UNIVERSITY OF EXETER

Theory-based emergent constraints on climate sensitivities

Submitted by
Femke Joanne Myrthe Milene Nijsse
to the University of Exeter as a thesis for the degree of
Doctor of Philosophy in Mathematics,
October 2020.

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that any material that has previously been submitted and approved for the award of a degree by this or any other University has been acknowledged.

Signed: 

Abstract

Earth system models show a wide variation of possible futures under climate change. To develop appropriate policy for curbing global warming and adapting to unavoidable change, better understanding of the climate system is crucial. One of the approaches to reduce uncertainty in climate models is the identification of emergent constraints. These are physically plausible empirical relationships between a particular simulated characteristic of the current climate versus future climate change from an ensemble of climate models, which can be exploited to reduce uncertainty using current observations.

This thesis discusses various interpretations of this technique and includes a comparison with other methods that combine models and observations in climate. A mathematical theory based on linear response theory is developed for an important subset of constraints showing how nonlinear relationships appear from an interplay of system and forcing time scales. Several statistical issues are examined, such as the best way to deal with internal variability.

This theory is applied to three emergent relationships. Decadal climate variability and climate sensitivity are found to be related in both conceptual climate models and in CMIP5 climate model ensemble. This results in compound risk: significant temperature surges on top of the long-term trend are more likely if climate sensitivity is high. Two additional emergent relationships are used to find a constraint on transient warming and climate sensitivity from observed warming. They exploit the fact that uncertain aerosol cooling is increasingly overshadowed by greenhouse gas warming. Many of the CMIP6 climate models show a warming inconsistent with observed trends. The thesis concludes with suggestions for future work.

Contents

List of Figures

List of Tables

Associated papers

1. Nijsse, F.J.M.M., Cox, P.M., Huntingford, C., and Williamson, M.S. (2019). Decadal global temperature variability increases strongly with climate sensitivity. *Nature Climate Change*, 9(8):598–601.

2. Nijsse, F.J.M.M., Cox, P.M., Williamson, M.S., Emergent constraints on transient climate response (TCR) and equilibrium climate sensitivity (ECS) from historical warming in CMIP5 and CMIP6 models. *Earth System Dynamics*, 11(3):737–750.

3. Nijsse, F.J.M.M. and Dijkstra, H.A. (2018). A mathematical approach to understanding emergent constraints. *Earth System Dynamics*, 9(3):999–1012.

   Work for this paper was started during my Master's research and finished after I started my PhD. From Equation 3.22 onwards equations are either corrected or completely new. The same is true for Figures 3.1– 3.6.

4. Cox, P.M., Williamson, M.S., Nijsse, F.J.M.M., and Huntingford, C. (2018). Cox et al. reply. *Nature*, 563(7729):E10–E15.

   For this paper I performed the statistical analysis specifically with regards to the intercept term of the emergent constraint. This is extended in Section 4.3.

5. Williamson, M.S., Thackeray, C.W., Cox, P.M., Hall, A., Huntingford, C. and Nijsse, F.J.M.M., Emergent constraints on climate sensitivities (resubmitted)

   I was responsible for writing the statistics section, which form the basis of Sections 4.4 and 4.7.

6. Williamson, M. S., Cox, P. M., and Nijsse, F. J. M. M. (2018). Theoretical foundations of emergent constraints: relationships between climate sensitivity and global temperature variability in conceptual models. *Dynamics and Statistics of the Climate System*, 3(1).

   I contributed to the design and discussion and verified the relationship between ECS and $1/\lambda+\gamma$, relevant for Section 4.3 and Chapter 5.

Acknowledgements

Throughout my PhD research I have received a great deal of support. I would first like to thank my supervisor Professor Peter Cox, who taught me to find the big questions and the simple answers. My second supervisor Mark Williamson was always there to help with technical aspects of the work. I would like to acknowledge all the colleagues in 901, who helped me find bugs in my code and were there to offer a listening ear every time I got stuck. Professor Chris Huntingford from the Centre for Ecology and Hydrology has also supported me throughout and helped me with model data.

I would further like to thank everybody in the office for making the last couple of years so lovely. It was always a joy to come in, especially with Rebecca making sure breaks were held regularly. Thank you Joe, Noah, Becky, Arthur, Simon, Paul and Sarah. From afar, thanks Evelien, Sebastian, pap, mam and Youri, who were always just a phone call away. Finally I would like to acknowledge Esmée, who helped me type throughout the writing of my thesis when I was struggling with RSI.

Chapter 1

Introduction

The climate crisis is one of the defining problems of our time. Global warming has surpassed 1 °C (World Meteorological Organization, 2020), with the Intergovernmental Panel on Climate Change (IPCC) indicating that warming of more than 1.5 °C has far-reaching consequences including the risk of passing irreversible thresholds in the climate system (Masson-Delmotte et al., 2018). Many climate impacts scale with global quantities such as temperature, for instance the chance of heat waves or reaching certain tipping points (see Figure 1.1).

The main driver of global warming since the preindustrial period has been the anthropogenic emissions of greenhouse gases (Hegerl et al., 2007; Allen et al., 2018). Humans emit greenhouse gases from the burning of fossil fuels, cement production and agriculture (Friedlingstein et al., 2019; Saunois et al., 2020). These emissions are disproportionally caused by rich people, whereas the impacts of climate change are disproportionally felt by the poor (Chancel and Piketty, 2015). To curb future warming emissions need to be reduced, reaching net-zero, in particular carbon dioxide and methane emissions (Rogelj et al., 2018).

In the Paris agreement, it was agreed by nearly all countries worldwide that global warming should not exceed 2.0 °C, and efforts should be made so that temperatures not surpass 1.5 °C (Fuglestvedt et al., 2018). Reaching that latter goal requires societal changes unprecedented in scale and speed (Masson-Delmotte et al., 2018). The mitigation challenge depends on the social economic pathway the world chooses: it is easier to transform society towards sustainability when countries work together and more difficult when countries opt for fossil-fuel driven development (Riahi et al., 2017).

How much countries can still emit to stay below these two targets is not precisely known. This carbon budget is the outcome of two scientific questions (Mengis et al.,

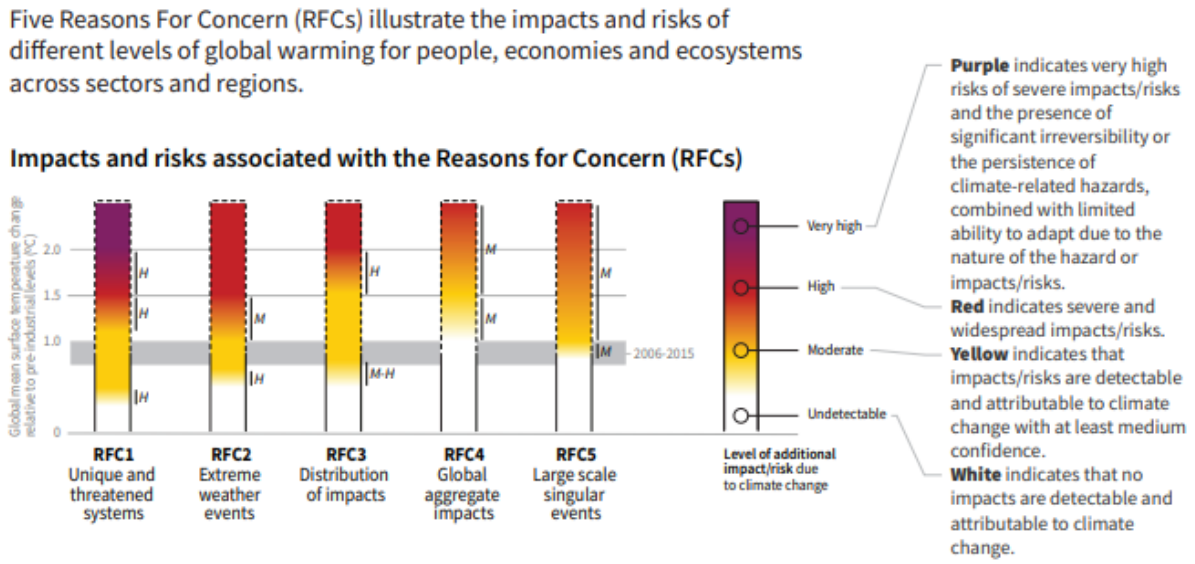


Figure 1.1: Key impacts and risks are summarised in five *Reasons for Concern*, based on expert assessment. Impacts scale with global mean temperatures. At low degrees of warming unique and threatened systems are already at risk and extreme weather is set to increase. Women and poor people have been impacted disproportionately, and this is expected to become more pronounced when global warming surpasses 1.5 °C. Simultaneously the risk of global aggregate impacts (f.i. economic downturn) and threshold behaviour rises. Reproduced from IPCC (2018), Figure SPM.2.

2018):

1. How much of our greenhouse gas emissions stay in the atmosphere rather than being absorbed by ocean and land or broken down chemically?

2. How much warming occurs when atmospheric greenhouse gas concentrations increase?

The first question requires a good understanding of the biochemical cycles in the climate system. The second question can be answered studying various physical processes on Earth. These questions are not easy to answer, because the initial response of the climate system is enhanced or balanced by climate change feedbacks. Greenhouse gases induce an initial warming by trapping heat, and this gets amplified by various processes. One self-reinforcing feedback involves snow melt: a warming Arctic will lose snow and ice cover, exposing darker surfaces which do not reflect as much sunlight as the lighter surfaces did before, spurring further temperature rises (Roe, 2009). The next section gives an overview of important feedback processes.

1.1 An (im)balance of feedbacks

Future warming depends not only on greenhouse gas emissions, but also on the balance of positive (self-reinforcing) and negative (balancing) feedbacks.

In terms of greenhouse gas concentration, most atmospheric carbon will eventually end up in the ocean drawing down temperatures, but causing ocean acidification that is nearly unprecedented in geological history (Caldeira and Wickett, 2003). Currently the land and the ocean absorb about a quarter each of yearly emissions, but this share is set to decrease in warmer climate (Walsh et al., 2017). A positive self-reinforcing feedback on land carbon involves wildfires. In a drying and warming climate, wildfires become more likely, releasing carbon into the atmosphere, further raising temperatures (Liu et al., 2014). Carbon dioxide is not the only greenhouse gas interacting with the climate system: methane for instance, is released from permafrost when it melts (Burke et al., 2018).

In addition to these biochemical feedbacks, which affect greenhouse gases, there are physical feedbacks that directly impact temperatures. For example, the water vapour content of the atmosphere increases with rising temperatures. As water vapour is itself a greenhouse gas, this acts as a positive feedback, approximately doubling the warming of well-mixed greenhouse gases such as CO_2. The darker ocean surface that replaces sea ice absorbs more heat, so that Arctic temperatures rise further. The net effect of a diverse set of changes in clouds probably also amplifies warming, but the strength of this feedback remains largely unknown. Increased infrared radiation by a warmer Earth surface (sometimes called the Planck "feedback"[1]) acts as the primary balancing mechanism (Fahey et al., 2017).

Figure 1.2 shows the strength of various feedbacks in two generations of climate models: the CMIP3 generation which informed the 2007 IPCC report, and the CMIP5 generation informing the 2013 IPCC report. The abbreviation CMIP stands for Coupled Model Intercomparison Project: an internationally coordinated program to evaluate climate models. In both generations, cloud feedbacks were the biggest source of disagreement. The same is true for the newest generation of climate models: CMIP6, where low clouds in the extra-tropics were identified as the culprits of model disagreement (Zelinka et al., 2020).

Uncertainty in these feedbacks explains why temperature projections under a single greenhouse gas concentration scenario show a wide envelope. Figure 1.3 demon-

[1] A feedback is defined as something that enhances or suppresses an initial signal. As the initial temperature signal is determined by the balance of increased greenhouse gas forcing, and inceased infrared radiation, the latter cannot also be a feedback in mathematical terms. (Roe, 2009)

strates the temperature response to different sets of greenhouse gas *concentration* pathways of models in the CMIP5 archive. The scenario with the highest concentrations (the representative concentration pathway 8.5) has rising global mean surface temperatures ranging between 3.0 and 5.0 ℃. In the CMIP6 archive, temperature rise in response to higher concentrations of greenhouse gases is more pronounced: 84% of these models indicate global warming surpasses 1.5 ℃ if we keep concentrations at the 2019 levels (Huntingford et al., 2020). These same physical feedbacks are also responsible for the uncertainty in climate sensitivity, which is the amount of warming to be expected for a doubling of CO_2.

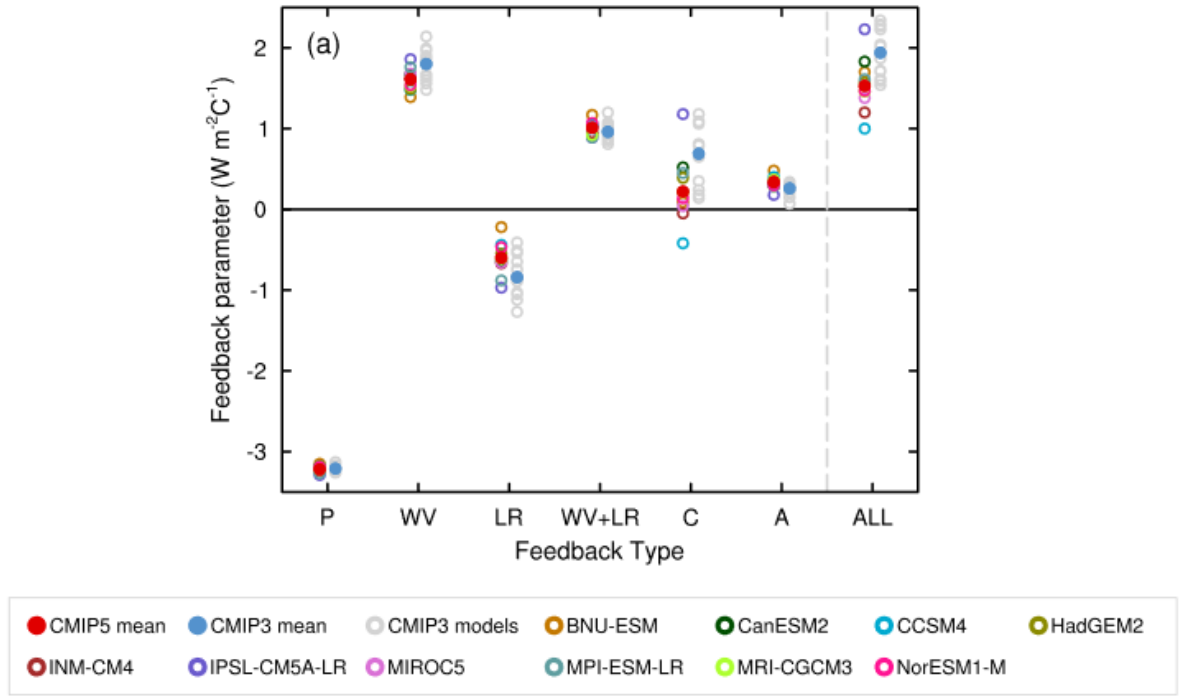


Figure 1.2: The strength of individual feedback the CMIP3 and CMIP5 ensemble. The P stands for the Planck "feedback", WV stands for water vapour, LR stands for the lapse rate feedback (changes in how fast the temperature falls with height), C is the net cloud feedback and finally A represents the albedo feedback. The spread in cloud feedbacks is largest. Reproduced from Flato et al. (2013, p. 819), Figure 9.43.

1.2 Estimating climate sensitivity

The amount of warming under a doubling of CO_2 is called climate sensitivity and is a sum of the physical feedbacks described above. Specifically, equilibrium climate sensitivity (ECS) is the temperature rise thousands of years after CO_2 has doubled,

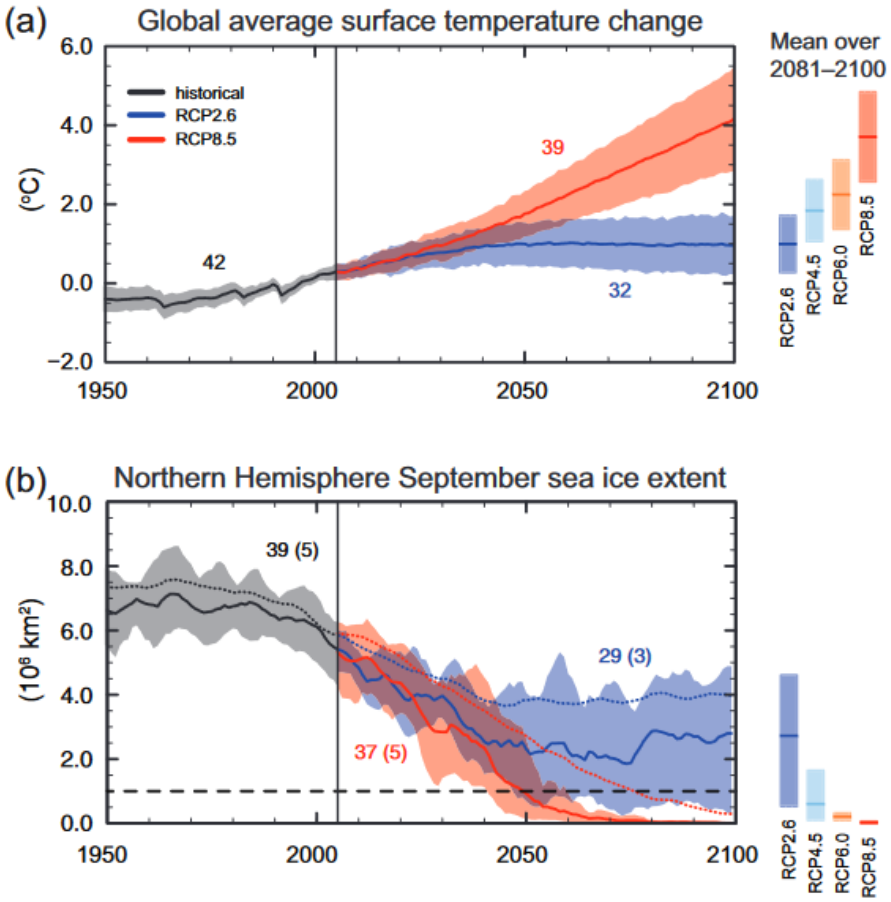


Figure 1.3: (a) Time series of global surface air temperature (relative to 1986-2005) from CMIP5 model simulations driven by concentration pathways. The thick line is the model mean and the shading represents the variation in models, specifically the 5–95 % confidence interval. (b) The sea ice extent at the end of the melt season on the Northern Hemisphere under different concentration pathways. Reproduced from IPCC (2013c), Figure SPM.7.

when the immense mass of the ocean has also finished warming to reach a new steady state. When talking about climate sensitivity, scientists often referred to the fast feedback sensitivity, sometimes called the Charney sensitivity (Previdi et al., 2013). This measure of climate sensitivity includes fast feedbacks such as water vapour and cloud feedbacks, but excludes slow acting feedback such as the albedo change from melting ice sheets in Antarctica and Greenland. In this thesis, we use this measure and compute it using the Gregory method unless specified differently. With only the first 140 years of a climate simulation used, this method slightly underestimates true equilibrium temperatures if self-reinforcing feedbacks increase over time (Rugenstein et al., 2020).

The IPCC assessed that (the fast feedback) equilibrium climate sensitivity is likely

between 1.5 ℃ and 4.5 ℃ (IPCC, 2013c). Tanaka and O'Neill (2018) estimated that if climate sensitivity is above 3.4 ℃, it would be unfeasible to decarbonise sufficiently fast to keep global warming under 2 ℃ for the entire 21st century, and temperatures may overshoot 2 ℃ at least temporarily. In 2012, the Earth reached its halfway point to a doubling of forcing from atmospheric CO_2 with respect to the levels before the Industrial Revolution (Myhre et al., 2017).

Without any feedbacks, climate sensitivity would be 1.2–1.3 ℃, determined by a balance between the increased radiative forcing of carbon dioxide and the Stefan-Boltzmann relation, the tendency of warmer bodies to emit more heat into space (i.e. the Planck "feedback"). With feedbacks in play, estimating climate sensitivity becomes more difficult (Roe, 2009). There are broadly speaking three strategies to constrain possible values of climate sensitivity: improving the process understanding of the various feedbacks, investigating the historical climate record and studying evidence from climate change in Earth's more distant past.

To understand individual feedback processes better, scientists study observations of the climate system, for instance from satellites. They also make physical models: Global Climate Models (GCMs, details in Figure 1.4), or the more extensive Earth System Models (ESMs), which represent the carbon cycle and other biochemical cycles in more detail. These models try to give a complete representation of the climate system by approximating physical laws. Approximations have to be made because computers, even supercomputers, do not have enough power to explicitly resolve processes at a small scale. These processes at small scales are instead replaced by *parametrizations*: simplified representations. Modelling centres use distinctive but often equally defensible parameterisations, leading to different projections of future climate. A different type of model, Process Resolving Models, have significantly higher resolutions and therefore do not need to make as many approximations and parametrisations for unresolved processes. They are used to assess processes in high detail in small regions (Sherwood et al., 2020).

A second line of evidence to constrain estimates of climate sensitivity comes from observed historical warming. This is combined with estimates of how much forcing (the energy imbalance at the top of the atmosphere) sunlight-reflecting aerosols and greenhouse gases have exerted. Using the fact that climate sensitivity is the amount of warming per forcing, these two numbers can be used to estimate sensitivity. Lastly, the distant past: two periods in Earth's past are particularly informative about climate sensitivity. The Last Glacial Maximum was $3\,℃ - 7\,℃$ colder than pre-industrial temperatures (often defined as a mean temperature between 1850 and 1900), whereas the mid-Pliocene warm period 3 million years ago was $1\,℃ - 5\,℃$ above pre-industrial

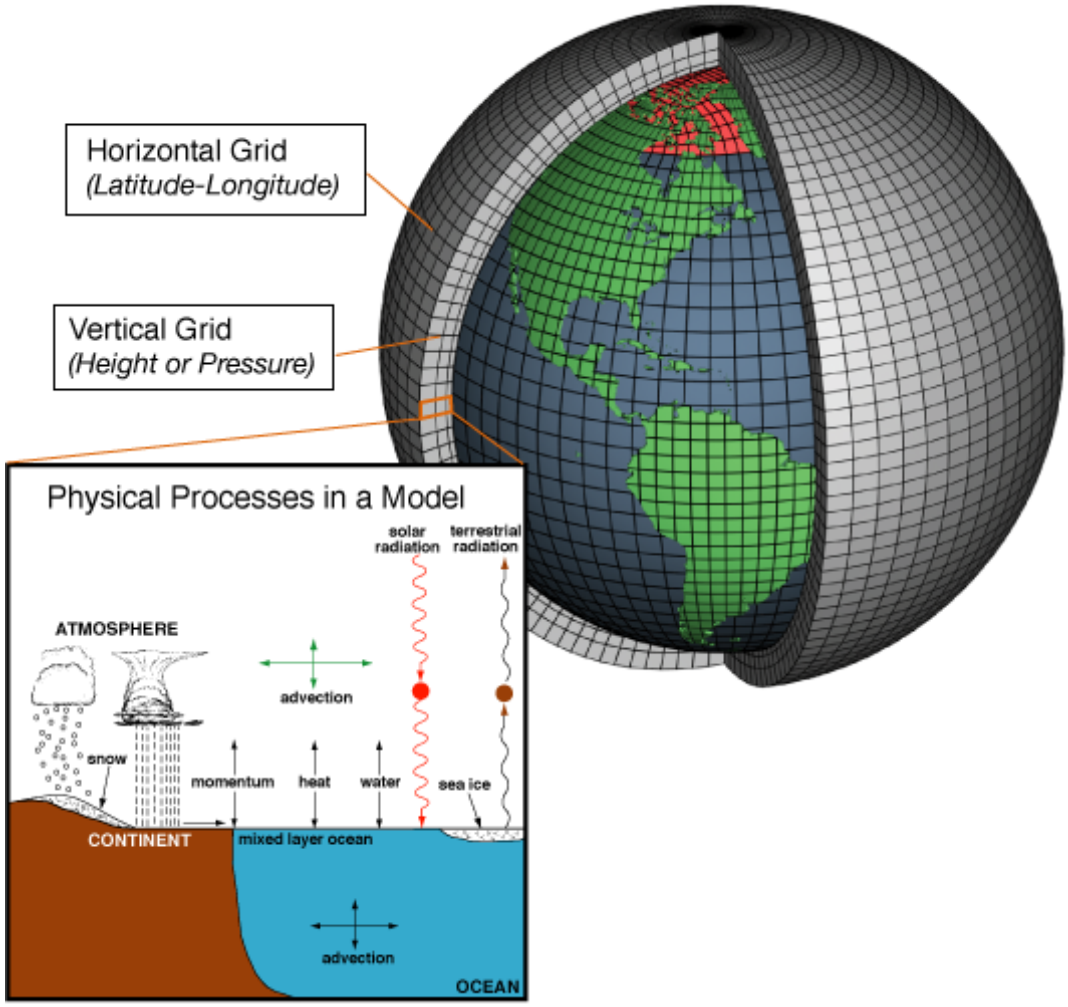


Figure 1.4: Climate models divide the Earth system into different grids: a high resolution model now has a typical resolution of 0.1 ° in latitude and longitude, and a typical model's grids are 10 times as big at 1 ° in latitude and longitude (Newsom et al., 2016). In each grid, physical laws are computed: the movement of air and water, the distribution of energy, and precipitation. Because many of these processes take place in a smaller scale than a grid, each climate model needs to find approximations on the sub-grid scale. Image source: NOAA

temperatures. In combination with estimates of the drivers behind these past changes, information about Earth's sensitivity to greenhouse gases can be obtained (Sherwood et al., 2020).

Improving the accuracy of climate projections is one of the most important challenges in climate modelling. Uncertainty might be reduced by the development of more sophisticated global climate models, capturing new processes and scales. This is a long process and many decisions on climate change adaptation and mitigation have to be made on the short term. Alternative approaches to better predictions are

therefore valuable. Emergent constraints introduced in the next chapter are one of these alternatives.

1.3 Thesis structure

The thesis is structured as followed. In Chapter 2, we give an overview of the technique of emergent constraints and how they fit within the broader suite of strategies to combine observations and models. Chapter 3 continues with a discussion of emergent constraints from a general system dynamics perspective. In Chapter 4, we detail several statistical considerations, some of which are illustrated with the emergent constraint in Cox et al. (2018a). Chapter 5 uses an emergent relationship not to constrain a future variable, but to show the present and future impact of the joint probability of sensitivity and variability. From the emergent relationship between decadal temperature variability and climate sensitivity, it is shown that a period without an increase in surface temperature may actually be more likely in high-sensitive models. Chapter 6 describes two emergent constraints on ECS and TCR and how to understand them using conceptual models. The final chapter includes a conclusion and discussion of possible future avenues for research: the impact of tuning, the opportunities of using emergent constraints in paleo-contexts, and which conceptual climate models can be revisited for use in emergent constraints research.

Chapter 2

Emergent constraints

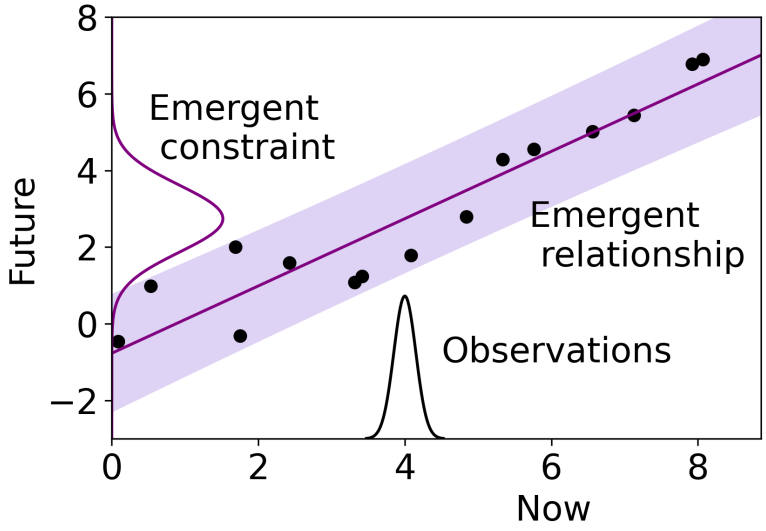


Figure 2.1: Schematic of an emergent constraint. Each dot represents a climate model. An emergent constraint combines uncertainty from observations with the uncertainty in the emergent relationship found in a set of climate models.

One of the proposed methods to accomplish reduced uncertainty in climate projections has been the use of so-called emergent constraints, where current observations are used to constrain model projections of the future (Klein and Hall, 2015). In multimodel ensembles of climate models, it is frequently possible to find a relationship between two variables, the first one observable in the current climate (e.g., seasonal variation in sea ice), the second one a future prediction (e.g. 21st century sea ice retreat). More credibility is attached to models that match the observed mean state, variability or trend well over the recent period. In this way, current observations provide a constraint to long term trends. The observable variable is called the predictor, and

the future variable the predictand (Klein and Hall, 2015). Figure 2.1 shows a schematic of an emergent constraint: each of the dots represent a (hypothetical) climate model.

To construct a regression line between a predictor X and a predictand Y, not only good models (consistent with observations), but curiously also the "bad models" (inconsistent with observations) are used (Brient and Schneider, 2016). It is a subjective assessment to what extent these bad models are similarly able to inform the relationship between predictor and predictand.

In recent years many emergent constraints have been found. Caldwell et al. (2018), for instance, evaluated 19 separate emergent constraints on climate sensitivity only. Further constraints have been found for for Arctic temperature change (Bracegirdle and Stephenson, 2013), snow albedo feedback (Hall and Qu, 2006), tropical carbon (Wenzel et al., 2014), and Antarctic ozone (Karpechko et al., 2013) amongst other variables.

In Section 2.1, we will discuss the difference between the qualitative and quantitative interpretation of emergent constraints. Section 2.2 discusses conditions under which emergent constraints are convincing. In Section 2.3, the technique of emergent constraints will be put in a broader perspective, comparing it to techniques from numerical weather prediction and general statistical modelling.

2.1 Interpretation: qualitative or quantitative

There are broadly two ways to interpret the relationships found in emergent constraints. The 'strong or quantitative' interpretation quantifies a probability distribution of the predictand using the (linear) regression between predictor and predictand. The first emergent constraint (Allen and Ingram, 2002) and many that followed use this interpretation (Hargreaves et al., 2012; Cox et al., 2013; Winkler et al., 2019, e.g.). Another proposed quantitative interpretation uses joint Gaussian distributions (Bretherton and Caldwell, 2020), as further discussed in Section 4.7. The weak interpretation, for instance by Hall and Qu (2006), simply states that models with a high or low predictand are more or less likely to be close to the truth given the observations.

A related point of contention is the value of models that lie far outside of the observations. In linear least squares regression, outliers influence the regression parameters markedly. To what extent can we trust a model that performs badly in terms of the observable of interest to show the correct correspondence between predictors and predictands?

In the theory of emergent constraints it is often a priori assumed that the relation-

ship between predictand and predictor is linear. This does not have to be the case. Effects of nonlinearities are discussed in Chapter 4 of this thesis (Section 4.4). The regression choice can partially be based on an analytically-derived relationship using reduced-form modelling, exemplified in Chapters 5 and 6. Scaling up to ESMs may alter alter this relationship. Particular care should be taken when the observed predictor is on the low or high end of the model range, as bias there is highest.

2.2 Why and when to trust emergent constraints?

Not all emergent constraints are equally well-supported. Klein and Hall (2015) propose to categorise emergent constraints in three groups: potential, promising and confirmed emergent constraints, depending on how reliable one judges the emergent constraint to be. They discussed three different reliability criteria: a strong physical basis, robustness to choice of model ensemble and no obvious multiple influences. Similarly, Hall et al. (2019) distinguish proposed (no physical intuition or mechanism) and confirmed emergent constraints.

Strong physical basis

To ensure emergent constraints are not simply found by data mining, every emergent constraint should have a strong physical basis. Searching for emergent constraints using blind data mining my lead to spurious and misleading correlations, as there always a pletora of correlations in a highly dimensional dataset by chance alone, some of which may seem physical. (Caldwell et al., 2014).

Klein and Hall (2015) argue that, ideally, the physical mechanism should point to specific parametrisations or physical parameters. They further state that with this physical explanation in hand, researchers should show that this mechanism is in fact the reason behind the emergent constraint. Unfortunately, model diagnostics are often not available to perform this check (ibid.).

Klein and Hall (2015) identify this as the most important reliability criterion. If it is only partially satisfied, two other criteria should be taken into account.

Robustness to choice of model ensemble

If the physical relation between predictor and predictand is robust, it should appear when testing out of sample (Hall et al., 2019), for instance in different CMIP model ensembles (CMIP3, CMIP5, CMIP6), different model versions or perturbed physics

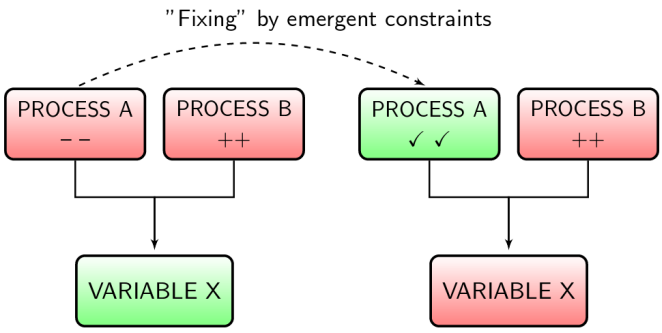


Figure 2.2: Schematic of compensating errors. When the effect of process A and B are both related to variable X it is possible to find an emergent relationship between A and X, but also between B and X. If this second relationship is ignored, one might end up with a worse estimate of X compared to the model mean.

ensembles. Failing this criterion does not automatically disqualify an emergent constraint. There are a small set of conditions under which a valid emergent constraint is only present in some ensembles. First of all, the identification of a certain emergent constraint or bias might have led some modelling groups to improve that process so that variation in the ensemble disappears. If an emergent constraint materialises due to newly added processes, older generations cannot show it. Lastly, a model ensemble may have improved or converged on the representation of a separate process that clouded the presence of an emergent constraint in a previous ensemble.

No obvious multiple influences

Establishing the reliability of an emergent constraint that is subject to many independent or dependent processes is difficult. One of the major problems in these types of emergent constraints is that these processes possibly have compensating errors as shown in Figure 2.2. If two compensating processes together determine the predictand, where one is overestimated and the other underestimated, an emergent constraint using but one process will make the prediction worse than taking the model mean.

2.2.1 Categorisation by dynamics

In Chapter 3, we will propose a different classification for emergent constraint, based on their dynamical characteristics. A distinction is made between dynamical and static emergent constraints. The first category consists of those emergent constraints that

use the system response to a periodic or random forcing of the system. The second category, static constraints, are those emergent constraints that find relationships between the background state or mean state of the climate system, and sensitivity on long time scales to forcing.

Both categories are further divided into direct and indirect emergent constraints. Direct constraints use the same observable for predictor and predictand, while indirect constraints are relationships between two different observables.

This framework complements the frameworks based on reliability in the sense that it helps formulating and showing a physical basis for the emergent constraint of interest. Direct and dynamical emergent constraints are considered more physically convincing as there are fewer possibilities for compensating errors and systematic errors respectively.

2.3 Comparing emergent constraints to other methods of combining models and observations

Many methods exist that combine models and observational data to help predict some future quantity. Emergent constraints are a simple technique that can be seen as "post-processing" climate model ensembles. In contrast, other techniques sometimes delve into the model code. In this section, we give an overview of techniques used in weather and climate prediction and contrast them with emergent constraints.

2.3.1 Data assimilation

Data assimilation combines information from measurements and dynamical (geophysical) models to obtain an optimal estimate of a variable field (Reichle, 2008).

In data assimilation, the state of a weather model is updated every n time steps using observations and taking into account the relative uncertainty of the model compared to the observations. The updates can be done using a Kalman filter, which is in essence a Bayesian update with the model output as a prior and the observations as a likelihood function (Kalman, 1960). The original Kalman filter assumed a linear model, but since, more complicated Kalman filters have been developed that work for nonlinear models and use an ensemble of models (Lei and Hacker, 2015).

