous published work on emergent constraints and discusses the promise and potential dangers of the approach. Most importantly, it argues that emergent constraints should be based on well-founded physical principles such as the fluctuation-dissipation theorem. This review will stimulate physicists to contribute to the rapidly developing field of emergent constraints on climate projections, bringing to it much needed rigor and physical insights.

CONTENTS

I. INTRODUCTION

Numerical methods have become a standard technique to simulate complex systems. The equations governing components of such systems may be well known. But their solutions cannot be solved analytically, creating a need for numerical approaches. Because of the discretization of time and space inherent in numerical techniques, modeling complex systems must involve ways to include effects of unresolved processes. Often there is no "first principles" approach to do this. Typically, the effects of unresolved processes are included by resorting to quasiempirical relationships between them and explicitly resolved variables, otherwise known as "parametrization." There are usually multiple defensible ways to parametrize unresolved processes. Thus, independently developed models of the same complex system might incorporate different parametrization choices. The more models that are independently developed, the greater the diversity of approaches for modeling the same natural system.

A classic example of this model diversity is the use of numerical models of the atmosphere to predict hurricane development. Initial conditions are imposed on a model at some time, and it is integrated forward in time to produce simulations of critical hurricane features such as intensity and track. This approach might be replicated for multiple models incorporating different parametrization choices, producing an "ensemble" of hurricane forecasts. The spread in the forecasts is a measure of the uncertainty in the future hurricane behavior, given the range of plausible approaches to atmospheric modeling. Over time, with enough hurricanes and associated predictions, the various models can be evaluated for their prediction skill. Certain parametrization choices may emerge as producing systematically better predictions. The models can then be rebuilt or recalibrated with the better choices. Over time the ensemble will become more skillful, with less spread. In fact, this is approximately the process that has resulted in dramatic improvements not only in hurricane prediction but also in weather forecasting generally over the past seven decades.

Earth's climate is another example of a complex system whose governing equations can be solved only through numerical methods. (In fact, the dynamical equations for the atmospheric component of a climate model are typically almost identical to those in the previously referred to hurricane models.) As expected, there are a variety of plausible approaches to parametrization in the components of Earth system models (ESMs).[1] Thus, modeling groups throughout the world have built a few dozen ESMs with different approaches to parametrization. Because of these differences, these models produce different future climate states, even when the same scenario of radiative forcing is imposed (associated, for example, with an increase in greenhouse gases).

A classic climate change experiment is to double CO_2 concentrations in the atmospheric component of an ESM and measure the surface warming that occurs after the simulation has equilibrated (Manabe and Wetherald, 1975), an important

---

[1]State-of-the-art climate models are also commonly called general circulation models (GCMs). This was particularly true in the past when they consisted of just an atmosphere and sometimes an ocean. As time has progressed and more processes have been included the term ESM has become more common (previously ESM may have referred to models of reduced complexity featuring a carbon cycle). Here we use the term ESM to mean a full complexity dynamical state-of-the-art climate model, although this could be used interchangeably with GCM.

number in climate science referred to as equilibrium climate sensitivity (ECS). Reaching a true equilibrium with a full complexity ESM requires long, computationally expensive simulations of thousands of years, so ECS is usually estimated from shorter duration CO_2 doubling experiments. When these experiments are done with contemporary ESMs, the spread in values across the ensemble of ESMs is large, between 1.5 and 6 °C (Forster *et al.*, 2020), although true equilibrium values are higher (median 17% higher) (Rugenstein *et al.*, 2020). The international climate science community has organized itself to generate scenarios of greenhouse gas (GHG) emissions that result in more realistic future radiative forcing than the CO_2 doubling experiment. These correspond to scenarios of controls (or lack thereof) on future GHG emissions (Moss *et al.*, 2010; Riahi *et al.*, 2017). The scenarios are designed so that the same radiative forcing time series is imposed on each ESM. This allows for intercomparison of ESM responses at some future specified time, say, the end of the 21st century. The ESM responses vary significantly across the ensemble, an indicator of deep uncertainty in Earth's climate future, simply due to the variety of plausible ways to handle parametrization within ESM components.

If each ESM is treated as an equally plausible analog of the real climate system, known as "model democracy," an undesirably large factor of 3 difference in future response to man-made GHG response currently results. Apart from the subject of this review, there are several techniques used to reduce this uncertainty and much research effort has been dedicated to this important task. One approach is to give ESMs that simulate the real world better (according to some chosen metric) more weight in future projections. There are also methods that do not use ESMs at all, by constraining future response from past observations and those that use a combination of all of these. Many more details were given by Eyring *et al.* (2019) and Sherwood *et al.* (2020) and in Chaps. 9 and 11–14 of IPCC (2013).

The approach taken to improve hurricane forecasting models, to evaluate their performance and adjust parametrizations over the course of multiple prediction cycles, is unfortunately impractical in the case of projecting future climate with ESMs. With future climate, we have only one realization of the real system's trajectory, and our mandate is to predict it as best we can now. To compound the problem, the climate system involves many more components than the atmosphere alone, including the ocean, the land surface, glaciers and ice sheets, and the marine and terrestrial biospheres. The engineering required to model each component and allow them to interact in a simulation is impressive. But the components that can interact with one another in information flows are so complex that it is not easy to predict how behavior in one component might affect behavior in another. For example, parametrization reformulation or adjustment in one component can significantly affect the simulated state of another component for reasons that are not always clear (Donner *et al.*, 2011).

To circumvent the impossibility of directly evaluating ESMs for their ability to simulate a future climate state that has not yet been observed, a new technique has emerged over the past decade and a half, known as the *emergent constraint* (EC) approach (Fig. 1 shows a schematic). The basic idea is to
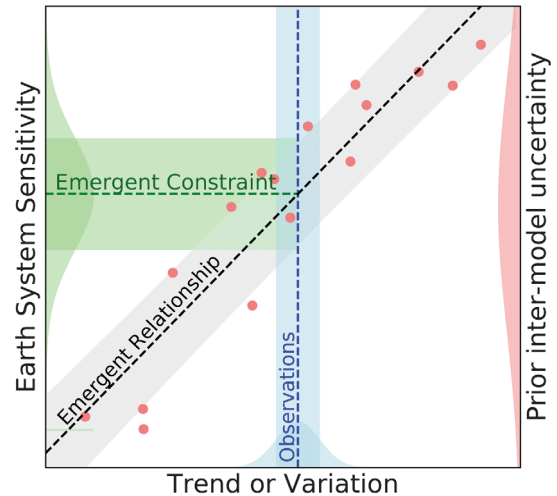

FIG. 1.   Schematic showing the most common procedure used to derive emergent constraints on Earth system sensitivities. An ensemble of ESMs (each red dot is an individual ESM) running the same experiment (the PDF on the right-hand y axis represents the spread in the ensemble) is used to identify an emergent relationship (black dashed line with gray uncertainty range) between an uncertain Earth system sensitivity Y (y axis) and an observed trend or variation X (x axis). An observation of the trend or variation (blue PDF on the x axis) can then be combined with the model-based emergent relationship to derive an emergent constraint on the Earth system sensitivity (green PDF on the left-hand y axis).

identify an element X of the observable climate that varies significantly across the ESM ensemble and that exhibits a statistically significant relationship with variations in some important variable Y describing the ESM's future simulated state. If we call this relationship f, then $Y = f(X) + \varepsilon$, where ε is a relatively small departure from f. Since X is an element of the observable climate, it is a quantity that can be measured. The relationship f may then place a useful constraint on Y, provided that the measurement uncertainty in X is small compared to the range of simulated values. This constraint is "emergent" because the *emergent relationship f* cannot be diagnosed from a single ESM. It becomes apparent only when the full ensemble is analyzed. If the relationship f arises from model physics or dynamics common to the ESMs, then reducing the spread in X through reformulation or adjustment of parametrizations ought to result in spread reduction in Y. A corresponding reduction in the spread in other future climate variables affected by Y should also occur. If the process is repeated for enough variables X and Y, then we can imagine that overall simulated spread in multiple aspects of the future climate would gradually be reduced. We also note that ECs can be used directly to ascertain the most likely values for a particular Y.

To further illustrate the concept, the subject of the first published EC well established across multiple generations of ESMs is used as an example (Hall and Qu, 2006). In this case, the future climate variable of interest Y is the snow albedo feedback (SAF), a climate mechanism characterized by the retreat of highly reflective snow cover under climate warming and the associated reduction in surface albedo, which amplifies warming and promotes further snow melt (Hall, 2004;
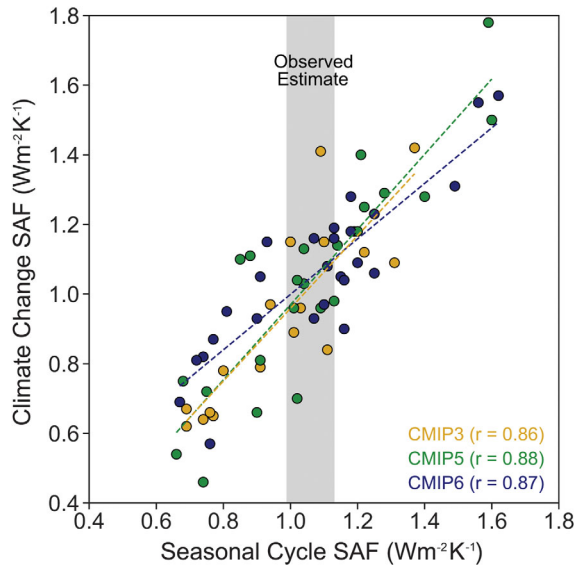
FIG. 2. Emergent relationship between springtime snow albedo feedback (SAF) across Northern Hemisphere land under climate change (Y) and an observable snow albedo feedback associated with the current climate's seasonal cycle (X). An observational estimate derived from satellite data is shown as a vertical bar. Each point represents an individual climate model from the three most recent generations (CMIP3, CMIP5, and CMIP6). Methodology for calculating SAF is adapted from Qu and Hall (2014) [further details were provided by Thackeray *et al.* (2021)].

Bony *et al.*, 2006; Thackeray and Fletcher, 2016). However, modern ESMs disagree on the strength of this feedback, exhibiting a nearly threefold spread (Qu and Hall, 2014; Thackeray, Qu, and Hall, 2018). SAF also occurs each spring in the current climate when Northern Hemisphere snow cover recedes from its winter peak to its summer minimum, enhancing seasonal warming in the process (X). A comparison between simulated SAF under climate change (Y) and in the seasonal cycle context (X) uncovers a strong linear relationship in three different generations of ESM ensembles (Fig. 2). The strength of this relationship tells us that seasonal SAF is likely highly predictive of climate change SAF, while the consistency across model generations illustrates robustness to out-of-sample testing. Furthermore, we can quantify the seasonal SAF using satellite-derived observations, thus allowing for model bias to be properly assessed. For example, an ESM that underestimates observed SAF in the seasonal context is likely to underestimate SAF in the future climate. Producing a more accurate SAF in ESMs is more than just an academic exercise, as variability in SAF can account for a significant portion of uncertainty in 21st century projections of warming across Northern Hemisphere extratropical land (Qu and Hall, 2014). Thus, by reducing variability in SAF we can expect similar reductions in projections of regional warming. Figure 3 highlights the first application of this and several other ECs from the literature (pertaining to the carbon cycle and climate sensitivity), all of which are further discussed in Sec. V.

In this review we synthesize what is now understood about ECs. We discuss how relationships f might emerge in ESM ensembles in Sec. II, as well as the theory that underlies ECs in Sec. III. In Sec. IV we note pitfalls associated with ECs and how to guard against them. In Sec. V we survey the ECs that have been found in the climate system and describe the statistics that can be used to apply them rigorously in Sec. VI. Finally, we discuss how techniques similar to ECs have been or might be used in other fields, and we consider the outlook for development of the technique going forward in Sec. VII before concluding in Sec. VIII.

II. HOW RELATIONSHIPS IN MODEL ENSEMBLES MIGHT "EMERGE"

ECs are possible because of emergent relationships appearing in an ensemble of ESMs. An emergent relationship $Y = f(X) + \varepsilon$ between an element of the observable climate X, something that can be measured now in the real world, and the unknown future response Y one wants to know can then be used to place a constraint on the real world value of Y via observations of X to give an EC. For example, Cox, Huntingford, and Williamson (2018) took as the observable the variability in global annual mean surface air temperature during roughly the last 100 years, and the unknown future response was ECS (defined as the magnitude of Earth's warming to doubled CO_2 levels), one of the oldest and most important numbers in climate science. Despite decades of research the latest ESMs still vary widely in their predictions of this number by a factor of 3 (Forster *et al.*, 2020; Zelinka *et al.*, 2020).

How might a relationship f between observable and response emerge in an ensemble of ESMs? In the EC on ECS given by Cox, Huntingford, and Williamson (2018), an analytically soluble energy balance model, much reduced in complexity compared to the ESMs in the ensemble, predicted a functional form between X and Y for the emergent relationship [although this provoked debate; see Brown, Stolpe, and Caldeira (2018), Cox *et al.* (2018), Po-Chedley *et al.* (2018), Rypdal *et al.* (2018), and Williamson, Cox, and Nijsse (2019)]. This mechanism from high- to low-dimensional relationship emergence and others are given later. First we discuss the differences in the types of climate model ensembles commonly used to find emergent relationships.

A. Commonly used ESM ensembles: Multimodel and perturbed physics ensembles

Model simulations are not perfect reproductions of the system that they are designed to emulate. Ensembles of models are used to get a handle on (i) initial condition uncertainty, (ii) parametric uncertainty, and (iii) structural uncertainty.

Uncertainty in the initial state is particularly important for numerical weather prediction. As the equations governing the weather are chaotic, forecast solutions with small differences in their initial states, equally likely to be the "real" initial state, can diverge strongly after just a few days. This sensitivity to initial conditions means running the weather model just once, even if the model is a perfect reproduction of the real world, may produce a significantly different forecast from the one actually experienced. By running the same weather model
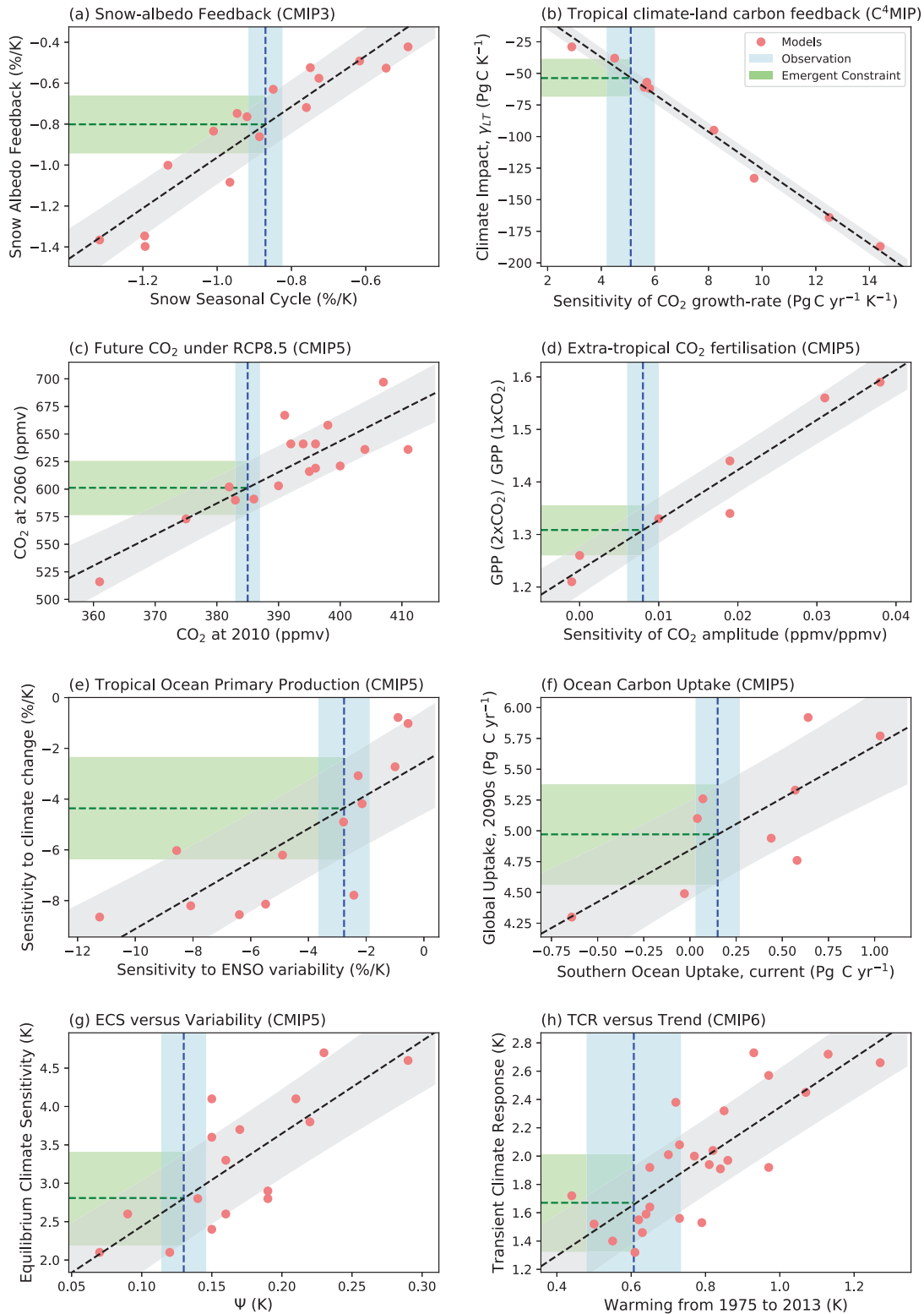
FIG. 3.   Emergent constraints on Earth system sensitivities based on some key examples published in the literature. (a) Snow albedo feedback derived from the snow seasonal cycle (Hall and Qu, 2006). (b) Sensitivity of tropical land carbon to global warming calculated from interannual variability in CO_2 (Cox *et al.*, 2013). (c) Atmospheric CO_2 concentration at 2060 projected from atmospheric CO_2 concentration at 2010 (Hoffman *et al.*, 2014). (d) CO_2 fertilization of plant photosynthesis calculated from changes in the seasonal cycle of CO_2 (Wenzel *et al.*, 2016). (e) Sensitivity of tropical ocean primary production to warming derived from interannual variability (Kwiatkowski *et al.*, 2017). (f) Global ocean carbon sink in the 2090s projected from the current day carbon sink in the Southern Ocean (Kessler and Tjiputra, 2016). (g) Equilibrium climate sensitivity calculated from interannual variability of temperature (Cox, Huntingford, and Williamson, 2018). (h) Transient climate response determined from the increase in global mean temperature (Nijsse, Cox, and Williamson, 2020). In each case the emergent constraint was reconstructed from data available in the literature or provided directly by the authors. The model ensemble used in each original study is shown in the brackets after the panel title.

many times with different, but equally plausible, initial states chosen to sample the most unstable and divergent regions of the model phase space, the likelihood of experiencing a particular weather forecast can be constructed. Ensemble forecasting is now a standard tool in numerical weather prediction (Epstein, 1969; Leith, 1974; Toth and Kalnay, 1993; Molteni *et al.*, 1996). Initial value ensembles in climate applications are usually used to assess natural climate variability (Kay *et al.*, 2015; Maher *et al.*, 2019).

Generally speaking, climatic variables are the long-term statistics of weather and their prediction is less impacted by initial state uncertainty relative to parametric and structural uncertainties (Hawkins and Sutton, 2009, 2011). Parametric uncertainty arises from uncertainty in the values of constants ("parameters") in quasiempirical relations (parametrizations) used to model the effects of unresolved, subgrid scale processes. Parametric uncertainty can be sampled by running the same model with different values of these parameters. ESM model ensembles that do this are known as perturbed physics or perturbed parameter ensembles (PPEs) (Murphy *et al.*, 2004). Examples of ESM PPEs are the climateprediction.net experiment (Stainforth *et al.*, 2005) and the Met Office's Quantifying Uncertainties in Model Projections (QUMP) ensemble (Collins *et al.*, 2011), which both used the HadCM3 ESM.

Structural uncertainty arises from uncertainty in the functional form of equations. Although many of the model equations in ESMs are well known, there are multiple, equally defensible parametrization schemes for unresolved processes (such as convective precipitation and cloud radiative properties), so different equations can appear in different ESMs. ESM ensembles that sample structural uncertainty are called multimodel ensembles (MMEs). They include the coupled model intercomparison project (CMIP) ensembles (Meehl, Covey *et al.*, 2007; Taylor, Stouffer, and Meehl, 2012; Eyring *et al.*, 2016) used to inform the Intergovernmental Panel on Climate Change (IPCC) reports (Cubasch *et al.*, 2001; Meehl *et al.*, 2007; Collins *et al.*, 2013) and are used extensively in much of the EC literature.

The CMIP MME ensembles consist of a number of different ESMs developed by different international groups and thus sample structural uncertainty. Some CMIP groups also submit multiple initial value runs and occasionally multiple physics parameter runs performed with the same model, sampling initial condition and parametric uncertainty to some degree. Numerical weather prediction ensemble forecasting has also incorporated more parametric and structural uncertainty into their forecasting as time has gone on (Palmer, 2019).

Most ECs have been derived from MME ensembles to date although there are some examples from PPEs (Knutti *et al.*, 2006). ECs derived from MMEs are the most believable. Each model in a MME could be thought of as that particular group's best guess of the Earth system. PPEs, on the other hand, generally vary only a handful of parameters and thus sample just a few dimensions of the model's phase space. Because of the few varied dimensions, relationships hard coded in the fixed equations between model variables become easy to find (based on our experience with the QUMP PPE) (Lambert *et al.*, 2013). These ECs need to be backed up with physical

arguments and ideally tested in MMEs. Collins *et al.* (2011) found that long wave cloud feedback was highly correlated to climate sensitivity in the QUMP HadCM3 PPE ensemble; however, this correlation was absent in smaller sample size MMEs. Yokohata *et al.* (2010) also found that strong correlations in a PPE with HadSM3 were not always present in the equivalent MIROC3.2 PPE. Further, Yokohata *et al.* (2010) found that the mechanism for variations in climate sensitivity was different.

Values of parameters chosen to correctly give the proper balance between opposing processes in MMEs could be skewed to one of the processes by simultaneously varying them in a PPE, thus reducing the realism of each ensemble member solution. However, PPEs can be useful for testing ECs [see Kamae *et al.* (2016), Wagman and Jackson (2018), and Sec. IV.I] and quantifying parametric uncertainty (Sec. VII.F).