In addition to these 'objective data assimilation techniques' there are empirical data assimilation techniques, such as nudging, also known as Newtonian relaxation. Here a weather or climate model is nudged towards an observational field. In contrast to Kalman filters the dynamical code of the model is directly changed by inserting an

additional feedback term proportional to the observation-model misfit, which makes the technique continuous. It is computationally less expensive than the objective data assimilation (Lei and Hacker, 2015).

Data assimilation is a technique designed to solve a different problem than emergent constraints, namely the initial value problem of numerical weather prediction. Data assimilation has also been applied to climate models for decadal predictions (Chikamoto et al., 2019, e.g.). In emergent constraints and climate projections more generally, we are instead interested in the average properties of the climate system: the transient state after decades to centuries or, more often, an equilibrium state, such as the carbon budget and climate sensitivity.

2.3.2 Calibration and history matching

Climate model developers use observations to calibrate or tune their models. The observations are used to find optimal parameters in uncertain parameterisations (Steele and Werndl, 2013).

In history matching, computationally cheap models (either physical models or statistical models of those models called emulators) are run with a large set of parameter values. The parameter space is usually quite high-dimensional (say 20 different parameters) and therefore sampled randomly using priors for all parameters. Combinations of parameters are accepted if they are able to produce the observed history with a certain degree of error (Goodwin et al., 2018).

History matching is often contrasted with full Bayesian calibration. The former is computationally more efficient, since it can rule out physically very unrealistic parameter values first by only using those model outputs for which it is easy to construct an emulator. In the next iterator of history matching the emulator uses a smaller parameter space, which makes it easier to emulate complex phenomena. Full Bayesian calibration immediately takes into account all sources of observations so that a lot of attention has to be focused on constructing a emulator which performs well in very nonphysical circumstances. An advantage of the latter technique is that there is a formal quantification of uncertainty for model uncertainty (Williamson et al., 2013).

History matching defines some implausibility measure that compares the emulator output (plus its uncertainty) to the observations. In the standard application of history matching, this implausibility measure does not take into account autocorrelation, which was identified in Cox et al. (2018a) to be an independent source of information informing plausibility of models. A more important difference with emergent constraints is that information from different model formulations is lost.

2.3.3 Post-processing

The output of weather models is usually post-processed. These statistical post-processing techniques compensate for known biases in the models. It also compensates for under- or overconfidence of the initial value ensemble, which may vary in space and time. The testing and development of post-processing relies heavily on hindcasts: using current models to predict old weather patterns (Van Schaeybroeck and Vannitsem, 2012). Again, one big difference between weather prediction and climate projections is that we are usually not interested in the temporal evolution just after an initial state, but longer-term trends, for which few hindcasts are available.

Because every model is imperfect, post-processing is sometimes also done using a multi-model ensemble. The set of techniques that are used in (hydro)-meteorology includes simple model average, multimodel superensemble, and Bayesian model averaging (BMA) (Li et al., 2017, and references therein).

In climate modelling, as with weather models, a common post-processing technique is using the multimodel mean as best estimate with the multi-model variance as a measure of uncertainty. This approach is called model democracy as all models are considered equally likely to be "true" and used with equal weight for the final prediction (Knutti, 2010).

Stott and Kettleborough (2002) developed a method to estimate future warming combining model with temperature observations. They assumed that the fractional error between model and observations would remain constant, which was consistent with historical model-observation discrepancy. Uncertainty was computed from estimated natural forcing and control-simulation derived internal variability. As the technique is applied to one model, information from other models is lost. Uncertainty may be underestimated, as to models showing the same historical warming can show different future warming for other reasons than internal variability.

Sanderson et al. (2015) propose a different method that takes into account model interdependency. Using observational data of surface temperature, precipitation, outgoing top-of-the-atmosphere (TOA) shortwave radiation, outgoing longwave radiation, sea level pressure, atmospheric temperature, and relative humidity, they define a distance function between models and between model and observations. This distance function allows interpolation between models with the idea of preventing very similar models having a large impact on the predictands.

Although sophisticated, this method judges the models on their overall performance, and not on the performance on those metrics that are of interest for certain predictands. If only those observations are used that have a high predictive power for,

say, climate sensitivity, one might compromise on the power of the method to account for model interdependency as a result.

Model weighting based on one relevant observation, as performed in Brient and Schneider (2016), circumvents this issue. They perform model weighting using the Kullback-Leibler divergence, which is the relative entropy between model and observation. Consequently no information is used from models inconsistent with observations. They argue that regression in emergent constraints can lead to an overly constrained estimate of Y if one of the models consistent with observations lies too far outside of the regression prediction interval. But conversely, models consistent with observations may show too little spread, and the spread around the emergent relationship outside of the observationally consistent region may provide useful information about this.

Chapter 3

A mathematical approach

This chapter is adapted from Nijsse and Dijkstra (2018), and its basis was developed as part of my Master thesis. See Associated papers for the PhD contributions.

3.1 Introduction

For individual emergent constraints physical processes have been identified explaining the relationship. An example is the relationship between the seasonal variation in the snow-albedo feedback (SAF) and the feedback under global warming (Hall and Qu, 2006). Hall and Qu (2006) elucidated the key physical process behind the emergent constraint. Models where the maximum albedo of snow is highest have the largest SAF on both time scales because the contrast between snow-covered and snow-free areas is highestf (Qu and Hall, 2007).

However, a more general dynamical picture on how emergent constraints occur in multi-model ensembles or even in a perturbed parameter ensemble of a single model has been lacking. Under which circumstances are these constraints expected to arise? Some emergent constraints may be spurious and could arise because of shared errors in a particular multimodel ensemble (Bracegirdle and Stephenson, 2013). A mathematical framework is desired to identify spurious constraints and to give an indication as to where new emergent constraints might arise.

In this chapter, we investigate how and under what conditions emergent constraints appear and what can be learned about the physics of the climate system from their existence. We will use linear response theory (LRT) to address the problem of forcing-response relations on different time-scales (Risken and Frank, 1996). Ruelle (1998) demonstrated that LRT can be extended to study the response of non-equilibrium systems to external forcing. As with the fluctuation-dissipation theory, Ruelle's LRT uses

the statistical properties of the unforced (equilibrium) state only, but it does not assume (quasi)-equilibrium. Recently, LRT has been proposed as a rigorous framework for computing the response of the climate system and its applicability has been tested on the Lorenz-96 model and on the idealised global climate model PlaSim (Lucarini and Sarno, 2011; Ragone et al., 2016).

The chapter is organised as follows. To obtain an understanding of emergent constraints we start by formulating a mathematical framework in terms of susceptibilities by making use of LRT (Section 3.2). This results in explicit expressions for the appearance of emergent constraints in terms of susceptibility functions. In Section 3.3, a classification scheme for emergent constraints is proposed which may provide insight into the question of reliability posed in the introduction. Then, in Section 3.4, applications are presented for conceptual climate models, such as Ornstein-Uhlenbeck processes in one and two dimensions, an energy balance model and a climate model of intermediate complexity (PlaSim). The results are summarised and discussed in Section 3.5.

3.2 Response functions

In this section explicit expressions are given for response functions of the state of a dynamical system which depends on a single parameter and which is subjected to a non-stationary forcing. Such response functions are used in the following section to classify the different emergent constraints. Rigorous results for linear response properties of large class of general stochastic systems were obtained by Hairer and Majda (2010). Linear response theory for nonequilibrium systems was developed by Ruelle (1998, 2009).

We illustrate the approach using the general one-dimensional forced stochastic differential equation (SDE)

$$dX_t = (-V'(X_t) + F(t))dt + \sqrt{\sigma}dW_t. \tag{3.1}$$

Here $V(x)$ is a smooth confining potential, meaning that an equilibrium solution exists for the unforced system (Pavliotis, 2014), and $F(t)$ is a prescribed forcing. Furthermore, σ is the noise amplitude and the associated Wiener process is indicated by W_t.

The probability density function of the unforced ($F(t) = 0$) system, say $\bar{p}$, satisfies the Fokker-Planck equation

$$\frac{\partial \bar{p}}{\partial t} = \frac{\partial (V'(x)\bar{p})}{\partial x} + \frac{\sigma}{2}\frac{\partial^2 \bar{p}}{\partial x^2} = \mathscr{L}^*(\bar{p}), \tag{3.2}$$

which defines the Fokker-Planck operator $\mathscr{L}^*$. The equilibrium distribution of the unforced system, here indicated by $\bar{p}_e$, is given as:

$$\bar{p}_e(x) = \frac{1}{Z} e^{\frac{-2V(x)}{\sigma}}; \qquad Z = \int_{-\infty}^{\infty} e^{\frac{-2V(x)}{\sigma}}\, dx. \tag{3.3}$$

Linear response theory (Ragone et al., 2016) provides an expression for the change in the expected value of an observable O (e.g. the temperature, ice extent or the standard deviation of either), say $\Delta O(t)$ when the system is forced, compared to the unforced case, i.e.

$$\Delta O(t) = E[O(X_t)] - E[O_e(X_t)], \tag{3.4}$$

where again the subscript e indicates the equilibrium of the unforced system. It follows that

$$\Delta O(t) = \int_0^t R_O(t-s)F(s)ds; \qquad R_O(t) = \mathcal{H}(t) \int_{-\infty}^{\infty} O(x)\, e^{\mathscr{L}^* t}(-\frac{\partial \bar{p}_e}{\partial x})dx, \tag{3.5}$$

where $R_O(t)$ is the response function, which is extended to be zero for $t < 0$ with a Heaviside function $\mathcal{H}(t)$ to ensure causality. When Equation 3.5 is Fourier transformed we find, using the convolution theorem,

$$\mathcal{F}(\Delta O(t))(\omega) = \chi(\omega)\hat{F}(\omega), \tag{3.6}$$

where the Fourier transform $\chi(\omega)$ of the response function $R_O(t)$ is the susceptibility. If we take a cosine forcing, i.e. $F(t) = F_0 \cos\omega_0 t$ then $\hat{F}(\omega) = F_0\pi(\delta(\omega-\omega_0) + \delta(\omega+\omega_0))$ so once we know $\chi(\omega)$, we can determine the response $\Delta O(t)$.

When we take the identity operator $O = x$ as the observable, thus taking the mean value of this variable, we obtain a response function and corresponding susceptibility:

$$R_O(t) = \frac{2}{\sigma}\sum_{l=1}^{\infty} \beta_l e^{-\lambda_l t}, \qquad \chi(\omega) = \frac{2}{\sigma}\sum_{l=1}^{\infty} \frac{\beta_l}{\lambda_l + i\omega}, \tag{3.7}$$

For $A = x$, we find from Equation 3.5 that

$$R_A(t) = \int_{-\infty}^{\infty} x\, e^{\mathscr{L}^* t}(-\frac{\partial \bar{p}_e}{\partial x})\, dx. \tag{3.8}$$

Using the expression for the equilibrium solution $\bar{p}_e$ from Equation 3.3, we find

$$-\frac{\partial \bar{p}_e}{\partial x} = \frac{2}{\sigma} V'(x)\bar{p}_e \tag{3.9}$$

and hence Equation 3.8 becomes

$$R_A(t) = \int_{-\infty}^{\infty} x \; e^{\mathscr{L}^* t} \left(\frac{2}{\sigma} V'(x) \bar{p}_e \right) \; dx. \tag{3.10}$$

With the standard L^2-inner product, the adjoint of $\mathscr{L}$ determined as $\langle \mathscr{L}^* g, h \rangle = \langle g, \mathscr{L} h \rangle$, where $\mathscr{L}$ is the generator of the OU process, is given by

$$\mathscr{L} u = V'(x) \frac{\partial u}{\partial x} + \frac{\sigma}{2} \frac{\partial^2 u}{\partial x^2}. \tag{3.11}$$

Using this property in Equation 3.10, we find

$$\langle x, e^{\mathscr{L}^* t} (V'(x) \bar{p}_e) \rangle = \langle e^{\mathscr{L} t} x, V'(x) \bar{p}_e \rangle \tag{3.12}$$

and hence

$$R_A(t) = \frac{2}{\sigma} \int_{-\infty}^{\infty} e^{\mathscr{L} t}(x) \; V'(x) \bar{p}_e \; dx. \tag{3.13}$$

Next an inner product $\langle g, h \rangle_{\bar{p}_e}$ is defined as

$$\langle g, h \rangle_{\bar{p}_e} = \int_{-\infty}^{\infty} g h \bar{p}_e \; dx. \tag{3.14}$$

As a next step, let λ_l and ϕ_l be the eigenvalues of the generator, i.e. solutions v of

$$\mathscr{L} \phi = -\lambda \phi. \tag{3.15}$$

For reversible processes, these eigenvalues are real, positive and discrete under the inner product $\langle \, , \rangle_{\bar{p}_e}$. The eigenfunctions form a complete orthonormal basis, such that $\langle \phi_n, \phi_m \rangle_{\bar{p}_e} = \delta_{nm}$ (Pavliotis, 2014). Now $e^{\mathscr{L} t}(x)$ represents solutions $u(x,t)$ of the problem

$$\frac{\partial u}{\partial t} = \mathscr{L} u \tag{3.16}$$

with initial condition $u(x,0) = x$. We can expand u into eigenfunctions as

$$u(x,t) = \sum_{l=1}^{\infty} \alpha_l \phi_l(x) e^{-\lambda_l t}. \tag{3.17}$$

From the initial condition, we find

$$\sum_{l=1}^{\infty} \alpha_l \phi_l(x) = x \tag{3.18}$$

and using the orthogonality of the ϕ_l under the inner product $\langle , \rangle_{\bar{p}_e}$, we find

$$\alpha_l = \langle x, \phi_l \rangle_{\bar{p}_e}. \tag{3.19}$$

On the other hand, substituting the expression for u into Equation 3.13 gives

$$\int_{-\infty}^{\infty} \sum_{l=1}^{\infty} \alpha_l \phi_l(x) e^{-\lambda_l t} V'(x) \bar{p}_e \, dx = \sum_{l=1}^{\infty} \beta_l e^{-\lambda_l t}, \tag{3.20}$$

where

$$\beta_l = \alpha_l \langle V'(x), \phi_l \rangle_{\bar{p}_e} = \langle x, \phi_l \rangle_{\bar{p}_e} \langle V'(x), \phi_l \rangle_{\bar{p}_e}. \tag{3.21}$$

Repeating the derivation above with a general observable $A = f(x)$ gives $\beta_l = \langle f(x), \phi_l \rangle_{\bar{p}_e} \langle V'(x), \phi_l \rangle_{\bar{p}_e}$. The first factor in β_l denotes the projection of the observable on the eigenfunctions and could intuitively be interpreted (for $l > 0$) as the amenability of the observable to change. The second projection term in β_l can be understood to be the amenability of the whole system to change under the influence of the forcing field. In Equation 3.24 those observables are Y_1 and Y_2, so that $g_l = \langle Y_1, \phi_l \rangle_{\bar{p}_e} \langle V'(x), \phi_l \rangle_{\bar{p}_e}$ and $h_l = \langle Y_2, \phi_l \rangle_{\bar{p}_e} \langle V'(x), \phi_l \rangle_{\bar{p}_e}$.

The amplitude A of the response to a periodic forcing $F(t) = F_0 \cos \omega_0 t$ is determined by the absolute value of the susceptibility

$$A(\Delta X(t))(\omega_0) = \frac{2F_0}{\sigma} \sum_{l=1}^{\infty} \frac{\beta_l}{\sqrt{\lambda_l^2 + \omega_0^2}}. \tag{3.22}$$

When the predictor is a response to a uniform frequency forcing, it can be expressed as in Equation 3.22. The predictand will generally also have this form, as a forcing with a long time scale can be approximated by a low-frequency forcing. The previous analysis can be generalised to more dimensions. In two dimensions, for example, with a state vector $Y_t = (Y_{1t}, Y_{2t})^T$, the SDE becomes

$$dY_t = (-\nabla V(Y_t) + F(t)\hat{\boldsymbol{\imath}})dt + \sqrt{\sigma} I_2 dW_t, \tag{3.23}$$

where the term $F(t)\hat{\boldsymbol{\imath}}$ denotes a forcing in the direction of the first variable and I_2 the identity matrix. As shown in Pavliotis (2014), the derivation of the response function follows the one-dimensional case closely, resulting in:

$$R_{Y_1}(t) = \frac{2}{\sigma} \sum_{l=1}^{\infty} g_l e^{-\lambda_l t}; \qquad R_{Y_2}(t) = \frac{2}{\sigma} \sum_{l=1}^{\infty} h_l e^{-\lambda_l t}, \tag{3.24}$$

where g_l and h_l are again projection coefficients, g_l and h_l containing a factor describing strength of the response in Y_1 and Y_2 respectively. The derivation of the exact terms is given above. Calculating the response is analogous to the one-dimensional case, so that the Fourier transforms of the response functions are given by

$$A\left(\Delta Y_{1t}\right)(\omega_0) = \frac{2}{\sigma}\sum_{l=1}^{\infty}\frac{g_l}{\sqrt{\lambda_l^2+\omega_0^2}}; \tag{3.25}$$

$$A(\Delta Y_{2t})(\omega_0) = \frac{2}{\sigma}\sum_{l=1}^{\infty}\frac{h_l}{\sqrt{\lambda_l^2+\omega_0^2}}. \tag{3.26}$$

In uncoupled multidimensional systems, the eigenfunctions are found to be the tensor products of the eigenfunctions in the one-dimensional case, while the corresponding eigenvalues are the sum of the eigenvalues in the one-dimensional case.

3.3 Classification of emergent constraints

Although a wide set of different emergent constraints have been found, no attempts have been made to classify them so far using dynamical criteria. Here, a classification is proposed based on the time-characteristics of the predictor and on the relationship between the predictor and the predictand. Using this classification, assessment of their applicability becomes easier. Furthermore, a classification is a prerequisite for a dynamical description of emergent constraints.

Firstly, an emergent constraint can be either direct or indirect. In the direct case, the predictor and predictand are the same observable, while in the indirect case they are not. In the latter case, the predictor variable and predictand variable have to be closely linked, for instance via a physical process. We make a further distinction between static and dynamic emergent constraints. In a dynamic emergent constraint a response to a known, or sometimes even unknown, forcing in the (present-day) predictor is linked to the response of the (future) predictand under the same (or a similar) forcing. For example: the forcing can be the annual cycle of solar radiation, but can also be caused by ENSO or historical climate change. In a static emergent constraint a relationship between the time-independent quantity of the unforced system in the present-day (predictor) is linked to the response in a quantity under climate change.

As an illustration, we apply our classification to examples of emergent constraints found in the literature in Table 3.1. Although this is not a complete overview, examples are found of the four types of emergent constraints. There are many examples of direct dynamical constraints, such as the one involving the snow-albedo feedback

	Reference	Climate predictor	Climate predictand
1	Knutti et al. (2006)	Seasonal cycle land temperature amplitude	ECS
2	Hall and Qu (2006); Qu and Hall (2014)	Springtime SAF	SAF
3	Boe et al. (2009)	Arctic sea ice extent trend	Arctic sea ice extent
4	Clement et al. (2009)	Sensitivity LLC to Pacific decadal variability	Sign LLC feedback
5	Trenberth and Fasullo (2010)	SH net radiation TOA	ECS
6	Fasullo and Trenberth (2012)	Mid-tropospheric RH over ocean in subsidence region	ECS
7	Bracegirdle and Stephenson (2013)	Arctic SAT	Arctic SAT
8	Gordon and Klein (2014)	Sensitivity of extra-tropical LLC optical depth to temperature	Extra-tropical LLC optical depth.
9	Qu et al. (2014)	Sensitivity of LLC cover to SST	LLC cover
10	Sherwood et al. (2014)	Cloud-scale and large-scale lower tropospheric mixing over oceans	ECS
11	Su et al. (2014)	RH & cloud fraction tropics	ECS
12	Wenzel et al. (2014)	Short-term sensitivity of atmospheric CO_2	Tropical land carbon storage
13	Tian (2015)	Precipitation and mid-tropospheric RH asymmetry bias (for ITCZ)	ECS
14	Kwiatkowski et al. (2017)	Tropical primary production under ENSO-driven SST variations	Tropical primary production
15	Cox et al. (2018a)	Function of autocorrelation of GMST	ECS
16	Donat et al. (2018)	Seasonal land-atmosphere feedback	Frequency heat extremes
17	Terhaar et al. (2020)	Present-day Arctic sea surface density	Future Arctic Ocean acidification

Table 3.1: Selection of emergent constraints from the literature to demonstrate the classification system. The classification is showcased in Figure 3.1

Abbreviations stand for RH: relative humidity, ITCZ: inter-tropical convergence zone, TOA: top of atmosphere, SH: southern hemisphere, ECS: equilibrium climate sensitivity, LLC: low-level cloud, SAF: snow-albedo feedback. SAT: surface air temperature. The emergent constraint found by Trenberth and Fasullo (2010) seems to be spurious: no physical mechanism was proposed and it did not appear in different ensembles, such as CMIP5 (Grise et al., 2015).

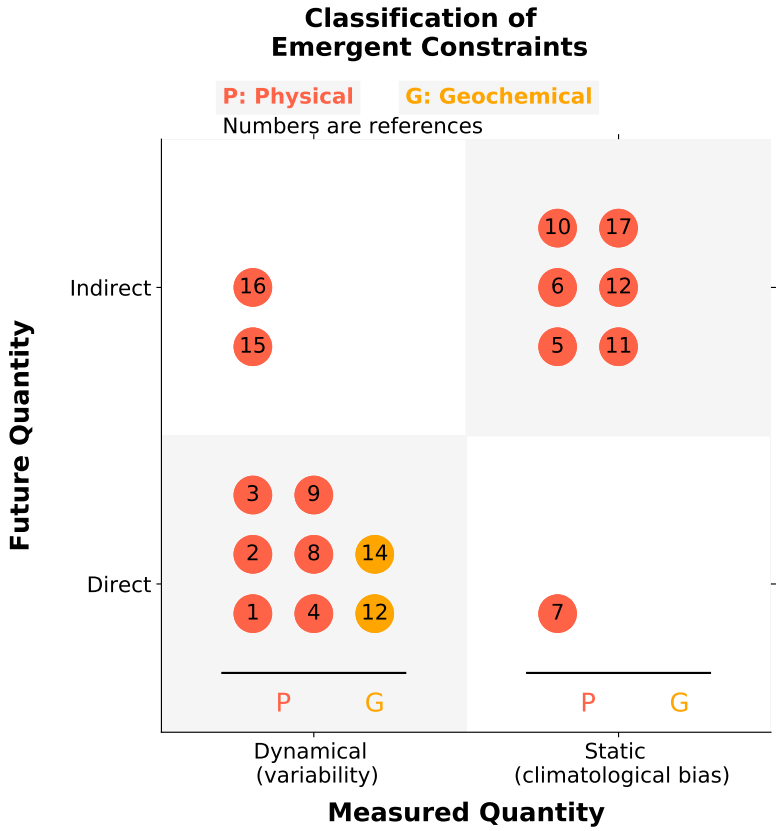


Figure 3.1: Classification of emergent constraints in four categories. Most ECs fall either in the category of dynamical direct constraints or static indirect constraints. Numbers correspond to Table 3.1.

shown in Figure 3.7 (Hall and Qu, 2006). Dynamic direct emergent constraints are the most intuitive. As long as the variations in the predictor are of a sufficient amplitude compared to those of the predictand, a correlation between the predictor and predictand automatically points towards a common physical basis, for example a common dynamical response to an external forcing. The direct static emergent constraint found by Bracegirdle and Stephenson (2013) makes use of spatial patterns. All of the indirect constraints involve equilibrium climate sensitivity as the predictand. Often the mean of some variable with a known bias in the model ensemble is linked to ECS. For instance, in Tian (2015) the asymmetry bias in ITCZ is linked to climate sensitivity. An example of a dynamical indirect emergent constraint is provided by Cox et al. (2018a), who relate a function of autocorrelation of global surface temperature to ECS. In this case the short-time forcing is assumed to be caused by internal variability.

Based on the response function theory in Section 3.2, we further elaborate on the classification and also discuss conditions for each type of constraint for a dynamical system with varying parameters (which defines the ensemble of models).

For a direct dynamical emergent constraint, in the standard case of a linear relationship, the relation has the following form: $\mathrm{Predictand} = C^{st}\,\mathrm{Predictor}$, where C^{st} is a constant independent of the parameter used to generate the ensemble of models. Rewriting this, the ratio of the responses to forcing of frequencies ω_1 and ω_2 should be constant over the (parameter) ensemble members e_i. For the simple case of two forcings that only differ in frequency, we find the condition from the ratio of the susceptibilities SR as

$$\mathsf{SR}(e) = \frac{A(\Delta O(t))(\omega_2)}{A(\Delta O(t))(\omega_1)} = \frac{\sum_{l=1}^{\infty} \frac{\beta_l}{\sqrt{\lambda_l^2+\omega_2^2}}}{\sum_{l=1}^{\infty} \frac{\beta_l}{\sqrt{\lambda_l^2+\omega_1^2}}} = C^{st}, \tag{3.27}$$

One variable (B) can act as forcing to a second variable (O), while being itself forced externally (F). The predictor and predictand are then given by the quotient of the response functions of O and B. A further complication is that often the forcing patterns are not exactly the same for the short (F_2) and long (F_1) periodic forcing. In this case Equation 3.27 has to be adjusted to:

$$\mathsf{SR}(e) = \frac{A(\Delta O(t)|_{F_2})(\omega_2)}{A(\Delta O(t)|_{F_1})(\omega_1)} \frac{A(\Delta B(t)|_{F_1})(\omega_1)}{A(\Delta B(t)|_{F_2})(\omega_2)} = C^{st}, \tag{3.28}$$

This is further discussed in the example of the idealised energy balance model.

Physically, we expect that the same mechanisms to be responsible for the response at a short and long time scale to obtain this type of emergent constraint. The system should have response times smaller than the time scale of the forcing or equivalently: the generator should have eigenvalues λ larger than the frequency of the forcing. Naturally, the response times $\frac{1}{\lambda}$ of the dominant processes are expected to be at least smaller than the time scale of the slow forcing $\frac{1}{\omega_1}$.

Mathematically, the ratio in Equation 3.27 becomes one in the case that all eigenvalues λ_l are much larger than the forcing frequencies. Interestingly, the linear relation breaks down in the case that the fast forcing has the same order of magnitude as the eigenvalues of the dominant terms in the susceptibility. Under the assumption of a single dominant term in the susceptibility and a slow forcing with frequency $\omega_2 \to 0$ the first correction term to the slope-one linear relation between predictand Y and predictor X, is cubic in X.

For a direct dynamical emergent constraint, a linear relationship with slope one is reached when the forcing frequencies are significantly lower than the eigenvalues of the system. This is a weak assumption for future forcing, which can often be approached by saying $\omega_{\mathsf{slow}} \to 0$. We then only have to be worried about the fast forcing. If this fast forcing and the system response are approximately equally fast; the relation

that is found might seem linear, but it in fact not linear.

Take a system in which the susceptibility is dominated by only one term in its sum and for which $\omega_{\text{slow}} \to 0$. The amplitude of the response Y will be $\frac{\beta_l}{\lambda}$. The fast variable X will have amplitude of the form:

$$\frac{\beta_1}{\sqrt{\lambda^2+\omega_2^2}} = \frac{\beta_1}{\lambda}\frac{1}{\sqrt{1+(\frac{\omega}{\lambda})^2}} \tag{3.29}$$

To simplify, we first divide X and Y by β_1. Then we invert the second equation by substituting $y = \frac{1}{\lambda}$.

$$\begin{aligned}
\frac{1}{\lambda}\frac{1}{\sqrt{1+(\frac{\omega}{\lambda})^2}} &= x \\
y\frac{1}{\sqrt{1+(\omega y)^2}} &= x \\
y^2(1-x^2\omega^2) &= x^2 \\
y &= x\sqrt{\frac{1}{1-x^2\omega^2}} \\
y &\approx x\left(1+\frac{1}{2}\omega^2x^2\right).
\end{aligned} \tag{3.30}$$

From this we can conclude that a correction term is added to the linear relationship proportional to $\frac{\omega}{\lambda}$, so that in the case the response is not sufficiently fast, there will only be an apparent linear relationship. Normally, the sum determining the amplitude of the response will consist of multiple terms. The higher eigenvalues are larger, so that the assumption $\omega < \lambda$ is more easily met, resulting in smaller deviation from linearity.

An underestimate or overestimate of uncertainty can be made if a linear least square fitting is performed in combination with observational data if the functional relation is in fact non-linear (see Section 4.4).

In the case of indirect dynamical emergent constraints, a relationship between a predictor Y_1 and a predictand Y_2 is found. Assuming the predictor Y_1 is again a response to some forcing, we can repeat the analysis above for direct constraints for a system of two dimensions, where a forcing is added in one direction. Mutatis mutandis, a condition very similar to Equation 3.27 is found, as

$$\frac{A(\Delta Y_1(t))(\omega_2)}{A(\Delta Y_2(t))(\omega_1)} = \frac{\sum_{l=1}^{\infty}\frac{g_l}{\sqrt{\lambda_l^2+\omega_2^2}}}{\sum_{l=1}^{\infty}\frac{h_l}{\sqrt{\lambda_l^2+\omega_1^2}}} = C^{st}, \tag{3.31}$$

where g_l and h_l are defined as in Equation 3.24. For an emergent constraints to exist,

the projection terms of the different observables should thus change in a similar fashion under the change in parameter.

Static direct constraints link the mean of an observable (predictor) to a change in the system under a specific forcing (predictand). Note that the susceptibility only contains information about the response to such forcing. Even in the limit of $\omega \to 0$, it denotes the linear response of the system, without any information on the mean state (Lucarini and Sarno, 2011). So, to derive the condition for a linear relationship the mean $E[O_e(X_t)] = \int_{-\infty}^{\infty} \bar{p}_e O(x)\, dx$ and the susceptibility at frequency ω_1 are used.

For static emergent constraints, the linear relationship between the predictand and the predictor is not expected to pass through the origin, since the predictor will in general be nonzero. Therefore, an additional term I is added to the ratio, denoting the intercept of the line between the predictor's mean state and the predictand. Instead, the susceptibility is compared to the mean state and the following condition is derived, where C^{st} should again be a constant independent of parameter(s) that is used to generate the ensemble:

$$\frac{E[O_{1t}] - I}{A(\Delta O_2(t))(\omega_1)} = \frac{\int_{-\infty}^{\infty} \bar{p}_e O_1(x)\, dx - I}{\sum_{l=1}^{\infty} \frac{2}{\sigma} \frac{h_l}{\sqrt{\lambda_l^2 + \omega_1^2}}} = C^{st}. \tag{3.32}$$

Again C^{st} can either be positive or negative, depending on the physics under consideration. This equation is both valid for direct and indirect static emergent relationships; in the case of a direct constraint $O_1 = O_2$ and the term h_l contains information about the response of O_1 to a forcing, while in the indirect case $O_1 \neq O_2$ and h_l contains information about O_2. As evident from Equation 3.32, our framework provides little insight on these types of constraints.

As an illustration of the theory from Section 3.2 and a direct dynamical emergent constraint, we take the Ornstein-Uhlenbeck process (OU process). Here $V'(x) = \gamma x$, where γ is a parameter that indicates the steepness of the potential. The eigenvalues and eigenfunctions of the generator are given by Pavliotis (2014)

$$\lambda_l = \gamma l\ ; \qquad \phi_l(x) = \frac{1}{\sqrt{l!}} H_n\left(\sqrt{\frac{2\gamma}{\sigma}} x\right), \tag{3.33}$$

where H_n are the Hermite polynomials. For the Ornstein-Uhlenbeck case, the ratio of

response amplitudes reduces to

$$SR(\gamma) = \frac{\beta_1/\sqrt{\lambda_1^2 + \omega_2^2}}{\beta_1/\sqrt{\lambda_1^2 + \omega_1^2}} = \frac{\sqrt{1 + (\omega_1/\gamma)^2}}{\sqrt{1 + (\omega_2/\gamma)^2}}, \tag{3.34}$$

since both the observable and the derivative of the potential are orthogonal to all eigenfunctions other than ϕ_1. This ratio is dependent on γ. In the case $\gamma \gg \omega_i$ for $i \in \{1, 2\}$ this ratio is nearly one and an emergent relationship is present for a model ensemble generated by varying γ.

3.4 Application to idealised climate models

From the previous sections, it appears that the computation of the eigensolutions of the generator of the dynamical system are central to determine whether an emergent constraint will appear or not. In this section, we will provide examples using idealised climate models.

The eigenvalues and eigenfunctions of the generator were numerically determined using the fact that the eigenvalues of the Fokker-Planck operator $\mathscr{L}^*$ are equal to those of the generator and that the eigenfunctions can be computed from the transformation: $\phi_l = \phi_l^*/\bar{p}_e$. The Fokker-Planck operator was discretised using the Chang-Cooper algorithm (Chang and Cooper, 1970). Eigenvalues and eigenvectors were determined using an Implicitly Restarted Arnoldi Method (Lehoucq et al., 1998). Explicit simulations of the SDEs were performed using a stochastic Runge-Kutta method (Kloeden and Platen, 1992).

3.4.1 Ornstein-Uhlenbeck cases

First, the one-dimensional Ornstein-Uhlenbeck process is considered with SDE

$$dX_t = (-\gamma X_t + F_i(t))dt + \sqrt{\sigma} dW_t, \tag{3.35}$$

forcing $F_i(t) = \sin 2\pi t\omega_i$ and frequencies $\omega_1 = 0.001$ and $\omega_2 = 0.1$. A parameter ensemble is created by varying γ. In this case, analytic solutions exist for the eigenvalues and eigenvectors of the generator. Eigenvectors and eigenvalues were determined using the Chang-Cooper scheme on a domain $[-25, 25]$ with $\Delta x = 0.25$. The numerically computed susceptibilities, as shown in Figure 3.2b, are in agreement with the analytic ones and capture the response (Figure 3.2a) well, as expected in this linear case. In the two-dimensional Ornstein-Uhlenbeck case, the same forcing $F_i(t)$ is added but

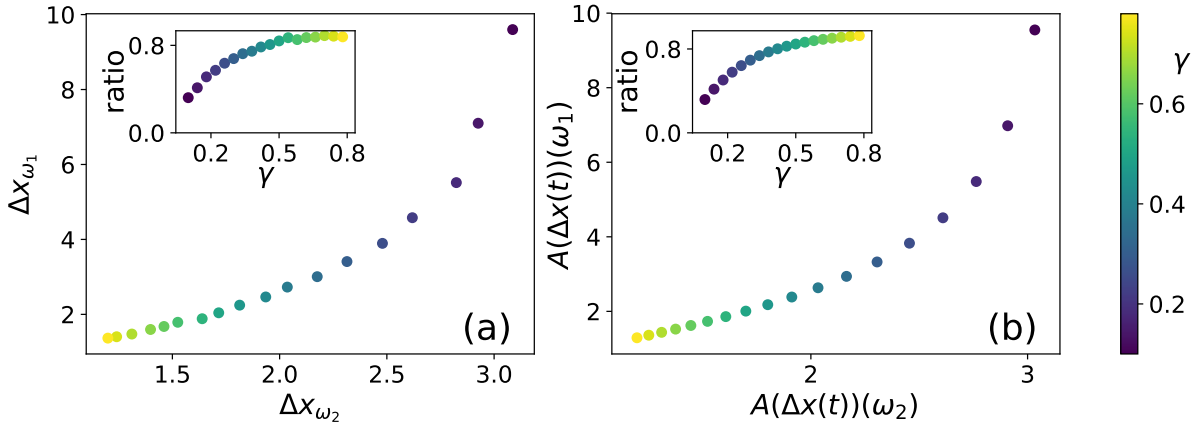


Figure 3.2: (a) Response to forcings at two different frequencies of the one-dimensional Ornstein-Uhlenbeck process. Shown is the average of a 500-member simulation of trajectories (b) The susceptibility at these frequencies, whose ratio is given in the inset figure. This is an example of a direct dynamical emergent relationship.

only in the first dimension. The governing SDE is given by

$$dX_t = \left[\begin{pmatrix} -\gamma_1 & \delta \\ \delta & -\gamma_2 \end{pmatrix} X_t + \begin{pmatrix} F_i(t) \\ 0 \end{pmatrix} \right] dt + \sqrt{\sigma} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} dW_t, \tag{3.36}$$

and a parameter ensemble is generated by changing the damping rate γ_1. Two ensembles are compared with $\delta = 0.2$ in the first ensemble and $\delta = 0.5$ in the second. The damping term γ_2 is held constant at $\gamma_2 = 0.6$.

In Figure 3.3, the eigenvalues and susceptibility ratios are plotted. In the case of a relatively weak coupling ($\delta = 0.2$) all nonzero eigenvalues are larger than the fast forcing frequency ω_2, so the system response time is smaller than the forcing time scales. On the other hand, the strong coupling ($\delta = 0.5$) leads to a slow down of the system, so that some eigenvalues now become smaller than ω_2. In these cases ($\gamma_1 < 0.5$) the system does not have time to portray the full response to a forcing, while for others ($\gamma_1 > 0.5$) it does. Consequently, the strength of the response actually decreases for $\gamma_1 < 0.5$. Directly calculating the expected value as the mean of 500 stochastic trajectories confirms this result (not shown). The results in Figure 3.4 show a large variation over the ensemble in the projection term of the predictor on the eigenfunctions (g_l). In contrast, the product of the two projection terms in the predictand (h_l) changes relatively little over the ensemble for both coupling strengths. Even though the projection terms now play a significant role in determining the response, the eigenvalues still determine whether the relation is linear (fast compared to forcing) or nonlinear (similar size to forcing frequency). In the weak-coupling system, the susceptibility ratio

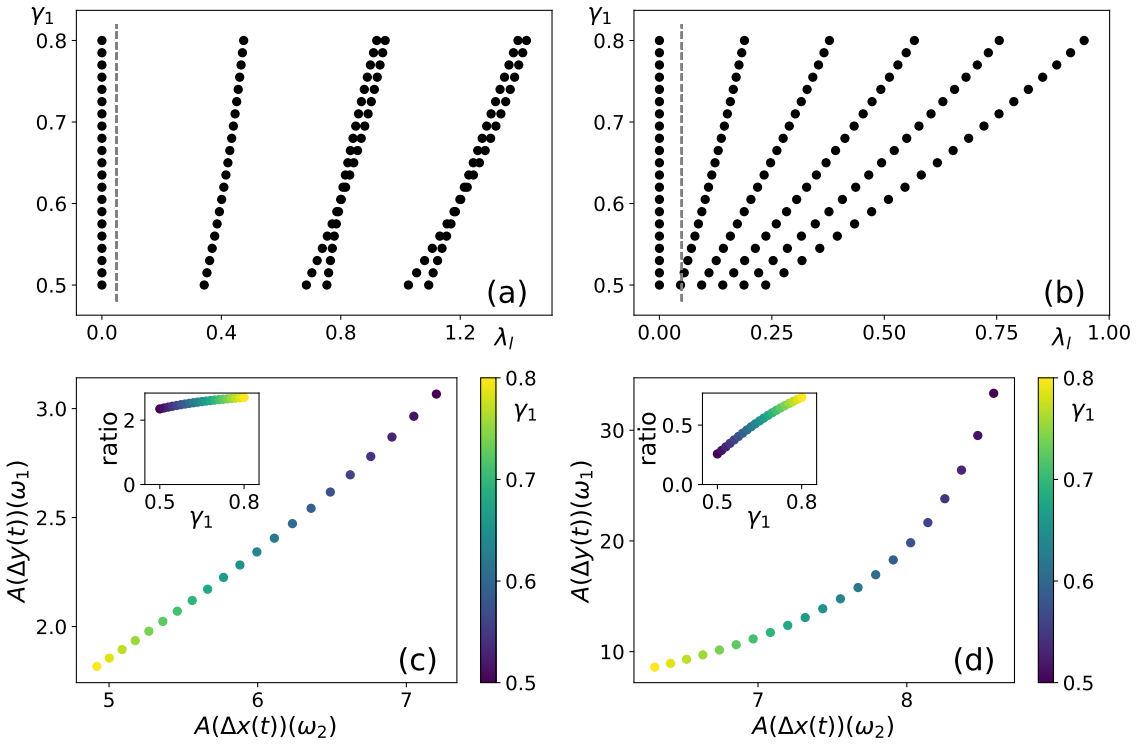


Figure 3.3: Eigenvalue spectrum for (a) $\delta = 0.2$ and (b) $\delta = 0.5$. The dashed line corresponds to the frequency ω_2 of the fast forcing. (c,d) Corresponding susceptibilities, with their ratio in the inset figures. This is an example of an indirect dynamical emergent relationship. Note that for reasons of numerical stability, the range of γ_1 is different than that of γ in Figure 3.2.

is almost constant and an emergent linear relationship is found. The strong-coupling system does only portray an emergent relationship for certain regimes (low or high γ_1). A case can be made though that the highly coupled system is the system for which finding an emergent constraint is more likely, because the strength of the response is substantially higher and a better signal-to-noise ratio can be obtained.

3.4.2 Energy balance model

In this section the snow-albedo feedback (SAF) first described by Hall and Qu (2006) is examined in more detail. They found a correlation between SAF on a seasonal scale and SAF under global warming. In models with a high snow albedo, the contrast between snow-covered and bare surfaces was largest and consequently the sensitivity to changes in temperature was largest (Qu and Hall, 2007). To study this emergent constraint we modify a simple energy balance model based on the seminal work by Budyko (1969) and Sellers (1969). The albedo is made temperature-dependent, following Fraedrich (1979) and a stochastic term is added following Sutera (1981). A

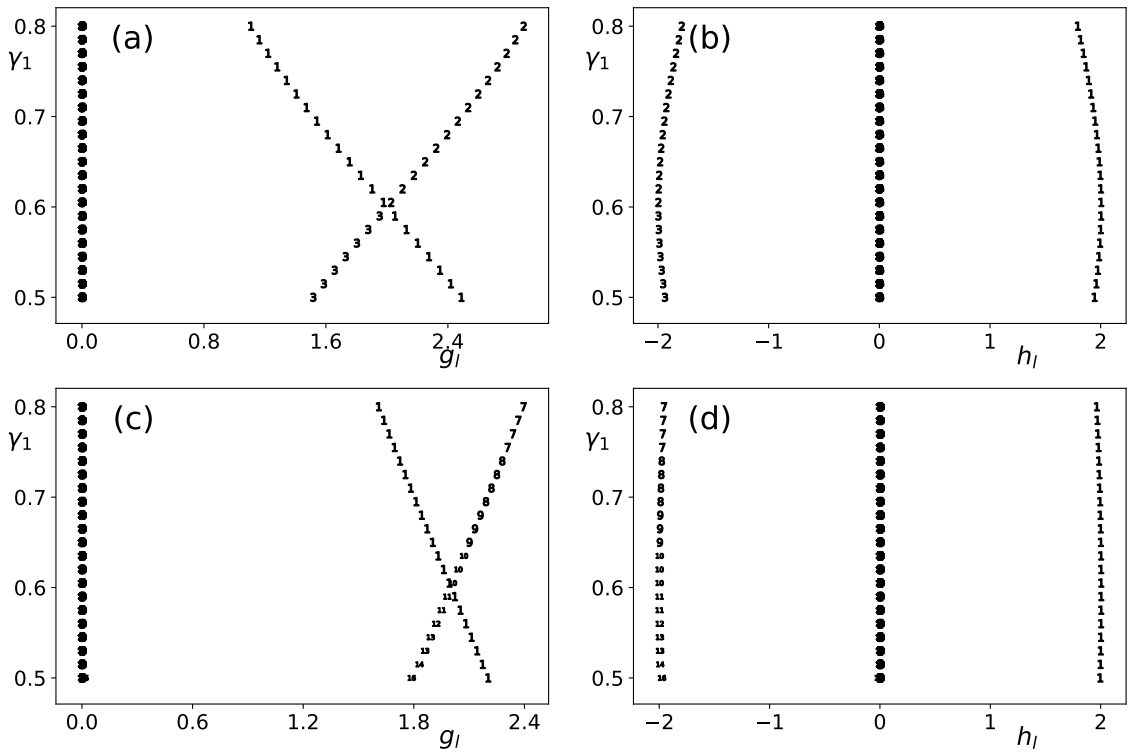


Figure 3.4: (a,b) Projections g_l (of predictor variable) and h_l (of predictand variable) for a weakly coupled two-dimensional OU system with $\delta = 0.2$, (c,d) same for $\delta = 0.5$.

parameter in the albedo function will be used to define a parameter ensemble.

With constant albedo, the energy balance model reads:

$$dT = \frac{1}{c_T}\left(Q(1-\alpha) + A\ln\frac{C}{C_{ref}} + G - \epsilon\sigma_B T^4\right)dt + \sqrt{\sigma_T}\,dW_t, \tag{3.37}$$

where dT is the temperature change, c_T the atmospheric heat capacity, Q the solar insolation, α the albedo, C the concentration of greenhouse gases, C_{ref} a reference concentration, G represents the radiative forcing due to the reference greenhouse gas concentration, σ_B the Stefan-Boltzmann constant and ϵ the emissivity of the Earth. The standard parameter settings for this model can be found in Table 3.2. The parameters of the albedo function are chosen to ensure that no bistability is present in the model, in which case LRT would break down.

Before examining the snow-albedo feedback, note that for some variables, notably the climate sensitivity, a simple EBM can react differently to forcing from solar insolation or greenhouse gases. This can be determined from, with $H = G + A\ln\frac{C}{C_{ref}}$ and

Constant	Value
c_T	5.0×10^8 J/m^2/K
A	20.5 W/m^2
Q_0	342 W/m^2
Q_s	115 W/m^2
G	150 W/m^2
C_{ref}	280 ppmv
τ_s	4.0×10^6 s
σ_α	1.0×10^{-5} s^{-1}

Constant	Value
ϵ	1.0
σ_B	5.67×10^{-8} W/m^2/K^4
α_{min}	0.2
α_{amp}	0.05–0.5
k	0.5
T_h	284 K
σ_T	2.0×10^{-7} K^2/s

Table 3.2: Constants for the energy balance model and the insolation function.

for a value of $\epsilon = 1$,

$$\begin{aligned} \frac{\partial}{\partial \alpha}\frac{\partial T}{\partial Q} &= \frac{\sigma^{1/4}}{4\left(Q(1-\alpha)+H\right)^{3/4}}\left(\frac{3Q(1-\alpha)}{Q(1-\alpha)+H}-1\right) < 0; \\ \frac{\partial}{\partial \alpha}\frac{\partial T}{\partial H} &= \frac{3Q\sigma^{1/4}}{16\left(Q(1-\alpha)+H\right)^{7/4}} > 0 \end{aligned} \tag{3.38}$$

Sensitivity to greenhouse forcing decreases when albedo increases, while sensitivity to solar insolation (seasonal sensitivity) increases for an increasing albedo, using typical values for Q and H.

To mimic the physical mechanism behind the emergent constraint, the albedo is taken to be temperature dependent, i.e., for low (high) temperatures, the albedo is high (low). A logistic function is used to model this effect,

$$\alpha(T) = \alpha_{\text{min}} + \frac{\alpha_{\text{amp}}}{1+\exp\left(k(T-T_h)\right)} \tag{3.39}$$

where α_{min} is the minimum albedo, α_{amp} is the amplitude, k is a steepness factor and T_h is the temperature at which half of the amplitude is reached. The amplitude α_{amp} is the parameter that is varied over the ensemble.

In the first case, the insolation forcing is given by $Q = Q_0(1 + Q_s \sin 2\pi t/\tau)$ where τ is one year and Q_s a seasonal modulation amplitude, with parameter values as shown in Table 3.2. The snow-albedo feedback term is then computed by dividing the amplitude of the albedo cycle by the amplitude of the temperature cycle. A second case is considered in which the greenhouse gas concentration C is increased 0.3% per year from 295 ppmv over a period of 300 year. Here the snow-albedo feedback is computed by dividing the total albedo response by the total temperature response. In each case, the variance of the noise σ_T in Equation 3.37 was chosen as 10^{-7} K^2/s.

Changing this parameter does not influence the eigenvalues as expected from the theory (Pavliotis, 2014). While the projections of the eigenvalues and eigenfunctions did change slightly, the susceptibility ratio was not influenced significantly by a variation of σ_T (halving and doubling of σ, not shown). In the computation of the solution of the Fokker-Planck equation using the Chang-Cooper scheme, we used a resolution of 1 K which is sufficient to accurately determine the eigenvalues and eigenfunctions of the generator.

As mentioned above, application of Equation 3.27 is not self-evident. Considering temperature to be a forcing ignores the fact that temperature responds differently to seasonal and greenhouse gas forcing, as shown in Equation 3.38. Secondly, using $d\alpha/dT$ as the observable directly does not work either. Linear response theory does not give the expected value of the observable, but the expected value of the deviation due to the forcing, while we are interested in the change due to a parameter change.

Instead, the SAF can be described by two variables: SAF is determined by taking the ratio of the susceptibilities of albedo to temperature. Therefore, we use the modified Equation 3.28:

$$\begin{aligned} \mathsf{RFS}(\alpha_{\mathsf{amp}}) =& \frac{A(\Delta\alpha(t)|_Q)(\omega_2)}{A(\Delta\alpha(t)|_C)(\omega_1)} \frac{A(\Delta T(t)|_C)(\omega_1)}{A(\Delta T(t)|_Q)(\omega_2)} \\ =& \frac{\sum_{l=1}^{\infty} \frac{\alpha_l}{\sqrt{\lambda_l^2+\omega_2^2}}}{\sum_{l=1}^{\infty} \frac{\gamma_l}{\sqrt{\lambda_l^2+\omega_1^2}}} \frac{\sum_{l=1}^{\infty} \frac{\delta_l}{\sqrt{\lambda_l^2+\omega_1^2}}}{\sum_{l=1}^{\infty} \frac{\beta_l}{\sqrt{\lambda_l^2+\omega_2^2}}} = C^{st}, \end{aligned} \tag{3.40}$$

where

$$\begin{aligned} \alpha_l &= \langle \alpha, \phi_l \rangle_{\bar{p}_e} \langle (1-\alpha(T))V'(T), \phi_l \rangle_{\bar{p}_e}, & \gamma_l &= \langle \alpha, \phi_l \rangle_{\bar{p}_e} \langle V'(T), \phi_l \rangle_{\bar{p}_e} \\ \beta_l &= \langle T, \phi_l \rangle_{\bar{p}_e} \langle (1-\alpha(T))V'(T), \phi_l \rangle_{\bar{p}_e}, & \delta_l &= \langle T, \phi_l \rangle_{\bar{p}_e} \langle V'(T), \phi_l \rangle_{\bar{p}_e}. \end{aligned} \tag{3.41}$$

In the case the susceptibilities are all dominated by one term with index l, this reduces to $C^{st} = (\alpha_l \delta_l)/(\beta_l \gamma_l) = 1$

In Figure 3.5 the sensitivity of temperature to varying amplitude of the albedo function is shown, as well as the sensitivity of the snow-albedo feedback and condition for the existence of an emergent constraint. As shown in Figure 3.5a, no emergent relationship is found for climate sensitivity, a feature that was analytically found in the case of constant albedo. In Figure 3.5b the emergent constraint on SAF is shown. In the warm regime (low albedo, lower line in the figure), the SAF becomes larger for larger α_{amp}. The larger the maximum albedo, the steeper the logistic albedo function. A second effect also takes place: the higher the maximum albedo, the warmer it gets. Consequently, sensitivity of the albedo function is smaller. This decrease in sensitivity

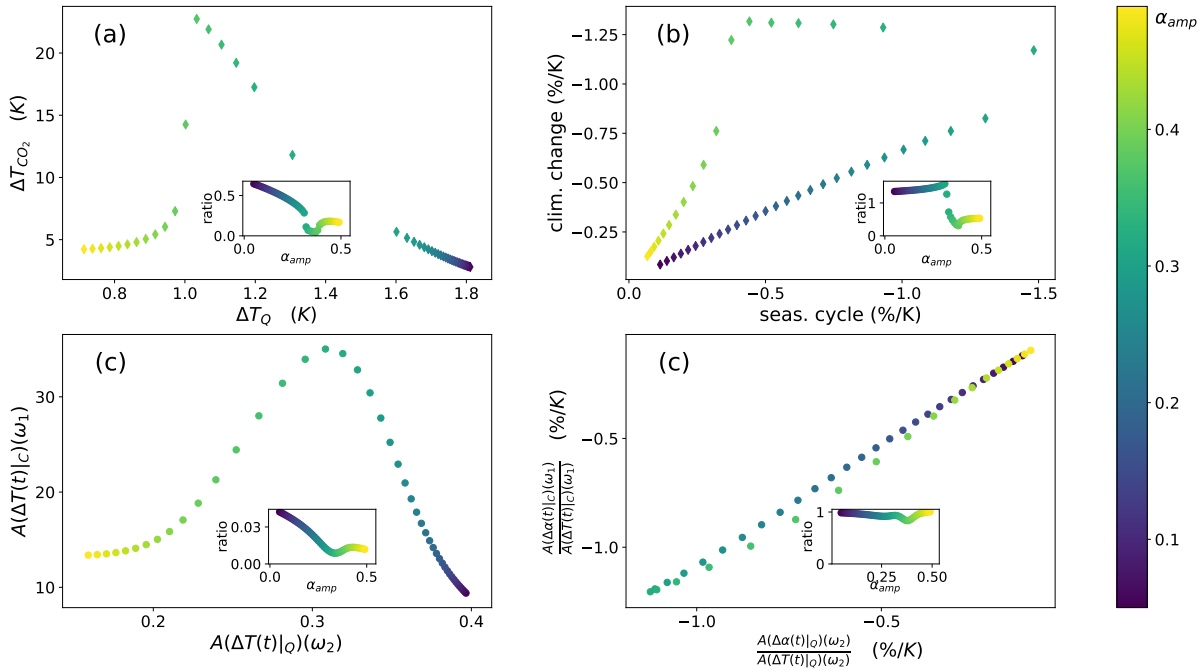


Figure 3.5: (a) The relation between temperature response to the seasonal cycle and the temperature response to greenhouse gas forcing. (b) The strength of the snow-albedo feedback to solar and greenhouse gas forcing on different time scales. In the inset: their ratio as a function of α_{max}. For clarity, (a,b) are shown without noise. (c) The susceptibilities for temperature as the observable (d) The ratio of albedo and temperature susceptibilities and their ratio (RFS).

also takes place in the cold regime; the colder it gets, the less sensitive the albedo gets. In the cold regime it is clear that this second mechanism dominates. The results can be reproduced by use of LRT, as shown in Figure 3.5c and 3.5d. The discrepancies disappear when forcing is small; the climate change forcing in particular is causing most of the differences.

One can extend the energy balance model by representing the response of snow and ice explicitly as a relaxation towards the logistic reference albedo function $\alpha(T)$ given in Equation 3.39. This gives the extended model

$$\begin{aligned} dT =& \frac{1}{c_T}\left(Q(1-\bar{\alpha}) + H - \epsilon\sigma T^4\right)\, dt + \sqrt{\sigma_T}\, dW_t \\ d\bar{\alpha} =& -\frac{1}{\tau_s}\left(\bar{\alpha} - \alpha(T)\right)\, dt + \sqrt{\sigma_{\bar{\alpha}}}\, dW_t, \end{aligned} \tag{3.42}$$

where $\tau_s = 4 \times 10^6$ s is the response time of the albedo. The drift term in the Fokker-Planck equation corresponding to Equation 3.42 is not the gradient of a potential but the eigensolutions of the generator can still be computed numerically.

Extending the model with an explicit albedo function does not change the dynamics

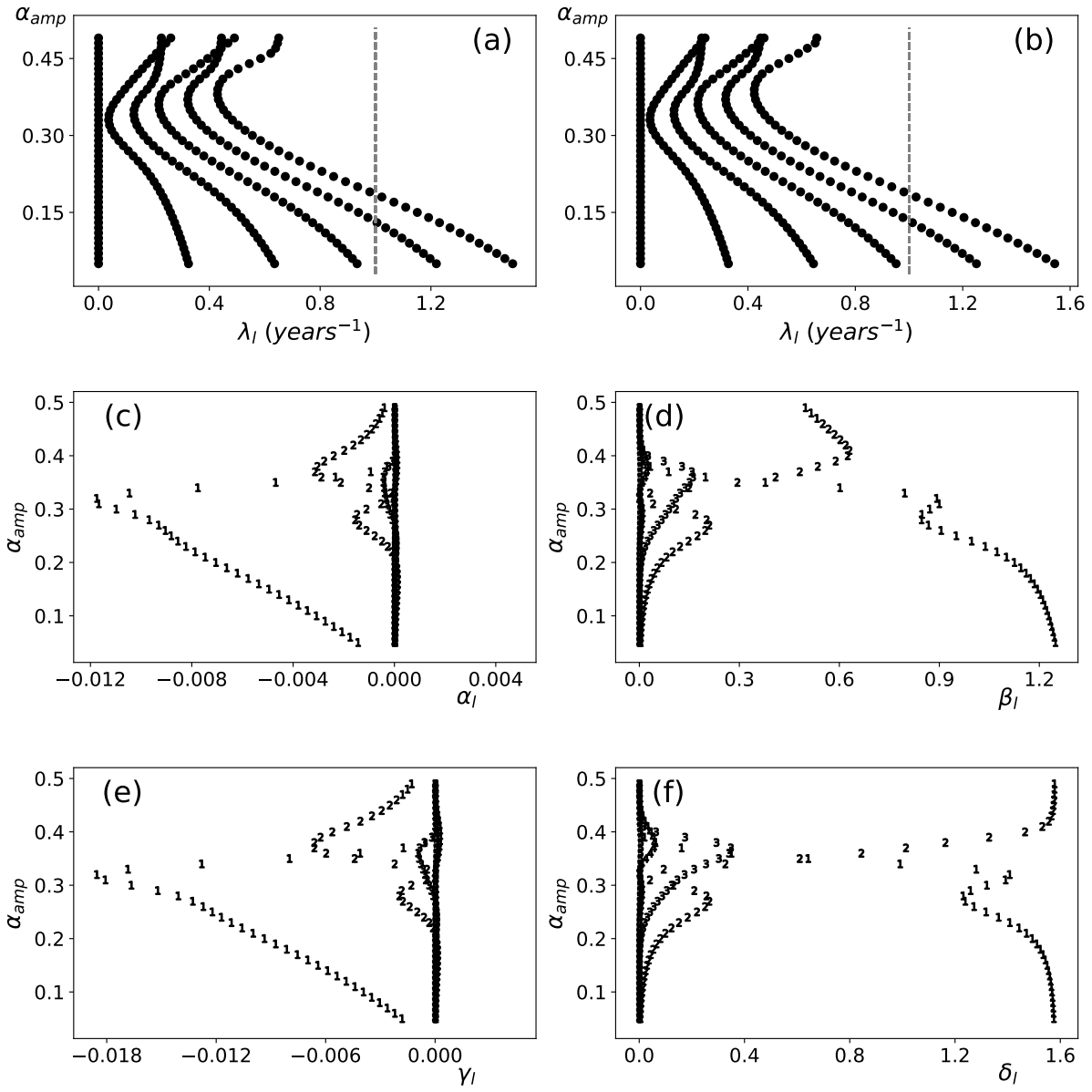


Figure 3.6: (a) Eigenvalues of the EBM depending on the amplitude of the albedo function for the simple EBM. The zero eigenvalues correspond to the invariant measure. (b) The extended EBM. (c) Albedo projection terms for solar forcing (α_l) as defined in Equation 3.41 where the markers denote l (d) Same for temperature β_l (e,f) projections terms for GHG forcing γ_l and δ_l respectively.

of the system significantly, nor the eigenvalues and eigenvectors. Figure 3.6b shows the eigenvalues of the extended EBM to be almost exactly equal to the eigenvalues of the original model, the imaginary parts continuing to be zero. The projection coefficients are very similar as well (not shown). Thus, the inclusion of a smaller time scale does not improve the response.

3.4.3 A model of intermediate complexity: PlaSim

To bridge the gap between perturbed parameter ensembles in simple dynamical systems and Earth System Models, the SAF emergent constraint is further examined in PlaSim. PlaSim is a numerical model of intermediate complexity, developed at the University of Hamburg to provide a fairly realistic present climate which can still be simulated on a personal computer (Fraedrich et al., 2005). The atmospheric dynamics are modelled using the primitive equations formulated for temperature, vorticity, divergence and surface pressure. Moisture is included by transport of water vapour. The equations are solved using the spectral method. A full set of parameterizations is used for unresolved processed such as long and shortwave radiation with interactive clouds, boundary layer fluxes of latent and sensible heat and diffusion.

In this climate model snow albedo is a function of surface temperature T_s, snow depth and vegetation cover. The bare soil snow albedo in PlaSim is described by:

$$A_{snow} = \begin{cases} A_{\text{max}}, & \text{if } T_s \leq 10 \text{ ℃.} \\ A_{\text{min}}, & \text{if } T_s > 0 \text{ ℃.} \\ A_{\text{min}} - (A_{\text{max}} - A_{\text{min}})\frac{T_s}{10} & \text{otherwise.} \end{cases} \tag{3.43}$$

This equation is modified in the presence of vegetation and in the case of shallow snow depth. See Lunkeit et al. (2011) for more details. A set of simulations was performed with A_{max} varying between 0.650 and 0.900. The historical forcing in PlaSim was approximated by a CO_2 increase from 295 ppm at a rate of 0.3% per year in the 20th century and 1% per year in the 21st century before it stabilised at 720 ppm; a 50-year spin-up corresponding to the period 1850–1900 was used.

In the right panel of Figure 3.7 the PlaSim results are shown which can be compared to the results from Hall and Qu (2006) in the left panel. Note that the variation in CMIP3 is significantly larger than the variation found in PlaSim, but that the PlaSim results fit on the relation found by Hall and Qu (2006). Variations in other parameterizations, such as the maximum snow albedo over forested regions, increase the spread in PlaSim SAF further (not shown). This simulation shows that the constraint that emerges in a multi-model ensemble with structurally different formulations of the snow response can to some extent also be reproduced using variations in one parameter. This provides the justification for simplifying further to energy balance models to examine the SAF emergent constraint.

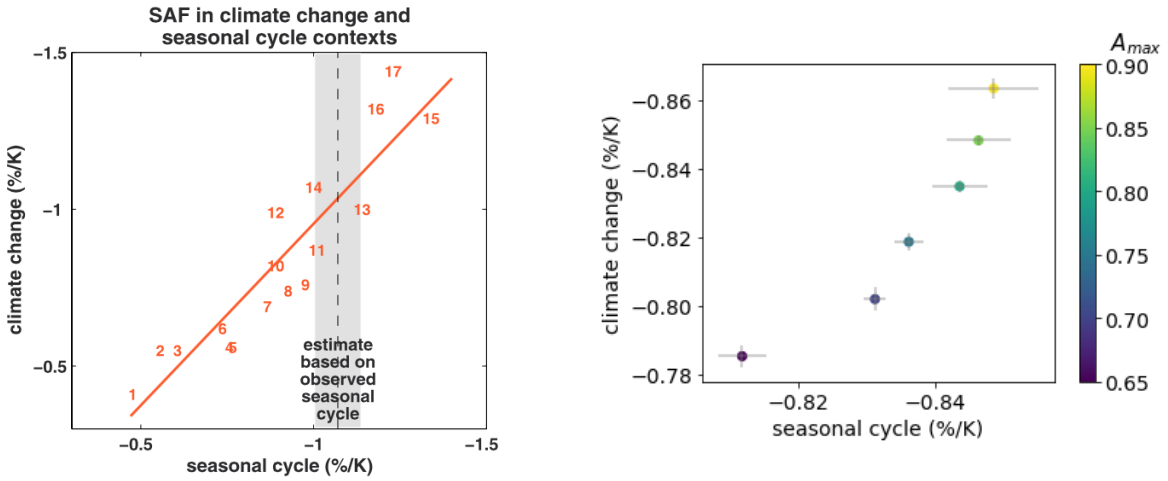


Figure 3.7: Left panel) The emergent constraint on snow-albedo feedback $\frac{\Delta\alpha_s}{\Delta T_s}$ (from Hall and Qu (2006), α_s given in units of %). This is an example of a direct emergent constraint (it links the SAF in both past and future time) and a dynamical emergent constraint (it uses a response to a seasonal forcing as its predictor). Right panel) Same as left, but now results from *PlaSim*

3.5 Summary, discussion and conclusions

In this chapter, we have presented a dynamical framework behind the occurrence of emergent constraints in parameter dependent stochastic dynamical systems. In these systems, emergent constraints are related to ratios of response functions which can be determined using linear response theory. It was shown that for a large class of systems, these ratios could be expressed in terms of eigenvalues and projections on eigenvectors of the generator of the system.

A classification of emergent constraints was given and several types could be distinguished depending on whether similar (direct) or different (indirect) variables are considered and whether a response in present-day climate (dynamical) or the time-independent part of present-day climate (static) is linked to a response of the future climate system. For a linear dynamical emergent constraint, the ratio of susceptibilities at the two frequencies under consideration should be a positive constant over the ensemble. When the response is computed with respect to an internal variable (in contrast to an external forcing), a condition is posed on the susceptibilities of the two variables in the system. Static constraints are encountered when a linear relationship is found between the expected value of the observable and the susceptibility at the frequency of the forcing.

Examples were given using several idealised climate models. In particular the emergent constraints involving the snow-albedo feedback was considered in detail.

We found that linear dynamical emergent relationships can occur when the time scale of the system, indicated by the eigenvalues, changes with the parameter and is smaller than the forcing time scales. This is of particular interest because differences in response size between climate models is often determined by feedbacks strength in climate systems. Larger feedbacks give rise to larger timescales (Roe, 2009), which is reflected in the eigenvalues of the generator. For an emergent constraint on a feedback quantity there is a more complicated constraint mechanism, where one has to take into account the response to two different variables, which typically have different time scales. When the condition of the predictor's time scale being smaller than the forcing time scale is not met, deviations from linearity occur. If linearity of the relation is exploited in further analysis, such as in the quantitative interpretation of emergent constraints by Wenzel et al. (2014), this might lead to a bias in the estimate of the predictand.

Modelling emergent constraints with conceptual models is justified when different ESMs are closely related and structural differences can be parametrized. This can for instance be tested using an intermediate complexity model with full parametrization of the process under consideration.

The classification of emergent constraints provided gives a hint to which kind of emergent constraints one can look out for in an ensemble of high-dimensional Global Climate Models (GCMs). Searching for emergent constraints using blind data mining my lead to spurious and misleading correlations (Caldwell et al., 2014). Using the susceptibilities to find new emergent constraints does not seem to have a direct advantage above directly looking for plausible correlations, but susceptibilities might provide additional information. For example, when a susceptibility shows a resonance at a certain frequency over the ensemble of models, this could suggest that the same feedback is present in all simulations.

In a high-dimensional dynamical system eigenfunctions and eigenvalues can be accessed with the help of transfer operators, associated with the propagation of probability densities associated with the Fokker-Planck operator. The eigenfunctions that lie on the invariant measure are then computed by making use of the ergodic properties of the climate system. To overcome the burden of high-dimensionality, a reduced transfer operator can be computed from a very long simulation, from which the eigenfunctions on the attractor are approximated (Tantet, 2016). However, a forcing on the system does not generally lie only on the attractor and should be split into a part parallel and perpendicular to the attractor. Consequently, the eigenvectors *off* the attractor cannot a priori be ignored (Lucarini and Sarno, 2011). Gritsun and Lucarini (2017) showed that indeed for some geophysical systems, specifically quasi-geostrophic flow

with orographic forcing, the response to the forcing may have no resemblance to the unforced variability in the same range of spatial and temporal scales. Lembo et al. (2020) successfully derived a response function for the global near-surface temperature the Atlantic Meridional Overturning Circulation the Atlantic Circumpolar current and the global heat uptake. They noted however that more than their available twenty initial value members are needed to reliably estimate the susceptibility.

In conclusion, while the current theoretical framework provides an understanding on how emergent constraints may arise in low-dimensional stochastic dynamical systems, its application to output from GCMs, in particular in finding novel and useful emergent constraints, is a challenging issue for future work.

In the next chapter we will discuss the second part of the theoretical underpinning of emergent constraints: the statistical foundations.

Chapter 4

Statistics of emergent constraints

A summary of Section 4.3 is included in Cox et al. (2018b). Adaptations of Section 4.4 and 4.7 are included in the Williamson et al. review paper, which I co-authored (item 5 of Associated papers)

This chapter details statistical considerations for emergent constraints. In essence, there are two "ingredients" in the statistical analysis: an estimate of the observational uncertainty and the prediction interval of the emergent relationship. In the following sections, the emergent relationship is denoted as $Y = f(X) + \epsilon$, with X the predictor, Y the predictand (future quantity), f the functional form (typically linear) and ϵ an error term. Emergent constraints on CMIP ensembles are necessarily done using a small sample of models. It is therefore of vital importance to use the proper statistical techniques, to prevent putting too much confidence in the results and to exploit all available information.

The chapter starts with a simple analytical derivation of the confidence interval of a linear emergent constraint with normally distributed uncertainty around the observation. The second section expands on this by introducing a hierarchical statistical model, which can include multiple simulations per model. The next section details why an intercept term may appear in an emergent constraint even if not physically expected. The chapter continues with notes on nonlinear constraints and on non-normality of the confidence interval of the observations. The chapter ends on a review about a potential discrepancy between models and reality, and a review and discussion of methods to combine multiple constraints. The statistics are exemplified using the emergent relationships found in Cox et al. (2018a) and Nijsse et al. (2019a), as well as generated synthetic data.

4.1 Propagation uncertainty

Both the observation and the emergent relationship itself carry uncertainty. This section describes the derivation of the combined confidence interval, subject to a weak approximation on the uncertainty in the emergent relationship. We start with an emergent relationship between a predictor X and predictand Y: $Y = a + bX$, with a the intercept and b the slope.

We take an observation with normally distributed uncertainty: $p(x) = \mathcal{N}(x \mid X_{\text{obs}}, \sigma_{\text{obs}})$, which is equivalent to:

$$p(x) = b\mathcal{N}(bx \mid bX_{\text{obs}}, b\sigma_{\text{obs}}). \tag{4.1}$$

In the quantitative interpretation, emergent relationships are assumed to represent a conditional probability of Y, given X:

$$p(y \mid x) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left\{-\frac{(y-a-bx)^2}{2\sigma_y^2}\right\} \tag{4.2}$$

with the prediction error σ_y given by $\sigma_y(x) = s\sqrt{1 + \frac{1}{N} + \frac{x-\bar{x}}{N\sigma_x^2}}$, with σ_x^2 the sample variance of X. The root mean square error s is computed as: $s^2 = \frac{1}{N-2}\sum_{n=1}^{N}(y_n - f(x_n))^2$ (Cox et al., 2018a). Rewriting the conditional probability:

$$p(y \mid x) = \mathcal{N}(y - bx \mid a, \sigma_y). \tag{4.3}$$

The convolution of two normal distributions $f(x) = \mathcal{N}(x \mid \mu_f, \sigma_f)$ and $g(x) = \mathcal{N}(x \mid \mu_g, \sigma_g)$ is given by:

$$\int f(x)g(y-x)dx = h(y), \tag{4.4}$$

where $h(y) = \mathcal{N}\left(y \mid \mu_f + \mu_g, \sqrt{\sigma_f^2 + \sigma_g^2}\right)$. This can be verified with a Fourier transform, as the convolution theorem states that the Fourier transform of the convolution of two functions is equal to the product of their Fourier transforms.