B. Null hypothesis: Emergent relationships occur by chance

A starting null hypothesis is that emergent relationships occur by chance and are not indicative of a deeper mechanistic relationship (Hall *et al.*, 2019). Data mining an ESM ensemble for high correlations between pairs of variables may fall into this category; see also Sec. IV.A. If an ESM ensemble is reasonably small and the number of variables output by each ESM is high, the expected number of variable pairs with high correlation is high purely by chance (Caldwell *et al.*, 2014).

More precisely, an ESM labeled by the index i in an ensemble of n ESMs $i = \{1, 2, \ldots, n\}$ calculates a large number m of different variables X^l, with each variable labeled with index l and $l = \{1, 2, \ldots, m\}$. In this data-mining scenario, a pair of variables (X_i^l, X_i^k) are chosen for the prospective emergent constraint, one for the observable X^l and the other for the response X^k one wants to know in the real world (previously written Y). In an indiscriminate data-mining approach, every possible pair of variables are used to create $(1/2)m(m-1)$ datasets, each labeled $\mathcal{S}^{kl}$. Each dataset has n elements, $\mathcal{S}^{kl} = \{(X_1^l, X_1^k), (X_2^l, X_2^k), \ldots, (X_n^l, X_n^k)\}$ with each element corresponding to one of the n ESMs. If the pair of variables making up $\mathcal{S}^{kl}$ are a good candidate for an emergent relationship and we assume that this relationship is linear, then the correlation r^{kl} in $\mathcal{S}^{kl}$ should be high. Correlation here is defined as $r^{kl} = \text{cov}(X^l, X^k)/\sigma_{X^l}\sigma_{X^k}$, and σ_X is the standard deviation of X.

The number of ESMs in a model ensemble n is typically small, around 10–40 in the state-of-the-art CMIP ensembles, although numbers increase with each successive generation (Meehl, Covey *et al.*, 2007; Taylor, Stouffer, and Meehl, 2012; Eyring *et al.*, 2016). If we calculate the correlation r^{kl} within every one of the $(1/2)m(m-1)$ possible datasets, some will have high correlations purely by chance. The likelihood increases as the number of models in the ensemble n gets smaller, giving fewer data points in each $\mathcal{S}^{kl}$. While the fraction of datasets with high correlations $|r| > |r_{\text{high}}|$ does not increase as the number of variables m gets larger, the total number of datasets with correlation above $|r_{\text{high}}|$ will increase, simply because there are more possible variable
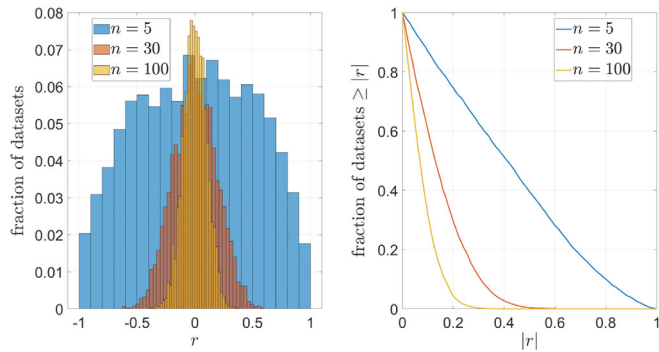
FIG. 4. High correlations between pairs of variables across small ESM ensembles are expected by chance. To illustrate this point, the fraction of all possible pairs of variables across a toy ESM ensemble of size n plotted against correlation r. Each ESM is not a real ESM: it is represented by a random variable drawn from a normal distribution of unit standard deviation, and each of the n ESMs produces $m = 100$ output variables, giving 4950 different possible variable pairs to correlate with each other. This example is designed to be a worst case scenario yet still shows that a calculated correlation even between uncorrelated objects can be large in small enough ensembles. Left panel: histograms of fraction of datasets with correlation falling in a particular interval. Right panel: total fraction of datasets with magnitude of correlation greater than $|r|$. Ensemble size is varied for toy ESM ensemble sizes of $n = 5$, 30, and 100.

pairs to correlate. Thus, one must be careful that any prospective emergent constraint makes physical sense, particularly in current, small ESM ensembles to avoid this pitfall.

To illustrate this point in a worst-case-type scenario where every ESM output variable is uncorrelated to any other output variable or any of the other ESMs, we can calculate the correlations expected between all possible pairs of variables as a function of ensemble size n. Figure 4 shows the fraction of all possible datasets with a particular correlation for a "toy" ESM ensemble of n models where each of the toy ESMs is a modeled as a random number generator. This is an extreme and unrealistic example, with each ESM producing m random output variables, the lth one labeled X_i^l, drawn from a normal distribution with unit standard deviation, $X_i^l \in \mathcal{N}(0, 1)$. Even though there are no mechanistic relationships in this example by construction, a fraction of the datasets have strong correlations, particularly in smaller ensemble sizes (small n). This is purely by chance.

C. Low-dimensional relationships emerge from high-dimensional ESMs

In the last example, even though all $n \times m$ toy ESM variables in the ensemble were uncorrelated from each other by construction, high correlations were still possible. This is more likely in small ensembles. In reality there are mechanistic relationships between X and Y in any particular ESM. Real ESMs are not random number generators: the output of an ESM is the numerical solution of a large set of coupled, nonlinear equations codifying relationships between variables representing the ocean and atmosphere dynamics

and thermodynamics and biogeochemical feedbacks.[2] However, the exact relationship between any pair of X and Y is not solely a function of only these two variables, it is generally also a function of many others that prescribe the exact state of the ESM at every point in space and time on the ESM grid. The minimal number of nonunique variables d required to describe the exact state of the ESM at time t can be thought of as a point in d-dimensional phase space with coordinates $(x_1, x_2, \ldots, x_d)$ and in the case of an ESM, the dimension of this space is extremely high.[3] As the ESM state changes with time, a trajectory is drawn between these phase space coordinates connecting the past and present states. Variables describing the precise state of the ESM such as temperature or precipitation, the X, and the Y may be functions of all the d phase space coordinates as well as the time $(d + 1)$, i.e., $X = f_X(t, x_1, x_2, \ldots, x_d)$ and $Y = f_Y(t, x_1, x_2, \ldots, x_d)$.

Although the mechanistic relationship between X and Y is generally a function of all $d + 1$ coordinates, climatic variables of interest are usually, although not exclusively, long-term temporal and/or spatial averages that can be thought of as time-invariant attractors of the phase space reducing the information relative to a precise configuration at any one time. In many circumstances the mechanistic relation between X and Y can be approximated to a good degree by a much smaller dimensional subset of the full phase space reducing the effective dimension.

The success of statistical mechanics and science in general[4] relies on this effective dimension reduction, that certain properties of systems with many degrees of freedom can be well approximated by fewer state parameters. For example, a container of gas is, as far as we know, most completely described by the equations of quantum mechanics. In principle, one could solve these equations for each of the N gas molecules, at each position and time in the container. The dimension of this space increases exponentially with N, requiring a large number of degrees of freedom to specify the state exactly even for just a handful of molecules. This could be useful if the question one wanted to answer required detailed knowledge of the full quantum state. However, if we only want to know the configuration of positions and momenta of the gas molecules, one can use an approximation of the quantum theory. Treating the gas as a bunch of identical

---

[2]Many of these equations are well known from other fields of physics such as the fluid dynamics Navier-Stokes equations. See Ghil and Lucarini (2020) and references within for the equations that are commonly featured in ESMs.

[3]To give an idea of a lower bound on the dimension of the phase space required to describe a typical CMIP6 era ESM, the atmosphere is typically cut into a total number of $\sim 200 \times 200 \times 100 = 4 \times 10^6$ spatial cubes. A single atmospheric prognostic variable will need at least this many phase space dimensions at each point in time to specify it exactly. A lower estimate for the number of prognostic variables required just to specify the physical state of the atmosphere is $\mathcal{O}(10)$. This already gives $d \sim 10^7$ even before including the ocean, land, and biogeochemistry.

[4]Low-dimensional models can often mimic the responses of more complex ones to a good approximation, and this is essentially why science is successful; i.e., we can comprehend and predict the high-dimension real world using relatively simple models.

spheres, each with a mass and a different position and momentum, the state can be described by a point in $6N$-dimensional phase space, a reduction in the phase space dimension yet still a massive number of dimensions for a realistic gas.

Generally, detailed knowledge of a precise system state is not desired or required. Bulk properties describing the mean overall state are a useful way of reducing information and making sense of complex systems. Going back to the container of gas, kinetic theory provides a way of mapping the many degrees of freedom from each of the molecules to the few bulk parameters of the ideal gas law under reasonable assumptions; see Schekochihin (2015). That is, the $6N$ dimensions effectively reduces to the three encoded by the relationship $pV = nRT$. The bulk variable pressure p felt as a force per unit area on the container wall and is related to how often and how vigorously on average molecules with mean square velocity $\langle v^2 \rangle$ hit the sides of the container ($p = (n/3V)\langle v^2 \rangle$, where n is the number of moles in the container of volume V), whereas the temperature T is a measure of the most probable kinetic energy of a molecule in the container. R is the ideal gas constant. This approach works well when N is large, as one finds that more of the $6N$ states correspond to the ideal gas law.

Effective dimension reduction in ESMs can also happen when we want answers to questions about bulk mean states of the climate. For instance, relations between bulk variables such as the annual global mean surface air temperature and the net incoming radiation are well modeled by the few degrees of freedom in a simple energy balance model [see Hasselmann (1976), Wigley and Raper (1990), and Gregory (2000)], even though the exact state of an ESM is given by all prognostic variables at every point and multiple times on its spatial grid. If one wanted to ask detailed questions about the state of an ESM such as, how many days in March does it rain more than x mm in Madagascar?, then the full ESM theory is best placed. However, if the question were, how much does the global mean temperature increase from a spatially uniform doubling of atmospheric CO_2?, a lot of that extra detail seems unnecessary. The analogy we attempt to present here is that ESMs are analogous to the solutions to the laws of quantum mechanics and the effective, reduced dimension relationship for a particular ESM i, $Y = f_i(X) + \epsilon$ is analogous to the ideal gas law. The analogy is far from perfect: ESMs are not as simple as containers of large numbers of identical molecules and there is as yet no such elegant route between the two solutions provided by the statistical mechanical or kinetic theory recipes. The best candidates for ESM ideal gas laws to serve as theoretical bases for emergent relationships are simplified, analytically soluble models of the climate or its subsystems such as the previously mentioned energy balance models (Hasselmann, 1976; Wigley and Raper, 1990; Gregory, 2000). These act more as testable, plausible hypotheses of the bulk behavior of ESMs than derivations from the basic laws that constitute them.

Thus far we have discussed how a physically plausible, mechanistic, low-dimensional relationship f_i between X and Y may appear from a particular ESM i, $Y = f_i(X) + \varepsilon$. We have not yet discussed how low-dimensional relationships, the emergent relationship, could appear across an ensemble of ESMs. Although each ESM is different, they should be equally plausible models of the (X, Y) relationship in the real world and it should not be a surprise that ESM solutions across a model ensemble should also be mechanistically related. This is because many of the equations are based on well-established and well tested physics (such as the Navier-Stokes equations) and are common among ESMs. Indeed, if all the equations were well known and could be numerically integrated exactly, the ESMs should produce identical outputs (and the problem of climate modeling would be solved and emergent relationships would disappear).

However, some equations cannot be exactly numerically integrated due to limitations on temporal and spatial resolution. One then has to parametrize subgrid scale processes and this may be a source of difference between ESM solutions, with each model giving a different (X, Y) pair; see Sec. II.A. Although they are our best attempts at understanding the climate, some ESM responses diverge on important questions. Such a scenario is where the emergent constraint approach may be helpful. To have a chance of finding an emergent relationship, the set of n (X_i, Y_i) data points should be different [i.e., $(Y_1, X_1) \neq (Y_2, X_2)$] and span a wide enough range for the emergent relationship to become apparent. The model relationship f_i also needs to be shared among the model ensemble [$f_i(X) \approx f_j(X) \ \forall \ i, j = \{1, 2, ..., n\}$]. Provided that such a relationship exists, the emergent relationship $f(X)$ can be determined. If this model relation is also shared by the real world, the real world's response can also be determined (along with its uncertainty) to give an emergent constraint; see Sec. II.E.

D. Range in response due to the same physical process having a wide range across ESMs

A related mechanism for a low-dimensional relationship to emerge between X and Y occurs when the magnitude of the same physical process, correlated to the observable X, differs appreciatively across the ESM ensemble, and the magnitude of this process largely determines the size of the response Y. This is also a case of effective dimension reduction: the range in response is dominated by the dimensions of phase space of just that physical process.

An example of this happening was provided by Caldwell, Zelinka, and Klein (2018), who evaluated ECs on ECS in the CMIP3 and CMIP5 ESM ensembles. That is, the targeted unknown response Y among these ECs was the same, ECS, although different observables X were found to be well correlated to it. Examples of a few of the X used in the different ECs were the strength of resolved-scale mixing between the boundary layer and lower troposphere in the tropical east Pacific and Atlantic (Sherwood, Bony, and Dufresne, 2014), error in the distribution of cloud-top pressure and optical thickness for regions between 60° N and 60° S (Klein *et al.*, 2013), the fraction of tropical clouds with tops below 850 mb whose tops are also below 950 mb (Brient *et al.*, 2016), and variability in global mean air temperature (Cox, Huntingford, and Williamson, 2018). Caldwell, Zelinka, and Klein (2018) showed that all of these observables were also highly correlated to the

shortwave (SW) cloud feedback, a strong and uncertain feedback on the resulting global temperature. The SW cloud feedback in turn was shown to be highly correlated to intermodel variations in ECS within the CMIP5 ensemble. To paraphrase Caldwell, Zelinka, and Klein (2018), "… intermodel variations in cloud feedback were so big that they left a strong imprint on intermodel variations in ECS. This means that fields that are strongly correlated with ECS are probably correlated with the SW cloud feedback (and vice versa)."

E. What is needed for an EC?

In addition to having an emergent relationship in a model ensemble, an EC also ideally requires a few other things to be useful.

1. Observable (X) range and uncertainty

The range of X in the ESM ensemble should be large relative to the uncertainty in the value of the real world observable. Uncertainty in each of the model X should also be small relative to the range in model X. Ideally the real world value of the observable X_{obs} should also fall well within the range of values in the ESM ensemble to avoid extrapolation issues.

2. Response (Y) range and uncertainty

One wants a large range in the values of the ESM ensemble responses relative to the uncertainty in each of the individual ESM responses. This reduces the uncertainty in the real world value of Y derived from the EC.

3. Relationship between X and Y

A particular ESM labeled i should have a physically plausible relationship f_i between the observable and response $Y = f_i(X) + \varepsilon$. It should be based on a simplified theory that predicts a functional response and has clear, testable assumptions that can be falsified independently. In addition, all models in the ensemble should share roughly the same relation between observable and response, i.e., $f_i = f_j$. The stronger this relationship is, the more useful the emergent constraint due to the potential uncertainty reduction in Y. This relationship should be shared by the real world.

4. Large ensemble size n

Ideally you would have a large number of independent ESMs in your ensemble (larger sample, lower error in estimates, less chance for high correlations by chance). See Sec. IV.E for further discussion of ESM ensemble independence.

III. UNDERLYING THEORY FOR EMERGENT CONSTRAINTS BASED ON TEMPORAL VARIABILITY

Many published EC studies relate the longer-term sensitivity of interest to aspects of the mean climate simulation in models; see Table I (Volodin, 2008; Kidston and Gerber, 2010; Massonnet *et al.*, 2012; Tian, 2015; Brient and Schneider, 2016; Simpson and Polvani, 2016; Lin *et al.*, 2017; Lipat *et al.*, 2017; Siler, Po-Chedley, and Bretherton,

TABLE I. Collection of existing ECs. Note that some of these ECs involve correlations that are lower than those portrayed in Fig. 1, with correspondingly less potential for uncertainty reduction. The ensemble or ensembles for which the EC appears to have value are also listed (note that many have been tested only on the ensemble that they were developed on).

	Future constrained quantity (Y)	Current climate quantity (X)	Ensemble	Reference
1	Equilibrium climate sensitivity	Seasonal cycle of temperature	CMIP1 PPE	Covey *et al.* (2000) and Knutti *et al.* (2006)
2	Equilibrium climate sensitivity	Difference in cloud fraction between the tropics and SH midlatitudes	CMIP3 CMIP5 CMIP6	Volodin (2008)
3	Equilibrium climate sensitivity	TOA radiation balance in the SH	CMIP3	Trenberth and Fasullo (2010)
4	Equilibrium climate sensitivity	Features of TOA radiation fluxes	CMIP3	Huber *et al.* (2011) and Tett *et al.* (2013)
5	Equilibrium climate sensitivity	Variability in climatological May–August relative humidity and cloud extent	CMIP3	Fasullo and Trenberth (2012)
6	Equilibrium climate sensitivity	Last Glacial Maximum cooling	CMIP3/PMIP2 CMIP5/ PMIP3	Hargreaves *et al.* (2012) and Schmidt *et al.* (2014)
7	Equilibrium climate sensitivity	Vertical mixing strength between the boundary layer and lower troposphere over tropical oceans	CMIP3 CMIP5 CMIP6	Sherwood, Bony, and Dufresne (2014)
8	Equilibrium climate sensitivity	Vertically resolved relative humidity and clouds between 45° S and 40° N	CMIP5	Su *et al.* (2014)
9	Equilibrium climate sensitivity	Precipitation in the double-ITCZ region	CMIP3 CMIP5	Tian (2015)
10	Equilibrium climate sensitivity	Seasonal sensitivity of low cloud to SSTs (20–40° latitude)	CMIP3 CMIP5 CMIP6	Zhai, Jiang, and Su (2015)
11	Equilibrium climate sensitivity	Height of tropical low clouds	CMIP5	Brient *et al.* (2016)
12	Equilibrium climate sensitivity	Sensitivity of subtropical low cloud albedo to SSTs	CMIP5 CMIP6	Brient and Schneider (2016)

(Table continued)

TABLE I. *(Continued)*

	Future constrained quantity (*Y*)	Current climate quantity (*X*)	Ensemble	Reference
13	Equilibrium climate sensitivity	Climatological latitude of the SH Hadley cell edge in winter	CMIP5	Lipat *et al.* (2017)
14	Equilibrium climate sensitivity	Statistics of interannual temperature variability	CMIP5	Cox, Huntingford, and Williamson (2018)
15	Equilibrium climate sensitivity	Cloudy-sky radiative flux sensitivity to temperature	CMIP5	Lutsko and Takahashi (2018)
16	Future global warming (midcentury or end of century)	Recent warming trend in CMIP models (1981–2017)	CMIP6	Tokarska *et al.* (2020)
17	Transient climate response and equilibrium climate sensitivity	Recent warming in CMIP models	CMIP5 CMIP6	Jiménez-de-la-Cuesta and Mauritsen (2019) and Nijsse, Cox, and Williamson (2020)
18	Low-cloud feedback sign	Low-cloud sensitivity to Pacific variability	CMIP3	Clement, Burgman, and Norris (2009)
19	Global mean cloud feedback and ECS	Climatological latitudinal gradient in the reflectivity of clouds	CMIP5	Siler, Po-Chedley, and Bretherton (2018)
20	Low-cloud optical depth change per degree climate warming	Low-cloud optical depth response to temperature anomalies	CMIP3 CMIP5	Gordon and Klein (2014)
21	Subtropical low-cloud cover change under climate warming	Subtropical low-cloud cover response to interannual temperature and stability anomalies	CMIP3 CMIP5	Qu *et al.* (2014, 2015)
22	Global mean cloud feedback under climate change	Global mean cloud feedback derived from interannual variability in current climate	CMIP5	Zhou *et al.* (2015)
23	Change in upper troposphere relative humidity and cloud fraction	Vertical gradient in climatological mean relative humidity and cloud fraction	CMIP5	Po-Chedley *et al.* (2019)
24	Snow albedo feedback	Seasonal snow albedo feedback (spring)	CMIP3 CMIP5 CMIP6	Hall and Qu (2006) and Qu and Hall (2014)
25	Sea-ice albedo feedback	Seasonal sea-ice albedo feedback (summer)	CMIP5	Thackeray and Hall (2019)
26	Timing of Arctic ice-free summer	Historical September sea-ice trend	CMIP3	Boe, Hall, and Qu (2009)
27	Timing of Arctic ice-free summer	Historical sea-ice characteristics	CMIP5	Massonnet *et al.* (2012)
28	Arctic thermal feedbacks	Different aspects of northern latitude regional temperatures	CMIP3 CMIP5	Bracegirdle and Stephenson (2013)
29	Sensitivity of tropical land carbon storage to warming	Short-term sensitivity of CO_2 to interannual temperature variability	C4MIP CMIP5	Cox *et al.* (2013) and Wenzel *et al.* (2014)
30	CO_2 fertilization of photosynthesis on the extratropics	Seasonal fluctuations in CO_2 concentrations	CMIP5	Wenzel *et al.* (2016)
31	Global land carbon feedback strength	Interannual variations in evapotranspiration, net biome productivity, and gross primary productivity	CMIP5	Mystakidis *et al.* (2017)
32	Change in tropical primary production to temperature anomalies	Sensitivity of tropical primary production to interannual SST anomalies	CMIP5	Kwiatkowski *et al.* (2017)
33	Future permafrost thaw	Relationship between mean annual air temperature and permafrost area	CMIP5	Chadburn *et al.* (2017)
34	Future CO_2 concentration	Simulated CO_2 concentration by 2010	CMIP5	Hoffman *et al.* (2014)
35	Change in Arctic gross primary productivity	Sensitivity of annual maximum leaf area index to increasing CO_2	CMIP5	Winkler *et al.* (2019)
36	Global ocean carbon uptake	Present-day Southern Ocean carbon uptake	CMIP5	Kessler and Tjiputra (2016)
37	Future Arctic Ocean acidification	Present-day Arctic sea-surface density	CMIP5	Terhaar, Kwiatkowski, and Bopp (2020)
38	Intensification of heavy rainfall across certain extratropical regions	Scaling of annual maximum daily precipitation with global land temperatures	CMIP5	Borodina, Fischer, and Knutti (2017b)
39	Change in tropical precipitation extremes under climate warming	Sensitivity of tropical precipitation extremes to temperature variability	CMIP3 CMIP5	O'Gorman (2012)
40	Indian summer monsoon rainfall increase with warming	Climatological mean precipitation in the western tropical Pacific	CMIP5	Li *et al.* (2017)
41	Future SST change over the Indian Ocean with application for East African precipitation	Present-day interannual SST–low level cloud sensitivity	CMIP5	Rowell (2019)

(Table continued)

TABLE I. *(Continued)*

	Future constrained quantity (*Y*)	Current climate quantity (*X*)	Ensemble	Reference
42	Future runoff projections over the western US	Historical runoff sensitivity to temperature and precipitation variations	CMIP5	Lehner *et al.* (2019)
43	Clear-sky shortwave absorption (global mean precipitation increase)	Sensitivity of shortwave radiative absorption to changes in column water vapor	CMIP5 CMIP6	DeAngelis *et al.* (2015)
44	Global hydrologic sensitivity	Variation between surface longwave cloud radiative effect and its sensitivity	CMIP5	Watanabe *et al.* (2018)
45	Global transpiration to evapotranspiration ratio	Local transpiration to evapotranspiration ratio	CMIP5	Lian *et al.* (2018)
46	North Atlantic subpolar gyre cooling	Stratification of the subpolar North Atlantic Ocean	CMIP5	Sgubin *et al.* (2017)
47	Summer warming over the central USA	Climatological summer temperature over the central USA	CMIP5	Lin *et al.* (2017)
48	Change in high-latitude temperature variability	Historical sea-ice metrics	CMIP5	Borodina, Fischer, and Knutti (2017a)
49	Future continental warming over Europe	Present-day climatological summer temperatures over Europe	CMIP5	Selten *et al.* (2020)
50	Frequency of heat extremes	Seasonal land-atmosphere feedbacks	CMIP5	Donat, Pitman, and Angélil (2018)
51	Poleward jet shifts under warming	Climatological position of jet stream	CMIP5	Simpson and Polvani (2016)
52	Poleward shift of Southern Hemisphere eddy-driven jet stream with climate warming	Climatological latitudinal position of Southern Hemisphere eddy-driven jet stream	CMIP3	Kidston and Gerber (2010)
53	Anthopogenic ozone radiative forcing	Tropospheric ozone effect on outgoing longwave radiation	ACCMIP	Bowman *et al.* (2013)

2018; Selten *et al.*, 2020). However, a growing number of papers are also now relating Earth system sensitivities to observable temporal variations, such as trends (Boe, Hall, and Qu, 2009; Jiménez-de-la-Cuesta and Mauritsen, 2019; Nijsse, Cox, and Williamson, 2020; Tokarska *et al.*, 2020), inter-annual variability (Clement, Burgman, and Norris, 2009; O'Gorman, 2012; Cox *et al.*, 2013; Wenzel *et al.*, 2014; Qu *et al.*, 2015; Kwiatkowski *et al.*, 2017; Cox, Huntingford, and Williamson, 2018), and seasonal cycles (Hall and Qu, 2006; Knutti *et al.*, 2006; Qu and Hall, 2014; Zhai, Jiang, and Su, 2015; Wenzel *et al.*, 2016; Thackeray and Hall, 2019). We discuss here the theoretical basis for emergent constraints based on such temporal variations.