The probability distribution around the predictand is found by marginalising out X. If we approximate σ_y as constant, reasonable for moderate to large values of N, this amounts to a convolution over bX with solution:

$$
\begin{aligned}
p(y) &= \int_{-\infty}^{\infty} p(y \mid x) p(x) dx \\
&= \int_{-\infty}^{\infty} b\mathcal{N}(bx \mid bX_{\text{obs}}, b\sigma_{\text{obs}})\mathcal{N}(y - bx \mid a, \sigma_y)\frac{1}{b}\mathrm{d}bx \\
&= \mathcal{N}\left(y \mid bX_{\text{obs}} + a, \sqrt{\sigma_y^2 + (b\sigma_{\text{obs}})^2}\right).
\end{aligned} \tag{4.5}
$$

4.2 Incorporating initial value members

Climate models have internal variability. Part of that variability can be explained by their chaotic nature: whenever you run the same model twice with small differences in initial condition there can be significant differences in the state of the climate system after decades to centuries (Ghil and Lucarini, 2020). To quantify the effects, many modelling groups choose to perform multiple simulations of their models with the same forcing, but a different initial (pre-industrial) climate state (Deser et al., 2020). Here we explore whether we can exploit this by developing a method using the information from all initial value simulations.

We chose a Bayesian modelling framework because of flexibility and transparency: assumptions are naturally made transparent. Including the initial value ensemble members as separate models would be incorrect, as they are strongly dependent. Instead, to include the initial value ensemble, we assume that each model has an unknown true X_T as its predictor. We further assume that every simulation of a model gives a value of X that drawn from a normal distribution with mean X_T and a standard deviation σ_x that is the same across all models. Every simulation of a model gives us more insight into X_T. Our Bayesian model then becomes:

$$
\begin{aligned}
X_{m,j} \mid X_m, \sigma_x &\sim \text{normal}(X_m, \sigma_x) && \text{// modelling } X_T \text{ using } N \text{ simulations} \\
Y_m \mid \alpha, \beta, \sigma_y &\sim \text{normal}(\alpha + \beta X_m, \sigma_y) && \text{// modelling } Y \text{ using } X_T \text{ from } N_g \text{ models}
\end{aligned}
$$

The second layer corresponds to normal linear regression with prediction error σ_y, while the first layer makes an estimate of the true X for each model. Note that for models with only one initial value member, this true X does not necessarily correspond with that the value of this single ensemble member.

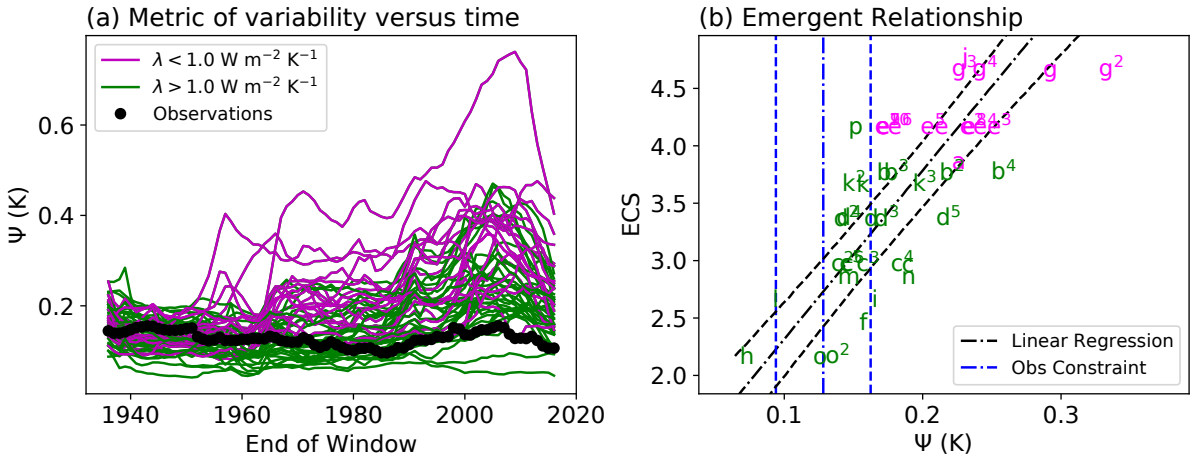


Figure 4.1: Linear regression as described above using all ensemble members. The prediction interval is relatively small, because it denotes the likely area of the Ψ_T, not each individual Ψ. The gradient is 14.5 K K^{-1}, while the intercept is reduced to 0.86 K.

$$Y_{\text{pred}} = \text{normal}\left(\alpha + \beta\,\text{normal}\left(X_{\text{obs}}, \sqrt{\sigma_x^2 + \sigma_{\text{obs}}^2}\right), \sigma_y\right)$$

As an example, we apply this statistical model to the data of Cox et al. (2018a). As priors we choose the following distributions:

$$\begin{aligned}
\alpha &\sim \text{normal}(0, 5) && \text{// prior intercept}\\
\beta &\sim \text{normal}(12, 10) && \text{// prior gradient}\\
\sigma_y &\sim \text{half-normal}(0.2, 10) && \text{// prior uncertainty regression}\\
\sigma_x &\sim \text{half-normal}(0.02, 0.1) && \text{// prior uncertainty independent variables}
\end{aligned}$$

The priors on α and β are weakly informed and stem from theoretical considerations (Cox et al., 2018a). The other priors carry very little information. We fit this model via MCMC using Stan (Carpenter et al., 2017). The method gives a slightly lower estimate of climate sensitivity compared to Cox et al. (2018a), with a likely range of 2.1–3.3 K of warming. The results of the regression are shown in Figure 4.1.

Ideally we would use a less computationally intensive method to do emergent constraints. Therefore, we have performed a set of experiments to investigate under which circumstances, if any, this method is superior to ordinary least squares using a single simulation per model or the mean value of multiple simulations. We also com-

pare it to a errors-in-variables technique called orthogonal distance regression (ODR) (Schennach, 2016). Orthogonal distance regression minimizes the error orthogonal to the regression line. We implement it like Jiménez-de-la Cuesta and Mauritsen (2019): first taking the average of all ensemble members per model, then performing ODR. For OLS and ODR, the uncertainty in X includes both a measuring uncertainty σ_{obs} and a model sample uncertainty corresponding to internal variability σ_{var}, so that $\sigma_{\text{tot}} = \sqrt{\sigma_{\text{obs}}^2 + \sigma_{\text{var}}^2}$.

We answer the following questions:

1. Which method predicts (median) Y with least error?
2. Which method has the best estimate of the prediction interval of Y?
3. Which method produces the best estimates of the slope and the intercept and under which conditions?

These questions are answered with synthetic data generated to resemble the distribution of data common in different ensembles used for emergent constraints. The synthetic data is generated as follows. First, for each model a true predictor X_m is drawn from $\mathcal{N}(0, 1)$. Then, the predictand Y is computed as $Y_m = X_m + \mathcal{N}(0, 0.4)$ for each model m. As X_m are not known due to measurement error and internal variability, in the next step we specify $X_{m,j} = X_m + \mathcal{N}(0, 0.4)$ for each initial value ensemble member j. Pseudo-observations are drawn from the same relationship: $Y_{\text{obs}} = X_{\text{m,obs}} + \mathcal{N}(0, 0.4)$. The observed X is then measured with error from internal variability and measurement: $X_{\text{obs}} = X_{\text{m,obs}} + \mathcal{N}(0, 0.4) + \mathcal{N}(0, 0.04)$.

First we investigate the effect of varying the number of initial value members per model. In both CMIP5 and CMIP6 there is a very skewed distribution of initial value members: many models only have one member while others have up to 50. For the test, we define four categories of increasing ensemble size. In category I, models have either one or two members with equal chance. Category II mimics CMIP5 by assigning each pseudo-model as many members as a random CMIP5 model (corresponding to Table A.1). Similarly for category III, each pseudo-model is assigned as many members as a random CMIP6 model (Table 6.1). In category IV each model has twenty ensemble members.

Figure 4.2 shows that ODR is typically best at getting the intercept (in this case the intercept with the line $x = -2$) and slope correct. However, ordinary least squares, which is optimized to reduce the error in Y, scores better in the predictand Y and also gives a significantly better confidence interval. ODR overestimates certainty in its prediction. The Bayesian hierarchical method typically scores comparably to linear

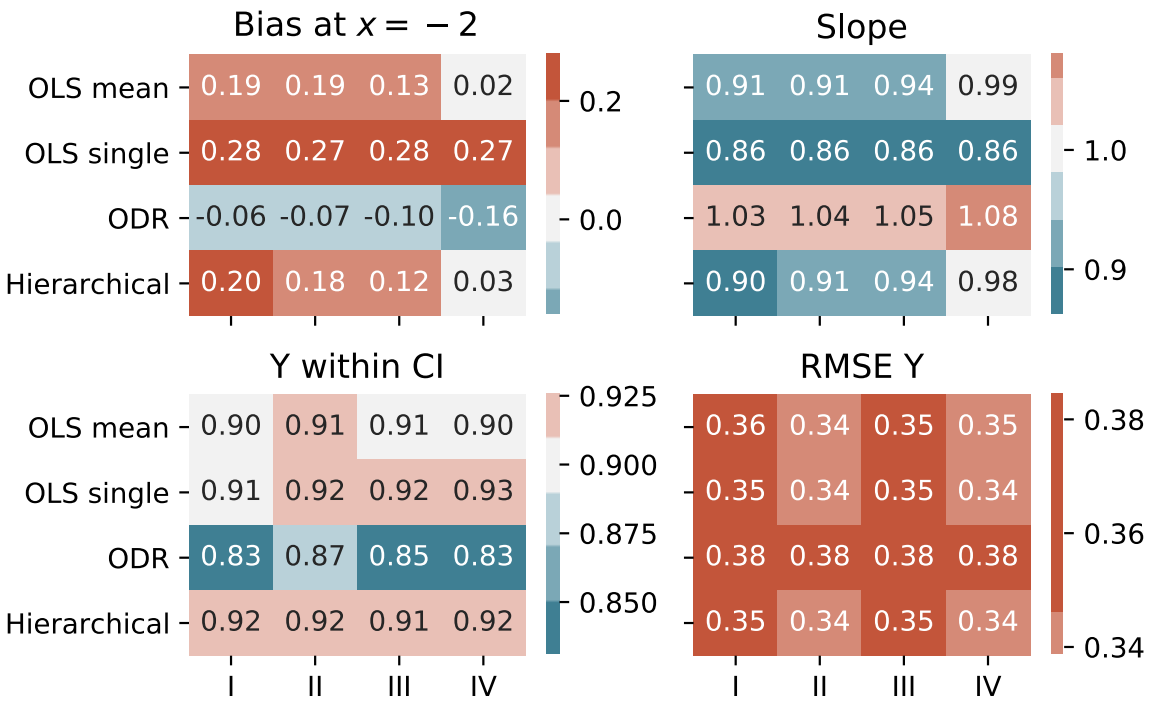


Figure 4.2: Comparison of accuracy linear regression with measurement error. a) The bias of estimating Y at $x = -2$, b) estimation of the slope, which should be 1. c) Fraction of Y_{obs} that lies in the 5–95% confidence interval (for linear regression and ODR) or credible interval (for Bayesian regression), d) The root mean square error of the median estimate of Y, compared to Y_{obs}.

regressions using the mean values, and is therefore useful if extensions are needed on the model, f.i., it can be easily extended to be non-Gaussian around the regression line. Normality around the regression line is a condition of OLS that some emergent constraints do not seem to meet perfectly (Bretherton and Caldwell, 2020). ECs on biochemical quantities often have fewer models available. Figure 4.3 shows the comparison in the case there are only eight models available. While the root mean square error is clearly larger in this case, the various biases are not significantly larger with fewer models. A small difference in performance between OLS (using mean values) and the hierarchical Bayesian method does become apparent, with the former performing better.

4.3 Presence of an intercept

In Cox et al. (2018a), an emergent constraint was found between equilibrium climate sensitivity (ECS) and a variable Ψ. The predictor Ψ includes information about tem-

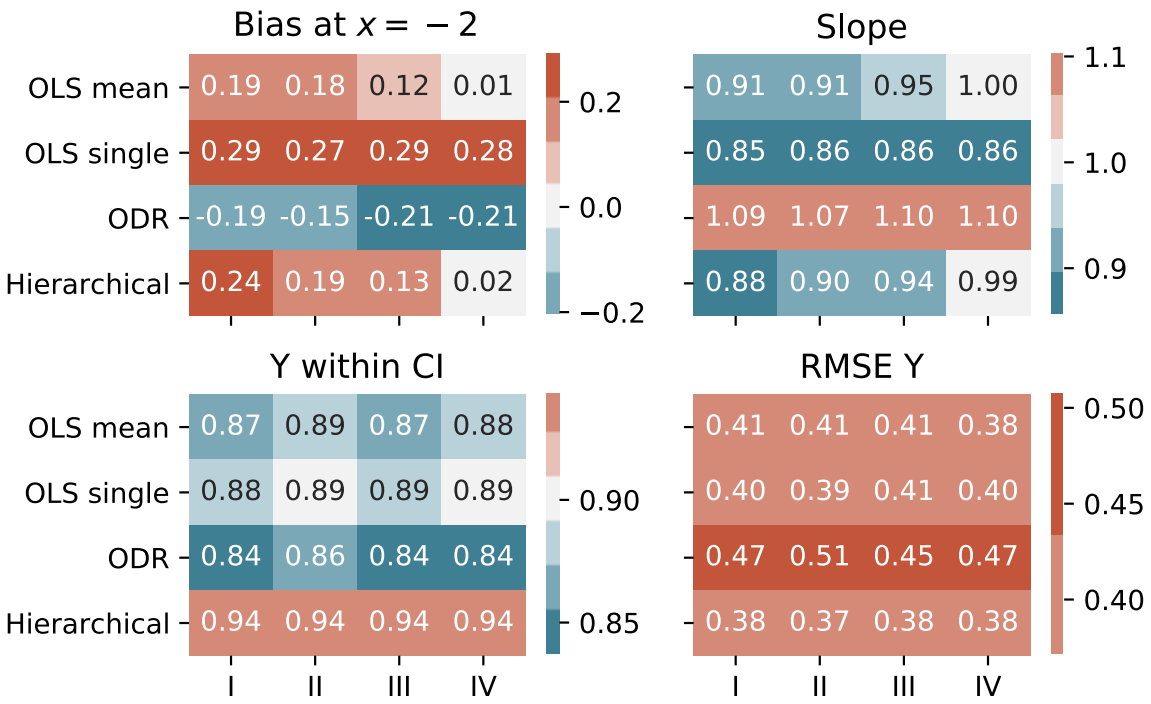


Figure 4.3: Comparison of accuracy linear regression with measurement error in an ensemble with eight models. Panels a)–d) are similar to Figure 4.2.

perature variability and autocorrelation: $\Psi = \sigma_T / \sqrt{-\ln \alpha_{1T}}$. The relationship $\mathrm{ECS} = \frac{F_{2\times}}{\sigma_Q} \frac{\sigma_T}{\sqrt{-\ln \alpha_{1T}}}$ was derived analytically using the Hasselmann model: a zero-order conceptual model of climate (Hasselmann, 1976):

$$C\frac{dT}{dt} = -\lambda T + Q, \tag{4.6}$$

with C the heat capacity, λ the climate feedback parameter, which is inversely related to equilibrium climate sensitivity, T surface temperature, and Q forcing (external and also representing internal forcing as white noise). Written as a stochastic differential equation, and without external forcing:

$$CdT = -\lambda T dt + \sigma_Q dW, \tag{4.7}$$

where σ_Q is the variance of the forcing noise and dW is the derivative of the Wiener process, representing white noise. The factor $\frac{F_{2\times}}{\sigma_Q}$ was found to be independent of ECS. It was excluded from the predictor as observational data was not available.

Importantly, as noted by Rypdal et al. (2018), there is no intercept in the analytically derived relationship between Ψ and ECS, whereas an intercept of 1.23 K was found in Cox et al. (2018a). The example described above is not an exception: the emergent

constraint between observed warming and transient climate response also features an intercept where the analytical solution showed none (see Section 6.2.6).

Five phenomena that give rise to this intercept term were identified, two of which are of a statistical nature. The magnitudes of each of these phenomena are estimated by using a large set of simulations of the Hasselmann model and a two-layer model, that will be introduced later. The results of this are shown in Table 4.1. Three of the contributions give rise to a positive intercept, while the others produce a negative intercept.

1. Regression dilution When there is variance in the independent variable, rather than only in the dependent variable, regression dilution takes place: the estimate of the slope is biased towards zero. Due to finite time effects, Ψ has substantial uncertainty, as is shown in Figure 4.1b.

In order to make predictions using regression (so not to determine the true physical relation between the variables), it is not necessary in general to correct the dilution bias unless the error in the observational constraint is unequal to the errors in the independent variable (Frost and Thompson, 2000).

Using an initial value ensemble, a first order correction for regression dilution can be performed. Using the hierarchical Bayesian regression model described in Section 4.2, we perform the regression with all the members of the initial value ensemble (N=49). The intercept is reduced from 1.23 K to 0.89 K.

2. Additional variance The theoretical relationship between Ψ and ECS has a factor $F_{2\times \mathrm{CO}_2}/\sigma_Q$. Similarly to classical regression dilution, this also leads to a dilution of the slope, as the extra variance is multiplicative, instead of an extra additional noise term on Y.

3. Finite time effects. Due to finite time effects the autocorrelation has a biased estimate (Foster et al., 2008), specifically an underestimation of time scales and the autocorrelation. The larger the timescale of the model system, the larger this underestimation is. This asymmetry in bias leads to a negative intercept in the linear fit.

4. Deep ocean heat uptake The one-layer Hasselmann model does not include a representation of deep ocean heat uptake. A second conceptual model of the climate system, a two-layer model, is used to study whether this simplification has an impact on the intercept.

	Simulation length	Ψ–$\frac{1}{\lambda}$	Ψ–ECS
1. Regression dilution	138 year	0.51 K W^{-1} m^2	2.18 K
2. Additional variance	5000 year	–	0.71 K
3. Finite time effects	138 year	–0.21 K W^{-1} m^2	0.57 K
4. Use of two-layer model	5000 year	–0.92 K W^{-1} m^2	–0.03 K
5. Whiteness assumption	5000 year	0.07 K W^{-1} m^2	0.98 K
Combined first four	138 year	0.40 K W^{-1} m^2	1.87 K

Table 4.1: Magnitude of intercept terms. All simulations use values of ECS, λ and γ corresponding to the 16 models used in Cox et al. (2018a). The detrending window was 55 year. Regression dilution is computed using the average of 200 simulations. Finite time effects and the effects of the two-layer model were computed by first accounting for regression dilution: for each value of λ, we used the average Ψ over 500 simulations before performing linear regression.

$$\begin{aligned} CdT &= (-\lambda T - \gamma(T - T_0))dt + \sigma_Q dW; \\ C_0 dT_0 &= \gamma(T - T_0)dt. \end{aligned} \tag{4.8}$$

The zero subscript denotes the deep ocean layer and γ is a heat exchange parameter.

Analysis of the two-layer model shows that Ψ may be better related to $\frac{1}{\lambda+\gamma}$, with the deep ocean heat uptake parameter γ playing a similar role as the feedback parameter λ (Williamson et al., 2018). By using $\frac{1}{\lambda}$ and a finite time series in the two-layer model, the intercept term turns strongly negative. For $\frac{1}{\lambda+\gamma}$ this becomes close to zero.

The effects of the additional variance term and the one-box simplification almost completely compensate. The regression does almost go through zero when Ψ is plotted against $\frac{1}{\lambda+\gamma}$. To get similar values of Ψ, we have increased the value of the noise σ_Q from 0.2 to 0.4 in the two-layer model compared to the single-layer model.

5: Reddening of the noise spectrum In the analytic derivation of Ψ–$\frac{1}{\lambda}$, it was assumed that the noise is white. Violation of this assumption leads to a small intercept term. Analysis of σ_N, the variation of the TOA net radiation, which can be considered a proxy for σ_Q, shows a positive autocorrelation. For Table 4.1, noise is generated with an autocorrelation time scale of one year.

Adding up the first four factors leads to a positive intercept. Using the two-layer

model, the last intercept cannot be ignored and is of similar order as the first three terms. All results are produced by using the λ, $F_{2\times CO_2}$ and γ values of the 16 models used in Cox et al. (2018a), with the latter values from Geoffroy et al. (2013b).

Overall, for the Ψ–$\frac{1}{\lambda}$ relationship, regression dilution and use of two-layer model contribute most to the intercept. In the case of the Ψ–ECS relationship, the additional variance caused by the radiative forcing parameters further contributes to a positive intercept. Again, regression dilution has a major impact. The finite time effects not related to internal variability, and the assumption of whiteness on the radiative forcing have smaller effects. The combination of the different factors over-explains the intercept of 1.23 K. There is a very large variation in the intercept, and the 1.23 K falls in the 5–95% confidence interval.

4.4 Nonlinear emergent constraints

As illustrated in Figure 4.4, an incorrect assumption of linearity can lead to incorrect estimates of the probability density function of the predictand. If in reality, the relationship were quadratic, the estimate of the confidence interval can be highly biased. For low (high) values of the independent variable, the true slope of the quadratic relationship is less steep (steeper) than the linear relationship. Consequently, there will be both a bias in the median prediction and a possible underestimate or an overestimate of uncertainty in the prediction. This well-known feature is illustrated in Figure 4.4. It is therefore important that a good statistical model is chosen. Reduced-form modelling of the emergent constraint under consideration can give clues what the proper statistical model is, but is not conclusive.

In Chapter 5 an example of an emergent constraint is shown that whose analytical reduced-form equations show a small nonlinearity. The emergent constraint details the relationship between the variation of temperature trends and climate sensitivity. It is examined in two different reduced-form models: the Hasselman model and the two-layer model. Both models show an emergent relationship, but the relationship in the two-layer model is slightly more nonlinear. The choice whether to include a nonlinear term in the statistical model can be made using information criteria. In Chapter 5, we use the Akaike information criterion (AIC) (Myung et al., 2009), which is a criterion for the trade-off between goodness of fit and the simplicity of model. The criterion is:

$$\mathrm{AIC} \;=\; 2k - 2\ln(\hat{L}), \tag{4.9}$$

where k is the number of fitted parameters and $\hat{L}$ the maximum of the likelihood func-

tion. The model with the smallest AIC is considered best. The inclusion of an additional term will improve a model, but each additional parameter increases the risk of overfitting.

4.5 Normality assumption on observation

In the previous sections, the distribution of the observational uncertainty due to internal variability was assumed to be normal. However, Ψ values computed from simulated T in the Hasselmann model portray a heavy-tailed distribution. To find a better distribution, we compare five distributions: the normal distribution, the lognormal distribution, the gamma distribution, the generalized extreme value distribution, and the inverse gamma distribution. For each of these distributions the parameters were computed using 2000 simulations of the Hasselmann model spanning the λ values of the CMIP5 ensemble. The best distribution was chosen using the Kolmogorov–Smirnov test (Massey Jr., 1951). The skewness of the distribution depends on the value of λ: for large λ, the distribution is closer to normal. There was no single best distribution, but the lognormal distribution scored high for most λ. The fitted shape parameter Ψ_{shape} of the lognormal for the smallest λ (i.e. $0.6\,\mathrm{K\,W^{-1}\,m^2}$) is 0.20. The influence of this is tested in the model below:

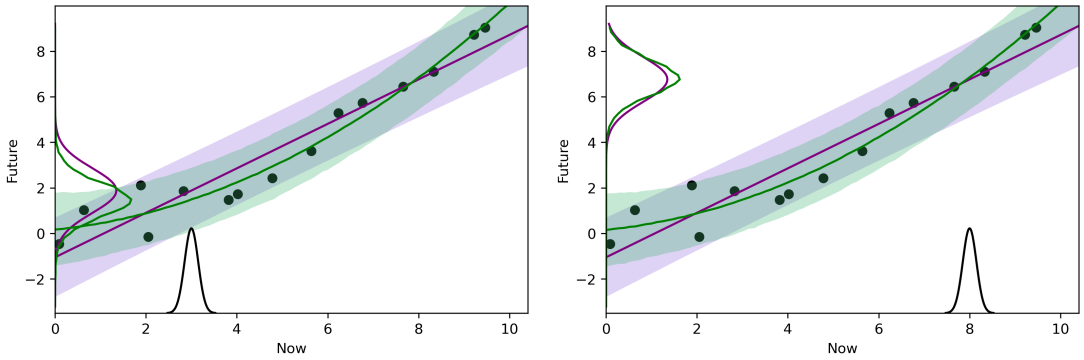


Figure 4.4: Schematic comparing a linear fit to a quadratic fit to compute a probability density distribution for Y consistent with observations. Synthetic data is created using a quadratic relationship between X and Y. (a) Using a linear fit results in an overestimation of the uncertainty when the observations line up with the shallow part of the quadratic. (b) This effect is less pronounced at the steep part of the graph. Note also that in the quadratic case, the probability density distribution is asymmetric and has skew.

$$\Psi_{\text{obs}} \sim \textbf{lognormal}(\Psi_{\mu}, \Psi_{\text{shape}}) \qquad \text{// model for observed } \Psi$$
$$\text{ECS} \sim \text{normal}(\alpha + \beta\Psi, \sigma) \qquad \text{// model for observed ECS}$$

$$\text{ECS}_{\text{pred}} = \text{normal}(\alpha$$
$$+ \beta\, \textbf{lognormal}(\Psi_{\mu}, \Psi_{\text{shape}}), \sigma) \qquad \text{// prediction ECS}$$

The priors are chosen similar to Section 4.2, with a very weak prior on the observational Ψ.

$$\alpha \sim \text{normal}(0, 10) \qquad \text{// prior intercept}$$
$$\beta \sim \text{normal}(12, 10) \qquad \text{// prior gradient}$$
$$\sigma \sim \text{half-normal}(0, 10) \qquad \text{// prior uncertainty regression}$$
$$\Psi_{\mu} \sim \text{normal}(0, 100) \qquad \text{// prior true observational } \Psi$$

As expected, this mostly influences the upper bound of the emergent constraint. Using an upper estimate for the skewness of the observational constraint, the new likely range is 2.1 to 3.6 K, compared to the range of 2.2 to 3.4 K when taking the observational uncertainty of Cox et al. (2018a).

4.6 Model discrepancy

A further assumption often employed in the use of emergent constraints, is that the real world is just a 'random sample' from the distribution of possible models, or in Bayesian terminology, that the real world is interchangeable with any single model and follows the linear regression found from analysing model output. Williamson and Sansom (2020) add two terms to their Bayesian linear regressing denoting the discrepancy between reality and the emergent relationship, one for the intercept and one for the slope. The priors for these values are chosen based on physical argumentation. For instance, one could argue that physical understanding of the emergent constraint is better if it does not change much between generations of climate models and that informative priors are justified.

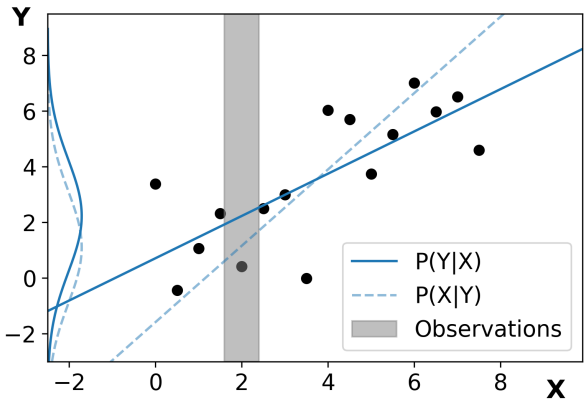


Figure 4.5: Standard, ordinary least squares linear regression $p(y \mid x)$ compared with the reverse regression $p(x \mid y)$. If the latter were to be used as the relationship on which the emergent constraint is founded, the final constraint would have a bias.

4.7 Combining multiple constraints

A very simple method to combine multiple constraints was used in Brient (2020). He used a Gaussian kernel density estimation (kde) on a histogram of the mean values of the posteriors of previous constraints. The variance of the posteriors were informing the weights of the Gaussian kernel density estimation. This method suffers from multiple drawbacks. Most prominently it does not capture increasing confidence from having multiple independent constraints. Imagine you add an identical EC with the same mean and variance, the posterior kde estimate would not change. Instead it is determined by the kernel bandwidth, which is subjectively chosen. It further does not take into account to what extent the different emergent constraints are related to each other.

In Bretherton and Caldwell (2020) multiple emergent constraints were combined 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. As there is strong collinearity in the eleven emergent constraints on climate sensitivity they evaluate, regularization was needed. A second 'method U' (for Uncorrelated) used a smaller subset of ECs, that were regarded as confirmed constraints. As collinearity was less important with fewer X this method simplifies C by dropping the covariance ECs. Both variants were extended to account for overconfidence in the EC: they scale the ratio of the explained to unexplained variance with a factor $\alpha^2 \leq 1$, reducing all correlation coefficients.

Renoult et al. (2020) proposed a further simple method to combine independent ECs to create a tighter estimate for Y. Where the regression is normally given as

$p(y \mid x)$, they propose to instead formulate the statistical model as $p(x \mid y)$, allowing for a prior $\pi(y)$ on Y to be integrated into the emergent constraint as

$$p(y) = \int_{-\infty}^{\infty} p(x \mid y)\pi(y)p(x)dx. \tag{4.10}$$

$\pi(y)$ here can be the posterior distribution of a previous EC. To make sure the two ECs are independent, the authors state that observations need to be independent, and that insofar 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 other methods described above. Linear regression is typically not symmetric; regression where X predicts Y, $p(y \mid x)$, describes a different function than regression where Y predicts X, $p(x \mid y)$ (Smith, 2009), as illustrated in Figure 4.5.

The basic statistical model used in emergent constraints is simple, and an analytical solution was derived in the chapter. Additional information from initial value ensemble members can improve estimates of the slope and intercept without sacrificing accuracy compared to ordinary least squares regression. A Bayesian framework can easily be adjusted to relax key assumptions on linearity and normality, which impact tail probabilities significantly. Further adjustments include explicitly modelling model discrepancy and including multiple constraints.

Chapter 5

Emergent relationships informing compound risk: a case of temperature variability and sensitivity

This chapter is adapted from the Nature Climate Change letter Nijsse et al. (2019a): "Decadal global temperature variability increases strongly with climate sensitivity", written in collaboration with Peter Cox, Mark Williamson and Chris Huntingford

5.1 Compound risks

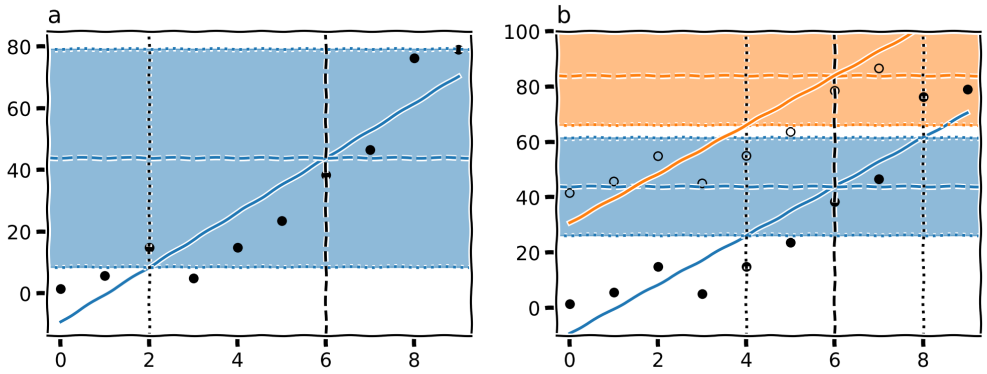


Figure 5.1: Two examples of emergent relationships that cannot be used as emergent constraints, but who can be used to assess compound risk. On the left, a case where the observational uncertainty is too large to effectively constrain the predictand. On the right, a case where quantitative interpretation of the emergent relationship is hampered by missing processes common to the models, leading to an unknown intercept.

The emergent constraint technique requires good observations for the predictor variable. These are not always available. There might not be a long enough record (Lutsko and Takahashi, 2018), it might be difficult to disentangle the record from a forced response (Brown et al., 2018; Cox et al., 2018b), or the observations might not be of sufficiently high quality (as in Figure 5.1a).

Alternatively, there might be no a good physical reason to constraint the intercept of the emergent constraint. Possibly, only the slope will have some strong physical basis, and no prior knowledge for the intercept is available. For instance, it may be known that certain long-term processes are not included in the model ensemble, which would only impact the predictand. In that case, the emergent constraint does not preclude any bias in the intercept and the predictand might be too high or too low (exemplified in Figure 5.1b).

In both cases, a quantitative use as an emergent constraint is not possible, but the relationship does contain information about the compound risk of the predictor and predictand: variability and sensitivity both contribute to risk. This chapter details an example of using an emergent relationship not to find a constraint, but instead to investigate such compound risk. It is shown that having a joint distribution of the variability and the final warming rate gives additional information important for climate risks.

5.2 Decadal global temperature variability increases strongly with climate sensitivity

Climate-related risks are not only dependent on the warming trend from greenhouse gases, but also on the variability about the trend. However, assessment of the impacts of climate change tend to focus on the ultimate level of global warming (United Nations General Assembly, 1992), only occasionally on the rate of global warming, and rarely on variability about the trend. In this chapter, it is shown that models which are more sensitive to greenhouse gas emissions (i.e. have higher equilibrium climate sensitivity (ECS)) also have higher temperature variability on time scales of several years to several decades.

Counterintuitively, high sensitivity climates, as well as having a higher chance of rapid decadal warming, are also more likely to have had historical “hiatus” periods than lower sensitivity climates. In this chapter, we show that cooling or “hiatus” decades over the historical period, which have been relatively uncommon, are more than twice as likely in a high ECS world ($\mathrm{ECS} = 4.5\,\mathrm{K}$) compared to a low ECS world

($\mathrm{ECS} = 1.5\,\mathrm{K}$). As ECS also affects the background warming rate under future scenarios with unmitigated anthropogenic forcing, the probability of a hyper-warming decade — over ten times the mean rate of global warming for the 20th century — is even more sensitive to ECS.

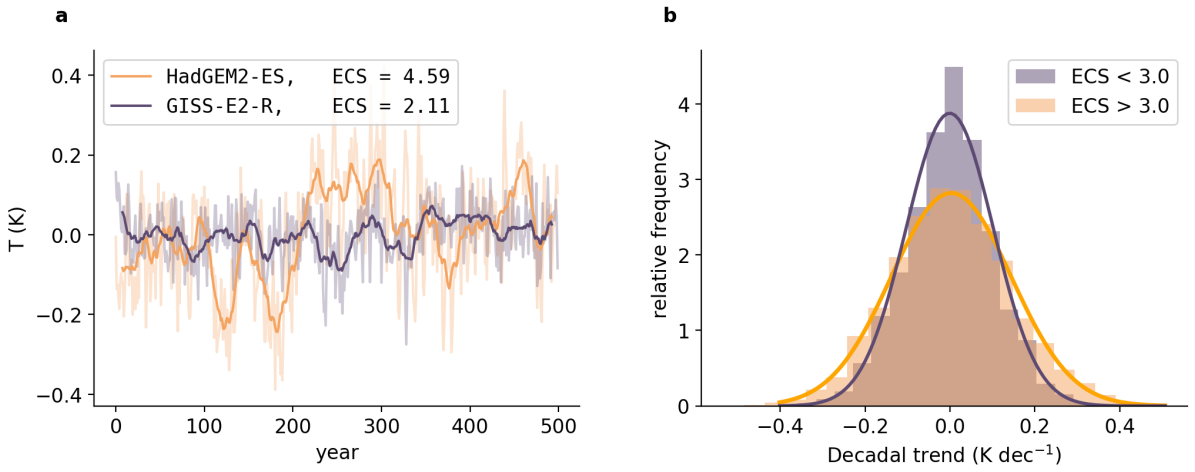


Figure 5.2: Decadal variability in global temperature. **a.** Global mean surface anomaly over a modelled 500-year period with no external forcing, for two control simulations in the CMIP5 database. HadGEM2-ES (brown line) is an example of a model with high climate sensitivity, while GISS-E2-R (purple line) has a low climate sensitivity. Heavy lines are 10-year running means. **b.** Multi-model histograms of decadal variability for low (purple) and high (brown) climate sensitivities in 500 year control simulations. Normal curves are fitted to histograms. The individual models are listed in Table 5.1 with their climate feedback parameter λ and equilibrium climate sensitivity ECS.

We look specifically at the combined effects of climate sensitivity and climate variability, which could stretch the ability of human and natural systems to adapt (Taylor et al., 2012; Stocker et al., 2013; Otto et al., 2013). Our approach is to study how decadal trends in global annual mean surface temperature vary with climate sensitivity across the CMIP5 multi-model ensemble (Taylor et al., 2012). The latter is partially motivated by the 2000-2012 slowdown of surface temperature increase, sometimes incorrectly dubbed the 'warming hiatus'. This slowdown has led some to suggest estimates of ECS below $1.5\,\mathrm{K}$ (Stocker et al., 2013; Otto et al., 2013). However, rather than making periods of no warming more likely for low climate sensitivities, we show the converse — that warming slowdowns can be expected more in high sensitivity climates.

The background to our claim is the well-known property that a more sensitive dynamical system responds to a perturbation more strongly and is slower to recover than a less sensitive one (Roe, 2009; Strogatz, 2000). Forcing from fast random, weather-

like perturbations, additional to slow anthropogenic forcing, can push the climate's temperature trend in both warm and cool directions. For more sensitive systems these excursions will be both larger and longer-lived, giving larger and longer-lived temperature trends.

We formalise this intuition by calculating the temperature trend b (K yr^{-1}) over a window of time W (yrs), usually a decade. How much the temperature trend varies (quantified as the standard deviation of b) with climate sensitivity is the main focus of study in this chapter. We follow the approach of using conceptual analytically solvable stochastic climate models to understand the climate system pioneered by Hasselmann and others (Hasselmann, 1976; Wigley and Raper, 1990). In particular, we solve for b and its standard deviation σ_b, using the Hasselmann model which describes the response of the annual global mean surface temperature anomaly ΔT (K) to forcing Q (W m^{-2}):

$$C\frac{d\Delta T}{dt} + \Delta T = Q \tag{5.1}$$

where Q parameterises fast, internally generated perturbations as a random variable. External driving factors such as anthropogenic forcing due to increases in greenhouse gases may also be included in this term.

The temperature response to the noise Q depends on the effective heat capacity C (W yr^{-1} m^{-2} K^{-1}) and the climate feedback parameter λ (W m^{-2} K^{-1}), the latter describing the net effect of all the individual negative and positive feedbacks within the climate. Climates with larger values of λ have a stronger overall negative (restoring) feedback on temperature anomalies and lower equilibrium climate sensitivity ECS. ECS is defined as the steady-state warming in response to the forcing from a doubling of atmospheric CO_2 and is inversely proportional to λ: $\mathrm{ECS} = Q_{2\times \mathrm{CO_2}}/\lambda$. Although the simple Hasselmann model is an imperfect representation of the climate system, it serves here to formulate a hypothesised relationship between variability and ECS, that we subsequently evaluate against the results from state-of-the-art Earth System Models.

Taking Q to be given only by stochastic forcing of magnitude σ_Q, analogous to an unforced, control climate model simulation, the Hasselmann model can be solved. To first approximation (the full expression is given in Section 5.3), this gives a relation for the standard deviation of b as a function of the trend length W and climate sensitivity:

$$\sigma_b = \frac{2\sqrt{3}\sigma_Q}{W^{2/3}\lambda} = \mathrm{ECS}\frac{2\sqrt{3}\sigma_Q}{W^{2/3}Q_{2\times CO_2}} \tag{5.2}$$

As expected, this equation predicts higher variability in warming trends (larger σ_b) for more sensitive climates (higher ECS). Although the single-layer Hasselmann

model is a poor representation of warming climates on long time-scales (Caldeira and Myhrvold, 2013), the two-layer model (Geoffroy et al., 2013b), which better describes oceanic heat storage, produces the same qualitative relationship (Equation 5.16). Wigley and Raper (1990) also noted the relation between long term temperature trends and climate sensitivity in numerical simulations of a stochastically forced upwelling diffusion model. Relating temperature variability to climate sensitivity can be thought of as a heuristic application of the Fluctuation-Dissipation theorem (Leith, 1975; Gottwald et al., 2016), a tool used in many fields of physics (Kubo, 1966). This way of modelling the response to a radiative forcing is complementary to methods that estimate λ as the sum of individual feedbacks. Metrics of variability derived from whole-system approaches can in principle be linked back to individual feedbacks (Lutsko and Takahashi, 2018; Caldwell et al., 2018).

5.3 Analytical relationship between ECS and the variability of trends.

A trend b of a W-yr time series is computed using an ordinary least squares fit of the timeseries. To obtain an analytical solution for the typical size of a trend b in the absence of external forcing, the standard deviation σ_b, we write the Hasselmann model as a stochastic differential equation (SDE) where Q is parametrized as a white noise process (the derivative of a Wiener process W) with standard deviation σ_Q.

$$CdT = -\lambda T dt + \sigma_Q dW \tag{5.3}$$

Using the Green's function associated with this Stochastic Differential Equation, namely $G_{Hasselmann} = \frac{1}{C} e^{-t/\tau}$, where τ is defined as $\tau = C/\lambda$, we write down the solution of temperature as a stochastic integral i.e.

$$T_t = \sigma_Q \int_0^t G_{Hasselmann}(t-s) dW_s. \tag{5.4}$$

Using this solution we find b and its standard deviation for a trend of W years. Multiple steps of algebraic manipulation, then lead to

$$\mathrm{Var}[b] = \frac{12\sigma_Q^2}{W^3\lambda^2}(1 - f(\tau, W)). \tag{5.5}$$

To first order, σ_b is linearly proportional to $1/\lambda$, which is in turn proportional to equilibrium climate sensitivity, defined as $\mathrm{ECS} = Q_{2\times CO_2}/\lambda$. The smaller W, however, the

larger the deviation towards nonlinearity. A similar result can be obtained when a deep ocean layer is added to give a two-layer model (Geoffroy et al., 2013b).

$$\begin{aligned} CdT &= (-\lambda T - \gamma(T - T_0))dt + \sigma_Q dW, \\ C_0 dT_0 &= \gamma(T - T_0)dt. \end{aligned} \tag{5.6}$$

Here the zero subscript denotes the deep ocean layer and γ is a heat exchange parameter. The Green's function for the top layer is (Geoffroy et al., 2013b):

$$G_T = \frac{1}{\lambda}\left(\frac{a_f}{\tau_f}e^{-t/\tau_f} + \frac{a_s}{\tau_s}e^{-t/\tau_s}\right), \tag{5.7}$$

with τ_f and τ_s denoting the fast and slow time scales and a_f and a_s the partial contribution of the fast and the slow mode to the response. Similarly to the Hasselmann model, σ_b is proportional to $1/\lambda$ to first order.

$$\mathrm{Var}[b] = \frac{12\sigma_Q^2}{\lambda^2 W^3}\left(1 - g(C, C_0, \lambda, \gamma, W)\right). \tag{5.8}$$

Proof

In mathematical terms, a trend is a simple linear regression between the variable of interest (here T) and time.

$$T = \alpha + bt + \epsilon, \tag{5.9}$$

with α the intercept, b the slope and ϵ an error term. The slope b (corresponding to the best estimate using ordinary least squares regression) is computed as:

$$b = \frac{\mathrm{Cov}(t, T)}{\mathrm{Var}(t)}. \tag{5.10}$$

Of interest is how the variance of b is linked to the variables: the feedback factor λ, the time scale τ (C/λ), the radiative forcing strength σ_Q and the window length W.

The covariance between t and T in a finite window is computed as:

$$\mathrm{Cov}[t, T] = \mathrm{E}[tT] - \mathrm{E}[t]\mathrm{E}[T], \tag{5.11}$$

where E denotes the expected value. To simplify the equation we estimate the parameter b in the time interval $[-\frac{W}{2}, \frac{W}{2}]$, so that the second term multiplying the means drops out.

The covariance of t and T for an unforced Hasselmnan model estimated over a

time window W is computed using the Green's function:

$$\mathrm{Cov}[t,T] = \mathrm{E}\left[\frac{1}{W}\int_{-\frac{W}{2}}^{\frac{W}{2}} tT_t dt\right]; \tag{5.12}$$

$$= \mathrm{E}\left[\frac{1}{W}\int_{-\frac{W}{2}}^{\frac{W}{2}} t\left[\frac{\sigma_Q}{C}\int_{-\infty}^{t} e^{-(t-s)/\tau} dW_s\right] dt\right]. \tag{5.13}$$

The variance of the time is given by:

$$\mathrm{Var}[t] = \frac{1}{W}\int_{-\frac{W}{2}}^{\frac{W}{2}} t^2 dt = \frac{W^2}{12} \tag{5.14}$$

Hasselmann model

Now computing the variance in the estimate of b over a period of W year:

$$\begin{aligned}
\mathrm{Var}[b] &= \mathrm{E}\left[\left(\frac{\mathrm{Cov}[tT_t]}{\mathrm{Var}[t]}\right)^2\right] \\
&= \mathrm{E}\left[\left(\frac{\frac{1}{W}\int_{-\frac{W}{2}}^{\frac{W}{2}} t\left[\frac{\sigma_Q}{C}\int_{-\infty}^{t} e^{-(t-s)/\tau} dW_s\right] dt}{\frac{W^2}{12}}\right)^2\right] \\
&= \left(\frac{12\sigma_Q}{W^3 C}\right)^2 \mathrm{E}\left[\left(\int_{-\frac{W}{2}}^{\frac{W}{2}} t\left[\int_{-\infty}^{t} e^{-(t-s)/\tau} dW_s\right] dt\right)^2\right].
\end{aligned}$$

To be able to make use of Itô's isometry, which states

$$\mathrm{E}\left[\left(\int_0^T X_t dW\right)^2\right] = \mathrm{E}\left[\int_0^T X_t^2 dt\right], \tag{5.15}$$

the order of integration has to be changed. The inner integral goes over s, so the factor t in the outer integral can be moved into the inner integral. The integration domain is split in two parts. The expectation value of the cross-product of these two parts equals zero because it is the product of two independent stochastic integrals whose individual

expectation value is zero.

$$
\begin{aligned}
&\mathrm{Var}[b] \\
&= \left(\frac{12\sigma_Q}{W^3 C}\right)^2 \mathrm{E}\left[\left(\int_{-\infty}^{-W/2}\int_{-W/2}^{W/2} te^{-(t-s)/\tau}dtdW_s + \int_{-\frac{W}{2}}^{\frac{W}{2}}\int_{s}^{W/2} te^{-(t-s)/\tau}dtdW_s\right)^2\right] \\
&= \left(\frac{12\sigma_Q}{W^3 C}\right)^2 \left(\mathrm{E}\left[\left(\int_{-\infty}^{-W/2}\int_{-W/2}^{W/2} te^{-(t-s)/\tau}dtdW_s\right)^2\right]\right. \\
&\qquad\qquad \left. + \mathrm{E}\left[\left(\int_{-\frac{W}{2}}^{\frac{W}{2}}\int_{s}^{W/2} te^{-(t-s)/\tau}dtdW_s\right)^2\right]\right) \\
&= \left(\frac{12\sigma_Q}{W^3 C}\right)^2 \left(\int_{-\infty}^{W/2}\left(\int_{-\frac{W}{2}}^{\frac{W}{2}} te^{-(t-s)/\tau}dt\right)^2 ds + \int_{-\frac{W}{2}}^{\frac{W}{2}}\left(\int_{s}^{W/2} te^{-(t-s)/\tau}dt\right)^2 ds\right) \\
&= \frac{12\sigma_Q^2}{C^2}\left(\frac{\tau^2}{W^3} - 3\frac{\tau^3}{W^4} + 12\frac{\tau^5}{W^6} - 3\frac{\tau^3}{W^6}(W+2\tau)^2 e^{-W/\tau}\right).
\end{aligned}
\tag{5.16}
$$

Writing as a function of τ/W as much as possible.

$$
\begin{aligned}
&\mathrm{Var}[b] \\
&= \frac{12\sigma_Q^2}{\tau^3\lambda^2}\left((\frac{\tau}{W})^3 - 3(\frac{\tau}{W})^4 + 12(\frac{\tau}{W})^6 - e^{-W/\tau}\left(3(\frac{\tau}{W})^4 + 12(\frac{\tau}{W})^5 + 12(\frac{\tau}{W})^6\right)\right).
\end{aligned}
$$

Getting $(\tau/W)^3$ in front of the expression:

$$
\frac{12\sigma_Q^2}{W^3\lambda^2}\left(1 - 3(\frac{\tau}{W}) + 12(\frac{\tau}{W})^3 - e^{-W/\tau}\left(3(\frac{\tau}{W}) + 12(\frac{\tau}{W})^2 + 12(\frac{\tau}{W})^3\right)\right).
$$

Writing the equation as a explicit function of λ and C:

$$
\frac{12\sigma_Q^2}{\lambda^2}\left((\frac{1}{W})^3 - 3(\frac{C}{\lambda W^4}) + 12(\frac{C^3}{\lambda^3 W^6}) - e^{-W\lambda/C}\left(3(\frac{C}{\lambda W^4}) + 12(\frac{C^2}{\lambda^2 W^5}) + 12(\frac{C^3}{\lambda^3 W^6})\right)\right).
$$

Figure 5.3 shows the relationship between σ_b and $1/\lambda$. For increasing values of W, the relationship becomes more linear. The relative range also improves for higher W.

Two-layer model

The analysis is repeated for the two-layer model of Equation 5.6. The Green's function for the upper layer temperature T can be derived by taking the time derivative of the the step-forcing response of the surface layer, which is given in Geoffroy et al. (2013b).

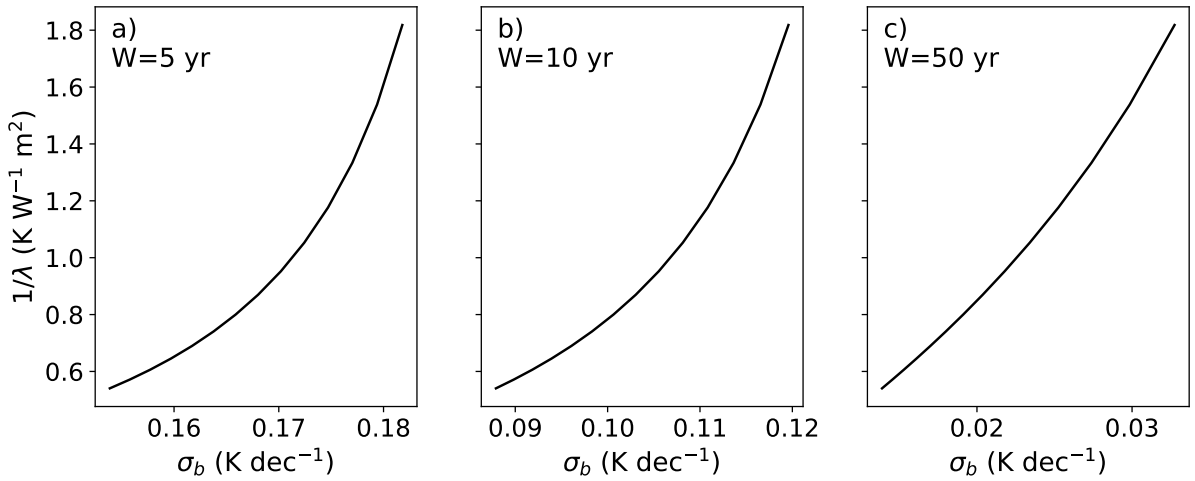


Figure 5.3: Assuming constant radiative forcing σ, variability σ_β for **a.** $W = 5$, **b.** $W = 10$ and **c.** $W = 50$. Heat capacity C is constant and set to $C = 7.5$, so that the time scale τ changes only for changing λ.

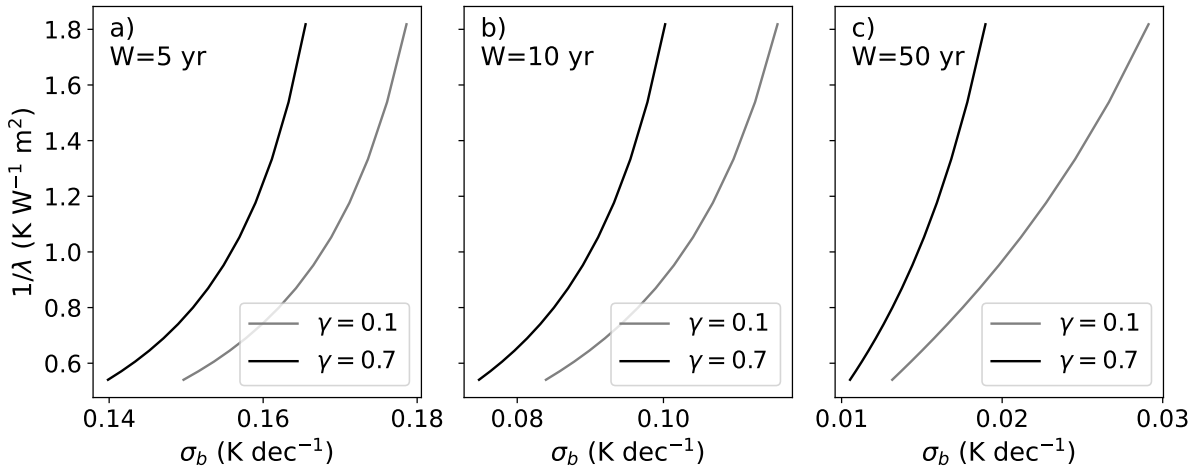


Figure 5.4: The relationship between decadal variability and $1/\lambda$ for a two-layer model as described in Section 5.3 under varying lengths of the trend (W). Assuming constant σ, decadal variability σ_b for **a.** $W = 5$, **b.** $W = 10$ and **c.** $W = 50$. C, C_0 and γ are constant at $C = 7.5$, $C_0 = 100$, and $\gamma \in \{0.1, 0.7\}$.

The step function response is, equivalently:

$$T(t) = T_{eq} - a_f T_{eq} e^{-t/\tau_f} - a_s T_{eq} e^{-t/\tau_s}; \tag{5.17}$$

$$T(t) = \frac{F}{\lambda}\left[a_f(1 - e^{-t/\tau_f}) + a_s(1 - e^{-t/\tau_s})\right]. \tag{5.18}$$

So that the Green's function is given by:

$$G_T(t) = \frac{a_f}{\tau_f} T_{eq} e^{-t/\tau_f} + \frac{a_s}{\tau_s} T_{eq} e^{-t/\tau_s}; \tag{5.19}$$

$$G_T(t) = \frac{1}{\lambda} \left[\frac{a_f}{\tau_f} e^{-t/\tau_f} + \frac{a_s}{\tau_s} e^{-t/\tau_s} \right]. \tag{5.20}$$

Using Mathematica and the techniques developed in the previous section, we derive the following equation for the variance of the trends in the two-layer model.

$$\begin{aligned}
\mathrm{Var}[b] =& \frac{12\sigma_Q^2}{\lambda^2 W^6(\tau_f+\tau_s)} e^{-W(\frac{1}{\tau_f}+\frac{1}{\tau_s})} \\
& \Bigg(a_f^2 e^{W/\tau_s}(-3\tau_f(W+2\tau_f)^2 + e^{W/\tau_f}(W^3 - 3W^2\tau_f + 2\tau_f^3))(\tau_f+\tau_s) \\
& + a_s^2 e^{W/\tau_f}(\tau_f+\tau_s)(-3\tau_s(W+2\tau_s)^2 + e^{W/\tau_s}(W^3 - 3W^2\tau_s + 12\tau_s^3)) \\
& + 2a_f a_s \Bigg[-3e^{W/\tau_s}\tau_f^2(W+2\tau_f)^2 - 3e^{W/\tau_f}\tau_s^2(W+2\tau_s)^2 \\
& + e^{W(\frac{1}{\tau_f}+\frac{1}{\tau_s})}(W^3(\tau_f+\tau_s) - 3W^2(\tau_f^2+\tau_s^2) + 12(\tau_f^4+\tau_s^4)) \Bigg] \Bigg).
\end{aligned} \tag{5.21}$$

$$\begin{aligned}
\mathrm{Var}[b] =& \frac{12\sigma_Q^2}{\lambda^2 W^6} \\
& \Bigg(a_f^2 e^{-W/\tau_f}(-3\tau_f(W+2\tau_f)^2 + (W^3 - 3W^2\tau_f + 12\tau_f^3)) \\
& + a_s^2 e^{-W/\tau_s}(-3\tau_s(W+2\tau_s)^2 + (W^3 - 3W^2\tau_s + 12\tau_s^3)) \\
& + 2\frac{a_f a_s}{(\tau_f+\tau_s)} \Bigg[-3e^{-W/\tau_f}\tau_f^2(W+2\tau_f)^2 - 3e^{-W/\tau_s}\tau_s^2(W+2\tau_s)^2 \\
& + W^3(\tau_f+\tau_s) - 3W^2(\tau_f^2+\tau_s^2) + 12(\tau_f^4+\tau_s^4) \Bigg] \Bigg).
\end{aligned} \tag{5.22}$$

Dividing by a factor of W^3 and writing out the brackets,

$$\begin{aligned}\mathrm{Var}[b] =& \frac{12\sigma_Q^2}{\lambda^2 W^3} \\ & \Bigg(a_f^2 \big\{ -e^{-W/\tau_f} \big(3\frac{\tau_f}{W} + 12(\frac{\tau_f}{W})^2 + 12(\frac{\tau_f}{W})^3\big) + 1 - 3\frac{\tau_f}{W} + 12(\frac{\tau_f}{W})^3 \big\} \\ & a_s^2 \big\{ -e^{-W/\tau_s} \big(3\frac{\tau_s}{W} + 12(\frac{\tau_s}{W})^2 + 12(\frac{\tau_s}{W})^3\big) + 1 - 3\frac{\tau_s}{W} + 12(\frac{\tau_s}{W})^3) \big\} \\ & + 2\frac{a_f a_s}{(\tau_f + \tau_s)} \Bigg[-e^{-W/\tau_f} \tau_f \big(3\frac{\tau_f}{W} + 12(\frac{\tau_f}{W})^2 + 12(\frac{\tau_f}{W})^3\big) \\ & \qquad - e^{-W/\tau_s} \tau_s \big(3\frac{\tau_s}{W} + 12(\frac{\tau_s}{W})^2 + 12(\frac{\tau_s}{W})^3\big) \\ & + (\tau_f + \tau_s) - 3\frac{(\tau_f^2 + \tau_s^2)}{W} + 12\frac{(\tau_f^4 + \tau_s^4)}{W^3} \Bigg] \Bigg).\end{aligned} \tag{5.23}$$

Figure 5.4 shows the relationship between σ_b and $1/\lambda$. In contrast to the one-box model, the dependence is not linear to first approximation. The sensitivity also seems lower, especially in the 5 year case, where σ_b increases a mere 15% over the full range of $1/\lambda$.

5.4 Methods

Observations have been used in combination with simple stochastic climate models to constrain long term variability (Wigley and Raper, 1990) and ECS (Schwartz, 2007). In contrast, we use an ensemble of state-of-the-art climate models (CMIP5 model ensemble (Taylor et al., 2012)) to first look for evidence of this relation in control simulations, before studying its implications in a climate perturbed by fossil fuel burning. Using the control, rather than historical or future simulations, allows for a cleaner test of the hypothesised link between internal variability and sensitivity. This is because historical simulations have additional external forcing and generally simulate shorter periods.

Models were included in our analysis if they had a control run spanning at least 500 years. Figure 5.2a shows the timeseries of annual global mean temperature of a high ECS model (HadGEM2-ES, brown line) and a low ECS model (GISS-E2-R, purple line). The thick line shows the 10-year running mean. The low sensitivity model shows shorter and smaller variation on the decadal timescale, in contrast to the longer and larger temperature trends in the high sensitivity model. Figure 5.2b shows composite distributions of decadal temperature trends for higher sensitivity ($\mathrm{ECS} > 3.0\,\mathrm{K}$, brown) and lower sensitivity models ($\mathrm{ECS} < 3.0\,\mathrm{K}$, purple). There is a clear distinction be-

tween high and low ECS models, the former having wider histograms indicating more variability in global temperature trends. Previous studies have noted a relationship between tropical decadal temperature variability and sensitivity in the CMIP5 ensemble (Colman and Power, 2018).

5.4.1 Data selection

We selected models based on a set of three criteria.

1. Maximum of one model per modelling group to avoid bias towards certain modelling centres.
2. Top of the atmosphere fluxes and forcing at $4\times CO_2$ should be available so that it can be tested that ECS is independent of internal forcing strength.
3. There must be at least 500 years of control data available.

For all models with more than 500 years, the last 500 years were chosen. Note that drift, if linear, does not affect the metric σ_b. Selected models are shown in Table 5.1.

5.4.2 Calculation of probabilities

The background warming for the historical period and future projections were computed using OLS linear regression between the temperature change and ECS. The temperature change itself was also computed using OLS linear regression between annual temperatures and time. Temperature time-series in models with multiple initial value members were averaged before a warming rate was computed.

In the second step, the emergent relationship between ECS and σ_b from the control simulations was used. Using a normal distribution for the decadal trend with the standard deviation dependent on ECS, probabilities are computed for either a period of cooling or a period of warming. In the case of an ECS-dependent background rate, the mean of the distribution is adjusted. This procedure is used for Figure 5.10.

	Model	λ	ECS	Initial value runs (rxi1p1)		
				RCP4.5	RCP8.5	Control
a	ACCESS1-0	0.79	3.83	1	1	1
b	CanESM2	1.04	3.69	1, 2, 3, 4, 5	1, 2, 3, 5,	1
c	CCSM4	1.23	2.89	1, 2, 3, 4, 5, 6	1, 2, 3, 4, 5	1
d	CNRM-CM5	1.14	3.25	1	1, 2, 4, 6, 10	1
e	CSIRO-Mk3-6-0	0.63	4.08	1 – 10	1 – 10	1
f	GFDL-ESM2M	1.38	2.44	1	1	1
g	HadGEM2-ES	0.64	4.59	1, 2, 3, 4	1, 2, 3, 4	1
h	inmcm4	1.43	2.08	1	1	1
i	IPSL-CM5A-LR	0.75	4.13	1, 2, 3, 4	1	1
j	MIROC-ESM	0.91	4.67	1	1	1
k	MPI-ESM-LR	1.13	3.63	1, 2, 3	1, 2, 3	1
l	MRI-CGCM3	1.25	2.60	1	1	1
m	NorESM1-M	1.11	2.80	1	1	1
n	bcc-csm1-1	1.14	2.82	1	1	1
o	GISS-E2-R	1.79	2.11	1, 2, 3, 4, 5, 6	1	1
p	BNU-ESM	1.0	4.1	1	1	1
q	FGOALS-g2	0.84	3.45			1
r	CESM1-WACCM	1.18	2.73			1
a^x	ACCESS1-3	0.86	3.45			1
f^x	GFDL-ESM2G	1.29	2.39			1
f^y	GFDL-CM3	0.75	3.97			1
i^x	IPSL-CM5B-LR	1.02	2.61			1
j^x	MIROC5	1.52	2.72			1
k^x	MPI-ESM-MR	1.20	3.44			1
o^x	GISS-E2-H	1.65	2.31			1
o^y	GISS-E2-H-CC	1.65	2.31			1
o^z	GISS-E2-R-CC	1.79	2.11			1
n^x	bcc-csm1-1-m	1.24	2.87			1
r^x	CESM1-CAM5	0.90	4.10			1

Table 5.1: List of CMIP5 climate models used. Values of ECS and λ are taken from Stocker et al. (2013) and Forster et al. (2013), except for CESM1-WACCM (Marsh et al., 2016) and CESM1(CAM5) (Meehl et al., 2013). ECS values for models with an added carbon cycle are from models with the carbon cycle turned off.

Finally, for the comparison of probabilities, i.e. the comparison of probability to have a decade of decreasing temperatures in a high ECS world, versus a low ECS world, a Bayesian linear regression was used for the emergent relationship using the STAN software. Weakly informative priors were used (Stan Development Team, 2018). This allowed us to get a collection of linear fits between σ_b and ECS. Note that in this collection, there are fits with a shallower and steeper slope compared to OLS linear regression. This translates into a having both high σ_b for low ECS and a low σ_b for high ECS in the shallow fits and visa versa for the steeper fits. From these pairs, pairs of probabilities of cooling decades are computed (as described in the previous paragraph), and these are divided to compute how much more likely a period of cooling is in a high ECS world compared to a low ECS world. Using pairs of regression lines leads to a larger estimate of uncertainty than a naive approach with OLS regression would have.

5.4.3 Analysis effect ENSO

The effect of ENSO was studied by regressing out the NINO3.4 index (Roberts et al., 2015). A linear regression between GMST and NINO3.4 was first performed, and then decadal trends were computed using the GMST residuals. For each trend length W, the correlation strength of emergent relationships was computed with GMST, and with GMST minus ENSO.

5.5 Results

In Figure 5.5a we plot decadal (W = 10 years) values of σ_b against ECS for each CMIP5 model control simulation. In Figure 5.5b standard deviations of decadal trends are plotted for the historical against the control simulations. Decadal trends in the historical simulations are larger due to a non-constant background trend. This causes differing means for the 1880–1950 period compared to the 1950–2012 one. Combining these two periods leads to a larger standard deviation of decadal trends, which explains the larger historical σ_b in Figure 5.5b. Using all 31 models and model variants in the CMIP5 archive, we find a similar but slightly weaker relationship (see Figure 5.6). While our theory predicts a weakly nonlinear relationship between the standard deviation of trends and ECS, we chose a linear regression between σ_b and ECS to prevent overfitting, based on the Akaike information criterion (Myung et al., 2009). Nonlinearities like this may be expected when the dominant time-scale of the climate system and the time-scale of the variability metric are of the same order of magnitude (see

Chapter 3).

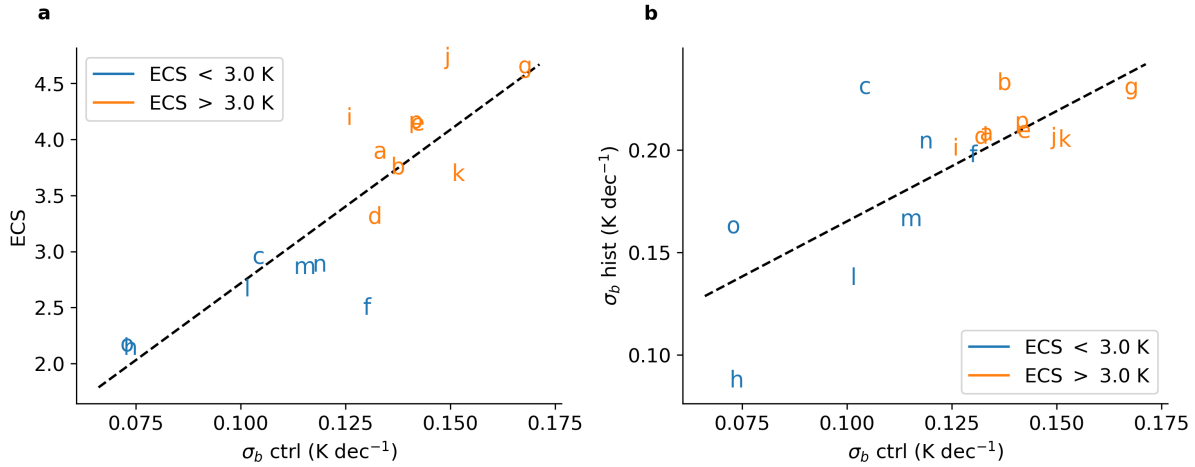


Figure 5.5: Emergent relationship between ECS and warming trends. **a)** Standard deviation of 10-year temperature trends in an ensemble of 500-yr control runs versus ECS. The dotted line is a linear ordinary least square fit with Pearson r=0.86. **b)** Ten-year variability in the control runs versus the 10-year variability in the historical period (1881 – 2017).

Figure 5.7a shows the variability of temperature trends of duration 3–50 years. Variability in trends of duration 5–25 years separate the low sensitivity (blue lines, lower variability) and high sensitivity models (orange lines, higher variability). The correlation between ECS and σ_b shown in Figure 5.7b is particularly strong for temperature trends of length 7 to 15 years ($r > 0.8$).

We have explored the possible impact of the El Niño-Southern Oscillation (ENSO) on our ECS versus σ_b correlation, and its dependence on trend length. To characterise ENSO, we use the NINO3.4 index which is based-on temperatures in the region between 120°W – 170°W and 5°S –5°N (Roberts et al., 2015). ENSO explains up to 40% of the variance in temperature trends (Figure 5.8b). By removing the ENSO signal, based on this index, it is shown that ENSO is not the dominant factor in our relationship (Figure 5.8a). It is notable however that the peak correlation at around 10 years disappears once the ENSO influence is removed, suggesting that the peak is mainly a consequence of ENSO variability. There may also be a smaller contribution to the peak correlation due to longer timescales in the climate response: When explicitly modelling the Hasselmann system with parameters corresponding to this CMIP5 ensemble, some simulations show a similar peak. (Figure 5.9). Excluding ENSO deteriorates the relationship between σ_b and ECS for all window lengths (Figure 5.8a). This is consistent with ENSO providing a useful additional stochastic forcing of the climate system, which helps to reveal ECS.

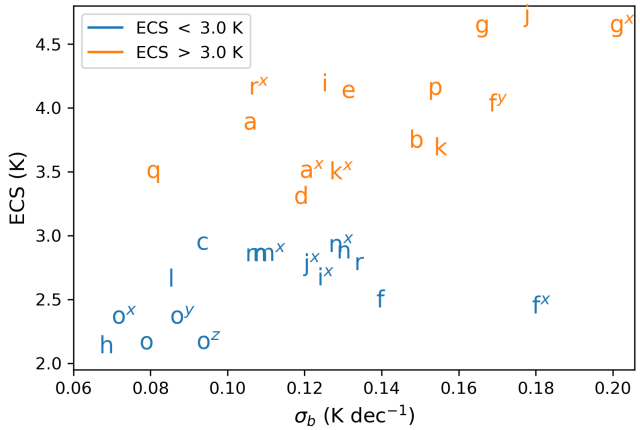


Figure 5.6: The emergent relationship for $W = 10$ with the inclusion of additional models. The last 200 years of each model were used to estimate the standard deviation of trends. The labels are explained in Table 5.1.

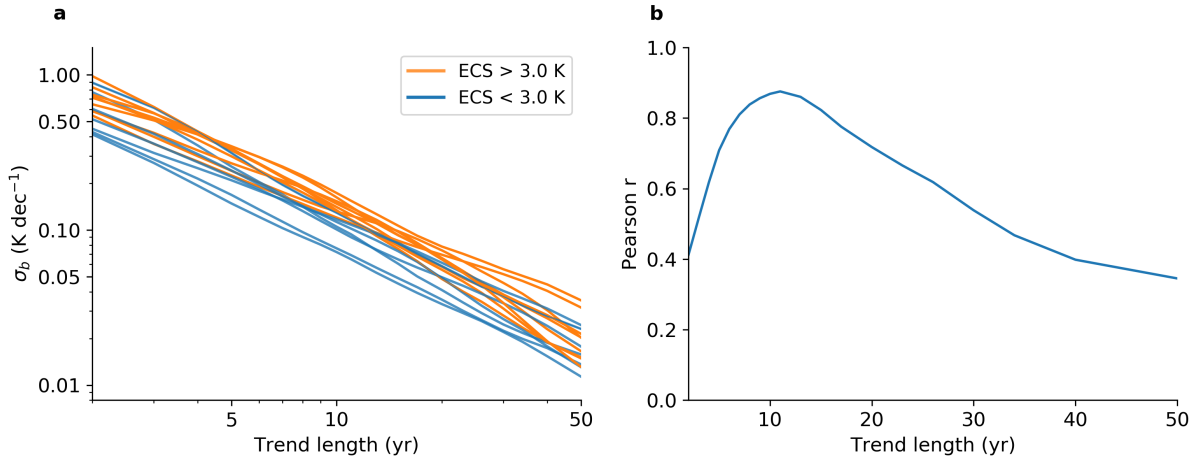


Figure 5.7: Varying window lengths. **a.** log-log plot of trend length versus standard deviation of the trend using the control simulations, differentiated in colour (as marked) between ECS value. **b.** Correlation (Pearson r) of the emergent relationship in **a**, between ECS and σ_b as a function of trend length.