In the case of changes in relatively fast variables, such as seasonal snow cover (Qu and Hall, 2007) or marine phytoplankton concentration (Kwiatkowski *et al.*, 2017), there may be a fairly straightforward near one-to-one relationship between the short-term variability and the longer-term sensitivity because the fast variable will be in a quasiequilibrium state even with short-term climate variations. For slower variables (such as the forest carbon storage), short-term variations are more likely to measure fluxes (or equivalently the rate of change of the store). In this case, finding a constraint on future changes in the store requires multiplying the flux sensitivity to short-term variations by a characteristic timescale for each model. In some cases, the characteristic timescale may be similar across the model spectrum, leading to a simple emergent relationship between the short-term flux sensitivity and the long-term sensitivity of the store

(Cox *et al.*, 2013; Wenzel *et al.*, 2014). In general, though, converting a flux sensitivity to a store sensitivity requires an independent estimate of the characteristic timescale of the store, which itself requires a theoretical basis (Williamson, Cox, and Nijsse, 2019).

The fluctuation-dissipation theorem (FDT) provides one such theoretical framework (Kubo, 1966). The FDT relates the sensitivity of a system to a small external forcing to the response of the same system to its own internally generated fluctuations. FDT is therefore a potential theoretical basis for ECs, as it links the natural variability of a system to its sensitivity. The size of the forcing is important because the theorem strictly applies only to near-equilibrium linear systems. Nevertheless, FDT-based approaches have had a huge impact in statistical physics, including Einstein's work on Brownian motion (Einstein, 1905), and the understanding of Johnson-Nyquist noise in electrical circuits (Johnson, 1928; Nyquist, 1928).

The first proposal to apply FDT to the climate system came almost 40 years ago (Leith, 1975), but there has been a recent resurgence of interest in this area fueled by methodological advances and a detailed comparison of FDT-derived estimates of climate change to climate model simulations (Majda, Abramov, and Gershgorin, 2010). In principle, it may be possible to use FDT to get good estimates of the response of the real climate system to small forcing (such as that due to doubling CO_2) purely from accurate long-term climate observations that reveal the full spectrum of natural fluctuations of the climate (Bell, 1980; Schwartz, 2007). Unfortunately,

though, the length of the detailed climate records required to achieve this are typically much longer than those available (Kirk-Davidoff, 2009) and include the contemporary period when the mean climate is changing.

In the EC technique model projections are instead used to define emergent relationships between observable variations and future climate (Hall *et al.*, 2019), and then specific observations provide a selection principle to constrain the range of future climate projections from these model-defined relationships (Eyring *et al.*, 2019). The EC technique has yielded many proposed constraints on aspects of the future climate (Hall *et al.*, 2019) and the carbon cycle (Cox, 2019). However, this power also comes with a danger: that blind data mining of multidimensional model outputs could lead to spurious and misleading constraints (Caldwell *et al.*, 2014). Protection from this risk can come from tests of robustness across different model ensembles, and from basing the search for ECs on firm theoretical and mathematical foundations. Theory-led ECs can also be considered as hypotheses that can be tested against the ensemble of complex ESMs. The remainder of this section aims to describe the reasons why relationships between Earth system variability and sensitivity are ubiquitous, and to provide some examples of the emergent relationships to be expected under different types of time-varying forcing (such as seasonal cycles and longer-term trends).

A. Relationships between variability of fluxes and the sensitivity of stores

We can write the time evolution of a dynamical system variable V in the general form

$$\frac{dV}{dt} = F(V, z_i), \qquad (1)$$

where F is the net flux into the system and z_i are the environmental variables that affect this net flux. Equation (1) represents a local conservation law, as it implies that the variable V can change only if the net flux is nonzero. When the environmental variables are in a steady state $z_i = z_i(0)$, the variable V has an equilibrium state $V = V(0)$, defined by

$$F(V(0), z_i(0)) = 0. \qquad (2)$$

In climate science we are typically interested in how a climate system variable, such as global mean temperature or land carbon storage in the tropics, varies with environmental factors, such as the atmospheric carbon dioxide concentration or the climate in the tropics. Where the associated perturbations Δz_i are small compared to $z_i(0)$, Eq. (1) can be linearly expanded about the equilibrium state:

$$\frac{d\Delta V}{dt} = \frac{\partial F}{\partial V}\Delta V + \frac{\partial F}{\partial z_i}\Delta z_i, \qquad (3)$$

where ΔV is the resulting perturbation to the state variable and the partial derivatives are calculated around the initial equilibrium $\{V(0), z_i(0)\}$. The initial equilibrium is stable if $\partial F/\partial V < 0$, and we can then define an effective timescale

to perturbations of V as $\tau = (-\partial F/\partial V)^{-1}$. Equation (3) can therefore be written as

$$\frac{d\Delta V}{dt} + \frac{\Delta V}{\tau} = \frac{\partial F}{\partial z_i}\Delta z_i. \qquad (4)$$

The right-hand side of Eq. (4) can be viewed as the "external" forcing factors that produce changes in V (such as the radiative forcing due to increasing atmospheric CO_2). For a permanent time-invariant change in the environmental variable Δz_i, V will be changed by an amount ΔV_{eq}, where

$$\Delta V_{eq} = \tau \frac{\partial F}{\partial z_i}\Delta z_i. \qquad (5)$$

Defining the sensitivity of V to the environmental variable z_i as $\xi_i = \partial V_{eq}/\partial z_i$, we can rewrite this equation as

$$\xi_i = \tau\mu_i, \qquad (6)$$

where $\mu_i = \partial F/\partial z_i$. Equation (6) implies that the sensitivity of V to a step change in z_i will be proportional to the sensitivity of the net flux F, with a constant of proportionality which is the characteristic lifetime of perturbations to V. In climate change research, the equilibrium sensitivities ξ_i are often the things we would most like to constrain (such as ECS).

B. Theoretical emergent relationships for idealized time-varying forcing

For emergent constraints we therefore need to find relationships between sensitivities ξ_i and observable variations in the climate system. Here we derive candidate emergent relationships for different time variations in the environmental variables z_i. First we rewrite Eq. (4) in the form

$$\frac{d\Delta V}{dt} + \frac{\Delta V}{\tau} = \mu_i \Delta z_i. \qquad (7)$$

As we limit ourselves here to this linear model with a single timescale τ, the following emergent relationships are intended to be illustrative or used as simple hypotheses to be tested against the outputs from complex models.

1. Sinusoidal forcing

As an idealized representation of the response of our system to diurnal and seasonal forcing, we first consider sinusoidal environmental variations of angular frequency ω and amplitude a_z, for which $\Delta z_i = a_z e^{i\omega t}$. The solution to Eq. (7) under these circumstances (after initial transients have died down) is

$$\Delta V = \mu_i a_z \frac{\tau}{\sqrt{1 + \omega^2 \tau^2}} e^{i(\omega t - \phi)}, \qquad (8)$$

where $\phi = \arctan(\omega\tau)$. Therefore, ΔV also has a sinusoidal variation, but with a phase lag relative to the forcing of ϕ that asymptotes to $\pi/2$ as $\omega\tau \to \infty$. Substituting $\Delta z_i = A_z e^{i\omega t}$. The amplitude of the sinusoidal variation in ΔV, a_v is proportional to a_v but also depends on the frequency of the sinusoidal forcing:

$$\frac{a_v}{a_z} = \mu_i \frac{\tau}{\sqrt{1 + \omega^2 \tau^2}}. \tag{9}$$

Using Eq. (6) we can therefore write the sensitivity ξ_i in terms of the ratio of the sinusoidal amplitudes:

$$\xi_i = \frac{a_v}{a_z} \sqrt{1 + \omega^2 \tau^2}. \tag{10}$$

In the limit of slowly varying forcing compared to the system timescale ($\omega\tau \to 0$) this formula reduces to

$$\xi_i \to \frac{a_v}{a_z} \quad \text{for } \omega\tau \to 0. \tag{11}$$

This is the limit of fast variables that are in a quasiequilibrium with the sinusoidal forcing, such as seasonal snow cover (Qu and Hall, 2007) or marine primary production (Kwiatkowski *et al.*, 2017). However, in the opposite high-frequency limit ($\omega\tau \gg 1$) the emergent relationship also depends linearly on the frequency of the forcing:

$$\xi_i \to \frac{a_v}{a_z} \omega\tau \quad \text{for } \omega\tau \gg 1. \tag{12}$$

2. Linearly increasing forcing

Contemporary climate change is largely being driven by an approximately exponential rate of increase in atmospheric carbon dioxide above the preindustrial level, which yields a radiative forcing that is approximately linear in time. This has motivated a long-running series of idealized climate model experiments that prescribe an exponential increase in carbon dioxide, including the *1% per year runs* which are used to define the concept of the *transient climate response* (TCR). In addition, many other Earth system changes can be approximated by linear trends. We therefore consider here emergent relationships under a linear increase in the environmental variable z_i:

$$\Delta z_i = \gamma t, \tag{13}$$

where γ is the linear rate of increase z_i that begins at time $t = 0$. Under this idealized forcing the solution to Eq. (7) is

$$\Delta V(t) = \tau \mu_i \gamma \{t - \tau(1 - e^{-t/\tau})\}. \tag{14}$$

From Eq. (5) we note that the sensitivity $\xi_i = \tau\mu_i$, so Eq. (14) can be rewritten as

$$\xi_i = \frac{\Delta V(t)}{\gamma} \frac{1}{t - \tau(1 - e^{-t/\tau})}. \tag{15}$$

The exponential decay term here is a transient response to the sudden switch on of the linear trend at $t = 0$. Once this term has died out the transient solution lags the quasiequilibrium solution by τ years such that

$$\xi_i = \frac{\Delta V(t)}{\gamma(t - \tau)}. \tag{16}$$

Equation (16) represents a potential emergent relationship between the sensitivity ξ_i and a transient change $\Delta V(t)$.

3. White-noise forcing

Emergent relationships have also been proposed between interannual variability and sensitivities (Schwartz, 2007; Cox *et al.*, 2013; Nijsse *et al.*, 2019), assuming that the environmental variable z_i is approximately Gaussian white noise. Under those circumstances Eq. (7) becomes the widely used Ornstein-Uhlenbeck equation of statistical physics (Uhlenbeck and Ornstein, 1930). Standard solutions relating the variance of V to the variance of z_i can be derived by integrating the sinusoidal solution given by Eq. (8) over all frequencies ω. For interannual variability this yields the following relationship for the ratio of standard deviations:

$$\frac{\sigma_v}{\sigma_z} = \mu_i \sqrt{\frac{\tau}{2}} = \xi_i \sqrt{\frac{1}{2\tau}}, \tag{17}$$

and an even simpler relationship for the lag-1 autocorrelation of Y:

$$\alpha_{v1} = e^{-1/\tau}. \tag{18}$$

Equations (17) and (18) can be combined to yield an equation for the sensitivity ξ_i without needing to know the timescale τ (Cox, Huntingford, and Williamson, 2018):

$$\xi_i = \frac{\sigma_v}{\sigma_z} \sqrt{\frac{2}{-\ln(\alpha_{v1})}}. \tag{19}$$

Similar formulas can be derived for more sophisticated representations, such as two-box and semi-infinite heat diffusion models of ocean heat uptake (Williamson, Cox, and Nijsse, 2019).

IV. HOW EMERGENT CONSTRAINTS MIGHT GO WRONG AND HOW TO GUARD AGAINST IT

Uncertainty in future projections makes adaptation planning difficult, so there is pressure on climate researchers to provide much more refined predictions of expected large-scale environmental change as atmospheric GHGs rise. However, in the absence of full knowledge of all climate processes and their parametrizations, such deficiencies will continue to cause substantial ESM differences. For this reason, the method of emergent constraints has attracted substantial attention. Such interest is because ECs offer a method to potentially "short cut" current deficiencies in process understanding by providing better estimates of change. Or, as a minimum, ECs offer more reliable estimates of bulk aggregated parameters of the climate system of interest to policy. We believe ECs do provide a route to reduce uncertainty, and we consider it appropriate for the technique to substantially underpin societal decisions regarding both adaptation planning and parallel mitigation programs to reduce GHG emissions. However, this brings tremendous responsibility, so the method must be reliable and robust. It is therefore prudent to discuss the multiple circumstances in which the method may fail, thus

making alterations when caution is needed. We now describe such situations.

A. Risks of solely using data mining

One approach with ECs is to simply "data mine" within climate model ensembles, such as CMIP5 and CMIP6, searching for two quantities in each model that when analyzed form a statistically significant inter-ESM regression. In these circumstances, the researcher is not led by process understanding, a hypothesis to be tested, or indeed intuition (see Sec. II.B for an extreme example). However, there is a logic that, in some circumstances, this approach might be valid. In the climate system, there may be links between different parts of the Earth system, and including between a contemporary measurable quantity and an attribute change in a future climate state, that are not immediately obvious as being connected.

One set of examples could be similar to the ongoing discovery of teleconnections in the climate system, where strong correlations are noted between variations in two parts that are a substantial geographical distance apart (Nigam and Baxter, 2015). In addition to spatial distances, discovered teleconnections can contain lags, although, unlike ECs that are designed to project decades ahead, the delays are months. These delays are often used to estimate expected meteorological conditions in the months ahead. For example, the status of El Niño can strongly weight the probabilities of particular weather features for land regions and across the tropics and for the season ahead. Although El Niño is an obvious system fluctuation to investigate for its impacts elsewhere, over the last couple of decades researchers have been investigating other candidates (Feldstein, 2000).

Given the sometimes diverse form of teleconnections, their discovery can require a data-mining approach, as comparisons between two different parts of the climate system that are additionally offset in space and time are unlikely to be found through intuition-led inspection of ESMs. If novel ECs exist between diverse parts of the Earth system, then such mining methods may be needed to aid their discovery. If ECs are discovered by statistical methods, then the subsequent process analysis may reveal that the connection has a strong physical basis. Hall *et al.* (2019) referred to this as "EC confirmation," although this may take many years to achieve, especially if additional ESM diagnostics are needed to confirm the size of any transport equation terms between different spatial locations.

However, while such an approach is expected to reveal new ECs, caution is needed. The concern is that statistically significant regressions are likely to be found in a small number of instances simply by chance, and the resultant EC is therefore not a robust indicator of future change (Sec. II.B). That is, presented loosely, if statistical mining finds a large number of new ECs, all with nonzero regressions forming them and at a 90% confidence level, then roughly one in ten will be invalid. This highlights the need to undertake a process-based confirmation of ECs found through statistical means.

B. The risk of *p* hacking and overconfidence

The term "*p* hacking" was introduced by Nuzzo (2014) and alerts one to the risk of self-selecting only scientific findings that are statistically significant. In the context here that would be ECs having their regressions at small *p* values, suggestive of a low probability of an *X*-*Y* relation occurring merely by chance. The concept has strong similarities to the dangers noted in data mining, where scanning across a range of potential ECs could lead investigators to concentrate on only a few with low *p* values, with the attendant risk that they might occur by chance. However, *p* hacking goes further and raises caution over other decisions that researchers may make in a rush to find a significant result. Behaviors could include a deliberate selection of one-sided tests and the stopping of sampling upon finding a low *p* value. It could also involve adjusting the parameters, relationships between data, or time frequencies considered until lower *p* values are discovered. In the case of the climate system and as an example, the assessment of attributes of the global hydrological cycle offers many timescales of interest. Precipitation statistics with potential to form different *X* and *Y* quantities of ECs range from short, intense rainfall events to seasons or even decades. To account for overconfidence Bretherton and Caldwell (2020) suggested that ECs be corrected by scaling up the unexplained variance by a user-defined factor across the board.

C. Missing process in all current models, measurement errors, and model compensating errors

ECs most frequently link a fluctuating quantity that is measurable for the contemporary period to either a future change (such as the extent of polar ice sheet cover) or an invariant system attribute that describes change (such as climate sensitivity). A potential concern for the validity of an EC is whether there is a missing process in every model that affects the regression forming the basis for the EC. If that error affected the *X*-axis fluctuating quantity, but the measured quantity were accurate, then this would introduce a bias into the projection of the future *Y*-axis quantity.

Such a missing process could itself impact ESM performance at all modeled levels of atmospheric GHG concentrations. One potential example is that many ESMs underestimate interception loss, by vegetation, of rainfall (Lian *et al.*, 2018; Yang *et al.*, 2018). Interception loss is the return of water from rainfall to the atmosphere that has not passed through soil but instead is temporarily held on leaves and branches. If an emergent constraint on a feature of the atmospheric part of the hydrological cycle that depended on the overall land-atmosphere exchange of water, then the EC could have an overly strong dependence on plant evaporation (i.e., to compensate for the low modeled values of interception loss). If, then, the true *X* value were taken from data of actual plant evaporation, this would introduce a bias into the estimate of the *Y* quantity. A straightforward corollary to this is that if an EC depends on an *X* quantity that is modeled well but measurements of it contain a bias, then this too would lead to inaccurate estimates of the value of the *Y* quantity.

Alternatively, other process omissions could become more important as GHG levels rise. Until recently, an example of a component frequently not modeled in ESMs was the terrestrial nitrogen cycle (Thomas, Jack Brookshire, and Gerber, 2015). Nitrogen limitation could become an increasingly important factor as CO_2 concentrations rise by suppressing levels of vegetation fertilization from the increased CO_2 in the atmosphere. The lack of nitrogen cycle inclusion in the land components of ESMs could therefore affect ECs linking current variations of the carbon cycle to future carbon stores, leading to potential overestimates of the capability of terrestrial ecosystems to offset future anthropogenic CO_2 emissions. That said, Wenzel *et al.* (2014) found that two ESMs that did have the nitrogen cycle included were close to an EC regression line in an ensemble of carbon cycle only ESMs. That analysis linked contemporary fluctuations in temperature and atmospheric CO_2 concentrations to the long-term stability of terrestrial carbon stores under global warming.

A corollary to single model errors is the case where two contemporary processes are defined but together allow a model to perform well on some metrics as a consequence of compensating errors. There is substantial anecdotal evidence that many ESMs have these issues, as model developers often discover that improvement of one component, by removing the compensation, actually results in less accurate contemporary climate simulations. Poor performance is then resolved only when the process representation of the second feature is also upgraded. Such compensations present a risk for ECs. Specifically, if the *X*-axis quantity of an EC relates to just one of the two processes that compensate, then that will introduce a bias. Returning to the interception example, if we consider the deficiency recognized but accounted for by parametrizing overly strong plant evaporation amounts, any EC dependent on plant evaporation *X* would introduce a bias in the prediction of *Y* if actual measurements of *X* were used.

Ultimately, climate science requires a set of ESMs that make similar projections and are highly accurate. Hence the ongoing attempt to include all relevant processes in ESMs, and to remove compensating errors, remains an appropriate activity to advance such predictive tools. However, a cobenefit is that such ESM enhancements will likely raise the accuracy of any ECs that are used to constrain understanding where uncertainty remains.

D. System passes through a tipping point

It is arguable that moderate human adjustment to the climate system is relatively linear, so the expected meteorological changes increase proportionally with changes to the combined radiative forcing of atmospheric GHG concentrations. If this same linearity extends to higher frequency system responses such as interannual variability, then observations of contemporary fluctuations may give a strong indication of the system's response to a more permanent forcing from increased GHGs. Cox *et al.* (2013) offered one example, where the change to tropical atmospheric CO_2 concentration in response to annual temperature variation may project any expected loss of rainforest carbon stores under sustained global warming. Hence, there is an assumption of linearity present in each ESM, or at least a relationship that is monotonic in increasing

temperature. When plotting the simultaneous variations in temperature and CO_2 (*X*) against response of future carbon stores in response to warming (*Y*), then a further linear inter-ESM relationship is found, and this is the EC itself.

A legitimate question, therefore, is what happens should the planetary system move beyond linear responses and, further, start to pass through tipping points (Lenton *et al.*, 2008)? Many examples of such potential nonlinear behaviors are conjectured, where a relatively small increase in radiative forcing could cause major system changes. Potential climate tipping points such as the collapse of the Atlantic meridional circulation (Stocker and Wright, 1991), Amazon dieback (Cox *et al.*, 2000), Greenland ice sheet melt (Toniazzo, Gregory, and Huybrechts, 2004), etc. (Drijfhout *et al.*, 2015) are presently thought to be low probability, high-risk events, although there is evidence that subsystems of the climate have "tipped" many times in the past (Bond *et al.*, 1992). Most of these "tips" have resulted in regional rather than global climate changes and have been emulated through simpler equation systems, amenable to bifurcation analysis to characterize the nonlinearity (Dijkstra, 2013).

Once past a large-scale tipping point, the Earth system might have radically different responses and feedbacks and thus behave differently. This is likely to impact the EC approach is two ways. First, for each ESM for sufficiently changed climate, the contemporary fluctuating quantity may no longer give accurate information in that model on future store size. Second, if there are system tipping points but an EC is validated only for moderate changes to the climate system (so the *Y*-axis quantity is derived only from ESMs operated under lower emission scenarios), then using it to project to more severe climate alteration might be inappropriate.

A more abstract but relevant question is whether information from pretipping point conditions has any validity at constraining climatic conditions beyond one? Do tipping points invalidate "information exchange" between different GHG levels when passing through one? Two ESMs may have similar responses up to a tipping point and thereby appear as points near each other on a standard EC *X-Y* plot, but markedly different responses beyond a tipping point, thus weakening an EC-based inter-ESM regression. Also in the abstract is a view that ECs are capitalizing on the "hard wiring" implicit in ESMs, as based on model parametrizations constrained from contemporary measurement datasets. Beyond any major tipping point represents a world that has not yet been witnessed and thus not measured for equation parametrization purposes.

We hope that climate researchers can be encouraged to investigate further whether the planet passing through a tipping point invalidates ECs, and possibly also the period just beforehand. This could be in the form of a conceptual model or by analysis of ESMs in the CMIP5 ensemble. ESM simulations with a tipping point could be split into a future period pretipping point, and one post-tipping point for a high-emissions scenario, testing the EC on both modeled time frames. One obstacle to undertaking such an analysis is that, although ESMs do exhibit tipping points, they vary markedly between models regarding their location, component of the Earth system affected, and level of global warming (Drijfhout *et al.*, 2015). Notable is that the strong regional features of

tipping points implies that any ECs that may be affected by their presence also have to be localized in application. Possibly of more concern is that if tipping points affect the predictive capability of ECs, there will be little inter-ESM consensus on when and how this might occur. This lack of understanding is due to major model differences in predictions of tipping point occurrence, or even existence (Drijfhout *et al.*, 2015), so an "EC-type" modulation factor to the original EC and common across models is unlikely to be discovered.

E. Problems with common code across many models and implications for "out-of-sample" testing

ECs rely on the statistics of regressions, which in turn assume an independence of data points. However, individual ESMs may not be completely independent. For parts of the climate system known to respond to well-established and well-understood physical processes, commonality in models is to be expected. The concern is for the model parts that are suspected as characterizing actual processes less well, and where ECs seek to constrain this uncertainty. The lack of independence may take the form of individual research centers offering multiple model versions but at different resolutions. Alternatively, some components of ESMs are shared between research centers or are coded in similar ways (Knutti, Masson, and Gettelman, 2013). The lack of independence could appear in the range of effective bulk parameters sampled, such as equilibrium climate sensitivity. Another possibility is there may be a common component existing between models, and that an EC seeks to reveal, but the ESM differences to reveal this are in the fluctuating forcings. These fluctuating quantities could have relatively low sampling due to similarities between how ESMs calculate their values.

As part of developing comprehensive process understanding of discovered ECs, Hall *et al.* (2019) encouraged out-of-sample testing. Such testing is where the EC relationship is checked to be valid in additional ESMs that are not part of the original set used to initially find the EC. New simulations from individual modeling centers can be analyzed as they become available. A more comprehensive test is to search within new ensembles, so an EC found in the CMIP5 climate model ensemble can be checked for its presence in the newly released CMIP6 set of ESM simulations. However, this too may not be a completely independent assessment if modeling centers retain substantial amounts of model code and parametrizations between ensemble contributions. We note that the out-of-sample testing performed by Schlund *et al.* (2020) assessed whether the ECs for climate sensitivity, ECS, found in the CMIP5 ensemble remain valid for the CMIP6 models. As commented elsewhere, they found that for CMIP6 the CMIP5-based ECs have less predictive capability and also give generally higher ECS values. Schlund *et al.* (2020) noted that the majority of emergent constraints for ECS are related to some extent to cloud feedbacks, and this is a major ongoing area of climate research to create numerical cloud schemes with strong predictive skill. As many cloud schemes are currently under development, then their newness implies that this is an example of little carrying of common code from CMIP5 to CMIP6. From that perspective, CMIP6 models have independent features from CMIP5, validating the use of this more recent ensemble for out-of-sample testing. However, if it is the new cloud schemes that cause the CMIP6-derived ECs to have less predictive capability, then this suggests that previously estimated EC-based uncertainties on ECS from the CMIP5 models are overly narrow. This form of error differs from that of Sec. IV.C, which considers how a missing process in all models may create a systematic bias in an EC projection. Here, instead, the suspected missing processes associated with cloud dynamics and feedbacks are such that there likely remains substantial uncertainty in how to model them accurately. Alternative cloud schemes may weaken ECs by causing a larger spread around their regression lines, expanding the bounds on predictions of quantity *Y* for quantity *X* and for which contemporary data also exist.

F. What to do when different ECs are found for the same quantity but differ in value, or differ between ensembles

In some instances, multiple and different contemporary measurements have been suggested, via ECs, to be able to predict the same quantity. This is particularly the situation where a broad range of climate attributes have been used to evaluate ECS. However, and notably, cloud-feedback-based (Sherwood, Bony, and Dufresne, 2014; Zhai, Jiang, and Su, 2015; Brient and Schneider, 2016; Brient *et al.*, 2016) or top-of-atmosphere (TOA) radiation-flux-based (Brown and Caldeira, 2017) emergent constraints tend to project higher ECS values than surface-temperature-based constraints (Cox, Huntingford, and Williamson, 2018). Investigations need to continue to ascertain why these ECs disagree. Some disagreements may be due to spurious correlations, affecting *Y* values. However, Caldwell, Zelinka, and Klein (2018) showed that these emergent relationships are highly correlated. With high correlation but different ECS estimates, this implies that differences may instead be due to measurement biases in some *X* value observations. Alternatively, the real world may not be sharing the same responses as the models, which could suggest a persistent error across ESMs for at least one of the ECs.

As ECs come under increasing scrutiny due to their growing widespread use, more needs to be understood about how they operate and any limitations. In that context, understanding the differences between projections of the same quantity will be a highly rewarding research path. However, to generate a single EC-based projection, Bretherton and Caldwell (2020) presented statistical methods to merge multiple ECs of the same quantity into a single range of uncertainty for the quantity being estimated. That analysis focuses on combining the multiple ECs used by other researchers to estimate ECS.

As noted in Sec. IV.E, new ensembles provide an out-of-sample test for existing ECs. The question then is how to use that new value, especially if it is substantially different than the value for an earlier ensemble. One possibility is to use the methods of Bretherton and Caldwell (2020) to merge the findings from two ensembles. A further possibility is to introduce a weighting to the Bretherton and Caldwell (2020) approach, but with a high influence for the more recent ensemble, corresponding to the hope that newer ESMs are better models.

G. ECs may cause future CMIP-type climate model ensembles to have much less spread in projections

In the most general terms, a reduction of spread of climate models is to be highly welcomed, especially if their convergence is onto projections that are accurate estimates of future change. Such a convergence is good if what is learned from ECs applied to previous ensembles is accurate, and this has caused climate modelers to make their new simulations achieve the constrained *Y* values; see Sec. VII.B. Convergence may also occur simply because previously uncertain parts of the Earth system have become better understood. However, the rapid development of ECs could cause a reduction in model spread that is potentially unwelcome, precluding further progress with the methodology.

As ECs are discovered and estimates of future change are refined, this may cause climate modeling groups to, either consciously or unconsciously, parametrize new ESMs so as to estimate changes that fall within those EC-based bounds. This convergence of projections could be because many modeling groups might not wish to estimate climatological changes that fall outside a consensus view based on previous EC-based estimates. However, this could have two detrimental side effects and could potentially cause future EC calculations to be less accurate. First, again noting that it is almost a paradox of ECs, as designed to constrain estimates of change, they can work well only with large inter-ESM differences. It is a substantial spread of ESM estimates that enable "strong" regressions. Hence any clustering of projections will reduce the capability to retest ECs for new ensembles of ESMs, and certainly make them less reliable. Second, it risks that if new processes are believed to be needed in the majority of new models, then other balancing processes could be tuned, incorrectly, to balance these changes in order to fit the earlier EC. For example, introducing the terrestrial nitrogen cycle more routinely into ESMs may suppress projections of future land carbon stores. But to remain in alignment with earlier EC estimates and as using ESMs mainly without nitrogen suppression, then instead in new ESMs this could encourage parametrization of an offsetting and incorrect overly strong CO_2 fertilization effect for vegetation.

Despite the aim of ECs being to reduce the effects of ESM uncertainty, climate science remains alert to these two concerns, where the existence of an EC may overly reduce the model spread in any new ensemble (although this is not something that we have encountered to date). Such a reduction in a new ensemble, where model developers may want to replicate the earlier EC, risks, in particular, the propagation of errors that might be prevalent in the earlier ensemble.

H. Inability to verify an EC

In Sec. IV.A we heed the dangers of ECs based on data mining, noting that Hall *et al.* (2019) encouraged the notion that ECs discovered statistically should first move to having potential, i.e., contain an element of intuition as their confirmation. Detailed process understanding, maybe even as far as analytical assessment of related key climate-based differential equations, ideally then move an EC from having "potential" to being "verified." At the same time, we suggest

a potential danger whereby it is not possible to follow this cascade of increased certainly in validity of an EC. For example, there may be insufficient saved diagnostics from ESMs to build and study off-line the process interactions that need understanding to achieve EC verification. The risk here is that important and informative potential ECs are dismissed, as their understanding cannot be developed. Although this may initially appear to be opposite to the suggestions in Sec. IV.A, we advise such ECs not to be cast aside, thereby with the danger of losing valuable insight into the climate system. Instead such potential ECs from data mining can generate requests to climate research centers to provide related additional ESM outputs, to allow additional testing.

Any additional ESM diagnostics may also reveal the extent to which a potential EC is dependent on one major physical process or is caused by an amalgamation of multiple interacting effects. If more than one major process is present, then this may make it more difficult to transition from "plausible" to verified.

I. Lack of perturbed physics experiments with ESMs

The operation of ESMs is extremely computationally expensive and often there is capacity in research centers only to perform simulations for standardized scenarios of changes in future atmospheric greenhouse gas concentrations [such as the representative concentration pathways (RCPs)] (Meinshausen *et al.*, 2011). This precludes performing simulations with varied parameters, especially for those where there is substantial uncertainty surrounding their true values (or effective values, if grid dependent). This lack of model investigation restricts further testing of ECs, in a way that may either preclude gaining extra confidence in their reliability or alternatively fails to alert to one that may not be as robust as initially believed. PPEs may require interpretation in two ways that are converse to each other. If a parameter value is changed and if it relates to a part of the Earth system believed to be independent of the underlying processes associated with a particular emergent constraint, then for that ESM the *Y* value should not change in the EC. Such invariance both confirms EC robustness and helps illustrate that the EC aligns to the parts of the model expected from intuition or process understanding. The alternative situation is where a parameter believed to relate to an EC is perturbed and potentially adjusted to the outer bounds of its expected value or beyond. The expectation then may be that for that particular ESM the point moves to the bounds of the EC regression. Such an anomaly confirms that the EC is affected by the part of the climate system the parameter pertains to and may indicate that outlying parameter values are not present in the other ESMs.

If spare computational cycles become available, perturbed physics experiments with ESMs will be beneficial to the development and testing of ECs.

V. EMERGENT CONSTRAINTS FOUND IN THE EARTH SYSTEM

Numerous ECs have been identified across physical and biogeochemical components of the Earth system (a handful of which have been discussed thus far), with a substantial

increase in their number over the past decade. To capture the breadth of its application in the geosciences, we discuss previously documented ECs based on their broad geophysical classification here. A list of EC studies is provided in Table I. Note that this list is not exhaustive and may be easily superseded. Moreover, some of the references in Table I offer more than one potential current climate quantity to explain Y, in which case we list only one example. For example, Sherwood, Bony, and Dufresne (2014) provided three metrics that describe various aspects of lower-tropospheric mixing to explain equilibrium climate sensitivity. Rather than list all three metrics, we discuss only the one that appears to be most robust (Caldwell, Zelinka, and Klein, 2018; Schlund *et al.*, 2020). An important criterion for establishing credibility of an EC is to show robustness to the choice of ensemble. Therefore, Table I also documents the model ensemble(s) for which each EC appears to have value (i.e., a statistically significant emergent relationship is present). In several cases this is only the ensemble from which the EC was derived as many of the more recent examples have not undergone out-of-sample verification at the time of this paper. We expect several upcoming publications to assess the validity of previously published ECs with the newer CMIP6 ensemble (Pendergrass, 2020; Schlund *et al.*, 2020).

A vast majority of ECs pertain to one of the following general topics: climate sensitivity, cloud feedbacks, cryospheric feedbacks and change, carbon cycle feedbacks, and the hydrologic cycle, but in theory the methodology can be used for any number of applications. There are also additional examples relating to topics such as radiative forcing (Bowman *et al.*, 2013), regional air and sea-surface temperature change (Lin *et al.*, 2017; Sgubin *et al.*, 2017; Selten *et al.*, 2020), frequency of temperature extremes (Donat, Pitman, and Angélil, 2018), and atmospheric circulation changes (Kidston and Gerber, 2010; Simpson and Polvani, 2016). For the purpose of this review, we focus primarily here on the main applications to date. Along with the differing applications, ECs can also be sorted by the type of constraint (Fig. 5). One way to think of this is to group ECs by the timescale of information that defines their X (such as from multidecadal to extreme events). For example, a rather simple class of ECs relates climatological biases or multidecadal trends in some quantity X to the future change in Y. Similarly, a collection of ECs use seasonal or interannual variations in X to constrain future changes in Y. In many of these instances, variability in some quantity to seasonal or interannual temperature variability is related to sensitivities of that same quantity to future warming. Constraints using climatological biases or short-term sensitivities are also common for attempting to constrain invariant properties such as equilibrium climate sensitivity. Applications to future changes in higher-order climate statistics (such as extreme events) are more rare. This highlights a relatively unexplored area of EC research.

A. Equilibrium climate sensitivity

By far the most popular application of the EC approach has been in relation to ECS (Table I). ECS is seemingly an ideal candidate for this technique because of its importance in

FIG. 5. Different types of emergent constraints proposed in the literature to date. Each number corresponds to a study listed in Table I. ECs are sorted by the timescale of their current climate quantity and the type of future constrained quantity. ECs are also sorted by whether the future quantity relates to an impact or a physical or geochemical quantity.

predicting future warming rates and its consistently large spread across various GCMs, which has been a persistent feature of all model generations. [The likely range of ECS according to the IPCC (1.5 to 4.5 K) has been largely unchanged for decades.] A series of contemporary quantities X have been identified as being closely linked to ECS, many of which relate to simulated cloud feedbacks or properties. In general, these cloud-based constraints tend to suggest relatively high ECS values (3.5–4.1 K) (Knutti, Rugenstein, and Hegerl, 2017; Brient, 2020). Relevant examples include the height of tropical low clouds (Brient *et al.*, 2016), the sensitivity of the reflection by subtropical low clouds to sea-surface temperature (Brient and Schneider, 2016), seasonal sensitivity of low clouds between 20–40° N and 20–40° S to sea-surface temperature (Zhai, Jiang, and Su, 2015), variability in relative humidity and cloud extent (Fasullo and Trenberth, 2012), vertically resolved zonally average relative humidity and clouds between 45° S and 40° N (Su *et al.*, 2014), and the climatological difference between tropical and Southern Hemisphere midlatitude cloud fraction (Volodin, 2008).

Another suite of studies relate ECS to intermodel differences in the characteristics of historical radiative fluxes, such as climatological TOA energy balance in the Southern Hemisphere (Trenberth and Fasullo, 2010), features of TOA radiation fluxes (Huber *et al.*, 2011; Tett *et al.*, 2013; Brown and Caldeira, 2017), and cloud-sky radiative flux sensitivity to temperature (Lutsko and Takahashi, 2018). Alternatively, studies have also used temperature characteristics defined over a variety of timescales to constrain ECS. This approach

tends to suggest relatively weaker ECS values than the others (Knutti, Rugenstein, and Hegerl, 2017). Examples of *X* include the seasonal cycle of temperature (Knutti *et al.*, 2006), 20th century warming (Annan and Hargreaves, 2006), and statistics of interannual variability (Cox, Huntingford, and Williamson, 2018). Along the same lines, there is a collection of studies that have evaluated temperature changes over much longer timescales (such as global mean cooling during the Last Glacial Maximum: 19–23 thousand years ago) as derived from paleoclimate data (Hargreaves *et al.*, 2012; Schmidt *et al.*, 2014). These paleoclimate-based constraints suggest ECS of 2.3 and 3.1 K, respectively. However, there is a large amount of uncertainty surrounding the model simulations of the Last Glacial Maximum, the proxies used to constrain the intermodel relationship, and the validity of using data from these time periods as proxies for anthropogenic warming (Harrison *et al.*, 2015; Brient, 2020). A variety of other metrics have also been proposed as being strongly related to ECS with large variability in their constrained predictands. The highest value (4.5 K) is derived from a constraint using vertical mixing between the boundary layer and the lower troposphere over tropical oceans (Sherwood, Bony, and Dufresne, 2014). Tian (2015) used climatological precipitation in the Intertropical Convergence Zone (ITCZ) region to suggest a slightly weaker ECS (4.1 K), while the weakest nonpaleo estimate uses the climatological latitude of the Southern Hemisphere (SH) Hadley cell edge in December–February to suggest an ECS of only 2.5 K (Lipat *et al.*, 2017).

As discussed in Sec. IV.F, having such a large number of ECs for the same quantity can complicate their interpretation when the constraints suggest differing *Y* values. In this instance, rigorous testing of various proposed ECs is necessary for better understanding their validity. Along these lines, ECs on ECS have been heavily scrutinized over the past few years with assessments tied to their plausibility and robustness across both multimodel ensembles (Grise, Polvani, and Fasullo, 2015; Caldwell, Zelinka, and Klein, 2018; Schlund *et al.*, 2020) and, to a lesser extent, perturbed physics ensembles (Kamae *et al.*, 2016; Zhao *et al.*, 2016; Wagman and Jackson, 2018). Caldwell, Zelinka, and Klein (2018) provided the most robust assessment to date on the credibility of various constraints on ECS. In doing so, they found that several of the aforementioned constraints lack a physically plausible mechanism connecting the *X* quantity with ECS (Volodin, 2008; Su *et al.*, 2014; Tian, 2015; Siler, Po-Chedley, and Bretherton, 2018). This limits how much faith can be put in the EC itself, as the statistical relationship could be fortuitous. Furthermore, many of the constraints are found to be closely related (Caldwell, Zelinka, and Klein, 2018). For example, one strongly correlated group relates aspects of present-day Southern Hemisphere cloud cover with ECS, suggesting they may all be capturing a single SH mechanism (Volodin, 2008; Trenberth and Fasullo, 2010; Lipat *et al.*, 2017; Siler, Po-Chedley, and Bretherton, 2018). Additional testing of EC robustness comes in the form of out-of-sample verification (i.e., evaluating the emergent relationship in a different model ensemble to the one from which the EC was originally derived). Although there are some known limitations to this approach related to the existence of common model code across generations (Sec. IV.E), it is still seen as a valuable exercise for evaluating ECs. For example, Grise, Polvani, and Fasullo (2015) showed that the Trenberth and Fasullo (2010) EC derived from CMIP3 is not valuable in CMIP5 because the emergent relationship exists in only a subset of models with unrealistic cloud properties in the Southern Hemisphere subtropics. Caldwell, Zelinka, and Klein (2018) extended this type of analysis to show that the Fasullo and Trenberth (2012) EC is also nonexistent in CMIP5. On the other hand, they found the Volodin (2008) EC to be present in both CMIP3 and CMIP5, despite lacking a physically plausible explanation. Schlund *et al.* (2020) recently assessed a number of the constraints on ECS documented here using the CMIP6 ensemble. They found that most of these constraints appear to be less skillful at predicting ECS in CMIP6, tied largely to differing representation of cloud processes. Three of the ECs discussed here (Volodin, 2008; Sherwood, Bony, and Dufresne, 2014; Brient and Schneider, 2016) still exhibit statistically significant skill, while two others no longer have value (Su *et al.*, 2014; Lipat *et al.*, 2017). Finally, two ECs are somewhat in between, what the researchers defined as "indeterminate"; they were discussed by Tian (2015) and Cox, Huntingford, and Williamson (2018) (Schlund *et al.*, 2020). Further testing of these ECs has utilized perturbed physics ensembles to better sample the parametric uncertainty space associated with various emergent relationships (Sec. II.A). Of particular interest to these studies is the EC proposed by Sherwood, Bony, and Dufresne (2014). Results from three different PPEs show that the physical mechanism proposed by this study may not be valid, while the emergent relationship between lower-tropospheric mixing and ECS exists only when certain convection schemes are used (Kamae *et al.*, 2016; Zhao *et al.*, 2016; Wagman and Jackson, 2018). Similar doubts were also reported for the Fasullo and Trenberth (2012) EC (Wagman and Jackson, 2018).