Figure 5.10a plots equal probability contours for anomalies in the decadal temperature change as a function of ECS over the historical period. The probabilities are computed by combining the relationship between the decadal variability and ECS as derived from the control run, with the background warming from the historical runs (Figure 5.11). Over the historical period (1960–2012) there is a small correlation between the background warming and ECS. Figure 5.10a is asymmetric, in contrast to Figure 5.11, because the probability of warming episodes is increased by the ECS-dependent background warming. Figure 5.9 shows that trends and variability are sep-

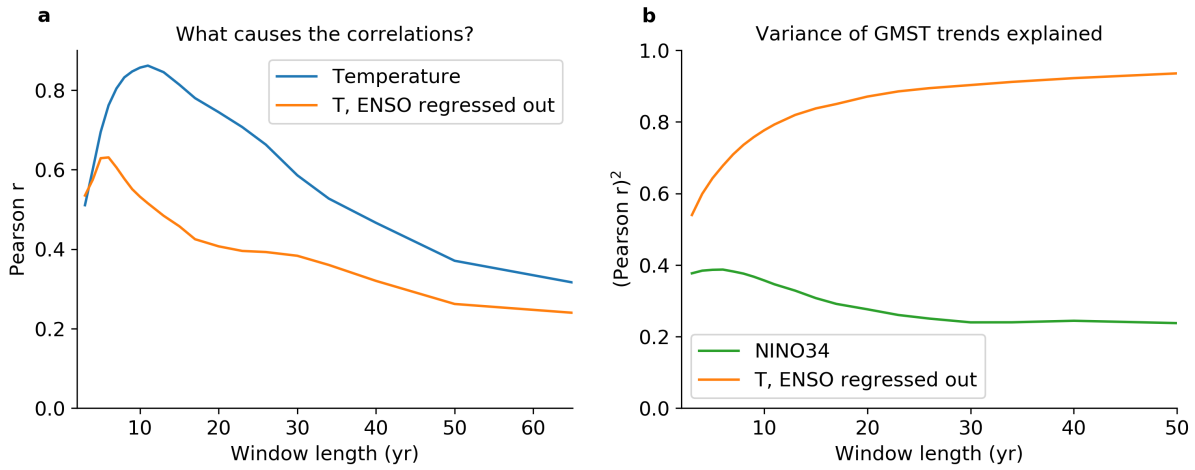


Figure 5.8: Effect of ENSO. The left figure shows the correlation as function of window length for the global temperature, and the global temperature with the NINO3.4 index regressed out. The right figure indicates the variance explained by ENSO as a function of window length.

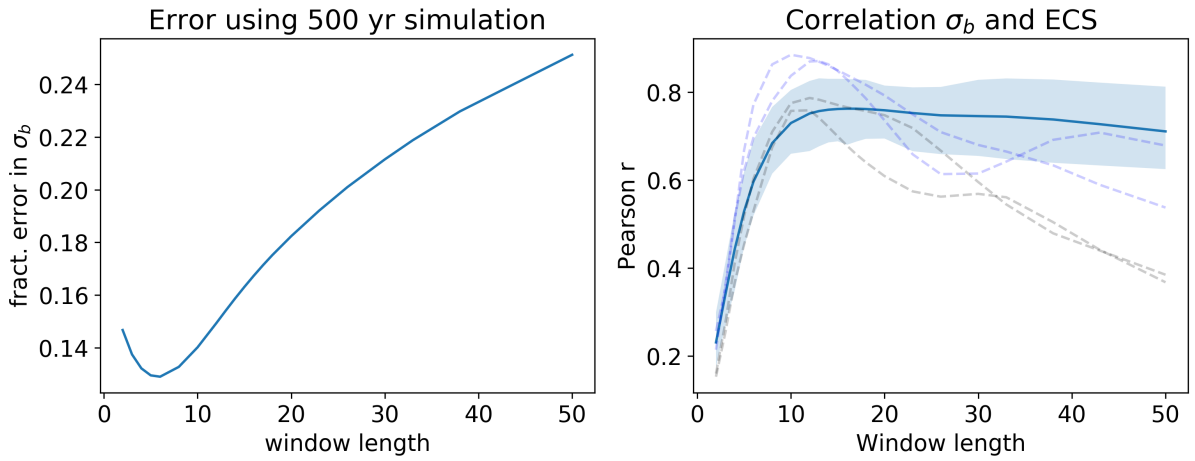


Figure 5.9: Correlation using two-layer model. Using the set of parameters that fits the two-layer model (Geoffroy et al., 2013b) corresponding to 16 of the models used in our study, this figure reconstructs possibilities of the correlation between ECS and σ_b. The left panel contains the mean fractional error, while the right panel contains the median (dark blue), the area between the 16th and 84th percentile (shaded). The dotted lines are individual simulations showing 5% most extreme behaviour in maximum peak (blue) and minimum correlation for at 50 years (grey).

arable: removing the forced trend by subtracting the mean of initial value ensembles (for those models with a sufficient amount of initial value runs), successfully retrieves the variability found in the control simulations.

Cooling or warming decadal episodes that occur only 5% of the time, show a large sensitivity to ECS. In Figure 5.10b, corresponding to the grey line in Figure 5.11a, we

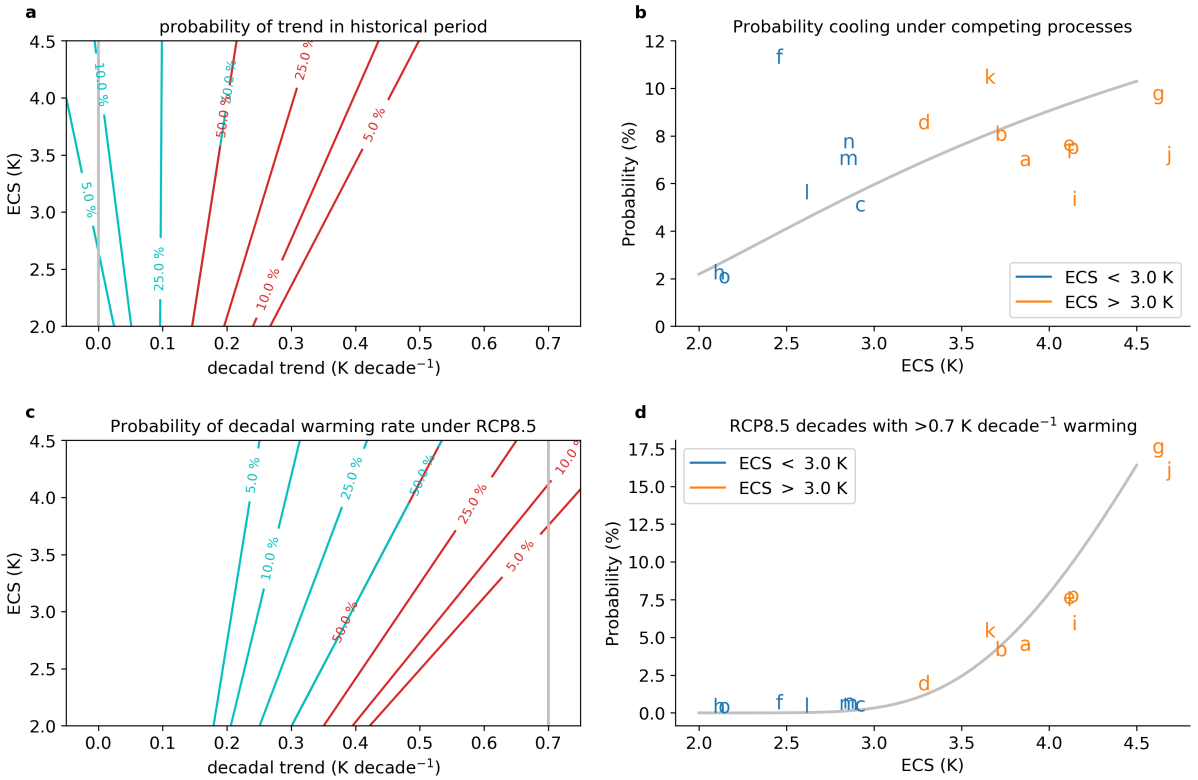


Figure 5.10: Probability of warming and cooling. **a.** Percentage of decades with a larger or equal warming (red) or cooling (blue) rate over the historical period, as computed from the normal distribution using the emergent relationship to estimate standard deviation. **b.**,The grey line indicates the probability of having a cooling event over the historical period. The background warming trend for each model was assumed linear in ECS, with the parameters estimated from linear regression between ECS and the modelled 1960–2012 warming rate. The probability of the individual models is computed separately, and marked as letters (key in Table 5.1). **c.** Dashed blue: chance of a period of cooling in RCP8.5 simulations, colour scheme the same as **a**. **d** Probability that a decade in RCP8.5 simulations shows hyperwarming: 0.7 K/decade or more (see also vertical gray line in **c**). Colours are the same as b. The calculations per model are again perfomed using a normal distribution with a standard deviation estimated from the control runs.

plot the probability of a cooling decade assuming a background warming rate consistent with the historical simulations for each model. This includes the weak increase in the warming trend with ECS, as well as the stronger increase in variability with increasing ECS. Even with these two opposing effects, the sensitivity of decadal variability to ECS implies that a 'hiatus' period was 2.2 [90% CrI 0.68 – 11] times as likely in a high ECS world ($\mathrm{ECS} = 4.5\,\mathrm{K}$) compared to a low ECS world ($\mathrm{ECS} = 1.5\,\mathrm{K}$). While some studies indicate that the recent slowdown can partially be explained by a decrease in forcing (Smith et al., 2016; Medhaug et al., 2017), our results show that even in the

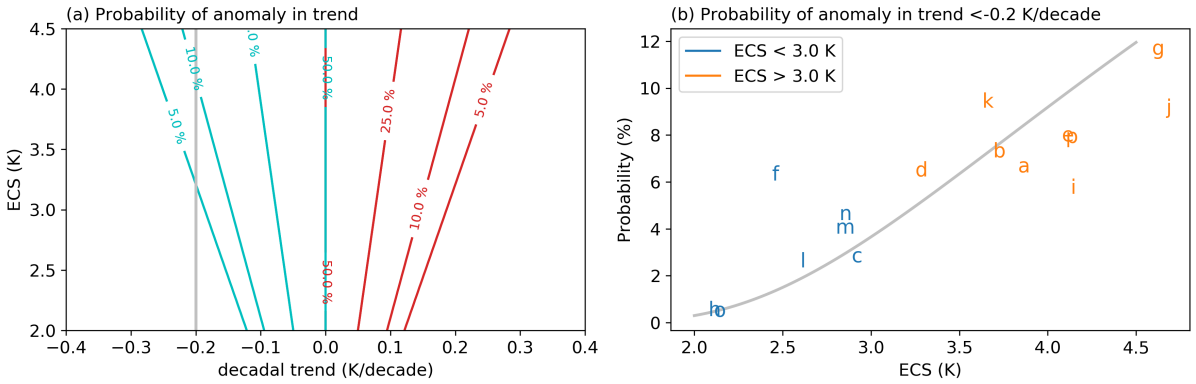


Figure 5.11: Similar to Figure 5.10a and b, but assuming a model-independent historical background warming rate of 0.2 K decade^{-1}

case forcing remained constant, a temporarily reduced trend does not imply ECS to be lower.

For a given future scenario of increasing anthropogenic forcing, ECS affects both the mean and the variability in the rate of global warming. Figure 5.10c plots probability contours for different absolute decadal warming rates as a function of ECS, under the RCP8.5 scenario. Figure 5.10d shows how ECS affects the probability of a 'hyper-warming decade' — which we define here as one with a warming-rate exceeding 0.7 K/decade (i.e. ten times faster than the mean rate of global warming over the 20th century). Whereas a hyperwarming decade very rarely occurs for $\mathrm{ECS} < 2.5\,\mathrm{K}$, it occurs 8% of the time for $\mathrm{ECS} > 3.5\,\mathrm{K}$.

To improve comparability between a global emergent constraint and various local emergent constraints, it is possibly useful to derive a spatial variant of our emergent constraint. Figure 5.12 shows the spatial pattern of the correlation. The average correlation above land is slightly lower than above sea. This is not surprising, as our theoretical derivation assumes the overall time scale is of a couple of years, an assumption not valid above land, which adjust to doubling of CO_2 or SST patterns is less than a month's time (Dong et al., 2009).

5.6 The last millennium

The PAGES 2k consortium has produced an analysis of the global mean surface temperature over the last two millennia (Neukom et al., 2019a). This subsection explores the emergent relationship in a paleo-context: are reconstructions and climate simulations sufficiently good to constrain ECS?

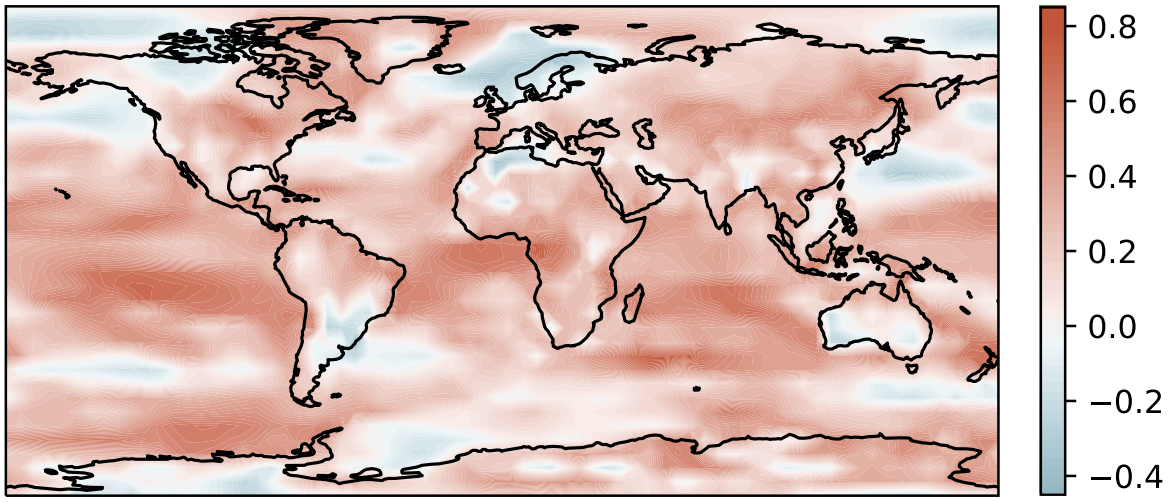


Figure 5.12: The spatial pattern of the Pearson r correlation between local decadal variability $\sigma_{b,\text{loc}}$ and ECS. Note that some regions have more internal variability and therefore weigh stronger in the global metric.

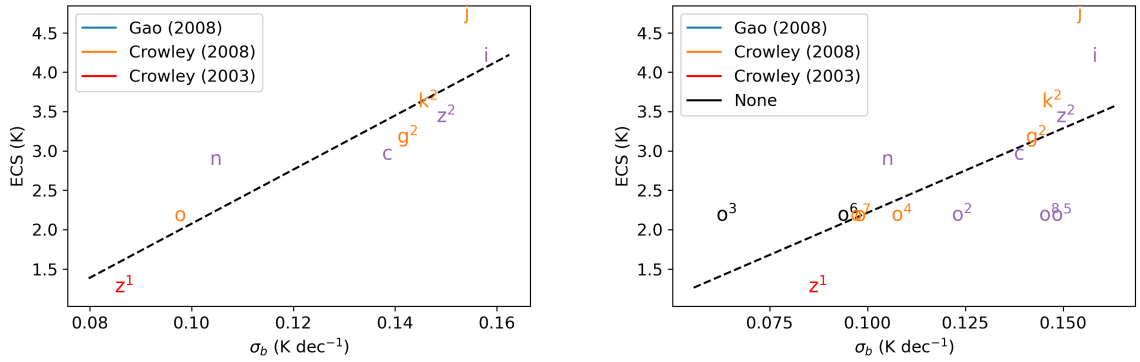


Figure 5.13: Emergent relationship between decadal variability and climate sensitivity for models simulating the last millennium. The left panels includes one simulation per model, whereas the right panel includes all available simulations, including those without volcanic forcing.

The emergent relationship between σ_b and ECS is found in CMIP5 data in the last millennium as well, see Figure 5.13. Only nine models were available covering this period, some of which are not in the ensemble described in Section 5.1. Otherwise, letters denote the same models (Table 5.2).

Years with large volcanoes are excluded from the analysis. Large volcanoes were selected by examining the outgoing longwave radiation. If this radiation is above a certain threshold (here, one standard deviation) above the average for that model, that year and n subsequent years (here ten years) are removed from the simulation. The inclusion of volcanoes significantly deteriorates the emergent relationship.

Removing large volcanoes from the analysis did not mean there was not any effect of eruptions. Two different reconstructions of radiative forcing from volcanoes were used by modelling groups: the Gao et al. (2008) and the Crowley et al. (2008) reconstruction. One model uses an older version the the Crowley reconstruction. The GISS model performed simulations using both reconstructions and simulation where volcanoes are turned off completely. From this analysis, it becomes clear that volcanoes introduce higher decadal variability and that the Crowley simulation introduces higher decadal variability compared to the Gao simulation.

Figure 5.14 shows a composition of temperature reconstructions of the last two millennia. Temperature reconstructions vary depending on the selection of proxies, geological or biological indirect indicators of past climates, and on the statistical methodology to interpret those proxies (Neukom et al., 2019c). The figure is based on the PAGES2k proxy dataset (Emile-Geay et al., 2017). Seven different reconstruction methods were used. For each of the methods, a thousand samples were taken incorporating uncertainty, so that in total 7000 individual reconstructions of the temperature of the last 2000 years were included. For each of these 7000 reconstructions, decadal variability was calculated. As for the models, years after major volcanic eruptions were excluded. The method employed for observations was slightly different: we used the Gao reconstruction of aerosol optical depth to select years to exclude, with a threshold of a 10 Tg sulphate aerosol injection.

The one standard deviation observational range encompasses the entire model range. Furthermore, choices with respect to the implementation of volcanic forcing play an almost equally important role explaining decadal variability as climate sensitivity. In conclusion, we cannot yet exploit the emergent relationship between decadal variability and climate sensitivity over the last millennium because of outstanding uncertainties in paleoclimate reconstructions.

	Model	Run	ECS (K)	Start year	Volcanic reconstruction
c	CCSM4		2.89	850	Gao
i	IPSL-CM5A-LR		4.13	850	Gao
g[2]	HadCM3		3.3	850	Cro
j	MIROC-ESM		4.67	850	Cro
n	bcc-csm1-1		2.82	850	Gao
o	GISS-E2-R	r1i1p121	2.11	850	Cro
z[1]	FGOALS-gl		1.2	1000	Cro2003
z[2]	FGOALS-s2		3.36	850	Gao
k[2]	MPI-ESM-P		3.55	850	Cro
o[2]	GISS-E2-R	r1i1p1221	2.11	850	Gao
o[3]	GISS-E2-R	r1i1p123	2.11	850	None
o[4]	GISS-E2-R	r1i1p124	2.11	850	Cro
o[5]	GISS-E2-R	r1i1p125	2.11	850	Gao
o[6]	GISS-E2-R	r1i1p126	2.11	850	None
o[7]	GISS-E2-R	r1i1p127	2.11	850	Cro
o[8]	GISS-E2-R	r1i1p128	2.11	850	Gao

Table 5.2: Decadal variability in the last millennium. The ECS values are from Forster et al. (2013), except for FGOALS-gl, which is from Zhang et al. (2013) and HadCM3 from Hunter et al. (2019). The Gao volcanic data set is from Gao et al. (2008), while the Cro data set is from Crowley et al. (2008) or in one case Crowley et al. (2003) denoting its previous version.

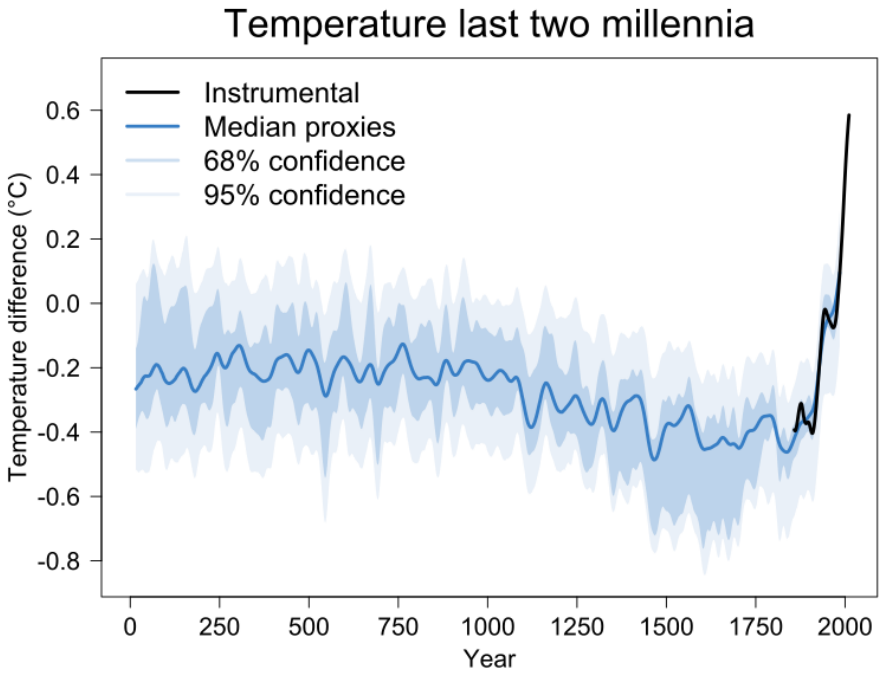


Figure 5.14: Temperature reconstruction over the past 2000 years, merging seven different reconstruction methods. Code was adapted from Neukom et al. (2019b).

5.7 Conclusion

Our findings indicate that the concept of equilibrium climate sensitivity (ECS) is relevant not only to the mean global warming at a given level of atmospheric CO_2, but also to temperature variability on decadal timescales. Counter-intuitively, this suggests that the slowdown in global warming from 2002–2012 was more likely in a high ECS world. It also means that decades of very rapid warming, which would stretch the adaptive capacity of ecosystems and society, are also much more likely if ECS is high. A previous constraint based on global temperature variability found a most likely value for ECS at 2.8 K (Cox et al., 2018a), which is lower than suggested by some other recent studies (Caldwell et al., 2018; Brown and Caldeira, 2017; Sherwood et al., 2014). Achieving a better consensus on the risk that we live in a high ECS climate is therefore of critical importance to both the climate mitigation challenge and also to inform efforts to build resilience to climate variability.

The associated code for this chapter is available at Code Ocean (Nijsse et al., 2019b).

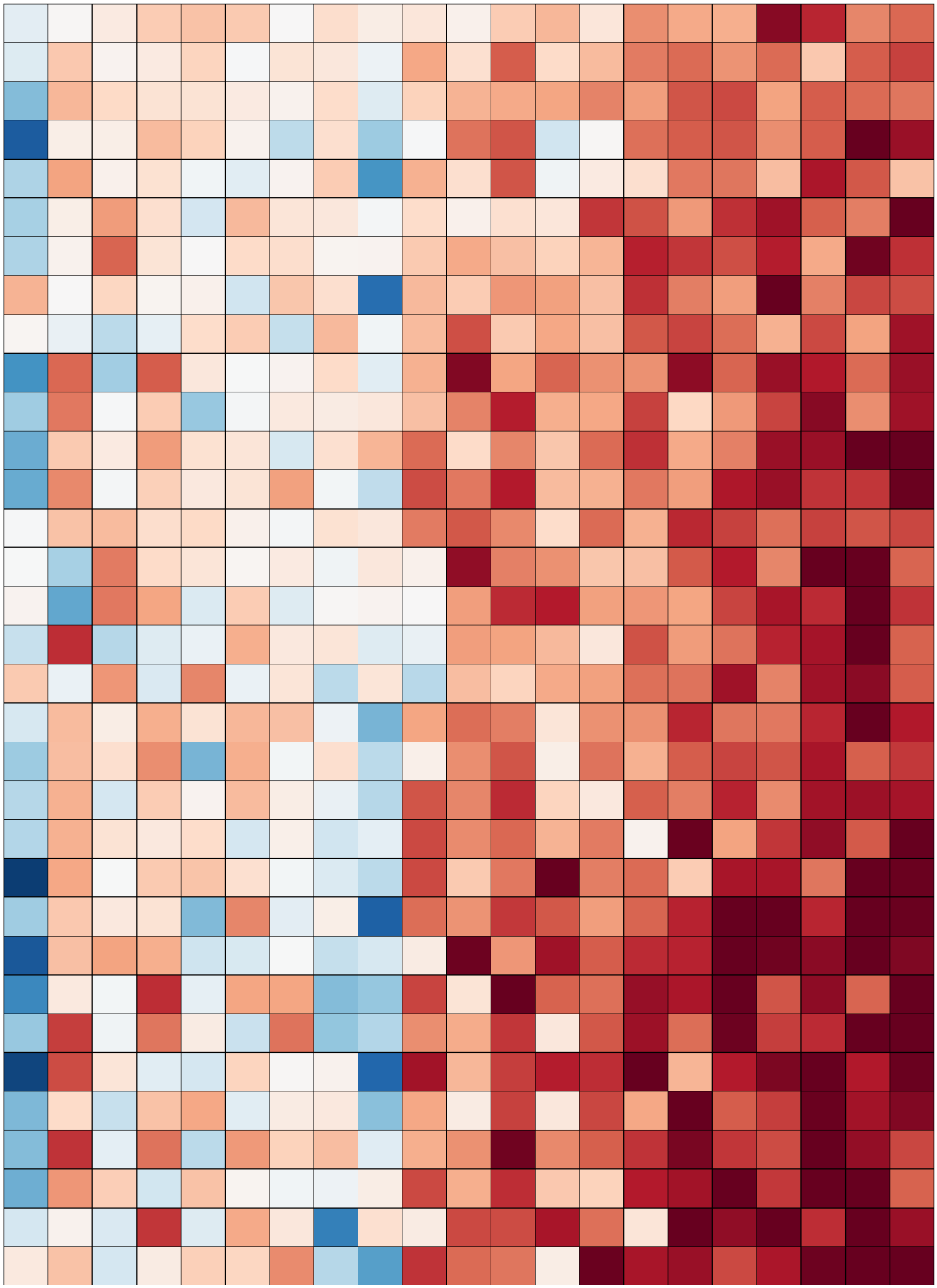


Figure 5.15: Warming blocks: a visualisation of the emergent relationship between 1880 and 2100. Each row denotes a climate model, each column a decade. Blue squares denote cooling decades, whereas red squares denotes warming decades. The rows are sorted from low to high decadal variability.

Chapter 6

Estimating TCR and ECS from observed warming

This chapter is adapted from the Earth System Dynamics paper Nijsse et al. (2020b): "Emergent constraints on TCR and ECS from historical warming in CMIP5 and CMIP6 models", co-authored by Peter Cox and Mark Williamson

Transient climate response (TCR) is the metric of temperature sensitivity that is most relevant to warming in the next few decades, and contributes the biggest uncertainty to estimates of the carbon budgets consistent with the Paris targets (Arora et al., 2020). Equilibrium climate sensitivity (ECS) is vital for understanding longer-term climate change and stabilization targets. In the IPCC 5th Assessment Report (AR5), the stated 'likely' ranges (17–83% confidence) of TCR (1.0–2.5 K) and ECS (1.5–4.5 K) were broadly consistent with the ensemble of CMIP5 Earth System Models available at the time. However, many of the latest CMIP6 ESMs have larger climate sensitivities, with 5 of 34 models having TCR values above 2.5 K, and an ensemble mean TCR of 2.0 $\pm$ 0.4 K. Even starker, 12 of 34 models have an ECS value above 4.5 K. On the face of it, these latest ESM results suggest that the IPCC likely ranges may need revising upwards, which would cast further doubt on the feasibility of the Paris targets.

In this chapter it is shown that, rather than increasing the uncertainty in climate sensitivity, the CMIP6 models help to constrain the likely range of TCR to 1.3–2.1 K, with a central estimate of 1.68 K. We reach this conclusion through an emergent constraint approach which relates the value of TCR linearly to the global warming from 1975 onwards. This is a period when the signal-to-noise ratio of the net radiative forcing increases strongly, so that uncertainties in aerosol forcing become progressively less problematic. We find a consistent emergent constraint on TCR when we apply the same method to CMIP5 models. Our constraints on TCR are in good agreement

with other recent studies which analysed CMIP ensembles. The relationship between ECS and the post-1975 warming trend is less direct and also non-linear. However, we are able to derive a likely range of ECS of 1.9–3.4 K from the CMIP6 models by assuming an underlying emergent relationship based on a two-layer energy balance model. Despite some methodological differences, this is consistent with a previously-published ECS constraint derived from warming trends in CMIP5 models to 2005. Our results seem to be part of a growing consensus amongst studies that have applied the emergent constraint approach to different model ensembles and to different aspects of the record of global warming.

6.1 Introduction

The key uncertainty in projections of future climate change continues to be the sensitivity of global mean temperature to 'radiative forcing', the change in Earth's energy budget induced by greenhouse gases and other atmospheric particles. This sensitivity is usually characterised in terms of the global mean temperature that would occur if the atmospheric carbon dioxide concentration was doubled, for which the radiative forcing is reasonably well-known.

Two related quantities are used to characterise the climate sensitivity of Earth System Models (ESMs). Equilibrium climate sensitivity (ECS) is an estimate of the eventual steady-state global warming at double CO_2. Transient climate response (TCR) is the mean global warming predicted to occur around the time of doubling CO_2 in ESM runs for which atmospheric CO_2 concentration is prescribed to increase at 1% per year. Across an ensemble of ESMs, TCR values are less than ECS values because of deep ocean heat uptake, which leads to a lag in the response of global temperature to the increasing CO_2 concentration (Hansen et al., 1985). The ratio of TCR over ECS tends to decrease with increasing ECS, and depends on spatial pattern effects (Armour, 2017).

Despite decades of advances in climate science, the Earth's ECS and TCR remain uncertain. The 'likely' range of ECS (66% confidence limit) has been quoted as 1.5 K to 4.5 K in all of the five Assessment Reports (ARs) of the Intergovernmental Panel on Climate Change (IPCC) starting in 1990, aside from the fourth AR which moved the likely lower range temporarily to 2 K. Similarly the likely range of TCR is given as 1 K to 2.5 K in the IPCC AR5, based on multiple lines of evidence.

There have been numerous attempts to constrain ECS using the record of historical warming or palaeoclimate data (Knutti et al., 2017), and more recently using emergent constraints which relate observed climate trends, variations or other variables to ECS

using an ensemble of models (Caldwell et al., 2018; Cox et al., 2018a). However, debate still rages about the likely range of ECS (Brown et al., 2018; Bretherton and Caldwell, 2020; Cox et al., 2018b; Gregory et al., 2019), in part because observed global warming is a rather indirect measure of global warming at equilibrium. On the other hand, TCR is more closely related to the rate of warming, and therefore ought to be more amenable to constraint by the record of global warming (Bengtsson and Schwartz, 2013; Gregory and Forster, 2008; Jiménez-de-la Cuesta and Mauritsen, 2019; Tokarska et al., 2020). Nevertheless, the accepted likely range of TCR has also resisted change (Knutti et al., 2017), for reasons we will discuss in this chapter. At the time of the AR5, the CMIP5 ESMs produced central estimates (mean $\pm$ standard deviation) of ECS (3.3 $\pm$ 0.7 K) and TCR (1.8 $\pm$ 0.3 K), that were broadly consistent with these IPCC likely ranges. However, there has been a general drift upwards towards higher climate sensitivities in the new CMIP6 ESMs, such that more than one third of the new CMIP6 models now have ECS values over 4.5 K (Forster et al., 2020), and five have TCR values over 2.5 K (Table 1). If the real climate system is similarly sensitive, the Paris climate targets will be much harder to achieve (Tanaka and O'Neill, 2018).

Therefore some key science- and policy-relevant questions arise:

(a) *Are such high climate sensitivities consistent with the observational record?*

(b) *If so, do the CMIP6 models demand an upward revision to the IPCC likely ranges for climate sensitivity?*

We address these questions in this chapter by evaluating the historical simulations of global warming from the CMIP6 models. In particular, we explore an emergent constraint on TCR based on global warming from 1975 onwards (Jiménez-de-la Cuesta and Mauritsen, 2019; Tokarska et al., 2020), but using the CMIP6 models and observational data up to 2019.

Emergent constraints are increasingly used to assess future change by exploiting statistical relationships in multimodel ensembles between an observable and a variable describing future climate (Cox et al., 2018a; Hall et al., 2019). In the work presented here, we use the latest CMIP6 multimodel ensemble to define an emergent relationship between historical warming (expressed in terms of GMST, the observable) and TCR (the variable related to future climate). In line with published recommendations (Hall et al., 2019; Klein and Hall, 2015), we check the robustness of the resulting emergent constraint against the CMIP5 ensemble, using exactly the same methodology as for CMIP6. We also follow the suggestion of Hall et al. (2019) in striving to base the emergent constraint on sound physical reasoning.

From physical principles, we expect values of TCR to be very well-correlated with simulated global warming across a model ensemble. By definition, TCR is a measure of warming from a simulation that is driven by an exponential 1.0% per year increase in CO_2. Historical global warming has been driven by a qualitative similar forcing, albeit somewhat less rapid. Instead of 1.0%, the atmospheric CO_2 concentration has increased at about 0.5% per year since 2000 (Dlugokenchy and Tans, 2019)), augmented by additional positive radiative forcing from other well-mixed greenhouse gases, and partially offset by the cooling effects of anthropogenic aerosols.

The radiative effects of the rise in greenhouse gas concentrations are relatively well-known (Myhre et al., 2013), and are broadly similar in different ESMs. By contrast, the radiative forcing due to changes in anthropogenic aerosols, especially indirect effects via changes in cloud brightness and lifetime, are poorly constrained (Myhre et al., 2013; Bellouin et al., 2020).

These uncertainties in aerosol forcing have hindered attempts to constrain TCR or ECS from the rate of warming, especially during the pre-1980 period when the burning of sulphurous coal led to increases in CO_2 concentrations and sulphate aerosols levels, that went up almost together (Andreae et al., 2005). As a result it has been difficult to distinguish, based purely on the observational record of global warming, between a model with high climate sensitivity and strong aerosol cooling, and a model with low climate sensitivity and weak aerosol cooling (Kiehl, 2007).

In order to minimise the effects of uncertainties in aerosol forcing, we need periods in which aerosol radiative forcing changes relatively little compared to the change in radiative forcing due to CO_2 and other well-mixed greenhouse gases. Fortunately, this applies to the decades after 1975 when total aerosol load from global SO_2 and NH_3 emissions were similar to values over the last decade (Stevens et al., 2017). For this reason, we focus on global warming since 1975. However, we also test the robustness of our conclusions to different start dates (see Figure 6.8(c)), including the start year of 1970 as used by Jiménez-de-la Cuesta and Mauritsen (2019) (hereafter JM19).

To establish an emergent constraint on ECS, we investigate the appropriate functional form between observed warming and climate sensitivity. Due to the slow response of the ocean, this is not expected to be linear, and using a set of assumptions, JM19 proposed an analytical form based on a two-layer box model. By computing the model parameters directly per model, we investigate the appropriateness of this analytical function, and use it to derive an emergent constraint.