This collection of analyses implies that the most robust constraints on ECS (determined by successful out-of-sample testing and plausible physical mechanisms) are tied to the present-day response of subtropical low clouds to sea-surface temperature (SST) variability (Zhai, Jiang, and Su, 2015; Brient and Schneider, 2016). However, one factor that may limit confidence in these constraints is the short observational record of low-cloud characteristics (Brient, 2020). ECs from Volodin (2008) and Sherwood, Bony, and Dufresne (2014) are also statistically significant in three ensembles (Table I), but they face credibility issues tied to their physical mechanism or lack thereof (Caldwell, Zelinka, and Klein, 2018; Wagman and Jackson, 2018). Despite these signs of robustness, the much weaker correlation values associated with these ECs in CMIP6 (Schlund *et al.*, 2020) supports the belief that it is unlikely that a single predictor will be able to physically explain a large amount of ECS variability (Caldwell, Zelinka, and Klein, 2018; Hall *et al.*, 2019). Rather, several constraints on various feedback components will likely be needed to make a difference for such complex processes. In this fashion, a recent comprehensive assessment of ECS gave little weight to ECs on ECS when attempting to determine its likely value, instead favoring observations of actual temperature change and climate feedbacks (Sherwood *et al.*, 2020). This is also believed to be true for the global cloud feedback (later

described in more detail) (Klein and Hall, 2015) given that the radiative effect of clouds in response to warming is expected to vary by cloud type and regime (Gettelman and Sherwood, 2016).

Similar research efforts have recently suggested that historical decadal warming trends are strongly predictive of the TCR (defined as the amount of warming that occurs at the time of atmospheric carbon dioxide doubling, having increased by 1% each year) (Jiménez-de-la-Cuesta and Mauritsen, 2019; Nijsse, Cox, and Williamson, 2020; Tokarska *et al.*, 2020). TCR is of interest because it more closely resembles the way carbon dioxide concentrations have changed in the past. These constraints suggest that the best estimate of TCR is 1.67 K in CMIP5 (Jiménez-de-la-Cuesta and Mauritsen, 2019) and between 1.60 and 1.68 K in CMIP6 (Nijsse, Cox, and Williamson, 2020; Tokarska *et al.*, 2020). The difference between the last two studies largely stems from the choice of evaluated historical time period (Nijsse, Cox, and Williamson, 2020). All three of these estimates are lower than the raw ensemble statistics (median of 1.95 K in CMIP6). Similar constraints on future warming have also been developed for regional temperature changes. Lin *et al.* (2017) and Selten *et al.* (2020) both found that present-day summer temperatures across the central USA and Europe, respectively, are strongly tied to future continental summer warming. They hypothesized that the same physical mechanisms operating in the current climate (tied to precipitation biases and soil hydrology) also govern the future climate response. Incorporating observations allows for the future spread in European summer warming to be reduced by nearly 50% (Selten *et al.*, 2020).

B. Cloud feedbacks

One of the primary contributors to intermodel spread in climate sensitivity is the cloud feedback, which encompasses changes in how clouds modulate the radiation budget in response to warming. [For a recent synthesis of cloud feedbacks see Sec. 3.3 of Sherwood *et al.* (2020).] The global mean cloud feedback is the most uncertain component of the total climate feedback (Bony *et al.*, 2006; Ceppi *et al.*, 2017; Sherwood *et al.*, 2020), varying from -0.36 to $1.17 \, \mathrm{W\,m^{-2}\,K^{-1}}$ in the latest generation of climate models (Zelinka *et al.*, 2020). Strengthening of ensemble mean cloud feedback strength in CMIP6 is also thought to be largely responsible for a recent ECS increase in many models (Zelinka *et al.*, 2020). Thus, many of the aforementioned constraints for ECS include cloud effects either directly or indirectly (Volodin, 2008; Fasullo and Trenberth, 2012; Su *et al.*, 2014; Zhai, Jiang, and Su, 2015; Brient and Schneider, 2016; Brient *et al.*, 2016; Qu *et al.*, 2018). This means that there is great interest in constraining cloud feedbacks as well (Brient and Bony, 2013; Gordon and Klein, 2014; Qu *et al.*, 2014; Zhou *et al.*, 2015; Siler, Po-Chedley, and Bretherton, 2018). For example, Siler, Po-Chedley, and Bretherton (2018) proposed a constraint on future global mean cloud feedback using the climatological latitudinal gradient in cloud reflectivity (a quantity that can be derived from satellite observations). In general, this EC suggests that ESMs with lower cloud albedo in warm-SST regions and higher cloud albedo in cool-SST regions will exhibit greater cloud feedback (and to a lesser extent ECS). It suggests a global cloud feedback in the upper range of ESM estimates ($0.58 \pm 0.31 \, \mathrm{W\,m^{-2}\,K^{-1}}$). However, this constraint lacks a testable physical explanation (Caldwell, Zelinka, and Klein, 2018). Another study that sought to constrain the global mean cloud feedback was conducted by Zhou *et al.* (2015). They highlighted strong similarities between the global mean cloud feedback and a present-day analog cloud feedback derived from interannual climate variability (Zhou *et al.*, 2015). This follows the assumption of several ECs that processes operating on seasonal to interannual timescales are also likely to be occurring on longer timescales. However, the usefulness of this EC is limited because the observational uncertainty in the present-day quantity is large relative to intermodel spread.

As previously mentioned, it is unlikely that a single current climate quantity can fully explain variability in the global cloud feedback (Klein and Hall, 2015). Thus, a number of studies have focused on isolating the individual components of the cloud feedback (high-cloud altitude, tropical marine low-cloud, tropical anvil cloud-area, land cloud amount, midlatitude marine low-cloud amount, and high-latitude low-cloud optical depth). For example, Po-Chedley *et al.* (2019) narrowed in on the changes to upper troposphere clouds and relative humidity across the tropics. They showed that future changes in the upper troposphere cloud fraction under climate warming are strongly tied to the vertical gradient in the climatological mean cloud fraction and relative humidity. Elsewhere, much of this research concerns tropical low clouds as ESMs struggle to represent them; they account for nearly half of the variance in the global mean cloud feedback (Zelinka, Zhou, and Klein, 2016; Klein *et al.*, 2017). Moreover, intermodel spread in the low-cloud feedback is closely tied to variability in ECS (Sherwood *et al.*, 2020). An early EC along these lines showed that the future low-cloud feedback sign Y is related to current low-cloud sensitivity as a result of the natural variability X in CMIP3 (Clement, Burgman, and Norris, 2009). Unlike most other ECs, this example simply suggests the sign of a feedback rather than specific values. Thus, it is not always mentioned in reviews of this nature (Klein and Hall, 2015; Brient, 2020). More traditional ECs relating to this topic include the link between low-cloud optical depth changes with warming and the optical depth response to temperature anomalies associated with natural variability (Gordon and Klein, 2014). This feedback is of limited importance globally but makes an important contribution at high latitudes (Klein and Hall, 2015). The Gordon and Klein (2014) EC also has a simple thermodynamic explanation, which makes it a promising example. Low-cloud amount changes under future warming play a much larger role in the global cloud feedback, so this has naturally been an area of focus (Qu *et al.*, 2014, 2015; Brient and Schneider, 2016). Qu *et al.* (2014) identified an EC on this low-cloud response through the low-cloud response to temperature and stability anomalies derived from interannual variability. Although weaker in CMIP5 than CMIP3 (Klein and Hall, 2015), this EC has value in both ensembles and features a strong physical basis explaining the relationship (Qu *et al.*, 2014, 2015). This type of process-based evidence can then be

used in combination to assemble the best estimate of the more complex global mean cloud feedback. Figure 7 of Sherwood *et al.* (2020) illustrates this concept to get a global cloud feedback of 0.45 ± 0.33 W m^{-2} K^{-1}. Going forward, it is imperative that we continue to improve our process-based understanding of individual cloud components, where potential physical mechanisms are more likely to be uncovered.

C. Carbon cycle

Elsewhere, the EC technique has been used extensively to constrain elements of climate-carbon cycle feedbacks (Table I, rows 28–36) (Cox *et al.*, 2013; Hoffman *et al.*, 2014; Wenzel *et al.*, 2014, 2016; Mystakidis *et al.*, 2017). Climate–carbon cycle feedbacks are characterized by changes in terrestrial and ocean carbon storage in response to climate change, which modify the atmospheric CO_2 concentration, thus enhancing warming. Intermodel differences in feedback strengths lead to highly variable projections of future atmospheric carbon dioxide [more than 300 ppmv of intermodel spread at the end of the 21st century for a high-emissions scenario (Cox, 2019)]. A rather simple relationship was uncovered between the midcentury atmospheric CO_2 concentration in ESMs and the simulated present-day concentration, but it was shown to weaken as the influence of varying climate-carbon feedbacks becomes more apparent later in the 21st century (Hoffman *et al.*, 2014). This EC suggests a much narrower likely range for atmospheric CO_2 concentrations at the end of the century (947 ± 35 ppm) than the full ensemble (980 ± 161 ppm).

Much of the model uncertainty in future atmospheric CO_2 has been shown to stem from differences in the land carbon sink (Cox *et al.*, 2013; Wenzel *et al.*, 2014). Therefore, significant effort has gone into constraining this feature of the carbon cycle. A prominent example uses the historical sensitivity of the annual atmospheric carbon dioxide growth-rate to temperature variability (which is strongly impacted by tropical land carbon storage fluctuations) to constrain the future loss of terrestrial tropical carbon (Cox *et al.*, 2013; Wenzel *et al.*, 2014). As with prior examples, the CO_2 growth-rate sensitivity is an observable quantity, thus allowing for the EC to suggest future tropical land carbon stability. This EC has been proven robust across both C4MIP and CMIP5 ensembles (Wenzel *et al.*, 2014)[5] but has yet to undergo testing pertaining to CMIP6. This is an important EC because there was a fourfold simulated spread in tropical land carbon reductions per degree of warming in C^{4}MIP (29–133 GtC K^{-1}) (Jones *et al.*, 2016), with the higher-end models suggesting potentially catastrophic dieback of the Amazon rainforest (Cox *et al.*, 2000, 2004). However, the Cox *et al.* (2013) EC suggests tropical land carbon reductions slightly weaker than the ensemble mean (53 ± 17 GtC K^{-1}).

---

[5]Note that Wang *et al.* (2014) also attempted to evaluate this EC in CMIP5, but their use of RCP8.5 to diagnose the sensitivity of tropical land carbon loss to tropical warming is not comparable to the work of Cox *et al.* (2013) and Wenzel *et al.* (2014). Given that there is land-use change in RCP8.5, they conflate changes in tropical land carbon due to climate change, with changes in land carbon due to deforestation (for which there is no physical reason to expect a correlation with X).

Another carbon cycle EC uses the change in seasonal atmospheric CO_2 amplitude to constrain CO_2 fertilization of photosynthesis on the extratropics (Wenzel *et al.*, 2016). On a more regional basis, Winkler *et al.* (2019) highlighted a strong relationship between future increases in terrestrial Arctic gross primary productivity and historical increases in leaf area index (greening). These two ECs both suggest that most models are underestimating future changes in gross primary productivity across the high latitudes (Wenzel *et al.*, 2016; Winkler *et al.*, 2019). While these ECs have yet to undergo out-of-sample verification, Winkler, Myneni, and Brovkin (2019) provided detailed evaluation of the factors contributing to uncertainty in the Winkler *et al.* (2019) constraint. They point to choices relating to the temporal period of the predictor variable, the choice of observational dataset, and the rate of CO_2 forcing as particularly large sources of uncertainty (Winkler, Myneni, and Brovkin, 2019). Finally, additional research identified constraints on terrestrial carbon cycle feedbacks using interannual variability in evapotranspiration, net biome productivity, and gross primary productivity. This EC suggests a 40% reduction in the climate-carbon feedback and a 30% reduction in the concentration-carbon feedback (Mystakidis *et al.*, 2017).

When it comes to understanding how the ocean carbon cycle will respond to climate change, one of the main challenges is to reduce uncertainty in estimates of tropical ocean primary productivity. Satellite observations of the interannual variability in ocean productivity resulting from El Niño–Southern Oscillation- (ENSO-) driven SST anomalies have been used to constrain highly uncertain projected changes in tropical marine primary productivity with warming (Kwiatkowski *et al.*, 2017). There is a strong inverse relationship between net primary productivity (NPP) and SST anomalies in the observational record, but the number of ENSO events in the satellite record is limited. This EC suggests a substantial reduction in the long-term tropical NPP sensitivity (-3.4% to -2.4% K^{-1}) from the unconstrained ensemble average ($-4.0 \pm 2.2\%$ K^{-1}). Another recently published EC by Kessler and Tjiputra (2016) relates future global ocean carbon uptake to the contemporary carbon uptake by the Southern Ocean. In this example, models with anomalously low uptake in the current climate project low global uptake over the course of the 21st century. Uncertainty in simulated ocean uptake also translates to uncertainty in projections of future ocean acidification, the impacts of which are expected to be greatest in the Arctic ecosystem (Terhaar, Kwiatkowski, and Bopp, 2020). Recent work uses CMIP5 models to identify an EC on future Arctic Ocean acidification using the simulated density of Arctic Ocean surface waters in the current climate (Terhaar, Kwiatkowski, and Bopp, 2020). Observations of sea-surface density, when coupled with this emergent relationship, imply that future Arctic Ocean acidification will be greater than previously expected.

D. High-latitude processes

Earth's high-latitude regions are rapidly warming and there is a great deal of uncertainty in how various components of the cryosphere (Earth's frozen surfaces) will evolve in the future (Meredith *et al.*, 2019; Mudryk *et al.*, 2020; Notz *et al.*, 2020).

Thus, ECs are an interesting option for reducing intermodel spread here. One of the earliest documented examples of an emergent constraint is for the snow albedo feedback (Hall and Qu, 2006), which was introduced in Sec. I. This is a leading example of the EC technique because of the simplicity behind its physical mechanism and its robustness across several generations of models (Fig. 2). An interesting point on the SAF EC is that despite substantial time since its initial publication, we have yet to see a meaningful reduction in its current climate quantity (X).

Following on from its application to SAF, recent work has identified that the emergent constraint approach can also be applied to a similar process over Arctic sea ice (Thackeray and Hall, 2019). The sea-ice albedo feedback (SIAF) is characterized by the enhancement of future warming through the reduction of surface albedo as a result of decreasing sea ice (Holland and Bitz, 2003; Hall, 2004). An analog of this process operates in the seasonal cycle during the transition from maximum to minimum ice extent (X). These two processes are found to be closely linked across the CMIP5 ensemble, forming the basis of an EC. However, given the projected rapid loss of sea ice in the 21st century, this constraint exhibits a regime dependence whereby the relationship between seasonal cycle and climate change SIAF begins to break down in the latter half of the 21st century, when most models exhibit an ice-free Arctic (Thackeray and Hall, 2019). This is similar to the concept of tipping points discussed in Sec. IV.D.

In addition to the aforementioned radiative feedbacks, there is a large amount of uncertainty when it comes to the fate of a number of high-latitude Earth system components, notably sea ice and permafrost (Stroeve *et al.*, 2012; Slater and Lawrence, 2013). Given this uncertainty and the potentially widespread implications associated with these changes, ECs can provide valuable information here. A number of studies have proposed constraints on projections of future Arctic sea ice (Boe, Hall, and Qu, 2009; Massonnet *et al.*, 2012; Liu *et al.*, 2013). Boe, Hall, and Qu (2009) used the observed historical trend in September Arctic sea-ice extent over the satellite era to constrain the time in the 21st century when the Arctic is likely to become ice free during the summer. Similar research by Massonnet *et al.* (2012) used a series of historical ice characteristics and trends in an attempt to constrain the time period when the Arctic is likely to become seasonally ice free. Their results suggest ice-free conditions for September to begin somewhere between 2041 and 2068 under a high-emissions scenario, a significant reduction from the full CMIP5 spread that spans nearly 100 years. The presence of constraints between historical ice properties and the future change in both CMIP3 and CMIP5 is a promising sign, but we are not aware of CMIP6 testing for this EC to date. Similar historical sea-ice metrics have also been used to inform future changes in high-latitude temperature variability (Borodina, Fischer, and Knutti, 2017a). Permafrost underlies a significant portion of the Northern Hemisphere high latitudes and is highly sensitive to warming (Lawrence, Slater, and Swenson, 2012), with its degradation expected to have detrimental climate and developmental effects (Teufel and Sushama, 2019). However, modern ESMs struggle to agree upon future changes in permafrost area (Slater and Lawrence, 2013).

Chadburn *et al.* (2017) used a relationship between mean annual air temperature and permafrost area during the historical period (X) to constrain projections of future permafrost thaw (Y). This constraint suggests that roughly 20% more permafrost will be lost per degree of warming than previously expected (Chadburn *et al.*, 2017).

E. Hydrologic cycle

Future changes to precipitation are of great interest to the climate science community because of wide-ranging impacts on natural and human systems. However, precipitation is highly variable in space and time, thus making it difficult to both observe and predict. These factors contribute to large intermodel differences in future projections of precipitation. Reducing this uncertainty is therefore vital for those making policy and infrastructure decisions based on model projections. ECs have been applied to multiple aspects of the hydrologic cycle across various scales (O'Gorman, 2012; DeAngelis *et al.*, 2015; Li *et al.*, 2017; Watanabe *et al.*, 2018; Rowell, 2019). First, we consider the globally averaged changes to the hydrologic cycle. In response to rising temperatures and increasing atmospheric water vapor, global mean precipitation is projected to increase. This is because the atmosphere radiatively cools to space and this radiative cooling must be balanced by latent heat release from precipitation, thus setting the radiative-convective balance (Pendergrass and Hartmann, 2014; DeAngelis *et al.*, 2015). Under a high-emission scenario, models project a future increase of between 2% and 10% in global mean precipitation (Kharin *et al.*, 2013). This spread can be partially explained by differing rates of future warming (Fläschner, Mauritsen, and Stevens, 2016), although a threefold spread still exists when the precipitation change is normalized by warming (Kharin *et al.*, 2013). In an attempt to minimize this spread, DeAngelis *et al.* (2015) used the sensitivity of clear-sky shortwave radiation absorption to changes in column water vapor to constrain the global mean precipitation increase with warming (termed hydrologic sensitivity). This constraint suggests a 35% reduction in the ensemble spread and a 40% reduction in the ensemble mean hydrologic sensitivity. This work was followed by Watanabe *et al.* (2018), who used the climatological global mean surface longwave cloud radiative effect (CRE) to constrain the future surface longwave CRE feedback. In combination with the most likely sensitivity of clear-sky shortwave radiative absorption to water vapor (DeAngelis *et al.*, 2015), this EC suggests a constrained estimate of 1.8% K^{-1} for global hydrologic sensitivity, down from an ensemble mean of 2.6% K^{-1} (Watanabe *et al.*, 2018). However, a recently published assessment of these constraints in CMIP6 revealed that although the DeAngelis *et al.* (2015) EC is still present, its strength is weakened (Pendergrass, 2020). Moreover, the Watanabe *et al.* (2018) EC on hydrologic sensitivity is found to be nonexistent in the new ensemble. This highlights the importance of applying out-of-sample testing to the remainder of the ECs on the hydrologic cycle components discussed later on.

Another key component of the global hydrologic cycle is terrestrial evapotranspiration (ET), which is made up of

evaporation and biological transpiration (T). The ratio of T to ET is used to estimate land water fluxes, while, because similar processes influence T and land-atmosphere carbon exchanges, better estimates of T/ET help to reduce uncertainty in global carbon cycle projections (Lian *et al.*, 2018). Climate models exhibit large variability in the global strength of T/ET spanning a factor of 3. Lian *et al.* (2018) found a strong relationship ($r^2 = 0.93$) between simulated local T/ET (averaged over grid cells corresponding to *in situ* measurements) and global mean T/ET. This EC reveals that the constrained global T/ET ratio is significantly higher than the CMIP5 ensemble mean.

Alternatively, several studies have attempted to use this technique to constrain the change in aspects of the hydrologic cycle over specific regions. Regional constraints have the potential for more policy and impacts relevance than globally averaged metrics. As a first attempt, O'Gorman (2012) proposed an EC on the change in future daily tropical precipitation extremes using the observed sensitivity of tropical precipitation extremes to temperature variability. They found a strong relationship between the sensitivity of heavy precipitation (99.9th percentile) to both seasonal and future temperature changes. The constrained estimate based on observations (6%–14% K^{-1}) was substantially narrower than the intermodel spread across the CMIP3 ensemble (2%–23% K^{-1}). Similarly, Borodina, Fischer, and Knutti (2017b) found that the scaling of annual maximum one day precipitation (R × 1 day) with warming global land temperatures can be used to constrain the future intensification of heavy rainfall across extratropical regions with sufficient data records. This EC suggests that the CMIP5 models are likely underestimating the future change in extreme rainfall where climatological rainfall intensity is high.

Another example links biases in simulated historical climate and future Indian summer monsoon (ISM) rainfall (Li *et al.*, 2017). In this case, Li *et al.* (2017) found that models overestimating historical precipitation across the tropical western Pacific also exhibit larger increases in monsoon rains in the future (Li *et al.*, 2017). Using observations of western Pacific rainfall reduces the projected increase in ISM rainfall by nearly 50%. CMIP5 models also differ greatly in their projections of future total East African long rains (March-April-May mean spanning a region from northwest Tanzania to southwest Ethiopia), from a 20% decrease to a 120% increase (Rowell and Chadwick, 2018). Rowell (2019) used an EC between present-day interannual SST–low level cloud sensitivity and future SST change over the Indian Ocean to assess the credibility of projections of East African precipitation. They found that one outlier model, which projects a doubling of seasonal precipitation, is likely unreliable because of unrealistic SST–low-cloud processes. This reduction in intermodel spread is another example of the value that an EC approach can provide. Along these lines, Lehner *et al.* (2019) recently suggested that the sensitivity of historical runoff to temperature and precipitation change across three watersheds in the western U.S. is closely tied to projections of future runoff across a series of ESMs. It is expected that regional applications of ECs to highly uncertain quantities like precipitation change will be a key area of future research.

VI. STATISTICAL UNDERPINNINGS

Accurately constraining the unknown future value of Y in the real world requires one to include and quantify all the possible sources of uncertainty in each step of the EC procedure. In this section, we give an overview of four types of uncertainty that have been incorporated into ECs so far: those stemming from uncertainty in the real world observation of X, uncertainty in X from internal variability, uncertainty in the functional form of the emergent relationship, and uncertainty from ESMs being imperfect replications of the real world. The section concludes with a discussion on how to combine those four types of uncertainty in the resulting EC. In addition, we discuss how to combine multiple emergent constraints of the same quantity, derived from alternative features of historical climate.

A. Uncertainty in observations

There are multiple sources of observational error in the real world value of X. First, a lack of spatial and/or temporal coverage can be present and this may lead to biases if not taken into account (Cowtan and Way, 2014). There are two ways to handle missing data: it can be interpolated and extrapolated from existing data or, alternatively, model output can be filtered to reflect only locations and times for which observational data are present (AchutaRao *et al.*, 2006; Durack *et al.*, 2014; Cox, Huntingford, and Williamson, 2018). Second, observational records are of finite length, introducing additional uncertainty from low sample size. Standard errors quantifying finite size effects can be computed, but care should be taken when time series are autocorrelated as this increases the standard error by effectively reducing the sample size (Trenberth, 1984). For stationary processes, standard equations for autocorrelation errors were given by Zhang (2006).

Estimates of errors in instrumentation and data gathering are often available in the literature (Hennermann, 2018). If multiple observational datasets are available, these can be used to infer uncertainty (Trenberth and Fasullo, 2010; Kwiatkowski *et al.*, 2017). The errors in these datasets might not be independent; different satellite products might for instance have the same biases. Independent sources of error σ_1 and σ_2 can be simply computed by $\sigma_{\text{total}}^2 = \sqrt{\sigma_1^2 + \sigma_2^2}$.

If observational uncertainty makes up a large percentage of overall uncertainty, care should be taken to assess whether to use a normal distribution to describe the probability density function (PDF). It may sometimes be possible to estimate a full PDF from measurements. Alternatively, stochastic reduced-form modeling of the system can be used to estimate the shape (Nijsse and Dijkstra, 2018; Williamson, Cox, and Nijsse, 2019).

Whatever method is used, it is important for this step of capturing observational error not to be neglected (Hall *et al.*, 2019).

B. Uncertainty from internal variability

Like the real world, climate models have internal variability. Because of the finite length of the simulation or the

observed climate record, internal variability can have a significant impact on the estimation of the predictor. One possibility is to use long model control simulations to estimate the size of internal variability if it is believed to be independent of forcing (Nijsse *et al.*, 2019). Variability may, however, be dependent on forcing and, consequently, estimating it from a forced initial value ensemble (see Sec. II.A) may be preferable (Tokarska *et al.*, 2020). For instance, global interannual variability is expected to decrease in the future (Huntingford *et al.*, 2013). Jiménez-de-la-Cuesta and Mauritsen (2019) used the 100-member historical ensemble of MPI-ESM1.1 to quantify the effect of internal variability, whereas Nijsse, Cox, and Williamson (2020) used all available historical initial value members from each CMIP6 model to estimate the mean model variability. Both used model estimates as a proxy for real internal variability.

C. Uncertainty in the functional form of the relationship

Reducing a high-dimensional climate model to a lower dimension brings some uncertainty. Not all variance will be explained with only two variables, and performing a regression is a tool to quantify this. While most emergent constraints thus far assume linear relationships between X and Y and use linear regression to infer the emergent relationship, the regression does not necessarily have to be linear (Bracegirdle and Stephenson, 2012; Nijsse and Dijkstra, 2018). If a linear relationship is imposed when, in reality, the relationship is nonlinear, additional errors will occur. A nonlinear emergent relationship leads mostly to a non-normal PDF for Y with a standard deviation that is potentially significantly larger or smaller than for a linear fit; see Fig. 6. More data or clear prior information on the parameters is needed when fitting additional parameters.

Regression dilution takes place as a consequence of errors in estimating model predictors (such as finite simulation length): when there is an error in the modeled explanatory variables, the slope of a linear regression fit will be smaller than without error and the intercept regresses toward the mean (Frost and Thompson, 2000). Multiple strategies to reduce this
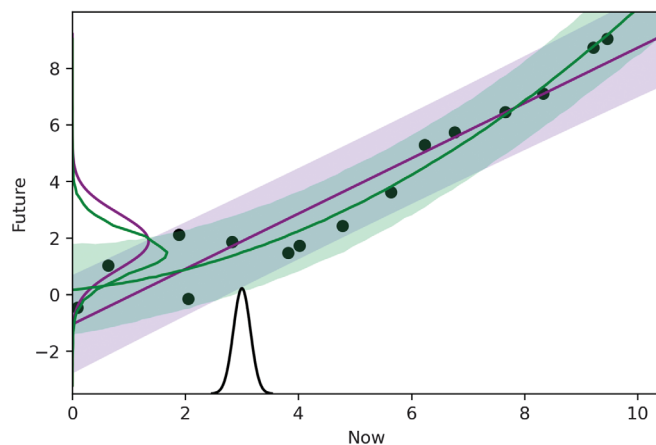

FIG. 6. Using a quadratic relationship decreases the uncertainty in Y compared to a linear fit provided that the observations line up with the shallow section of the quadratic function. Illustration with synthetic data.

can be employed: taking the mean of a set of initial value simulations (Jiménez-de-la-Cuesta and Mauritsen, 2019) using orthogonal distance regression, which takes into account errors in both the dependent and independent variables (Jiménez-de-la-Cuesta and Mauritsen, 2019), or using a hierarchical Bayesian method that assumes that the true independent variable is unknown (latent) and using the realizations to infer this true value simultaneously while performing the regression (Sansom, 2014; Nijsse, Cox, and Williamson, 2020).

Regression confidence intervals also depend on sample size. As climate models often have shared computer code, effective sample size is probably lower than the total number of models (Pennell and Reichler, 2011; Masson and Knutti, 2013; Herger *et al.*, 2018), effectively reducing its significance (Knutti, Masson, and Gettelman, 2013). This commonality can be partially addressed by selecting only one model per modeling center (Cox, Huntingford, and Williamson, 2018; Sansom, Stephenson, and Bracegirdle, 2019), where centers frequently offer multiple versions of alternative spatial resolutions or other small differences in model physics. However, even then effective sample size may be overestimated. As strongly related models in terms of code may have significantly different values for X and Y, it is not always clear how important corrections are (Nijsse, Cox, and Williamson, 2020).

D. Uncertainty from imperfect models

Almost paradoxically, the technique of emergent constraints relies on errors in models. Hence some type of errors are useful, while others simply contribute to widening the confidence interval. Model error comes from parametric and from structural errors; see also Sec. II.A. Parameter error is often examined by PPEs, an ensemble where a single climate model is run repeatedly with systematically varied parameters designed to span a large range of model responses. Care should be taken to include only the physically adequate parameter choices but to not restrict the parameter space too much (Wagman and Jackson, 2018).

Usually, due to the expense of running climate models, an *ad hoc* ensemble of different models is exploited to establish emergent constraints, the so-called ensemble of opportunity. In this case, models should be chosen in such a way that they are still comparable: If half the models contain a nitrogen cycle while the other half does not, they are unlikely to fit on a single regression (Wenzel *et al.*, 2014). Williamson and Sansom (2019) argued that, due to structural uncertainty, the derived regression between X and Y should be regarded not as the real relationship but instead as only informing the real relationship.

E. Combining sources of uncertainty in an EC

Williamson and Sansom (2019) argued on theoretical grounds that emergent constraints should be performed using a Bayesian framework, instead of the more common frequentists' framework using ordinary least squares fitting (Wenzel *et al.*, 2014), in which priors are set for the regression parameters but not for the predictand Y. Bowman *et al.*

(2018) formulated a Bayesian hierarchical statistical framework that combines uncertainty in observations with uncertainty related to the regression itself. An explicit formula for the probability distribution can be approximated incorporating these uncertainties: assuming normality and performing a linear regression with intercept a, slope b, and confidence around the regression of σ_f, the probability density function for the predictand $p(y)$ is computed by integrating the conditional probability function $p(y|x)$ with the observational probability density $p(x)$. This formula for $p(y)$, rewritten in the notation of other statistical frameworks and under Gaussian assumptions, turns into

$$p(y) = \int_{-\infty}^{\infty} p(y|x)p(x)dx$$
$$= \mathcal{N}\left(y|a + bX_{\text{obs}}, \sqrt{\sigma_f^2 + b^2\sigma_O^2}\right). \qquad (20)$$

Here X is the predictor, Y is the predictand, and X_{obs} is the best estimate of the observation. The observational uncertainty is denoted by σ_O.

To better estimate the regression and incorporate internal climate variability explicitly, a second hierarchical model was developed independently by Sansom (2014) and Nijsse, Cox, and Williamson (2020). Internal climate variability is incorporated both as part of the regression and as an additional term in the uncertainty around observations by using all initial value simulations $X_{m,j}$ of each model m and each initial value member j. Here they are formulated using Gaussian distributions

$$X_{m,j} \sim \mathcal{N}(X_m, \sigma_I), \qquad (21)$$

$$Y_m \sim \mathcal{N}(a + bX_m, \sigma_f). \qquad (22)$$

The internal variability is denoted by σ_I and is assumed to be independent of Y and X (and therefore model independent). The probability density function is sampled from estimates of the observed X_{obs} and the parameters of the emergent relationship and gives a similar equation to Eq. (20):

$$p(y) = \mathcal{N}\left(a + b\mathcal{N}\left(X_{\text{obs}}, \sqrt{\sigma_I^2 + \sigma_O^2}\right), \sigma_f\right).$$

Hall *et al.* (2019) queried the extent to which we should trust in only partially confirmed ECs. This can be addressed formally, by inclusion of a researcher's assessment of how reliable a certain EC is. Williamson and Sansom (2019) described a method explicitly incorporating the trust that is put in any particular EC, including for instance by how much the EC changes in different ensembles. They introduced an additional parameter σ_R that represents the uncertainty even after having taken all model information into account. This parameter is set subjectively, judged on the degree of physical trust the researcher has in the EC and features as $p(y) = \mathcal{N}(\beta x, \sigma_f^2 + \sigma_R^2)$, where β is a covariance matrix of the regression parameters, for which priors are also provided with information about possible biases or errors in the regression parameters.

F. Combining multiple constraints

A simple method of combining multiple constraints was used by Brient (2020), who utilized a Gaussian kernel density estimation of a histogram of the best values of previous constraints. To account for the variance, the posterior variance of the different ECs was included in a weighted Gaussian kernel density estimation. This method suffers from multiple drawbacks. Most prominently it cannot benefit from the increased confidence that results from multiple independent constraints. Adding a new constraint with the same mean does not automatically lead to a narrower combined constraint, with the choice of kernel bandwidth remaining subjective. It additionally does not take into account to what extent the different emergent constraints are related to each other.

Bretherton and Caldwell (2020) combined multiple emergent constraints using a multivariate Gaussian PDF, which can be viewed as a form of multilinear regression. Their "method C" (for correlated) includes information about correlations between different emergent relationships. Regularization was applied to deal with the strong collinearity between the emergent relationships. A second "method U" (for uncorrelated) used a smaller subset of ECs, those regarded as confirmed constraints. As collinearity is less important with fewer ECs this method simplifies C by dropping the covariance ECs. Both variants were extended with a transparent method to account for overconfidence in the EC: they scale the ratio of the explained to unexplained variance with a factor $\alpha^2 \leq 1$, thereby reducing all correlation coefficients.

Renoult *et al.* (2020) proposed a simple method to combine independent ECs to create a tighter estimate for Y. Where the regression is normally given as $p(y|x)$, they proposed to instead formulate the statistical model as $p(x|y)$, allowing for a prior on $\pi(y)$ to be integrated into the emergent constraint as

$$p(y) = \int_{-\infty}^{\infty} p(x|y)\pi(y)p(x)dx. \qquad (23)$$

$p(y)$ here is the posterior distribution of a previous EC. To make sure the two ECs are indeed independent, Renoult *et al.* (2020) stated that observations need to be independent and that, insofar as it is possible, the errors in models should also stem from different sources. Their example involved a warm and cold climate state for which temperature change was reconstructed. Temperature change is dominated by different processes in this case, so that model error can be considered independent to first order. This method is not consistent with the other previously described methods. Linear regression is typically not symmetric; regression where X predicts Y, $p(y|x)$ describes a different function than regression where Y predicts X, $p(x|y)$ (Smith, 2009), as illustrated in Fig. 7.

VII. OUTLOOK

Here we give some promising or exciting directions for future EC research.
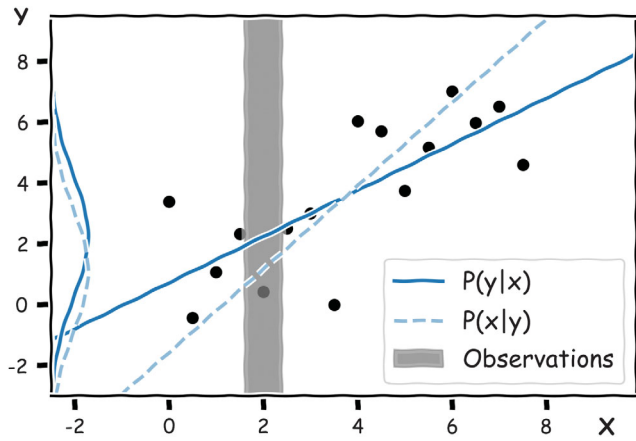
FIG. 7. Standard, ordinary least squares linear regression $p(y|x)$ compared with the reverse regression $p(x|y)$. If the latter were to be used as the relationship on which the emergent constraint is founded, the final constraint has a bias.

A. Key gaps in ECs to date

The collection of ECs described in Sec. V highlight key areas of focus to date, but this also demonstrates where gaps exist, thus highlighting where the potential for EC growth is greatest going forward. For the most part, ECs have been traditionally focused on globally aggregated quantities related to the mean state of climate (climate feedbacks, hydrologic sensitivity, etc.). However, recent applications to more regional features of the climate system are particularly encouraging (such as informing future monsoonal precipitation change over India) (Li *et al.*, 2017). Since these regional applications have the potential for greater policy impact than most globally averaged metrics, we anticipate much more focus in this area going forward. Climate extremes are also of great societal importance, but from Fig. 5 we can see that few studies have attempted to constrain the uncertainty in their future changes. We anticipate a greater focus on these and other higher-order moments of climate statistics going forward. This advance is spurred on by an ever-lengthening satellite record that is now long enough to robustly sample extreme precipitation events. Similarly, improving paleoclimate records offer great potential for constraining emergent relationships pertaining to aspects of the climate system that vary on long timescales. In terms of their Earth system components, proposed constraints are heavily skewed toward atmospheric and biogeochemical fields, while constraints pertaining to oceanography and ice sheets are lacking. The latter is a recent addition to ESMs, but as observational records continue to improve there could be great potential for reducing uncertainty in critical metrics like sea level rise. Finally, most ECs pertain to local relationships, but there are likely many undiscovered remote relationships in the climate system, where the current variability in one region is strongly tied to the future change in another region through atmospheric teleconnections (Rowell, 2019). It is also believed that biases in the position of various climate features may be systematically tied to how the features respond to future forcing (Hall *et al.*, 2019). Utilization of these spatial characteristics of climate for ECs has begun (Kidston and

Gerber, 2010; Simpson and Polvani, 2016), but better detection of these remote constraints likely requires improved analysis techniques (Barnes, Hurrell, and Uphoff, 2019).

B. Targeted model development

An appealing yet underused aspect of ECs is targeted model development for bias reduction in a particular aspect of climate change (Y), although there are dangers to be aware of too; see Sec. IV.G. If an EC has strong physical underpinnings, then we can expect that taking steps to reduce the spread in X will result in a corresponding spread reduction in Y. First, we must assess in detail how the structural and parametric makeup of the ESMs influence the spread in X. Through this analysis the best parametrizations or parameters for simulating X may be uncovered, thus providing guidelines for modeling centers on how they can efficiently reduce bias in X going forward. This type of analysis is encouraged for all well-established ECs but has been completed only for ECs on hydrologic cycle intensification (DeAngelis *et al.*, 2015), the sensitivity of extratropical cloud reflectivity to temperature (Gordon and Klein, 2014), and the snow albedo feedback (Thackeray, Qu, and Hall, 2018). It is also possible that these targeted development activities will lead to a reduction in spread for related attributes of climate change affected by Y.

C. Use of conceptual models as the basis of emergent relationships and understanding of more complex ESMs

In the earlier years of ECs there may have been a tendency to data mine for variables with high correlations within ESM ensembles, potentially resulting in bad ECs; see Sec. II.B. However, data mining is not necessarily bad if it guides the search for independently testable mechanisms that may not have been obvious otherwise (Sec. IV.A).

One way of guarding against fortuitous correlations and hypothesizing new mechanisms is using conceptual, analytically soluble models as the basis of the emergent relationship. Building and solving conceptual models requires one to make assumptions about the real world. These assumptions are often clear, simple to understand, and, ideally, testable against real world and ESM ensemble data (Williamson *et al.*, 2018). Provided that the assumptions survive these tests, the solution to the conceptual model can form the basis of the emergent relationship. This testing of simplified theory not only aids understanding and intuition of what might (or might not) be going on in the complex ESMs but also adds a greater degree of rigor to ECs. There are potential dangers one should be aware of, however: (i) *p* hacking might occur (see Sec. IV.B) and (ii) it might be that a process in the ESM one is trying to constraint is closer to the conceptual model in the spectrum of complexity than it is to reality. The use of conceptual models could therefore be used to justify the relationships arising from slightly less conceptual parametrizations that are implemented in the ESMs.

State-of-the-art ESM spatial resolution has continually increased and more physical and biogeochemical processes have been incorporated, making these models ever more comprehensive representations of Earth as time has gone on. They have become the main theoretical tool in climate

science. However, owing to their ever increasing complexity, their behavior is difficult to understand and predict without the models themselves being run. They are also frequently treated as one-stop shops and oracles for any question about the contemporary or future climate.

In the past, numerical simulation was more limited by computing power and researchers had to think carefully about which bits of the Earth system were important to include in their models to answer a particular question. This led to wider use of conceptual models specific to that question, analytical techniques, and better intuition due to the reduction of model complexity (Budyko, 1969; Hasselmann, 1976; Wigley and Raper, 1990; Saltzman, 2002; Dijkstra, 2013).

Today, there is even more of an opportunity for the top down insights of specific conceptual models to meet and complement the comprehensive, bottom up approach of from state-of-the-art ESMs; there are many more high quality observations from satellites (Yang *et al.*, 2013), ocean floats (Roemmich *et al.*, 2019), and continuous temperature records; the global warming signal has also become clearer over time; and there is also a large archive of past and present ESM simulations. With all the new data, it would be worth revisiting these simple, understandable models. A promising way forward to incorporate all the extra information from ESMs, conceptual models, and observations is the EC approach.

D. Multidimensional ECs and nonlinear emergent relationships

Much of the work on ECs has used just two scalar variables, a predictor X and a response or predictand Y, related linearly. This could be extended to two or more predictors or predictands (Renoult *et al.*, 2020). The predictand space to constrain will become larger, however, and this might be difficult with small ESM ensembles.

Most of the work to date has also assumed a linear emergent relationship, with exceptions being the constraints on ECS using global mean temperature change in Jiménez-de-la-Cuesta and Mauritsen (2019) and Nijsse, Cox, and Williamson (2020). This is also not necessary if theoretical reasoning suggests that a different relation is more suitable.

Equivalent to the use of multiple predictors or predictands, one could imagine using vectors for X and Y rather than scalars if theory suggests such a relationship. These vectors could encode spatial fields of a climate variable (Brown and Caldeira, 2017), spatial normal modes of variance, referred to as empirical orthogonal functions in meteorology (von Storch and Zwiers, 1999) or the leading linear dynamical modes, also known as principal oscillation patterns in meteorology (Hasselmann, 1988; Williamson and Lenton, 2015).

E. Continued improvement of climate projections and impact-led requirements

One prediction we make is that the discovery of new ECs will be led more by those asking questions as to how climate impacts of concern may adjust in the future. Many altered meteorological features expected to have the most detrimental effect on societal safety, livelihoods, and well-being are the magnitude and frequency of extreme weather events. A focus

on extremes may result in ECs that are different in composition than those discovered to date. Rather than a fluctuating and observable contemporary quantity X projecting a "bulk" climate property Y of future importance, a different format is required. That is, summary climate properties of large-scale features (or trends in them), as an observable quantity X, require investigation to determine if they can estimate the future extreme frequency of quantity Y. In addition, a search for such ECs will focus attention on three further challenges. First, attributes of extremes and their changes might show substantial geographical variation. Second, by definition extremes are rare, so there are few data points available for investigation in ESM diagnostics where meteorological values are above thresholds that may be unsafe. Third, the future connection between observable X and extreme feature Y may involve a teleconnection between locations, and possibly also with seasonal inertia.

F. Better understanding of the effects of parametric and structural uncertainty on ECs

Presently, the most commonly used ESM ensembles in climate as well as EC research are MMEs (see Sec. II.A) allowing structural uncertainty and, in more recent MMEs such as the CMIP datasets, initial condition uncertainty to be explored. Parametric uncertainty is also captured to some degree, although not in a systematic way (unlike in PPEs). It is therefore unclear whether all plausible model parametric configurations have been explored. See also Secs. IV.I and VI.D.

The strengths of MMEs and PPEs could be combined in a "superensemble" of PPEs generated from structurally different models. A better assessment of confidence intervals could be made, and questions about the necessity of plausible ESM simulations for good ECs could be answered, for example, is it necessary to filter out unrealistic simulations to strengthen the EC based on an observation unrelated to the predictor? Statistical emulators (Sacks *et al.*, 1989) of ESMs may help give enough effective simulations after filtering unrealistic simulations from a PPE.

G. Machine learning

Use of machine learning (ML) techniques in ECs may prove fruitful in future research. ML (Webb and Copsey, 2011) holds a wide definition but encompasses ECs, particularly supervised-learning-type algorithms. In a supervised ML approach, the algorithm builds a mathematical model from a set of training data that contain both inputs and desired outputs. In the case of ECs, these training data are from an ESM ensemble, with the inputs being the observable X and the outputs being the future projection Y. The mathematical model is f, i.e., $Y = f(X) + \epsilon$. The mathematical model that has been a popular choice in ECs thus far has been $f(X) = aX + b$, where a and b are constants fitted via linear regression.

Phrasing linear regression as an example of supervised ML could be seen to be an overcomplicated generalization. However, viewing ECs within this framework potentially allows one to take advantage of the existing tools within

ML. For example, the mathematical models fitted by the supervised ML algorithms could be more elaborate with more physically or statistically based parameters that give better explanatory power or more realism. The inputs and outputs equally could be more elaborate, with extra spatial dimensions and/or extra variables (Sec. VII.D). This could get as complicated as reproducing the original models that the inputs and outputs derive from. However, one ideally would like to walk the line between the physical plausibility and understandability of the model and explanatory power to derive a minimal description.

One also has to be aware of the amount of data available to fit to a particular mathematical model. Presently, MMEs consist of around 30 ESMs, thus limiting defensible fits of mathematical models to a few parameters. However, one may be able to leverage the extra information in the ESMs spatial and temporal degrees of freedom to increase data size and therefore mathematical model complexity.

H. Building connections to other fields

Though our expertise is in the area of climate and Earth system dynamics, we have found a few examples of EC-like approaches in other fields.