The remainder of this chapter is organised as follows: in Section 6.2 we describe our methodological choices and more technical details concerning the regression methods; Section 6.3 contains the emergent constraints on TCR and ECS and Section 6.4

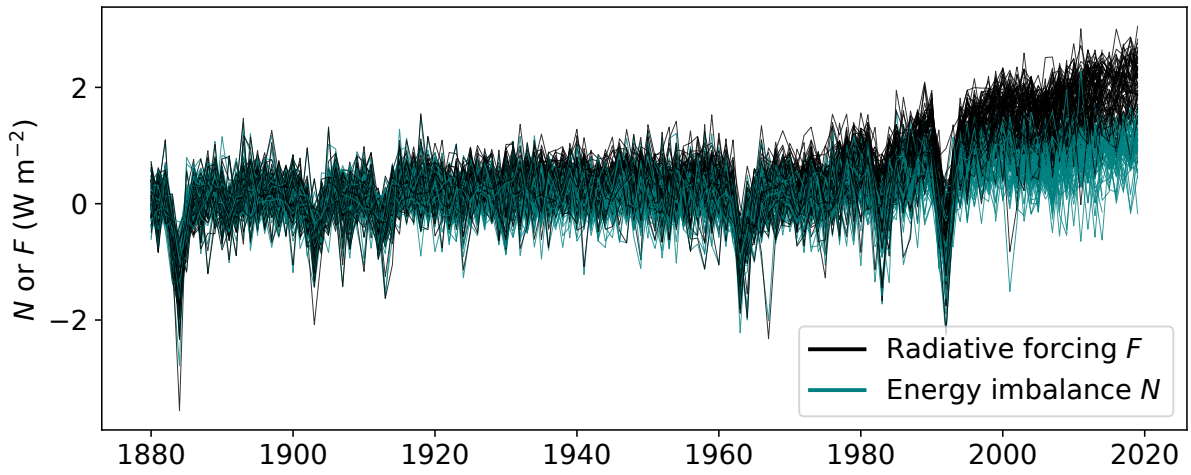


Figure 6.1: Effective radiative forcing F and the top-of-the-atmosphere energy imbalance N over the historical period, calculated from 22 CMIP6 models as $F = \Delta N + \lambda \Delta T$. A maximum of five simulations per model are shown.

contains the discussion and conclusions.

6.2 Methodology

6.2.1 Choice of period over which to calculate warming trends

To constrain climate sensitivity using observed warming, we seek a period for which the forcing is relatively similar across models. In order to identify such a period we compute the effective radiative forcing F (ERF) for each model run using

$$F = \Delta N + \lambda \Delta T \tag{6.1}$$

following Forster et al. (2013). Here ΔN is the difference in net top of the atmosphere radiative flux and ΔT is the difference in near-surface temperature, both computed as global annual-mean anomalies relative to the initial state. We calculate the signal-to-noise ratio of F at each time as the model mean F divided by the standard deviation of F across the model ensemble.

Figure 6.2 shows how the signal to noise ratio of the ERF varies from 1880 to 2010. It is notable that the signal-to-noise ratio increases rapidly from around 1975, as relatively well-known greenhouse gas forcing continues to increase but the uncertain aerosol forcing begins to saturate. We have therefore focused our analysis on the post-1975 warming, but we also performed a sensitivity analysis by varying the start year between 1960 and 2005.

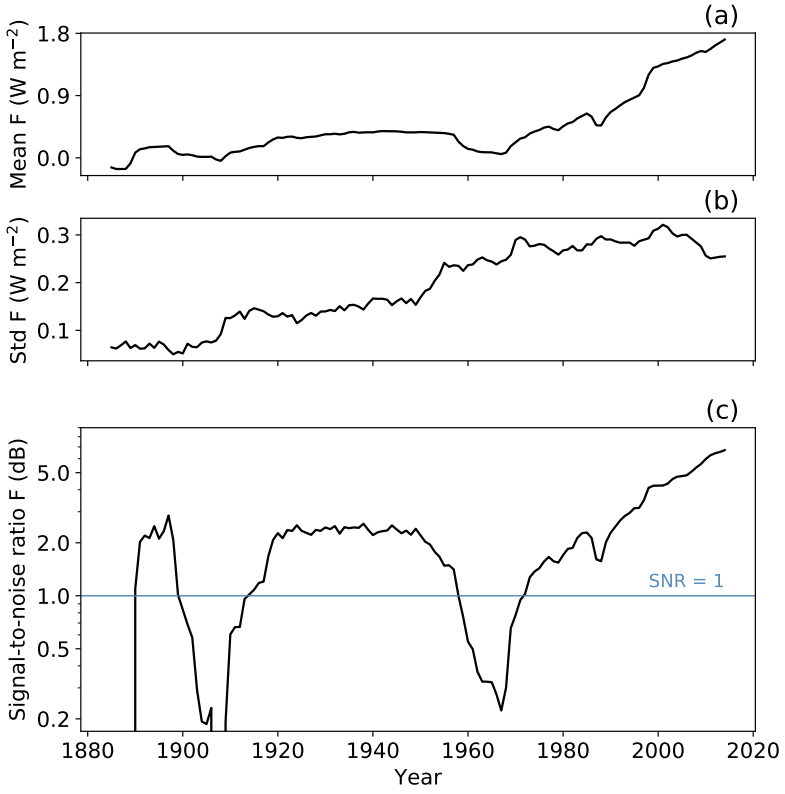


Figure 6.2: Effective radiative forcing over the historical period, calculated from 22 CMIP6 models as $F = \Delta N + \lambda \Delta T$: (a) ensemble mean; (b) ensemble standard deviation; (c) signal-to-noise ratio. Model means are calculated first, and then the ensemble mean is calculated.

6.2.2 Selection of model simulations

We use all currently available CMIP6 models that have control (piControl), historical, a Shared Socioeconomic Pathway simulation (SSP1-2.6, SSP2-4.5, SSP3-7.0 or SSP5-8.5) and one percent CO_2 increase per year (1pctCO2) experiments. We extend the historical simulations from 2014 to 2019 using the Shared Socioeconomic Pathways (SSPs) scenario runs. Additional warming over this 5 year period varies very little across the SSPs, so we use SSP2-4.5 as this has the largest number of participating models at the time of writing. Selection of CMIP5 models was similar, where instead of shared Socio-Economic Pathways, the RCP8.5 was chosen, which corresponds best with observational greenhouse gas concentrations (Schwalm et al., 2020).

6.2.3 Model sensitivity

From the 1pctCO2 experiment TCR is determined as the average temperature difference from the corresponding piControl run between 60 to 80 years after the start of the simulation (IPCC, 2013a). ECS is computed using the Gregory method (Gregory, 2004) on the first 150 year of the abrupt-4xCO2 simulations. The values of ECS and TCR that we derived are given in Table 6.1.

Centre	Model	$F_{2\times}$	λ	ECS	TCR	n	ΔT	SD
BCC	BCC-CSM2-MR	3.01	0.98	3.07	1.59	1	0.64	
CAMS	CAMS-CSM1-0	3.95	1.71	2.31	1.72	1	0.44	
CAS	FGOALS-f3-L	3.95	1.31	3.03	2.01	1	0.70	
CCCma	CanESM5	3.63	0.64	5.66	2.66	50	1.27	0.10
CNRM	CNRM-CM6-1	3.54	0.72	4.94	2.08	10	0.73	0.11
CNRM	CNRM-ESM2-1	3.09	0.66	4.66	1.92	5	0.65	0.15
CSIRO	ACCESS-CM2	3.21	0.67	4.81	2.00	1	0.77	
CSIRO	ACCESS-ESM1-5	2.71	0.68	3.97	1.91	3	0.84	0.10
EC-Earth-C.	EC-Earth3-Veg	3.32	0.77	4.34	2.57	2	0.97	0.23
EC-Earth-C.	EC-Earth3	3.30	0.78	4.22	2.38	10	0.72	0.16
INM	INM-CM4-8	2.61	1.42	1.84	1.32	1	0.61	
INM	INM-CM5-0	2.88	1.49	1.93	1.40	1	0.55	
IPSL	IPSL-CM6A-LR	3.32	0.72	4.63	2.32	6	0.85	0.10
MIROC	MIROC-ES2L				1.55	1	0.62	
MIROC	MIROC6	3.76	1.47	2.56	1.52	3	0.50	0.04
MOHC	HadGEM3-GC31-LL	3.38	0.60	5.62	2.45	4	1.07	0.19

MOHC	UKESM1-0-LL	3.56	0.66	5.41	2.72	5	1.13	0.13
MPI-M	MPI-ESM1-2-HR	3.58	1.20	2.99	1.64	2	0.65	0.07
MRI	MRI-ESM2-0	3.36	1.07	3.14	1.56	5	0.73	0.06
NCAR	CESM2-WACCM	3.08	0.63	4.90	1.92	3	0.97	0.15
NCAR	CESM2	3.13	0.59	5.30	2.04	3	0.82	0.01
NCC	NorESM2-LM	3.06	1.13	2.69	1.46	3	0.63	0.18
NOAA-GFDL	GFDL-CM4	2.91	0.71	4.09	1.97	1	0.86	
NOAA-GFDL	GFDL-ESM4	3.51	1.31	2.68	1.53	2	0.79	0.15
NUIST	NESM3	3.73	0.78	4.76	2.73	2	0.93	0.17
UA	MCM-UA-1-0				1.94	1	0.81	
Mean		3.69	0.95	3.90	1.96	4.9	0.78	0.12
Standard deviation		0.40	0.34	1.18	0.42	9.4	0.19	0.06
AS-RCEC	TaiESM1				2.34			
BCC	BCC-ESM1	3.03	0.89	3.39	1.74			
E3SM-Project	E3SM-1-0	3.23	0.60	5.38	2.99			
NASA-GISS	GISS-E2-1-G	3.89	1.43	2.71	1.68			
NASA-GISS	GISS-E2-1-H	3.55	1.14	3.12	1.89			
MOHC	HadGEM3-GC31-MM	3.36	0.61	5.52	2.37			
MPI-M	MPI-ESM1-2-HR	3.58	1.20	2.99	1.64			
SNU	SAM0-UNICON	3.83	1.02	3.76	2.25			
Mean CMIP6		3.70	0.95	**3.91**	**2.01**			
Standard deviation CMIP6		0.39	0.33	1.17	0.42			
Mean CMIP5		3.58	1.06	**3.31**	**1.79**			
Standard deviation CMIP5		0.22	0.29	0.76	0.34			

Table 6.1: List of CMIP6 models used in this chapter and their effective radiative forcing at CO_2 doubling $F_{2\times}$, the climate feedback parameter λ, equilibrium climate sensitivity (ECS) and transient climate response (TCR). Mean values are reported for models with multiple realisations. The values of $F_{2\times}$, ECS and λ are computed using the Gregory method (Gregory, 2004). Models above the horizontal line were used in the extended simulations to 2019. Models below the line did not have SSP simulations available at time of writing. Consistently derived values for CMIP5 are displayed in the Table A.1

6.2.4 Warming trend

Historical warming (our observable) is found from the historical and SSP simulations using the global annual mean surface air temperature (GMSAT) smoothed with a

equally weighted running mean. Some of these models have multiple runs starting from different initial conditions, forcing time series or parameter settings. We use all available runs.

We use smoothed GMSAT to calculate warming. This is to limit the random effect of internal variability on the forced change we wish to constrain. We smooth using a centred 11-year running mean to remove shorter interannual and mid-term variability from sources such as ENSO, as well as reducing the effect of longer period modes of natural variability. We have tested the robustness of the constraint on TCR to the length of the running mean. It remains relatively invariant past a length of 8 years, suggesting most of the internal variability in GMSAT resides in shorter periods.

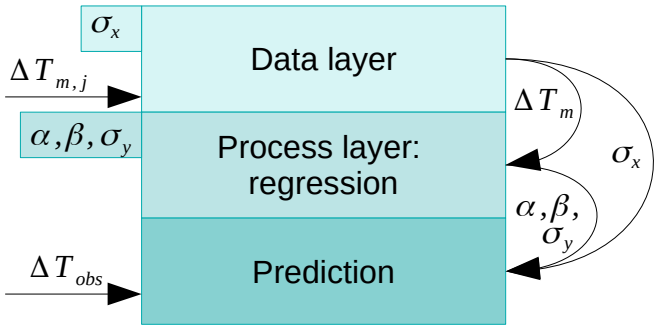
Warming ΔT is calculated as the difference in GMSAT between two periods, typically the years 1975–1985 and 2009–2019 (or equivalently, the difference in smoothed temperature between 2014 and 1980). We have chosen the end year to be 2019 to maximise the chances of discrimination between high and low sensitivity models. As the forcing from CO_2 increases with time, the warming in more sensitive models is more likely to diverge from less sensitive ones as we extend the period over which we calculate the trend. Extending to 2019 also allows us to include the most recent observational data and to eliminate possible effects from the warming slowdown between 2000 and 2012. This slowdown has been attributed to a combination of internal variability and decreased forcing, amongst other things (Medhaug et al., 2017). We assess the impact of the slow down by comparing emergent constraints derived from time series truncated to have different end years.

6.2.5 Regression

When only one realisation per model is used for ordinary least square regression, regression dilution takes place in which the slope is underestimated (Cox et al., 2018b). This has the potential to lead to a slight overestimation of TCR (Figure 6.8d), as the observed warming is on the lower end of the model range. JM19 used the average warming for models with multiple simulations. As not all models provide a sufficient amount of simulations, they state that this leads to a minor inflation of the estimation of uncertainty.

To systematically include the information from all model realisations, we use a hierarchical Bayesian model (Sansom, 2014). This model includes two layers: the normal linear regression (process layer) and a layer that computes the expected warming per model from all its initial value realisations (data layer). To include the initial value ensemble members, we assume that each model m has a ”true” or ”best” value for warm-

Figure 6.3: Schematic of the hierarchical Bayesian model. The data layer models a best estimate of historical warming for each model. With this estimate, a regression is performed between historical warming and TCR in the process layer. Using information from both layers and observed warming, a probability density function is estimated for TCR is the final step.



ing over the last decades denoted by ΔT_T. We further assume that every realisation j of a model gives a value of ΔT that is drawn from a normal distribution with mean ΔT_T and a standard deviation σ_x that is the same across all models. Our hierarchical model consists of two steps: for each model the best estimate of historical warming is computed and with this value a simple linear regression is performed:

$$\begin{aligned} \Delta T_{m,j} \mid \Delta T_m, \sigma_x &\sim \text{normal}(\Delta T_m, \sigma_x) \\ \mathrm{TCR}_m \mid \alpha, \beta, \sigma_y &\sim \text{normal}(\alpha + \beta \Delta T_m, \sigma_y) \end{aligned}$$

The second layer corresponds to normal linear regression, while the first level makes an estimate of the true ΔT_m. Note that especially for models with only few initial value member, the "best" ΔT_m does not necessarily correspond with the mean value of these ensemble members, but will instead lie closer to the regression line.

The probability density function for TCR is then sampled from the observed warming between 1975–1985 and 2009–2019 ΔT_{obs} using the emergent constraint. The observational uncertainty σ_{obs} is taken as the sample standard deviation of the four observational datasets.

$$\begin{aligned} &\mathrm{TCR}_{\text{pred}} = \\ &\quad \text{normal}\left(\alpha + \beta\, \text{normal}\left(\Delta T_{\text{obs}}, \sqrt{\sigma_x^2 + \sigma_{\text{obs}}^2}\right), \sigma_y\right) \end{aligned}$$

As no warming is expected if climate sensitivity were zero, we expect the regression to pass through the intercept and chose a prior for the intercept α of normal(0, 1). Weakly informative priors are chosen for the slope β, the uncertainty of the regression

σ_y and the internal variability σ_x:

$$\begin{aligned}
\alpha &\sim \text{normal}(0, 1) \\
\beta &\sim \text{normal}(2, 10) \\
\sigma_y &\sim \text{half-normal}(0.5, 10) \\
\sigma_x &\sim \text{half-normal}(0.2, 0.5)
\end{aligned}$$

6.2.6 Theoretical basis

Transient Climate Response (TCR)

Once choices of length of running mean, start and end years for calculation of ΔT are fixed (our observable), we can fit an emergent relationship between the observable and values of TCR via linear regression. Linear regression is performed using the hierarchical Bayesian model described in the previous section, which can take into account all the different simulations per model: models with more simulations have a better-constrained post-1975 warming. This results in a set of 127 simulations from 26 different models. The choice of linear regression is justified by considering a two-layer energy balance model (Winton et al., 2010; Geoffroy et al., 2013b):

$$\begin{aligned}
C\frac{dT}{dt} &= F - \lambda T - \epsilon\gamma(T - T_0) \\
C_0\frac{dT_0}{dt} &= \gamma(T - T_0).
\end{aligned} \tag{6.2}$$

Here T is the top layer temperature anomaly, T_0 the deep ocean temperature anomaly, λ is the climate feedback parameter, ϵ is the ocean heat uptake efficacy (reflecting a pattern effect), and γ is the ocean heat uptake parameter (Winton et al., 2010). The parameters C and C_0 are the heat capacity of the upper ocean and deep ocean, respectively. We will refer to this model as EBM-ϵ, or EBM-1 if ϵ is set to 1. We follow the approximations in Williamson et al. (2018) and JM19 in assuming no change in deep ocean temperature ($T_0 = 0$), and assuming the upper ocean to be in equilibrium ($dT/dt = 0$). These assumptions are reasonable for timescales larger than a decade, but smaller than a century (see JM19), and lead to the following relationship:

$$\mathrm{TCR} = s\Delta T \tag{6.3}$$

Here s is a forcing parameter, defined as $F_{2\times}/F$, and ΔT is the difference in temperature between two periods. For fitting, we include an offset η, so that $\mathrm{TCR} = s\Delta T + \eta$,

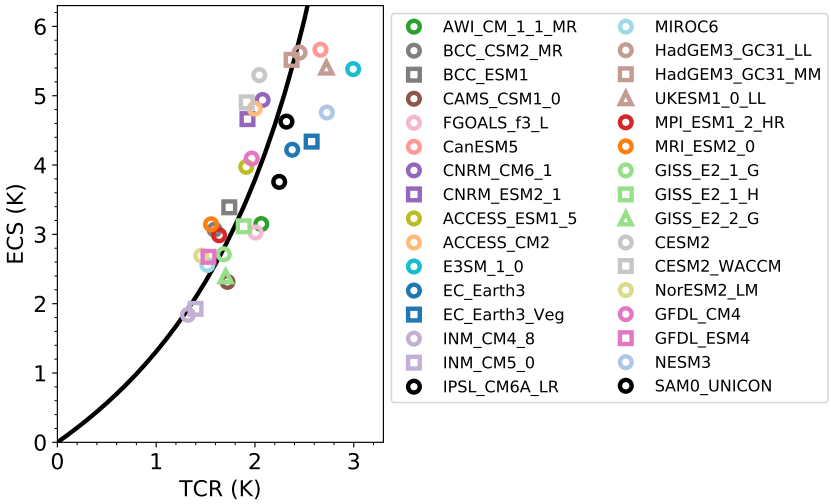


Figure 6.4: Scatter plot of TCR values plotted against ECS values for all CMIP6 models with both available at the time of submission. Models from the same modelling group are plotted with the same colour. Plot markers differentiate models from the same modelling centre. Black line uses the average ocean heat uptake parameters as fitted from the Geoffroy et al. (2013a) procedure, listed in Table A.3 and related ECS and TCR via: $\mathrm{ECS} = \mathrm{TCR}/(1 - e'\mathrm{TCR})$, with $e' = 0.24$, the model mean.

allowing for a possible model mis-specification and regression dilution (Hahn, 1977). A hierarchical linear regression was adopted which includes both uncertainty in ΔT and TCR. The choice of 1975 for the starting period minimises the uncertainty in our estimate of TCR. However, uncertainty is relatively flat for starting periods between 1975 and 1990. We also investigated the sensitivity of our TCR constraint to the final year, the length of the running mean, the model selection, and the method of regression.

Equilibrium climate sensitivity (ECS)

As with TCR, we use the warming between 1975–1985 and 2009–2019 to find an emergent constraint on ECS. The relationship between climate sensitivity and observed warming or TCR is not expected to be linear, as a smaller fraction of equilibrium warming is typically realised in models with high climate sensitive within the first decades of warming (Hansen et al., 1985; Rugenstein et al., 2020). Using Equation 6.2, $\mathrm{ECS} = F_{2\times}/\lambda$, and again assuming the upper ocean to be in equilibrium and the deep ocean temperature to not change, TCR and ECS are related via:

$$\mathrm{ECS} = \mathrm{TCR}/(1 - e'\mathrm{TCR}). \tag{6.4}$$

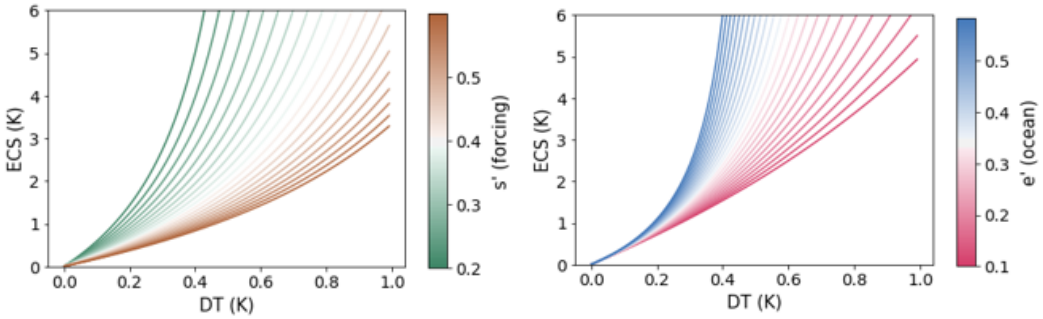


Figure 6.5: Variation of parameters of the function of ECS versus historical temperature rise. The left figure shows the variation caused by the forcing parameter, whereas the right picture shows the ocean parameter.

Figure 6.4 plots model TCR versus ECS, related via Equation 6.4, using the ensemble mean of fitted ocean parameters e' of the CMIP6 models. We fit the ocean heat uptake and forcing parameters, following the two algorithms described in Geoffroy et al. (2013b,a), with slight modifications to ensure solutions exist for all models, described in Appendix A.

From Equation 6.4, the relationship between ECS and ΔT ends up as:

$$\mathrm{ECS} = \frac{\Delta T}{s' - e'\Delta T}. \tag{6.5}$$

The forcing parameter is denoted by s', defined as $\Delta F/F_{2\times}$ and e' is the ocean heat uptake parameter defined as $\epsilon\gamma/F_{2\times}$. The function has an asymptote at $s' - e'\Delta T = 0$, and turns negative for larger ΔT values. As negative ECS values are unphysical, we modify the equation by keeping ECS at infinity for $\Delta T > s'/e'$. The appearance of negative ECS for high ΔT is an artefact of the no-deep-ocean-temperature-rise assumption: it corresponds to an equilibrium between the heating effect of $F - \lambda\Delta T$, balanced by $-\epsilon\gamma\Delta T$. In reality, this last term cancels completely when ΔT_0 reaches equilibrium and ocean heat uptake stops. Similarly, the negative ECS branch for $\Delta T < 0$ is also taken to be unphysical and therefore all negative ECS values are rejected during sampling.

This two-layer model, first proposed by Held et al. (2010), has a parsimonious description of the ocean pattern effects, with the extra term $(\epsilon - 1)\gamma(T - T_0)$ compared to a standard two-layer model, which changes over time as the deep ocean heats up, acting as a modification of the feedback parameter. Figure 6.5 shows how the function of ECS from ΔT changes depending on its parameters. The ocean parameter e' does not impact the function much for low values of ΔT. This means that it is more influenced by models with high ECS, which typically have a stronger pattern effect.

To test the validity of the function, we perform two checks. Firstly, using the fitted model parameters, we investigate the physical basis of Equation 6.5 with the EBM-ϵ and EBM-1 models. If this function derived from the two-layer model is a faithful representation, $\Delta T/(s' - e'\Delta T)$ should be better related to ECS with individual model parameters than with the bulk fitted parameter.

Secondly, by explicitly simulating the two-layer model, we investigate to what extent the analytical functional form deviates from the true functional form. We are especially interested in the upper region of this functional form, which, if too steep, could lead to an upper estimate of ECS biased high.

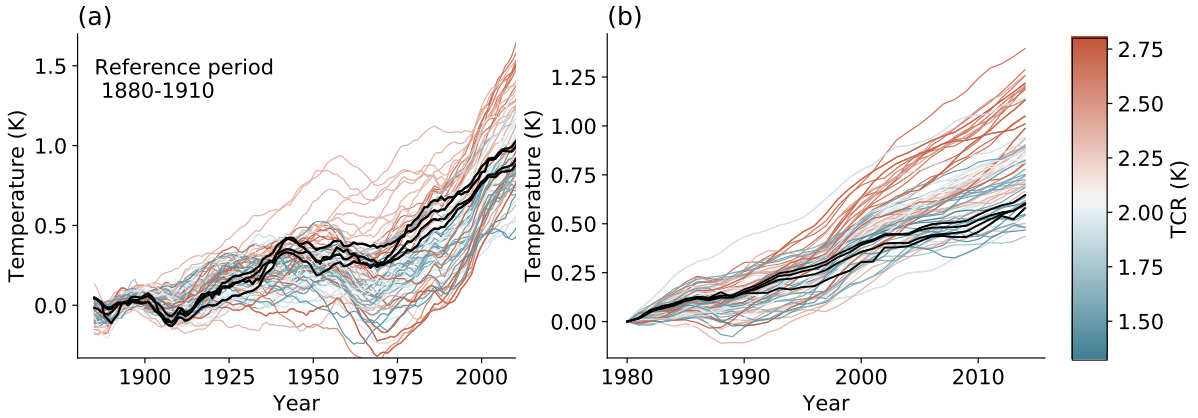
The parameters in Equation 6.5 are fitted with Orthogonal Distance Regression (ODR) using the Python implementation of the `ODRPACK` Fortran package (Boggs et al., 1989). We use Monte Carlo sampling to compute the confidence interval, generating a joint distribution of the model parameters and the error in the estimation. We draw from this joint distribution to generate a prediction interval. ECS is estimated by generating a second joint distribution, which also includes the observational uncertainty. As with TCR, internal variability is included in the observational uncertainty as estimated from model internal variability. Given the fact that not all models have multiple ensemble members, there is still some internal variability influencing the fit. This leads to a small overestimation of uncertainty.

6.3 Results

6.3.1 Transient climate response

Figure 6.6a shows the temperature anomaly over the period 1880 to 2019 simulated by 26 different CMIP6 models running a total of 127 simulations smoothed with a 11-year running mean. The reference period in this case is 1880–1910. Model runs have been colour coded by their TCR value, with darker red indicating models with higher TCR, and darker blue indicating lower TCR. Black lines are observational global warming datasets over the same period (Morice et al., 2012; Rohde et al., 2013; Lenssen et al., 2019; Zhang et al., 2019). Models with higher TCR either show large warming at the end of the period, or portray a strong aerosol cooling over the 20th century, particularly visible as a dip around 1960–1970 (notably CNRM-ESM1, UKESM1-0-LL and EC-Earth-Veg). Figure 6.6b shows the same information for the end of the historical period although the reference period is now chosen to be 1975–1985, after the temperature dip. The positive correlation intuitively expected between TCR and temperature increase ΔT is much clearer for this time interval.

Figure 6.6: Global mean surface temperature of the 26 CMIP6 models named in Table 6.1. To avoid visual over-representation, a maximum of ten realisations per model are plotted. An 11-year running mean was used. (a) Temperature anomaly ΔT using 1880–1910 as reference period. (b) Temperature anomaly ΔT relative to the 1975–1985 mean.

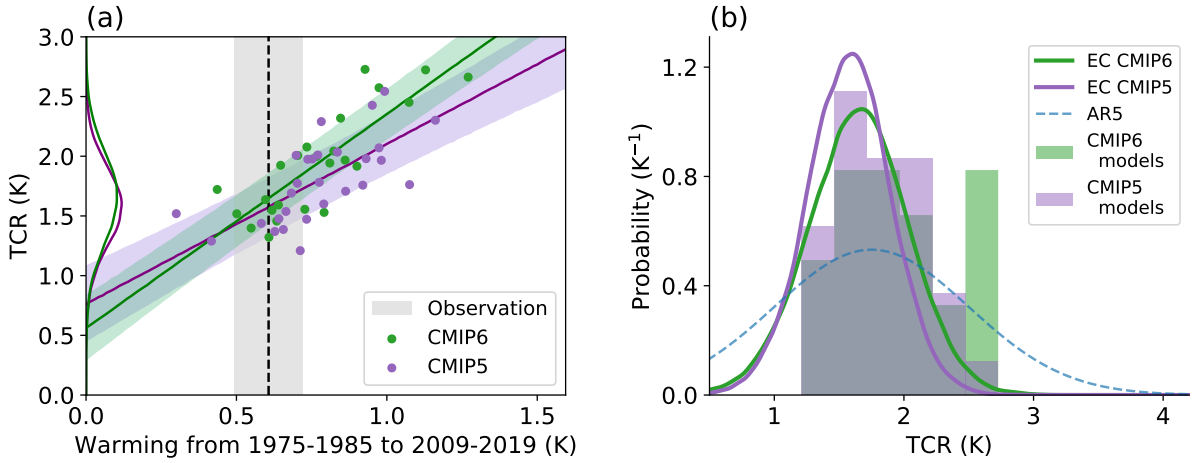


The ΔT for each model simulation in Figure 6.6b is used for the emergent constraint on TCR in Figure 6.7a. Observational warming (black vertical dashed line) is the mean of HadCRUT4 (Morice et al., 2012), Berkeley Earth (Rohde et al., 2013), GISSTEMP4 (Lenssen et al., 2019) and NOAA v5 (Zhang et al., 2019). The 90% observational confidence interval (grey shaded vertical area) is a combination of the observational uncertainty (conservatively estimated as the sample variance of the four data sets) and the internal variability.

To avoid double-counting observational uncertainty, the 90% regression confidence interval details the uncertainty of the "best" estimate of ΔT_m versus TCR. The models from the previous CMIP5 generation generally fall within the prediction interval of the CMIP6 emergent constraint: the emergent constraint is robust across generations (Klein and Hall, 2015). The best estimate (1.68 K) from this emergent constraint is higher than the best estimate using the larger set of models that have historical simulations up to 2014, but no future scenarios (median: 1.54 K, 5–95% range: 0.76–2.30 K). This can mostly be explained by the fact that 2004–2014 overlaps with the slow-down in surface temperature increase over the 2000–2012 period, but the larger selection of models over the historical period also impacts the regression.

Figure 6.7b shows the probability density functions (pdf) of TCR derived from the emergent constraint for both CMIP6 and the earlier CMIP5 model ensembles. For comparison, the raw model range in each CMIP is plotted as a histogram, as well as the reported IPCC AR5 likely range (assuming a normal distribution). Both CMIP5 and CMIP6 pdfs are very similar (central estimates differ by 0.1 K) even though CMIP6

Figure 6.7: (a) Emergent constraint on TCR against historical warming. The predictor ΔT is calculated from the difference between 1975–1985 and 2009–2019 of a timeseries of GMSAT. Linear regression is performed with all CMIP5 and CMIP6 simulations. Shaded areas indicate a 90% model prediction interval (so not including internal variability). The vertical dotted line is the mean value of the observations and y-axis shows the probability distribution of both generations of ensembles. **(b)** Comparison of probability distributions for the transient climate response using post-1975 warming using CMIP5 and CMIP6 simulations. The probability distribution in the fifth IPCC assessment is not fully specified, so the figure shows a normal distribution with the same likely range as IPCC.



contains many more high TCR models. As a continuation of the historical CMIP5 simulation, RCP8.5 is chosen. The tighter constraint in CMIP5 is mostly a consequence of differences in internal variability, which is 42% larger in CMIP6 than in CMIP5, in line with the findings of Parsons et al. (2020).

Period selection

Estimates of TCR depend on the final year chosen for the emergent constraint. Uncertainty in the estimate of TCR reduces as time increases and the central estimate converges as shown in Figure 6.8a. Later end years are favoured as the signal-to-noise ratio of the net radiative forcing increases monotonically after 1975 (see Figure 6.2). In the 21st century, the climate impact of volcanoes has been dominated by smaller eruptions (Stocker et al., 2019). The scenarioMIP simulations used for 2015–2019 include a small background forcing from volcanoes (O'Neill et al., 2016). We estimate errors from a potential mismatch between model and real forcing to be relatively small.

To mitigate the effect of internal variability, we use a running mean of GMSAT. Figure 6.8b shows the likely range of TCR as a function of the length of the running mean.

Since we use all available simulations including multiple realisations of the same model in the emergent constraint, the effect of internal variability is already reduced and the length of the running mean on the estimate of TCR is small — the central estimate and the likely range remain relatively invariant past a window length of 8 years.

Figure 6.8c shows the effect of the start year on the emergent constraint. Uncertainty in the estimated value of TCR is relatively flat between start years of 1975 and 1990. Uncertainty for start years from 1990 onwards increases until the estimate and the uncertainty revert towards the raw CMIP6 ensemble statistics (no predictive power) at later years.

Regression method

In addition to hierarchical Bayesian regression, we have investigated three other regression methods used in the emergent constraint literature: ordinary least squares (OLS) with only one realisation per model, OLS on the mean warming per model and orthogonal distance regression (Figure 6.8d). While the first three give very similar results, orthogonal distance regression gives a somewhat lower estimate of TCR, consistent with the results found in Section 4.2. Orthogonal distance regression assumes that there are both errors in the predictor and in the predictand, which leads to a steeper slope. As our observation lies under the average, a steeper slope results in a smaller predicted TCR value. Orthogonal distance regression is known to sometimes overcompensate for errors in the independent variable, for instance in the case the statistical model is not perfectly known; if the model deviates from being a perfectly straight line (Carroll and Ruppert, 1996).

Model selection

Model selection can prevent double counting of very similar models (Sanderson et al., 2015; Cox et al., 2018a). As models from the same centre can have very dissimilar climate sensitivities (Chen et al., 2014; Jiménez-de-la Cuesta and Mauritsen, 2019) and sensitivity can change drastically with only small adjustments to parameters (Zhao et al., 2016), we initially use all available models in the CMIP5 and CMIP6 ensemble. Figure 6.8e shows that this choice does not significantly change the best estimate of the transient response, and that using one model per modelling centre only very slightly increases the variance, even as models from one modelling centre are relative similar (Figure 6.4).

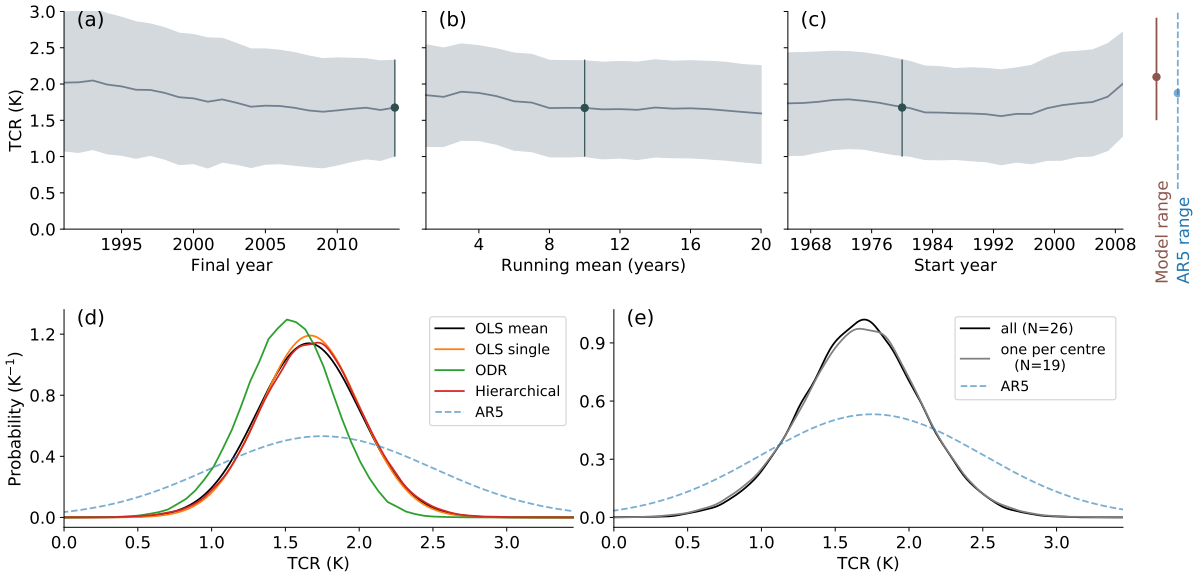


Figure 6.8: Robustness of the result to various parameter choices and the choice of regression method. Unless stated differently, start year is 1975, all years up to 2018 are used, and the length of the running mean is 11 years. For comparison, the 5–95% model range range and IPCC range are shown, both assuming normal distributions. **(a)** 5–95% TCR range as a function of the final year (blue line central estimate). **(b)**: 5–95% TCR range as a function of length of running mean. **(c)** 5–95% TCR range as a function of start year. **(d)** Pdf of TCR from different regression methods: the hierarchical Bayesian model is compared to three other linear regression methods used in the emergent constraint literature: ordinary least squares (OLS) with only one realisation per model and OLS on the mean warming per model and orthogonal distance regression (ODR). **(e)** Resulting pdfs on TCR from stricter model selection (one model per modelling centre) compared to regression using all models and the IPCC AR5 range.