In the closely related field of weather prediction, an approach known as "model output statistics" is used to enhance the quality and relevance of weather forecasts produced by numerical models (Veenhuis, 2013). As models are repeatedly used to make forecasts, an archive of past simulated data accumulate. This archive can be analyzed to diagnose systematic statistical relationships between model output and observed quantities of interest. (An example of a quantity of interest might be temperature at a particular weather station.) When the models are then used to make real-time forecasts, these relationships can be applied to the forecast output, generating values for the quantities of interest. This is similar to the EC approach in that the structure of model biases is ascertained and exploited to produce future predictions that are likely to be more realistic.

We find another example in the field of cosmology, where models have been created to simulate galaxy development. Like climate models, these models have tunable parameters. When the model parameters are varied to produce an ensemble of simulations, statistical relationships among galaxy components across the ensemble can be analyzed. The galaxy model of Terrazas *et al.* (2020) produced strong statistical relationship between the black hole mass and the associated stellar mass; cf. Fig. 7 of that paper. Under certain conditions, observations of the black hole and stellar mass mirror the simulated relationship. This suggests that given an observation of black hole mass for a galaxy without a corresponding stellar mass observation, the model could be used to make a meaningful prediction of stellar mass. This example is similar to ECs in the sense that the intrinsic model (and real world) physics connecting key quantities of interest can be ascertained only through analysis of a multimodel ensemble.

In the field of economics, past forecasts of economic performance using multiple economic models have been analyzed and compared to the actual ensuing economic performance. Felix, Kraeussl, and Stork (2020) found that the skewness of the ensemble forecast's distribution is a strong predictor of economic surprises. The implication is that if a current ensemble of economic models produces a skewed forecast distribution, the likelihood of an economic surprise is elevated. Such an outcome might not be detectable from more conventional predictions, such as the ensemble mean forecast. This example has some similarities to the model output statistics example from weather forecasting: Model biases are ascertained from multiple realizations of past forecast performance and used to improve the current forecast.

This brief survey of a few close cousins to the EC approach reveals that ECs and their variants are most likely to be useful where we are trying to understand complex systems in which experimentation with the real system is not possible or is cumbersome. In such systems we often need to simulate dynamical evolution, initializing from a snapshot in time or at best a relatively short period of observation. Multiple models of the system are possible, and which model is the correct one is unknown. Emergent relationships across model variants are a way to reveal the deeper connections between observable elements of simulations and model outcomes of interest. There are likely many more examples of EC-like research being done on complex systems. These disparate communities appear to be unaware of one another and could benefit from organized efforts to convene in workshops and meetings.

VIII. CONCLUSION

The spread in climate change projections has not been reduced substantially despite major advances in model resolution and process understanding and inclusion over recent decades. The lack of progress represents a disappointment for climate science and hinders society's ability to plan for future impacts. These challenges cannot be overcome by simply increasing the resolution of ESMs and including more physical and biogeochemical processes. A balanced interplay of observations as well as top down, conceptual and bottom up, comprehensive modeling and theory is required to achieve the required progress and scientific understanding. In this review, we have argued that the EC approach offers a promising way to incorporate all of these elements and ultimately reduce key uncertainties in the future climate.

The development of ECs has arisen from the requirement to reduce often substantial inter-ESM differences in projections of climate. Large model differences make adaption planning difficult and risk expenditure of funds set aside to help societies cope with climate change. For those tasked with formulating mitigation plans that will reduce greenhouse gas emissions in order to contain global warming, uncertainty prevents the formation of accurate trajectories of "allowable" fossil fuel burning. In this review, we have shown existing applications of ECs, discussed theoretical issues surrounding them that will likely require further investigation, and characterized a few circumstances in which they may start to fail. Nevertheless, although this seemingly points to more required research, ECs are a key methodology to distill the often conflicting information from climate modeling centers across the world. For this reason, we anticipate that the application of ECs will continue to grow, and results from their application

will play an increasing role in projecting future change, with strong representation in future UN reports by the IPCC.

This review has looked at how ECs have become a standard methodology, used by climate researchers to synthesize substantial inter-ESM prediction differences, into projections of change with lower levels of uncertainty. We have made a critical assessment of the EC technique, and many of the open questions raised here may lead to exciting avenues of research in the years ahead. It may also be true that other research disciplines will find a role for the EC approach. However, it is always worth recalling that the main scientific issue remains pressing, as without substantial reductions in emissions, the climate is expected to change substantially and to a dangerous state for a large fraction of society. For this reason, intense scrutiny of the EC method is highly welcome and appropriate. The relentless need for climate predictions is also likely to trigger the discovery of new ECs, caused instead by an impact-led requirement for information taking precedence over more curiosity-driven investigations.

---

[6]See http://www.agci.org/event/17s2.

REFERENCES

AchutaRao, K. M., B. D. Santer, P. J. Gleckler, K. E. Taylor, D. W. Pierce, T. P. Barnett, and T. M. L. Wigley, 2006, "Variability of ocean heat uptake: Reconciling observations and models," J. Geophys. Res. Oceans **111**, C05019.

Annan, J. D., and J. C. Hargreaves, 2006, "Using multiple observationally-based constraints to estimate climate sensitivity," Geophys. Res. Lett. **33**, L06704.

Barnes, Elizabeth A., James W. Hurrell, and Imme Ebert Uphoff, 2019, "Viewing forced climate patterns through an AI lens," Geophys. Res. Lett. **46**, 13389–13398.

Bell, Thomas L., 1980, "Climate sensitivity from fluctuation dissipation: Some simple model tests," J. Atmos. Sci. **37**, 1700–1707.

Boe, Julien, Alex Hall, and Xin Qu, 2009, "September sea-ice cover in the Arctic Ocean projected to vanish by 2100," Nat. Geosci. **2**, 341–343.

Bond, Gerard, *et al.*, 1992, "Evidence for massive discharges of icebergs into the North Atlantic Ocean during the last glacial period," Nature (London) **360**, 245–249.

Bony, Sandrine, *et al.*, 2006, "How well do we understand and evaluate climate change feedback processes?," J. Clim. **19**, 3445–3482.

Borodina, Aleksandra, Erich M. Fischer, and Reto Knutti, 2017a, "Emergent constraints in climate projections: A case study of changes in high-latitude temperature variability," J. Clim. **30**, 3655–3670.

Borodina, Aleksandra, Erich M. Fischer, and Reto Knutti, 2017b, "Models are likely to underestimate increase in heavy rainfall in the extratropical regions with high rainfall intensity," Geophys. Res. Lett. **44**, 7401–7409.

Bowman, K. W., *et al.*, 2013, "Evaluation of ACCMIP outgoing longwave radiation from tropospheric ozone using TES satellite observations," Atmos. Chem. Phys. **13**, 4057–4072.

Bowman, Kevin W., Noel Cressie, Xin Qu, and Alex Hall, 2018, "A hierarchical statistical framework for emergent constraints: Application to snow-albedo feedback," Geophys. Res. Lett. **45**, 13,050–13,059.

Bracegirdle, Thomas J., and David B. Stephenson, 2012, "Higher precision estimates of regional polar warming by ensemble regression of climate model projections," Clim. Dyn. **39**, 2805–2821.

Bracegirdle, Thomas J., and David B. Stephenson, 2013, "On the robustness of emergent constraints used in multimodel climate change projections of Arctic warming," J. Clim. **26**, 669–678.

Bretherton, Christopher, and Peter Caldwell, 2020, "Combining emergent constraints for climate sensitivity," J. Clim. **33**, 7413–7430.

Brient, F., and T. Schneider, 2016, "Constraints on climate sensitivity from space-based measurements of low-cloud reflection," J. Clim. **29**, 5821–5835.

Brient, F., T. Schneider, Z. H. Tan, S. Bony, X. Qu, and A. Hall, 2016, "Shallowness of tropical low clouds as a predictor of climate models' response to warming," Clim. Dyn. **47**, 433–449.

Brient, Florent, 2020, "Reducing uncertainties in climate projections with emergent constraints: Concepts, examples and prospects," Adv. Atmos. Sci. **37**, 1–15.

Brient, Florent, and Sandrine Bony, 2013, "Interpretation of the positive low-cloud feedback predicted by a climate model under global warming," Clim. Dyn. **40**, 2415–2431.

Brown, Patrick T., and Ken Caldeira, 2017, "Greater future global warming inferred from Earth's recent energy budget," Nature (London) **552**, 45.

Brown, Patrick T., Martin B. Stolpe, and Ken Caldeira, 2018, "Assumptions for emergent constraints," Nature (London) **563**, E1–E3.

Budyko, M. I., 1969, "The effect of solar radiation variations on the climate of the Earth," Tellus **21**, 611–619.

Caldwell, P. M., M. D. Zelinka, and S. A. Klein, 2018, "Evaluating emergent constraints on equilibrium climate sensitivity," J. Clim. **31**, 3921–3942.

Caldwell, Peter M., Christopher S. Bretherton, Mark D. Zelinka, Stephen A. Klein, Benjamin D. Santer, and Benjamin M. Sanderson, 2014, "Statistical significance of climate sensitivity predictors obtained by data mining," Geophys. Res. Lett. **41**, 1803–1808.

Ceppi, Paulo, Florent Brient, Mark D. Zelinka, and Dennis L. Hartmann, 2017, "Cloud feedback mechanisms and their

representation in global climate models," Wiley Interdisc. Rev. Clim. Change **8**, e465.

Chadburn, S. E., E. J. Burke, P. M. Cox, P. Friedlingstein, G. Hugelius, and S. Westermann, 2017, "An observation-based constraint on permafrost loss as a function of global warming," Nat. Clim. Change **7**, 340–344.

Clement, A. C., R. Burgman, and J. R. Norris, 2009, "Observational and model evidence for positive low-level cloud feedback," Science **325**, 460–464.

Collins, M., *et al.*, 2013, "Long-term climate change: Projections, commitments and irreversibility," in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. F. Stocker, D. Qin D, G. K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex V, and P. M. Midgley (Cambridge University Press, Cambridge, England).

Collins, Matthew, Ben B. B. Booth, B. Bhaskaran, Glen R. Harris, James M. Murphy, David M. H. Sexton, and Mark J. Webb, 2011, "Climate model errors, feedbacks and forcings: A comparison of perturbed physics and multi-model ensembles," Clim. Dyn. **36**, 1737–1766.

Covey, C., *et al.*, 2000, "The seasonal cycle in coupled ocean-atmosphere general circulation models," Clim. Dyn. **16**, 775–787.

Cowtan, Kevin, and Robert G. Way, 2014, "Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends," Q. J. R. Meteorol. Soc. **140**, 1935–1944.

Cox, P. M., R. A. Betts, M. Collins, P. P. Harris, C. Huntingford, and C. D. Jones, 2004, "Amazonian forest dieback under climate-carbon cycle projections for the 21st century," Theor. Appl. Climatol. **78**, 137–156.

Cox, P. M., R. A. Betts, C. D. Jones, S. A. Spall, and I. J. Totterdell, 2000, "Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model," Nature (London) **408**, 184–187.

Cox, Peter M., 2019, "Emergent constraints on climate-carbon cycle feedbacks," Curr. Clim. Change Rep. **5**, 275–281.

Cox, Peter M., Chris Huntingford, and Mark S. Williamson, 2018, "Emergent constraint on equilibrium climate sensitivity from global temperature variability," Nature (London) **553**, 319.

Cox, Peter M., David Pearson, Ben B. Booth, Pierre Friedlingstein, Chris Huntingford, Chris D. Jones, and Catherine M. Luke, 2013, "Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability," Nature (London) **494**, 341.

Cox, Peter M., Mark S. Williamson, Femke J. M. M. Nijsse, and Chris Huntingford, 2018, "Cox *et al.* reply," Nature (London) **563**, E10–E15.

Cubasch, U., G. A. Meehl, G. J. Boer, R. J. Stouffer, M. Dix, A. Noda, C. A. Senior, S. Raper, and K. S. Yap, 2001, "Projections of future climate change," in *Climate Change 2001: The Scientific Basis. Contribution of WG1 to the Third Assessment Report of the IPCC (TAR)* (Cambridge University Press, Cambridge, England), pp. 525–582.

DeAngelis, Anthony M., Xin Qu, Mark D. Zelinka, and Alex Hall, 2015, "An observational radiative constraint on hydrologic cycle intensification," Nature (London) **528**, 249–253.

Dijkstra, Henk A., 2013, *Nonlinear Climate Dynamics* (Cambridge University Press, Cambridge, England).

Donat, Markus G., Andrew J. Pitman, and Oliver Angélil, 2018, "Understanding and reducing future uncertainty in midlatitude daily heat extremes via land surface feedback constraints," Geophys. Res. Lett. **45**, 10,627–10,636.

Donner, Leo J., *et al.*, 2011, "The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL global coupled model CM3," J. Clim. **24**, 3484–3519.

Drijfhout, Sybren, Sebastian Bathiany, Claudie Beaulieu, Victor Brovkin, Martin Claussen, Chris Huntingford, Marten Scheffer, Giovanni Sgubin, and Didier Swingedouw, 2015, "Catalogue of abrupt shifts in intergovernmental panel on climate change climate models," Proc. Natl. Acad. Sci. U.S.A. **112**, E5777–E5786.

Durack, Paul J., Peter J. Gleckler, Felix W. Landerer, and Karl E. Taylor, 2014, "Quantifying underestimates of long-term upper-ocean warming," Nat. Clim. Change **4**, 999–1005.

Einstein, A., 1905, "Über die von der molekularkinetischen Theorie der Würme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen," Ann. Phys. (Berlin) **322**, 549–560.

Epstein, Edward S., 1969, "Stochastic dynamic prediction," Tellus **21**, 739–759.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016, "Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization," Geosci. Model Dev. **9**, 1937–1958.

Eyring, Veronika, *et al.*, 2019, "Taking climate model evaluation to the next level," Nat. Clim. Change **9**, 102–110.

Fasullo, J. T., and K. E. Trenberth, 2012, "A less cloudy future: The role of subtropical subsidence in climate sensitivity," Science **338**, 792–794.

Feldstein, Steven B., 2000, "The timescale, power spectra, and climate noise properties of teleconnection patterns," J. Clim. **13**, 4430–4440.

Felix, Luiz F. F., Roman Kraeussl, and Philip A. Stork, 2020, "Predictable biases in macroeconomic forecasts and their impact across asset classes," https://doi.org/10.2139/ssrn.3008976.

Fläschner, Dagmar, Thorsten Mauritsen, and Bjorn Stevens, 2016, "Understanding the intermodel spread in global-mean hydrological sensitivity," J. Clim. **29**, 801–817.

Forster, Piers M., Amanda C. Maycock, Christine M. McKenna, and Christopher J. Smith, 2020, "Latest climate models confirm need for urgent mitigation," Nat. Clim. Change **10**, 7–10.

Frost, Chris, and Simon G. Thompson, 2000, "Correcting for regression dilution bias: Comparison of methods for a single predictor variable," J. R. Stat. Soc. Ser. A Stat. Soc. **163**, 173–189.

Gettelman, A., and S. C. Sherwood, 2016, "Processes responsible for cloud feedback," Curr. Clim. Change Rep. **2**, 179–189.

Ghil, Michael, and Valerio Lucarini, 2020, "The physics of climate variability and climate change," Rev. Mod. Phys. **92**, 035002.

Gordon, N. D., and S. A. Klein, 2014, "Low-cloud optical depth feedback in climate models," J. Geophys. Res. Atmos. **119**, 6052–6065.

Gregory, J. M., 2000, "Vertical heat transports in the ocean and their effect on time-dependent climate change," Clim. Dyn. **16**, 501–515.

Grise, Kevin M., Lorenzo M. Polvani, and John T. Fasullo, 2015, "Reexamining the relationship between climate sensitivity and the Southern Hemisphere radiation budget in CMIP models," J. Clim. **28**, 9298–9312.

Hall, A., and X. Qu, 2006, "Using the current seasonal cycle to constrain snow albedo feedback in future climate change," Geophys. Res. Lett. **33**, L03502.

Hall, Alex, 2004, "The role of surface albedo feedback in climate," J. Clim. **17**, 1550–1568.

Hall, Alex, Peter Cox, Chris Huntingford, and Stephen Klein, 2019, "Progressing emergent constraints on future climate change," Nat. Clim. Change **9**, 269–278.

Hargreaves, J. C., J. D. Annan, M. Yoshimori, and A. Abe-Ouchi, 2012, "Can the Last Glacial Maximum constrain climate sensitivity?," Geophys. Res. Lett. **39**, L24702.

Harrison, S. P., P. J. Bartlein, K. Izumi, G. Li, J. Annan, J. Hargreaves, P. Braconnot, and M. Kageyama, 2015, "Evaluation of CMIP5 palaeo-simulations to improve climate projections," Nat. Clim. Change **5**, 735–743.

Hasselmann, K., 1976, "Stochastic climate models. Part I. Theory," Tellus **28**, 473–484.

Hasselmann, K., 1988, "PIPs and POPs: The reduction of complex dynamical systems using principal interaction and oscillation patterns," J. Geophys. Res. Atmos. **93**, 11015–11021.

Hawkins, Ed, and Rowan Sutton, 2009, "The potential to narrow uncertainty in regional climate predictions," Bull. Am. Meteorol. Soc. **90**, 1095–1108.

Hawkins, Ed, and Rowan Sutton, 2011, "The potential to narrow uncertainty in projections of regional precipitation change," Clim. Dyn. **37**, 407–418.

Hennermann, Karl 2018, "ERA5 data documentation, ERA5: Uncertainty estimation v14," https://confluence.ecmwf.int/display/CKB/ERA5%3A+uncertainty+estimation.

Herger, N., G. Abramowitz, R. Knutti, O. Angélil, K. Lehmann, and B. M. Sanderson, 2018, "Selecting a climate model subset to optimise key ensemble properties," Earth Syst. Dyn. **9**, 135–151.

Hoffman, F. M., *et al.*, 2014, "Causes and implications of persistent atmospheric carbon dioxide biases in Earth system models," J. Geophys. Res. Biogeosci. **119**, 141–162.

Holland, M. M., and C. M. Bitz, 2003, "Polar amplification of climate change in coupled models," Clim. Dyn. **21**, 221–232.

Huber, Markus, Irina Mahlstein, Martin Wild, John Fasullo, and Reto Knutti, 2011, "Constraints on climate sensitivity from radiation patterns in climate models," J. Clim. **24**, 1034–1052.

Huntingford, Chris, Philip D. Jones, Valerie N. Livina, Timothy M. Lenton, and Peter M. Cox, 2013, "No increase in global temperature variability despite changing regional patterns," Nature (London) **500**, 327–330.

IPCC, 2013, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, Cambridge, England).

Jiménez-de-la-Cuesta, Diego, and Thorsten Mauritsen, 2019, "Emergent constraints on Earth's transient and equilibrium response to doubled CO_2 from post-1970s global warming," Nat. Geosci. **12**, 902–905.

Johnson, J. B., 1928, "Thermal agitation of electricity in conductors," Phys. Rev. **32**, 97–109.

Jones, C. D., *et al.*, 2016, "C4MIP—The coupled climate-carbon cycle model intercomparison project: Experimental protocol for CMIP6," Geosci. Model Dev. **9**, 2853–2880.

Kamae, Y., H. Shiogama, M. Watanabe, T. Ogura, T. Yokohata, and M. Kimoto, 2016, "Lower-tropospheric mixing as a constraint on cloud feedback in a multiparameter multiphysics ensemble," J. Clim. **29**, 6259–6275.

Kay, J. E., *et al.*, 2015, "The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability," Bull. Am. Meteorol. Soc. **96**, 1333–1349.

Kessler, A., and J. Tjiputra, 2016, "The Southern Ocean as a constraint to reduce uncertainty in future ocean carbon sinks," Earth Syst. Dyn. **7**, 295–312.

Kharin, V. V., F. W. Zwiers, X. Zhang, and M. Wehner, 2013, "Changes in temperature and precipitation extremes in the CMIP5 ensemble," Clim. Change **119**, 345–357.

Kidston, J., and E. P. Gerber, 2010, "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. **37**, L09708.

Kirk-Davidoff, D. B., 2009, "On the diagnosis of climate sensitivity using observations of fluctuations," Atmos. Chem. Phys. **9**, 813–822.

Klein, Stephen A., and Alex Hall, 2015, "Emergent constraints for cloud feedbacks," Curr. Clim. Change Rep. **1**, 276–287.

Klein, Stephen A., Alex Hall, Joel R. Norris, and Robert Pincus, 2017, "Low-cloud feedbacks from cloud-controlling factors: A review," Surv. Geophys. **38**, 1307–1329.

Klein, Stephen A., Yuying Zhang, Mark D. Zelinka, Robert Pincus, James Boyle, and Peter J. Gleckler, 2013, "Are climate model simulations of clouds improving? An evaluation using the ISCCP simulator," J. Geophys. Res. Atmos. **118**, 1329–1342.

Knutti, R., D. Masson, and A. Gettelman, 2013, "Climate model genealogy: Generation CMIP5 and how we got there," Geophys. Res. Lett. **40**, 1194–1199.

Knutti, R., G. A. Meehl, M. R. Allen, and D. A. Stainforth, 2006, "Constraining climate sensitivity from the seasonal cycle in surface temperature," J. Clim. **19**, 4224–4233.

Knutti, Reto, Maria A. A. Rugenstein, and Gabriele C. Hegerl, 2017, "Beyond equilibrium climate sensitivity," Nat. Geosci. **10**, 727–736.

Kubo, R., 1966, "The fluctuation-dissipation theorem," Rep. Prog. Phys. **29**, 255.

Kwiatkowski, Lester, Laurent Bopp, Olivier Aumont, Philippe Ciais, Peter M. Cox, Charlotte Laufkötter, Yue Li, and Roland Séférian, 2017, "Emergent constraints on projections of declining primary production in the tropical oceans," Nat. Clim. Change **7**, 355.

Lambert, F. Hugo, Glen R. Harris, Matthew Collins, James M. Murphy, David M. H. Sexton, and Ben B. B. Booth, 2013, "Interactions between perturbations to different earth system components simulated by a fully-coupled climate model," Clim. Dyn. **41**, 3055–3072.

Lawrence, David M., Andrew G. Slater, and Sean C. Swenson, 2012, "Simulation of present-day and future permafrost and seasonally frozen ground conditions in CCSM4," J. Clim. **25**, 2207–2225.

Lehner, Flavio, Andrew W. Wood, Julie A. Vano, David M. Lawrence, Martyn P. Clark, and Justin S. Mankin, 2019, "The potential to reduce uncertainty in regional runoff projections from climate models," Nat. Clim. Change **9**, 926–933.

Leith, C. E., 1974, "Theoretical skill of Monte Carlo forecasts," Mon. Weather Rev. **102**, 409–418.

Leith, C. E., 1975, "Climate response and fluctuation dissipation," J. Atmos. Sci. **32**, 2022–2026.

Lenton, T. M., H. Held, E. Kriegler, J. W. Hall, W. Lucht, S. Rahmstorf, and H. J. Schellnhuber, 2008, "Tipping elements in the Earth's climate system," Proc. Natl. Acad. Sci. U.S.A. **105**, 1786–1793.

Li, G., S. P. Xie, C. He, and Z. S. Chen, 2017, "Western Pacific emergent constraint lowers projected increase in Indian summer monsoon rainfall," Nat. Clim. Change **7**, 708.

Lian, Xu, *et al.*, 2018, "Partitioning global land evapotranspiration using CMIP5 models constrained by observations," Nat. Clim. Change **8**, 640–646.

Lin, Yanluan, Wenhao Dong, Minghua Zhang, Yuanyu Xie, Wei Xue, Jianbin Huang, and Yong Luo, 2017, "Causes of model dry and warm bias over central U.S. and impact on climate projections," Nat. Commun. **8**, 881.

Lipat, Bernard R., George Tselioudis, Kevin M. Grise, and Lorenzo M. Polvani, 2017, "CMIP5 models' shortwave cloud radiative response and climate sensitivity linked to the climatological Hadley cell extent," Geophys. Res. Lett. **44**, 5739–5748.

Liu, Jiping, Mirong Song, Radley M. Horton, and Yongyun Hu, 2013, "Reducing spread in climate model projections of a September ice-free Arctic," Proc. Natl. Acad. Sci. U.S.A. **110**, 12571.

Lutsko, N. J., and K. Takahashi, 2018, "What can the internal variability of CMIP5 models tell us about their climate sensitivity?," J. Clim. **31**, 5051–5069.

Maher, Nicola, *et al.*, 2019, "The Max Planck Institute Grand Ensemble: Enabling the exploration of climate system variability," J. Adv. Model. Earth Syst. **11**, 2050–2069.

Majda, A. J., R. Abramov, and B. Gershgorin, 2010, "High skill in low-frequency climate response through fluctuation dissipation theorems despite structural instability," Proc. Natl. Acad. Sci. U.S.A. **107**, 581–586.

Manabe, Syukuro, and Richard T. Wetherald, 1975, "The effects of doubling the CO_2 concentration on the climate of a general circulation model," J. Atmos. Sci. **32**, 3–15.

Masson, David, and Reto Knutti, 2013, "Predictor screening, calibration, and observational constraints in climate model ensembles: An illustration using climate sensitivity," J. Clim. **26**, 887–898.

Massonnet, F., T. Fichefet, H. Goosse, C. M. Bitz, G. Philippon-Berthier, M. M. Holland, and P. Y. Barriat, 2012, "Constraining projections of summer Arctic sea ice," Cryosphere **6**, 1383–1394.

Meehl, G. A., C. Covey, T. Delworth, M. Latif, B. McAvaney, J. F. B. Mitchell, R. J. Stouffer, and K. E. Taylor, 2007, "The WCRP CMIP3 multimodel dataset," Bull. Am. Meteorol. Soc. **88**, 1383–1394.

Meehl, G. A., *et al.*, 2007, "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 S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller (Cambridge University Press, Cambridge, England).

Meinshausen, Malte, *et al.*, 2011, "The RCP greenhouse gas concentrations and their extensions from 1765 to 2300," Clim. Change **109**, 213.

Meredith, M., *et al.*, 2019, "Polar regions," in *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*, edited by H.-O. Pörtner *et al.* (in press), Chap. 3, pp. 203–320.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996, "The ECMWF ensemble prediction system: Methodology and validation," Q. J. R. Meteorol. Soc. **122**, 73–119.

Moss, Richard H., *et al.*, 2010, "The next generation of scenarios for climate change research and assessment," Nature (London) **463**, 747–756.

Mudryk, Lawrence, María Santolaria-Otín, Gerhard Krinner, Martin Ménégoz, Chris Derksen, Claire Brutel-Vuilmet, Mike Brady, and Richard Essery, 2020, "Historical Northern Hemisphere snow cover trends and projected changes in the CMIP6 multi-model ensemble," Cryosphere **14**, 2495–2514.

Murphy, James M., David M. H. Sexton, David N. Barnett, Gareth S. Jones, Mark J. Webb, Matthew Collins, and David A. Stainforth, 2004, "Quantification of modelling uncertainties in a large ensemble of climate change simulations," Nature (London) **430**, 768–772.

Mystakidis, Stefanos, Sonia I. Seneviratne, Nicolas Gruber, and Edouard L. Davin, 2017, "Hydrological and biogeochemical constraints on terrestrial carbon cycle feedbacks," Environ. Res. Lett. **12**, 014009.

Nigam, S., and S. Baxter, 2015, "General circulation of the atmosphere—Teleconnections," in *Encyclopedia of Atmospheric Sciences*, 2nd ed., edited by Gerald R. North, John Pyle, and Fuqing Zhang (Academic Press, Oxford), pp. 90–109.

Nijsse, Femke J. M. M., Peter M. Cox, Chris Huntingford, and Mark S. Williamson, 2019, "Decadal global temperature variability increases strongly with climate sensitivity," Nat. Clim. Change **9**, 598–601.

Nijsse, Femke J. M. M., and Henk A. Dijkstra, 2018, "A mathematical approach to understanding emergent constraints," Earth Syst. Dyn. **9**, 999–1012.

Nijsse, F. J. M. M., P. M. Cox, and M. S. Williamson, 2020, "Emergent constraints on transient climate response (TCR) and equilibrium climate sensitivity (ECS) from historical warming in CMIP5 and CMIP6 models," Earth Syst. Dyn. **11**, 737–750.

Notz, Dirk, *et al.* (SIMIP Community), 2020, "Arctic sea ice in CMIP6," Geophys. Res. Lett. **47**, e2019GL086749.

Nuzzo, Regina, 2014, "Statistical errors: p values, the 'gold standard' of statistical validity, are not as reliable as many scientists assume," Nature (London) **506**, 150–152.

Nyquist, H., 1928, "Thermal agitation of electric charge in conductors," Phys. Rev. **32**, 110–113.

O'Gorman, Paul A., 2012, "Sensitivity of tropical precipitation extremes to climate change," Nat. Geosci. **5**, 697–700.

Palmer, Tim, 2019, "The ECMWF ensemble prediction system: Looking back (more than) 25 years and projecting forward 25 years," Q. J. R. Meteorol. Soc. **145**, 12–24.

Pendergrass, A. G., 2020, "The global-mean precipitation response to CO_2-induced warming in CMIP6 models," Geophys. Res. Lett. **47**, e2020GL089964.

Pendergrass, Angeline G., and Dennis L. Hartmann, 2014, "The atmospheric energy constraint on global-mean precipitation change," J. Clim. **27**, 757–768.

Pennell, C., and T. Reichler, 2011, "On the effective number of climate models," J. Clim. **24**, 2358–2367.

Po-Chedley, Stephen, Cristian Proistosescu, Kyle C. Armour, and Benjamin D. Santer, 2018, "Climate constraint reflects forced signal," Nature (London) **563**, E6–E9.

Po-Chedley, Stephen, Mark D. Zelinka, Nadir Jeevanjee, Tyler J. Thorsen, and Benjamin D. Santer, 2019, "Climatology explains intermodel spread in tropical upper tropospheric cloud and relative humidity response to greenhouse warming," Geophys. Res. Lett. **46**, 13399–13409.

Qu, X., and A. Hall, 2014, "On the persistent spread in snow-albedo feedback," Clim. Dyn. **42**, 69–81.

Qu, X., A. Hall, S. A. Klein, and A. M. DeAngelis, 2015, "Positive tropical marine low-cloud cover feedback inferred from cloud-controlling factors," Geophys. Res. Lett. **42**, 7767–7775.

Qu, Xin, and Alex Hall, 2007, "What controls the strength of snow-albedo feedback?," J. Clim. **20**, 3971–3981.

Qu, Xin, Alex Hall, Anthony M. DeAngelis, Mark D. Zelinka, Stephen A. Klein, Hui Su, Baijun Tian, and Chengxing Zhai, 2018, "On the emergent constraints of climate sensitivity," J. Clim. **31**, 863–875.

Qu, Xin, Alex Hall, Stephen A. Klein, and Peter M. Caldwell, 2014, "On the spread of changes in marine low cloud cover in climate model simulations of the 21st century," Clim. Dyn. **42**, 2603–2626.

Renoult, Martin, James Douglas Annan, Julia Catherine Hargreaves, Navjit Sagoo, Clare Flynn, Marie-Luise Kapsch, Uwe Mikolajewicz, Rumi Ohgaito, and Thorsten Mauritsen, 2020, "A Bayesian framework for emergent constraints: Case studies of climate sensitivity with PMIP," Clim. Past **16**, 1715–1735.

Riahi, Keywan, *et al.*, 2017, "The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview," Global Environ. Change **42**, 153–168.

Roemmich, Dean, *et al.*, 2019, "On the future of Argo: A global, full-depth, multi-disciplinary array," Front. Mar. Sci. **6**, 439.

Rowell, David P., 2019, "An observational constraint on CMIP5 projections of the East African long rains and southern Indian Ocean warming," Geophys. Res. Lett. **46**, 6050–6058.

Rowell, David P., and Robin Chadwick, 2018, "Causes of the uncertainty in projections of tropical terrestrial rainfall change: East Africa," J. Clim. **31**, 5977–5995.

Rugenstein, Maria, *et al.*, 2020, "Equilibrium climate sensitivity estimated by equilibrating climate models," Geophys. Res. Lett. **47**, e2019GL083898.

Rypdal, Martin, Hege-Beate Fredriksen, Kristoffer Rypdal, and Rebekka J. Steene, 2018, "Emergent constraints on climate sensitivity," Nature (London) **563**, E4–E5.

Sacks, Jerome, William J. Welch, Toby J. Mitchell, and Henry P. Wynn, 1989, "Design and analysis of computer experiments," Stat. Sci. **4**, 409–423.

Saltzman, B., 2002, *Dynamical Paleoclimatology: Generalized theory of global climate change* (Academic Press, New York).

Sansom, Philip G., David B. Stephenson, and Thomas J. Bracegirdle, 2019, "On constraining projections of future climate using observations and simulations from multiple climate models," arXiv: 1711.04139v3.

Sansom, Philip George, 2014, "Statistical methods for quantifying uncertainty in climate projections from ensembles of climate models," Ph.D. thesis (University of Exeter).

Schekochihin, A. A., 2015, "University of Oxford 2nd year undergraduate physics: Lectures on kinetic theory of gases and statistical physics," http://www-thphys.physics.ox.ac.uk/people/AlexanderSchekochihin/A1/2014/A1LectureNotes.pdf.

Schlund, M., A. Lauer, P. Gentine, S. C. Sherwood, and V. Eyring, 2020, "Emergent constraints on equilibrium climate sensitivity in CMIP5: Do they hold for CMIP6?," Earth Syst. Dyn. **11**, 1233–1258.

Schmidt, G. A., *et al.*, 2014, "Using palaeo-climate comparisons to constrain future projections in CMIP5," Clim. Past **10**, 221–250.

Schwartz, Stephen E., 2007, "Heat capacity, time constant, and sensitivity of earth's climate system," J. Geophys. Res. Atmos. **112**, D24S05.

Selten, F. M., R. Bintanja, R. Vautard, and B. J. J. M. van den Hurk, 2020, "Future continental summer warming constrained by the present-day seasonal cycle of surface hydrology," Sci. Rep. **10**, 4721.

Sgubin, Giovanni, Didier Swingedouw, Sybren Drijfhout, Yannick Mary, and Amine Bennabi, 2017, "Abrupt cooling over the North Atlantic in modern climate models," Nat. Commun. **8**, 14375.

Sherwood, S., *et al.*, 2020, "An assessment of Earth's climate sensitivity using multiple lines of evidence," Rev. Geophys. **58**, e2019RG000678.

Sherwood, S. C., S. Bony, and J. L. Dufresne, 2014, "Spread in model climate sensitivity traced to atmospheric convective mixing," Nature (London) **505**, 37–42.

Siler, N., S. Po-Chedley, and C. S. Bretherton, 2018, "Variability in modeled cloud feedback tied to differences in the climatological spatial pattern of clouds," Clim. Dyn. **50**, 1209–1220.

Simpson, I. R., and L. M. Polvani, 2016, "Revisiting the relationship between jet position, forced response, and annular mode variability in the southern midlatitudes," Geophys. Res. Lett. **43**, 2896–2903.

Slater, Andrew G., and David M. Lawrence, 2013, "Diagnosing present and future permafrost from climate models," J. Clim. **26**, 5608–5623.

Smith, Richard J., 2009, "Use and misuse of the reduced major axis for line-fitting," Am. J. Phys. Anthropol. **140**, 476–486.

Stainforth, D. A., *et al.*, 2005, "Uncertainty in predictions of the climate response to rising levels of greenhouse gases," Nature (London) **433**, 403–406.

Stocker, Thomas F., and Daniel G. Wright, 1991, "Rapid transitions of the ocean's deep circulation induced by changes in surface water fluxes," Nature (London) **351**, 729–732.

Stroeve, Julienne C., Vladimir Kattsov, Andrew Barrett, Mark Serreze, Tatiana Pavlova, Marika Holland, and Walter N. Meier, 2012, "Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations," Geophys. Res. Lett. **39**, L16502.

Su, H., J. H. Jiang, C. X. Zhai, T. J. Shen, J. D. Neelin, G. L. Stephens, and Y. L. Yung, 2014, "Weakening and strengthening structures in the hadley circulation change under global warming and implications for cloud response and climate sensitivity," J. Geophys. Res. Atmos. **119**, 5787–5805.

Taylor, Karl E., Ronald J. Stouffer, and Gerald A. Meehl, 2012, "An overview of CMIP5 and the experiment design," Bull. Am. Meteorol. Soc. **93**, 485–498.

Terhaar, Jens, Lester Kwiatkowski, and Laurent Bopp, 2020, "Emergent constraint on Arctic Ocean acidification in the twenty-first century," Nature (London) **582**, 379–383.

Terrazas, Bryan A., *et al.*, 2020, "The relationship between black hole mass and galaxy properties: Examining the black hole feedback model in IllustrisTNG," Mon. Not. R. Astron. Soc. **493**, 1888–1906.

Tett, S. F. B., D. J. Rowlands, M. J. Mineter, and C. Cartis, 2013, "Can top-of-atmosphere radiation measurements constrain climate predictions? Part II: Climate sensitivity," J. Clim. **26**, 9367–9383.

Teufel, B., and L. Sushama, 2019, "Abrupt changes across the Arctic permafrost region endanger northern development," Nat. Clim. Change **9**, 858–862.

Thackeray, C. W., X. Qu, and A. Hall, 2018, "Why do models produce spread in snow albedo feedback?," Geophys. Res. Lett. **45**, 6223–6231.

Thackeray, Chad W., and Christopher G. Fletcher, 2016, "Snow albedo feedback: Current knowledge, importance, outstanding issues and future directions," Prog. Phys. Geogr. **40**, 392–408.

Thackeray, Chad W., and Alex Hall, 2019, "An emergent constraint on future Arctic sea-ice albedo feedback," Nat. Clim. Change **9**, 972–978.

Thackeray, Chad W., Alex Hall, Mark D. Zelinka, and Christopher G. Fletcher, 2021, "Assessing prior emergent constraints on surface albedo feedback in CMIP6," J. Clim. **34**, 3889–3905.

Thomas, R. Quinn, E. N. Jack Brookshire, and Stefan Gerber, 2015, "Nitrogen limitation on land: How can it occur in Earth system models?," Global Change Biol. **21**, 1777–1793.

Tian, Baijun, 2015, "Spread of model climate sensitivity linked to double-intertropical convergence zone bias," Geophys. Res. Lett. **42**, 4133–4141.

Tokarska, Katarzyna B., Martin B. Stolpe, Sebastian Sippel, Erich M. Fischer, Christopher J. Smith, Flavio Lehner, and Reto Knutti, 2020, "Past warming trend constrains future warming in CMIP6 models," Sci. Adv. **6**, eaaz9549.

Toniazzo, T., J. M. Gregory, and P. Huybrechts, 2004, "Climatic impact of a Greenland deglaciation and its possible irreversibility," J. Clim. **17**, 21–33.

Toth, Zoltan, and Eugenia Kalnay, 1993, "Ensemble forecasting at NMC: The generation of perturbations," Bull. Am. Meteorol. Soc. **74**, 2317–2330.

Trenberth, K. E., and J. T. Fasullo, 2010, "Simulation of present-day and twenty-first-century energy budgets of the southern oceans," J. Clim. **23**, 440–454.

Trenberth, Kevin E., 1984, "Some effects of finite sample size and persistence on meteorological statistics. Part I: Autocorrelations," Mon. Weather Rev. **112**, 2359–2368.

Uhlenbeck, G. E., and L. S. Ornstein, 1930, "On the theory of the Brownian motion," Phys. Rev. **36**, 823–841.

Veenhuis, Bruce A., 2013, "Spread calibration of ensemble MOS forecasts," Mon. Weather Rev. **141**, 2467–2482.

Volodin, E. M., 2008, "Methane cycle in the INM RAS climate model," Izv. Atmos. Oceanic Phys. **44**, 153–159.

von Storch, H., and F. W. Zwiers, 1999, *Statistical Analysis in Climate Research* (Cambridge University Press, Cambridge, England).

Wagman, B. M., and C. S. Jackson, 2018, "A test of emergent constraints on cloud feedback and climate sensitivity using a calibrated single-model ensemble," J. Clim. **31**, 7515–7532.

Wang, Jun, Ning Zeng, Yimin Liu, and Qing Bao, 2014, "To what extent can interannual CO_2 variability constrain carbon cycle sensitivity to climate change in CMIP5 Earth system models?," Geophys. Res. Lett. **41**, 3535–3544.

Watanabe, Masahiro, Youichi Kamae, Hideo Shiogama, Anthony M. DeAngelis, and Kentaroh Suzuki, 2018, "Low clouds link equilibrium climate sensitivity to hydrological sensitivity," Nat. Clim. Change **8**, 901–906.

Webb, A. R., and K. D. Copsey, 2011, *Statistical Pattern Recognition*, 3rd ed. (Wiley, New York).

Wenzel, Sabrina, Peter M. Cox, Veronika Eyring, and Pierre Friedlingstein, 2014, "Emergent constraints on climate-carbon cycle feedbacks in the CMIP5 Earth system models," J. Geophys. Res. Biogeosci. **119**, 794–807.

Wenzel, Sabrina, Peter M. Cox, Veronika Eyring, and Pierre Friedlingstein, 2016, "Projected land photosynthesis constrained by changes in the seasonal cycle of atmospheric CO_2," Nature (London) **538**, 499–501.

Wigley, T. M. L., and S. C. B. Raper, 1990, "Natural variability of the climate system and detection of the greenhouse effect," Nature (London) **344**, 324–327.

Williamson, Daniel B., and Philip G. Sansom, 2019, "How are emergent constraints quantifying uncertainty and what do they leave behind?," Bull. Am. Meteorol. Soc. **100**, 2571–2588.

Williamson, M. S., M. Collins, S. Drijfhout, R. Kahana, J. V. Mecking, and T. M. Lenton, 2018, "Effect of AMOC collapse on ENSO in a high resolution general circulation model," Clim. Dyn. **50**, 2537–2552.

Williamson, M. S., and T. M. Lenton, 2015, "Detection of bifurcations in noisy coupled systems from multiple time series," Chaos **25**, 036407.

Williamson, Mark S., Peter M. Cox, and Femke J. M. M. Nijsse, 2019, "Theoretical foundations of emergent constraints: Relationships between climate sensitivity and global temperature variability in conceptual models," Dyn. Stat. Clim. Syst. **3**, dzy006.

Winkler, Alexander J., Ranga B. Myneni, Georgii A. Alexandrov, and Victor Brovkin, 2019, "Earth system models underestimate carbon fixation by plants in the high latitudes," Nat. Commun. **10**, 885.

Winkler, Alexander J., Ranga B. Myneni, and Victor Brovkin, 2019, "Investigating the applicability of emergent constraints," Earth Syst. Dyn. **10**, 501–523.

Yang, Hui, Shilong Piao, Chris Huntingford, Philippe Ciais, Yue Li, Tao Wang, Shushi Peng, Yuting Yang, Dawen Yang, and Jinfeng Chang, 2018, "Changing the retention properties of catchments and their influence on runoff under climate change," Environ. Res. Lett. **13**, 094019.

Yang, Jun, Peng Gong, Rong Fu, Minghua Zhang, Jingming Chen, Shunlin Liang, Bing Xu, Jiancheng Shi, and Robert Dickinson, 2013, "The role of satellite remote sensing in climate change studies," Nat. Clim. Change **3**, 875–883.

Yokohata, Tokuta, Mark J. Webb, Matthew Collins, Keith D. Williams, Masakazu Yoshimori, Julia C. Hargreaves, and James D. Annan, 2010, "Structural similarities and differences in climate responses to CO_2 increase between two perturbed physics ensembles," J. Clim. **23**, 1392–1410.

Zelinka, Mark D., Timothy A. Myers, Daniel T. McCoy, Stephen Po-Chedley, Peter M. Caldwell, Paulo Ceppi, Stephen A. Klein, and Karl E. Taylor, 2020, "Causes of higher climate sensitivity in CMIP6 models," Geophys. Res. Lett. **47**, e2019GL085782.

Zelinka, Mark D., Chen Zhou, and Stephen A. Klein, 2016, "Insights from a refined decomposition of cloud feedbacks," Geophys. Res. Lett. **43**, 9259–9269.

Zhai, C. X., J. H. Jiang, and H. Su, 2015, "Long-term cloud change imprinted in seasonal cloud variation: More evidence of high climate sensitivity," Geophys. Res. Lett. **42**, 8729–8737.

Zhang, Nien Fan, 2006, "Calculation of the uncertainty of the mean of autocorrelated measurements," Metrologia **43**, S276.

Zhao, Ming, *et al.*, 2016, "Uncertainty in model climate sensitivity traced to representations of cumulus precipitation microphysics," J. Clim. **29**, 543–560.

Zhou, Chen, Mark D. Zelinka, Andrew E. Dessler, and Stephen A. Klein, 2015, "The relationship between interannual and long-term cloud feedbacks," Geophys. Res. Lett. **42**, 10,463–10,469.