Study	CMIP	Period	Percentile		
			Median	5–95%	16–84%
JM19	CMIP5	1970 - 2005	1.7 K	1.2 – 2.2 K	
Nijsse et al. (2020)	CMIP5	1970 - 2005	1.7 K	1.1 – 2.3 K	1.4 – 2.1 K
Tokarska et al. (2020)	CMIP6	1981 - 2017	1.6 K		1.2 – 2.0 K
Nijsse et al. (2020)	CMIP6	1975 - 2019	1.7 K	1.0 – 2.3 K	1.3 – 2.1 K

Table 6.2: Emergent constraint on TCR depending on choices of ensemble and time period. Results from JM19 and Tokarska et al. (2020) are also shown for comparison.

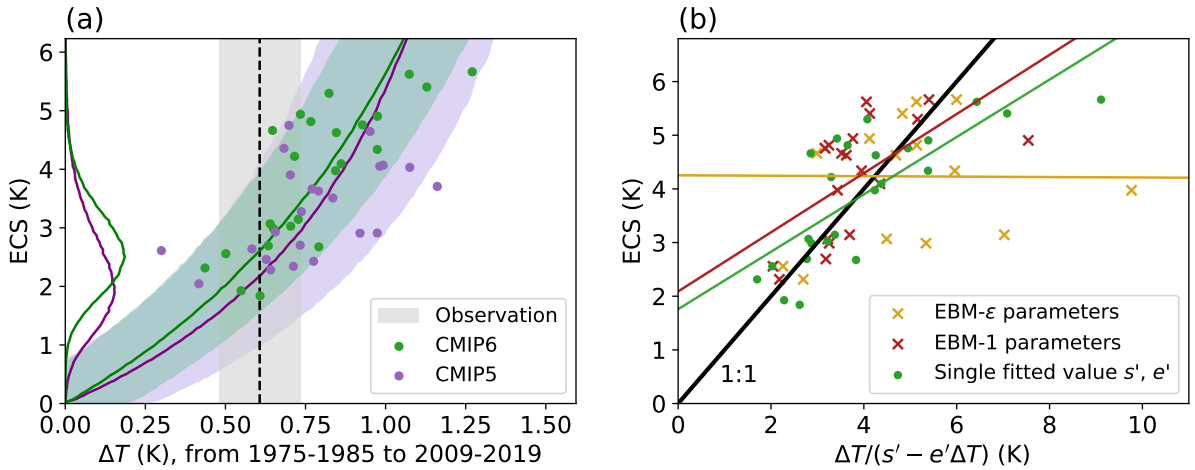


Figure 6.9: a) Emergent constraint on ECS, using the functional form of Equation 6.5. The shaded area includes the 5–95% confidence interval. b) Comparison of emergent constraint fitted parameters, with using model values for s' and e'. The coloured lines are OLS fits for the three cases, and the black line indicates the 1:1-line. Three values for the EBM-ϵ model are not shown as their $\Delta T/(s' - e'\Delta T)$ are between 75 K and 90 K.

Ensemble	Median	5–95% range
CMIP5 1970–2005	2.3 K	0.7 – 8.4 K
CMIP5 hist + RCP8.5	2.2 K	1.0 – 4.1 K
CMIP6 1970–2005	2.5 K	1.0 – 8.6 K
CMIP6 historical	1.9 K	1.0 – 3.3 K
CMIP6 hist + SSP2-4.5	2.6 K	1.5 – 4.0 K

Table 6.3: Emergent constraint on ECS depending on the choice of ensemble and time period.

6.3.2 Equilibrium climate sensitivity

Figure 6.9a shows the emergent constraint on ECS. For CMIP5, the 5–95% confidence interval lies between 0.96–4.09 K. The constraint is stronger for CMIP6, with the 5–95% confidence interval spanning 1.52–4.03 K. Further results are shown in Table 6.3.

The results are highly dependent on the time interval chosen. For shorter intervals, the theoretical functional form shows an increased steepness for higher values of ΔT, making it more difficult to constrain. For instance, taking the time period in line with JM19, i.e. 1970–1989 versus 1994–2005, we obtain a 5–95% interval of 0.70–8.41 K for CMIP5, significantly wider than found in JM19, which reported a 5–95% confidence interval of 1.72-–4.12 K. The major differences lie in the definition of the theoretical function, where we have cut off the unphysical branch, and a correction of a coding error in the computation of the prediction interval.

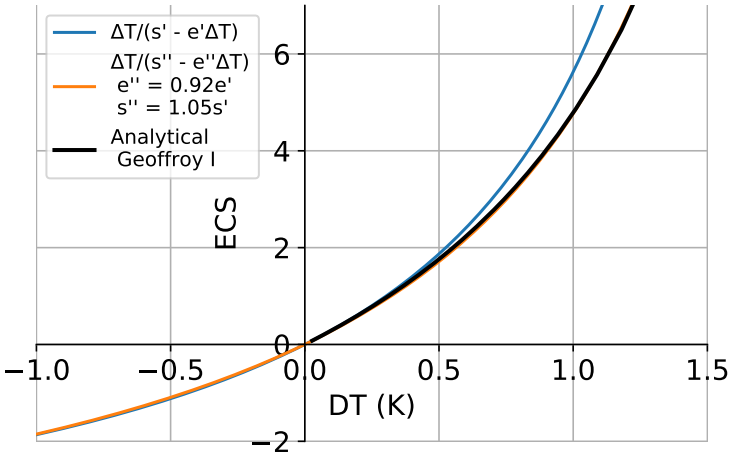


Figure 6.10: Comparison of two functions describing the relationship between observed warming and climate sensitivity. The analytical function (black) corresponding to Equation 14 in Geoffroy et al. (2013b), not only depends on γ and the forcing, but also on the heat capacity of the two layers, for which typical values were chosen of $C = 8$ and $C_0 = 100$. The linear forcing was $1/140\, F_{2\times}$, and the function was evaluated between year 20 and 70, to cancel transient effects. The parameter s' was chosen to be 5/14, and e' is 0.18. The blue and orange lines are both drawn using $\mathrm{ECS} = \Delta T/(s' - e'\Delta T)$, where in the latter case, the parameters are chosen to approximate the full equation as closely as possible (see legend).

In Figure 6.9b the dark green dots represent expected ECS from observed warming (using Equation 6.5) and true ECS, using the fitted parameters from Figure 6.9a. The light green dots denote the same, but now every model uses its own ocean parameters, $F_{2\times}$ and model forcing computed using Equation 6.1. The yellow data shows the expected ECS computed from the EBM-1 model. Full parameter fits for both models are found in Tables A.2 and A.3.

The EBM-ϵ model performs poorly for large values of the ocean heat uptake efficacy parameter ϵ. Models with ϵ around 1.8 in particular show an expected ECS far above a realistic range, with one expected ECS reaching a value of 89 K. Equation 6.5 is nonlinear and small errors in parameter estimation quickly lead to large errors in ECS. For the EBM-ϵ model in particular, high internal variability may skew the parameter estimate upwards.

The EBM-1 fit leads to an improved estimation of ECS compared to the Equation 6.5 fit in 53% of the cases, whereas the EBM-ϵ model leads to an improvement in 34% of cases. This pattern in similar in the case only historical models are used, with 66% and 42% improved respectively.

Functional form

Explicitly simulating the relationship between ΔT and ECS using the two-layer model shows that the steepness of the function used for the emergent constraint (Equation 6.5) is overestimated for high ΔT: assuming no deep ocean temperature rise ($T_0 = 0$) dampens the temperature response of the upper ocean. Geoffroy et al. (2013b) derived an analytical solution to the two-layer model of Equation 6.2 under the weaker assumption of a linearly increasing forcing (so, allowing the deep ocean to take up heat), which also showed a less steep increase of ECS with ΔT for high values of ΔT.

This leads the question whether the upper range of ECS is overestimated. In Figure 6.10, we show this is not the case: by using a decreased ocean heat uptake parameter e' and forcing, the two analytical solutions do overlap, which demonstrates that using the approximated Equation 6.5 in the regression should not lead to biased results in the emergent constraint, but simply that the fitted parameters will be slightly different from the model parameters. This also explains why the regression using model parameters in Figure 6.9b is not significantly better than using the overall fitted parameters of Figure 6.9a.

6.4 Discussion and conclusion

The emergent constraint found on TCR in this chapter is very similar to the one found in JM19 and Tokarska et al. (2020). The most important determinant of the constraint is the periods taken. We have slightly expanded on the amount of models compared to a Tokarska et al. (2020), taking a different period, and we compared further regression choices.

Major unknowns in determining climate sensitivity relate to aerosol forcing and, for equilibrium sensitivity, the possibility of feedback increasing over time. We have moderated the impact of aerosol forcing by our choice of time period. We assume that the remaining uncertainty is captured by model diversity, and therefore included in our confidence interval. Time-varying feedbacks are modelled via the inclusion of ocean pattern effects in our functional form for the emergent relationship, where the diversity in model efficacy again informs the confidence interval.

Our best estimate for TCR from the CMIP6 models is 1.68 K, which remains close to the centre of the likely range (1–2.5K) given in the IPCC AR5 (IPCC, 2013b). The emergent constraint on TCR from the CMIP6 models is however strong enough to indicate a much tighter likely range of TCR (16–84%, 1.29–2.05 K).

We find a consistent emergent constraint from the CMIP5 models against observed global warming from 1975 to 2019 (16–84%, 1.27–1.88 K). Furthermore, both of these likely ranges overlap strongly with the emergent constraint on TCR derived by Jiménez-de-la Cuesta and Mauritsen (2019) using a similar method, but only considering global warming from 1970 to 2005 (5–95%, 1.17–2.16 K). In terms of the classification proposed by Hall et al. (2019), we therefore now have a *confirmed* emergent constraint on TCR, with consistency across generations and a sound theoretical framework.

Equilibrium climate sensitivity is found to be likely between 1.9 K and 3.4 K (16–84% percentile). This finding strengthens previous evidence that ECS very unlikely above 4.5 K (Cox et al., 2018a; Jiménez-de-la Cuesta and Mauritsen, 2019; Goodwin et al., 2018). For instance, Goodwin et al. (2018) used history matching, a simple emulator, and observations of surface temperature, ocean heat uptake, and carbon fluxes to estimate climate sensitivity and concluded upon a 5–95% range of 2.0 K to 4.3 K. Renoult et al. (2020) used a combined emergent constraint of the last glacial maximum and mid-Pliocene Warm Period to constrain ECS to 1.1–3.9 K, with the same best estimate of 2.6 K.

Does the presence of many models with ECS over 4.5 K mean that the CMIP5 generation was better or more useful for understanding climate sensitivity than CMIP6? From the point of view of emergent constraints the answer is clearly no, as model spread helps capture the shape of the emergent relationship.

In the future, we hope that this TCR constraint will become the basis for constraints also on TCRE (transient climate response to emissions), but this will require the inclusion of additional constraints on land and ocean carbon uptake.

However, we are now in a position to answer the questions that we posed in Section 6.1:

(a) *Are such high climate sensitivities consistent with the observational record?*

No, models with high ECS (> 4.5 K) and high TCR (< 2.5 K) do not appear to be consistent with observed global warming since 1975 (Figure 6.6b).

(b) *If so, do the CMIP6 models demand an upward revision to the IPCC likely ranges for climate sensitivity?*

No, instead emergent constraints on TCR (Figure 6.7) and ECS (Figure 6.9) suggest narrower likely ranges for TCR (1.3–2.1 K) and ECS (1.9–3.4 K).

The associated code is available at Code Ocean (Nijsse et al., 2020a).

Chapter 7

Conclusion and outlook

7.1 Conclusion and discussion

Emergent constraints have proven to be a simple technique that can shed light on key questions in climate science, such as the strength of important feedbacks and how much warming will occur under a doubling of CO_2. In this thesis, we have sought to improve theoretical understanding of emergent constraints in general and showcase several examples of theory-led emergent constraints.

Not all emergent constraints have the same general mechanism. We have proposed classifying emergent constraints into four classes, according to their mathematical underpinning. Direct dynamical constraints correlate variability in a variable to its response under a warming climate. On the other side, static indirect emergent constraints relate a climatological bias in one variable to response in a different variable. The former can be well understood from linear response theory, and so confidence in this type of constraint is more easily established.

Conceptual simple models form the basis of theory-led emergent constraints. Their use was explored for the analysis of the snow-albedo feedback using a modification of the energy balance model by Fraedrich (1979), and we used the Hasselmann model (Hasselmann, 1976) and generalisations for the relationship between decadal variability and climate sensitivity. The same models were the basis for an emergent constraint between the trend in observed warming and climate sensitivity. This type of analysis suggests that the term “emergent constraints” may in fact be a misnomer. If done correctly, hypotheses should come first based on physical reasoning and multimodel relationships should be expected, instead of catching the researcher by surprise by emerging.

A good physical understanding also helps us choose the proper statistical model.

Often, ordinary linear regression between the predictor and the predictand is chosen without good justification. The model data itself is often insufficient to decide on the functional form. Other statistical models with few free parameters may be equally valid and better fit our physical understanding. Regression can also be improved by using all the available data. Specifically, initial value members can be exploited to improve the prediction confidence interval.

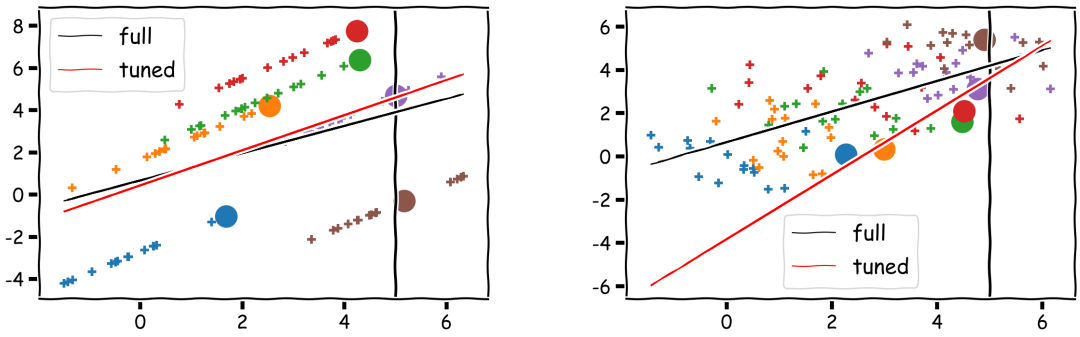
Three specific emergent relationships were studied in this thesis. The first one concerned the relationship between decadal temperature trends and climate sensitivity. It was found that within the CMIP5 model ensemble, these two quantities were strongly correlated. As a consequence, counter-intuitively, models with high climate sensitivity are more likely to display a period without any warming under current greenhouse gas concentration trends. Furthermore, the compound risk of a decade of high warming and background global warming was found to be significant: decades with as much warming as the entire 20th century can not be excluded under high climate sensitivity.

A second set of emergent constraints were found linking surface temperature rise to transient warming and climate sensitivity. The newest generation of climate models, CMIP6, contains more models showing very high sensitivity to greenhouse gases. These models are typically not consistent with warming trends observed since the 1970s, a period with comparable aerosol-induced cooling as the current day, revealing greenhouse gas heating. These results indicate that the Paris goal of limiting global warming to well below 2℃ is still within reach.

7.2 Future research

7.2.1 Tuning and its influence on emergent constraints

A possible complication in the interpretation of emergent constraints is to what extent the observational data has been used in the formulation of the model. In short; all models have been implicitly and sometimes explicitly tuned to the observational record (Jebeile and Crucifix, 2020). Climate scientists have in the past expressed concern about data being used both for calibration and confirmation of climate models. In response, Steele and Werndl (2013) have shown that it is often defensible to use observational data for tuning (which version of models A and B is best) and relative confirmation (whether model A is better than model B). It could be argued that emergent constraints are akin to model confirmation, but the mathematical treatment is different and it is not yet known to what extent calibration has an impact on the conclusions of emergent constraints.



(a) (b)

Figure 7.1: The effect of tuning on a hypothetical emergent constraint. In both panels, each colour corresponds to one climate model. 'Tuning' is performed by choosing the model variant which is closest to the observations, and the chosen variant is denoted by a circle instead of a small cross. (a) each model deviates from the 1:1-emergent relationship in X and Y, but calibration happens on a perfect slope-1 line. (b) Models lie on the 1:1-emergent relationship and each version has an independent random offset in both the independent and the dependent variable.

Possibly, some emergent relationships only appear because the state space has been reduced sufficiently with calibration to display a simple relationship. In other cases, the tuning might actually give a biased estimate of the gradient of the linear regression. It is vital to try and identify the consequence of calibration on emergent constraints. An example of when this can pose a problem is when the observations of the predictor are already used in the tuning process. In Figure 7.1, two quick examples of this type of tuning are presented, showing two potential ways that a climate model may change as a function of its parameters. Each model has a range over which it can be calibrated: even with an optimal parameter, models with an inadequate structure will not be able to reproduce observations (cf. Figure 3.7, which shows the snow albedo feedback range in the PlaSim climate model over a range of parameter values. The range was smaller than in the multi-model ensemble). These two examples mainly impact the gradient, and do not give a (large) bias in the expected predicted value, but many more examples might exist.

7.2.2 Targeted model development

An underused aspect of emergent constraints is targeted model development. If an EC has a strong physical foundation, we can expect that taking steps to reduce bias in X will result in a corresponding convergence in Y. First, model developers and users

should work together to assess how the structural and parametric design of the climate model affects the differences in the predictor. With this collaborative analysis the best parameterizations or parameters for simulating X may be revealed, thereby providing guidance for modelling centres on how they can improve their model if they have a bias in X. If the error is parametric, calibration with the identified predictor may be used to find a new optimal parameter value. If the error is caused by the parametrizations, more significant change in the ESM may be necessary.

As far as we are aware, this type of analysis has only been performed for emergent constraints on the sensitivity of extra-tropical cloud optical depth to temperature (Gordon and Klein, 2014), hydrological cycle intensification (DeAngelis et al., 2015), and snow-albedo feedback (Thackeray et al., 2018).

7.2.3 Paleo-climate

Paleo-records have often been used constrain climate sensitivity (Knutti et al., 2017). Hargreaves et al. (2012) and Schmidt et al. (2014) use the emergent constraints technique on this data to better understand temperature. Paleo-records also contain a treasure of information about other uncertain future quantities, such as sea-level rise and precipitation.

Conversely, emergent constraints may be used to "predict" past climate where paleo-data is insufficient. This application offers the possibility of testing the technique of emergent constraints in an independent manner: in contrast to future climate change which may be centuries away, paleo-climate has already happened and, for many quantities of interest, better proxies are only years or decades away.

7.2.4 Multiple constraints and nonlinearity

Most of the work to date has assumed linear emergent relationships, exceptions being the constraints on climate sensitivity using global surface temperature rise in the current work and Jiménez-de-la Cuesta and Mauritsen (2019). Improved theoretical support of emergent constraints, for instance using a hierarchy of models, can show when nonlinear emergent constraints are more appropriate.

A majority of the work on emergent constraints has further used two linearly related scalar variables, a predictor X, and a response Y. This has been extended to two or more predictors (Renoult et al., 2020; Bretherton and Caldwell, 2020, e.g.). When using multiple predictors, a logical extension to simple linear regression is multilinear linear regression, considering possible collinearity.

When combining these two generalizations of emergent constraints (multiple inputs and physical relationships), the technique may produce statistical models of the global climate. With more of the information being derived from observations as predictors, the impact of the climate models may even be reduced significantly, for instance to confirming the theoretical relationship instead of informing the fitting parameters.

Equivalent to the use of multiple predictors, the scalar X can be replaced by a multidimensional field X and even Y, rather than scalars. This could for example be the spatial field of a climate variable (Brown and Caldeira, 2017), or spatial normal modes of variance, referred to as empirical orthogonal functions (EOFs) in meteorology (Hannachi et al., 2007). With stronger statistical pre-processing, it may however be more difficult to find a theory. Curiously, without theory, emergent constraints may do more right by their name: emerging from model data instead of being sought based on physical intuition and understanding.

7.2.5 Use of conceptual models for establishing a physical basis

One way to form hypotheses of new mechanisms for emergent relationships is using conceptual models that are (often) analytically solvable. Conceptual models require (sometimes strong) assumptions about the real world. These assumptions are ideally testable against the real world and model data (Williamson et al., 2018, e.g.). If the assumptions are adequately met, the solution to the conceptual model can form the physical basis of the emergent relationship. This testing of simplified theory also helps our understanding and intuition of complex climate models.

Using transparent hypotheses helps us guard against two potential dangers in the application of emergent constraints. (i) p-hacking, which might occur when too many hypotheses are tested and (ii) overestimation of correlation by fudging potential free parameters (see Section 4.7).

Model spatial and temporal resolution is continually increasing, and more physical and biochemical processes have been incorporated, resulting in more comprehensive representations of the climate system. They have become the primary theoretical tool in climatology. However, as a result of their growing complexity, it has become increasingly difficult to get a good understanding of their behaviour.

In the past, numerical climate modelling was more limited by computer resources, forcing scientists to think methodically about which parts of the climate system were essential to include to answer specific questions. This led to wider use of conceptual models geared to each question, analytical techniques and a better process intuition (Budyko, 1969; Hasselmann, 1976; Wigley and Raper, 1990; Saltzman, 2002; Raper

and Braithwaite, 2009).

There is an opportunity for the top-down insights of specific conceptual models to complement the comprehensive, bottom-up approach of the latest generation of ESMs; there are many satellite observations (Yang et al., 2013), oceanic profiling floats (Roemmich et al., 2019), and longer surface records. The collection of past and present climate model simulations has grown at the same time. With all the new data, it may be fruitful to revisit these simple intuitive models in the context of emergent constraints.

In conclusion, emergent constraints have already been used to gain a better understanding of many processes in the climate system including the snow-albedo feedback, climate sensitivity, precipitation and changes in the carbon cycle. With the suggestions above, a wider range of aspects of climate change can be explored using emergent constraints.

Appendix A

Additional tables

Parameters in Table A.2 were determined following the algorithm described in Geoffroy et al. (2013b), and parameters in Table A.3 following Geoffroy et al. (2013a). For models that come close to equilibrium within the first 150 years, we made two slight adjustments to the algorithm, allowing parameters to be estimated even in the case where the temperature in a certain year exceeded $2 \times \mathrm{ECS}$. The long time scales (normally estimated from $\log\left(1 - \frac{T}{T_{eq}}\right) \approx \log a_s - \frac{1}{\tau_s} t$, the second estimation step described in Geoffroy et al. (2013b)), were estimated dropping the years after the temperature first exceeds $2 \times \mathrm{ECS}$ to avoid taking the logarithm of negative values. Similar problems occurred for a limited number of models when estimating the fast time scale (normally estimated as $\tau_f = t / \log a_f - \log\left(1 - \frac{T}{T_{eq}} - a_s e^{-t/\tau_s}\right)$). Here instead, we used direct least-squares curve-fitting (using the scipy package `curve_fit` function) over the first ten years of Equation 5.17 to estimate the parameter when taking a logarithm was not possible.

Table A.1: CMIP5 models included in analysis Chapter 6. Values were determined using the Gregory method, consistent with the CMIP6 values.

Centre	**Model**	$F_{2\times}$	$\boldsymbol{\lambda}$	ECS	TCR	n	$\boldsymbol{\Delta T}$	SD
BCC	bcc-csm1-1-m	3.36	1.15	2.91	2.07	1	0.97	
BCC	bcc-csm1-1	3.70	1.27	2.91	1.76	1	0.92	
BNU	BNU-ESM	3.77	0.93	4.07	2.54	1	0.99	
CCCma	CanESM2	3.76	1.01	3.71	2.30	5	1.16	0.05
CMCC	CMCC-CM		1.01		1.98	1	0.75	
CNRM	CNRM-CM5	3.32	1.01	3.28	1.97	5	0.74	0.12
CSIRO-BOM	ACCESS1-0	4.26	1.09	3.90	1.77	1	0.70	
CSIRO-BOM	ACCESS1-3	3.97	1.09	3.63	1.60	1	0.79	
CSIRO	CSIRO-Mk3-6-0	4.76	1.09	4.36	1.69	10	0.68	0.08
INM	inmcm4	1.48	0.72	2.05	1.29	1	0.42	
IPSL	IPSL-CM5A-LR	3.13	0.77	4.05	1.97	4	0.98	0.08
IPSL	IPSL-CM5A-MR		0.54		1.98	1	0.93	
IPSL	IPSL-CM5B-LR	1.43	0.54	2.64	1.44	1	0.58	
MIROC	MIROC-ESM	2.58	0.54	4.75	2.01	1	0.70	
MIROC	MIROC5	1.47	0.54	2.70	1.47	3	0.73	0.08
MOHC	HadGEM2-ES	2.52	0.54	4.64	2.43	4	0.95	0.14
MPI-M	MPI-ESM-LR	5.48	1.49	3.66	2.01	3	0.77	0.04
MPI-M	MPI-ESM-MR	2.74	0.78	3.51	2.03	1	0.84	
MRI	MRI-CGCM3	2.04	0.78	2.61	1.52	1	0.30	
NASA-GISS	GISS-E2-H	1.89	0.78	2.43	1.78	5	0.78	0.08
NASA-GISS	GISS-E2-R	2.18	0.96	2.28	1.48	5	0.64	0.11
NCC	NorESM1-ME		1.56		1.54	1	0.66	
NCC	NorESM1-M	4.57	1.56	2.93	1.39	1	0.66	
NOAA-GFDL	GFDL-CM3	2.45	0.61	4.03	1.76	1	1.08	
NOAA-GFDL	GFDL-ESM2G	1.42	0.61	2.34	1.21	1	0.71	
NOAA-GFDL	GFDL-ESM2M	1.49	0.61	2.46	1.37	1	0.63	
NSF-DOE-NCAR	CESM1-BGC		0.61		1.71	1	0.86	
NSF-DOE-NCAR	CESM1-CAM5		1.09		2.29	3	0.78	0.07
	Mean	2.98	0.90	3.30	1.80	2.3	0.78	0.08
	SD	0.24	0.31	0.79	0.34	2.1	0.18	0.03

Table A.2: Fitted parameters two-layer model for 34 CMIP6 models. Equilibrium climate sensitivity (ECS), climate feedback parameter λ, radiative forcing at doubling of CO_2 $F_{2\times}$, upper ocean/land/atmosphere heat capacity C, deep ocean heat capacity C_0, the heat exchange parameter γ, the fast time scale τ_f, and the slow time scale τ_s.

Model	ECS	λ	$F_{2\times}$	C	C_0	γ	τ_f	τ_s
AWI-CM-1-1-MR	3.13	1.21	3.80	6.34	51.4	0.55	3.56	138.1
BCC-CSM2-MR	3.02	1.03	3.11	7.16	74.1	0.78	3.88	169.7
BCC-ESM1	3.34	0.94	3.12	6.84	96.2	0.66	4.22	250.6
CAMS-CSM1-0	2.30	1.76	4.05	7.79	52.9	0.64	3.21	114.1
FGOALS-f3-L	3.00	1.35	4.04	7.16	71.0	0.71	3.44	154.3
CanESM5	5.66	0.64	3.61	6.69	73.9	0.55	5.54	255.9
CNRM-CM6-1	4.90	0.73	3.59	6.20	111.1	0.58	4.69	348.7
CNRM-ESM2-1	4.75	0.64	3.03	6.39	114.3	0.56	5.28	388.7
ACCESS-CM2	4.70	0.71	3.34	6.98	86.9	0.60	5.23	271.5
ACCESS-ESM1-5	3.86	0.74	2.85	6.31	84.2	0.72	4.24	234.9
E3SM-1-0	5.31	0.63	3.34	6.83	39.7	0.43	6.29	160.6
EC-Earth3-Veg	4.29	0.80	3.42	6.12	38.4	0.52	4.51	124.5
EC-Earth3	4.18	0.81	3.40	6.58	39.8	0.50	4.90	132.4
INM-CM4-8	1.83	1.48	2.71	4.39	27.9	0.80	1.89	54.8
INM-CM5-0	1.91	1.55	2.97	6.85	45.2	0.57	3.19	109.6
IPSL-CM6A-LR	4.53	0.77	3.49	6.37	58.9	0.47	5.06	205.5
MIROC6	2.56	1.46	3.75	7.48	171.6	0.66	3.51	378.7
HadGEM3-GC31-LL	5.55	0.62	3.45	6.77	72.4	0.55	5.64	252.6
HadGEM3-GC31-MM	5.44	0.63	3.44	7.10	70.4	0.65	5.40	225.7
UKESM1-0-LL	5.37	0.67	3.61	6.27	75.0	0.54	5.08	257.4
MPI-ESM1-2-HR	2.95	1.25	3.70	6.46	81.1	0.72	3.23	178.7
MRI-ESM2-0	3.15	1.07	3.36	5.72	89.7	1.08	2.62	169.5
GISS-E2-1-G	2.71	1.44	3.91	5.44	142.3	0.88	2.33	261.5
GISS-E2-1-H	3.08	1.18	3.62	7.23	82.5	0.66	3.89	197.4
GISS-E2-2-G	2.40	1.58	3.79	7.78	908.6	0.51	3.71	2340.9
CESM2-WACCM	4.71	0.69	3.26	5.88	79.4	0.85	3.73	212.9
CESM2	5.22	0.64	3.34	5.93	72.2	0.84	3.91	204.3
NorESM2-LM	2.56	1.33	3.42	2.94	100.6	1.09	1.21	168.5
GFDL-CM4	3.91	0.80	3.15	4.87	80.3	0.72	3.15	214.4
GFDL-ESM4	2.66	1.35	3.58	7.17	125.8	0.62	3.63	298.7
NESM3	4.69	0.82	3.85	4.93	98.4	0.48	3.77	328.6
SAM0-UNICON	2.83	1.06	3.01	3.26	21.5	0.75	1.75	50.4

Table A.3: Fitted parameters two-layer model with ocean efficacy for 34 CMIP6 models. Values were determined following the algorithm described in Geoffroy et al. (2013a).

Model	ECS	λ	$F_{2\times}$	C	C_0	γ	τ_f	τ_s	ϵ
AWI-CM-1-1-MR	3.28	1.27	4.17	7.03	51.35	0.55	3.65	177.07	1.33
BCC-CSM2-MR	3.19	1.07	3.42	7.95	74.09	0.78	3.97	205.55	1.26
BCC-ESM1	3.63	0.93	3.39	7.46	96.15	0.66	4.27	323.42	1.29
CAMS-CSM1-0	2.38	1.86	4.43	8.67	52.86	0.64	3.32	146.40	1.33
FGOALS-f3-L	3.29	1.42	4.68	8.46	70.97	0.71	3.57	230.34	1.56
CanESM5	5.77	0.66	3.80	6.83	75.11	0.55	5.40	274.38	1.08
CNRM-CM6-1	5.08	0.73	3.69	6.39	111.14	0.58	4.70	385.93	1.10
CNRM-ESM2-1	4.40	0.65	2.87	6.04	114.27	0.56	5.26	315.70	0.82
ACCESS-CM2	5.56	0.68	3.79	7.95	86.93	0.60	5.29	410.60	1.48
ACCESS-ESM1-5	4.76	0.73	3.46	7.77	84.18	0.72	4.38	391.60	1.66
E3SM-1-0	5.83	0.63	3.68	7.57	39.65	0.43	6.44	222.54	1.41
EC-Earth3-Veg	4.57	0.85	3.89	7.04	38.38	0.52	4.67	162.52	1.41
EC-Earth3	4.41	0.85	3.76	7.34	39.84	0.50	5.03	167.79	1.34
INM-CM4-8	1.85	1.67	3.10	5.35	27.89	0.80	2.08	61.91	1.28
INM-CM5-0	1.98	1.63	3.22	7.58	45.23	0.57	3.31	140.28	1.31
IPSL-CM6A-LR	4.91	0.77	3.79	6.95	58.95	0.47	5.12	271.14	1.33
MIROC6	2.70	1.44	3.87	7.73	171.58	0.66	3.51	458.47	1.18
HadGEM3-GC31-LL	6.00	0.62	3.71	7.29	72.38	0.55	5.69	310.24	1.23
HadGEM3-GC31-MM	5.64	0.64	3.59	7.45	70.44	0.65	5.44	249.61	1.12
UKESM1-0-LL	5.65	0.67	3.80	6.61	75.96	0.54	5.11	297.04	1.16
MPI-ESM1-2-HR	3.21	1.33	4.26	7.55	81.08	0.72	3.32	247.80	1.47
MRI-ESM2-0	3.33	1.18	3.94	6.95	89.72	1.08	2.76	202.70	1.32
GISS-E2-1-G	2.76	1.46	4.02	5.64	142.31	0.88	2.35	278.99	1.07
GISS-E2-1-H	3.19	1.19	3.80	7.62	82.46	0.66	3.93	225.62	1.16
GISS-E2-2-G	2.08	1.79	3.72	7.65	908.63	0.51	3.71	969.30	0.50
CESM2-WACCM	5.61	0.73	4.08	7.54	79.42	0.85	3.92	321.55	1.60
CESM2	6.44	0.67	4.30	7.86	72.19	0.84	4.15	331.02	1.71
NorESM2-LM	2.95	1.69	4.98	4.47	100.55	1.09	1.29	261.79	1.93
GFDL-CM4	4.84	0.85	4.10	6.50	80.33	0.72	3.31	372.34	1.84
GFDL-ESM4	2.73	1.35	3.67	7.36	125.75	0.62	3.63	332.29	1.11
NESM3	4.69	0.82	3.85	4.93	98.38	0.48	3.77	329.87	1.00
SAM0-UNICON	2.84	1.12	3.19	3.31	21.53	0.75	1.63	49.86	1.09

Bibliography

Allen, M., Dube, O., W. Solecki, F. A.-D., Cramer, W., Humphreys, S., M. Kainuma, J. K., Mahowald, N., Mulugetta, Y., R. Perez, M. W., , and Zickfeld, K. (2018). Chapter 1: Framing and Context. In Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., J. Skea, P. S., A. Pirani, W. M.-O., Péan, C., Pidcock, R., Connors, S., Matthews, J., Y. Chen, X. Z., Gomis, M., Lonnoy, E., Maycock, T., Tignor, M., and Waterfield, T., editors, *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*, page 59.

Allen, M. R. and Ingram, W. J. (2002). Constraints on future changes in climate and the hydrologic cycle. *Nature*, 419(6903):224–232.

Andreae, M. O., Jones, C. D., and Cox, P. M. (2005). Strong present-day aerosol cooling implies a hot future. *Nature*, 435(7046):1187–1190.

Armour, K. C. (2017). Energy budget constraints on climate sensitivity in light of inconstant climate feedbacks. *Nature Climate Change*, 7(5):331–335.

Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., Schwinger, J., Bopp, L., Boucher, O., Cadule, P., et al. (2020). Carbon–concentration and carbon–climate feedbacks in CMIP6 models and their comparison to CMIP5 models. *Biogeosciences*, 17(16):4173–4222.

Bellouin, N., Quaas, J., Gryspeerdt, E., Kinne, S., Stier, P., Watson-Parris, D., Boucher, O., Carslaw, K. S., Christensen, M., Daniau, A.-L., Dufresne, J.-L., Feingold, G., Fiedler, S., Forster, P., Gettelman, A., Haywood, J. M., Lohmann, U., Malavelle, F., Mauritsen, T., McCoy, D. T., Myhre, G., Mülmenstädt, J., Neubauer, D., Possner, A., Rugenstein, M., Sato, Y., Schulz, M., Schwartz, S. E., Sourdeval, O., Storelvmo, T.,

Toll, V., Winker, D., and Stevens, B. (2020). Bounding global aerosol radiative forcing of climate change. *Reviews of Geophysics*, 58(1):e2019RG000660.

Bengtsson, L. and Schwartz, S. E. (2013). Determination of a lower bound on Earth's climate sensitivity. *Tellus, Series B: Chemical and Physical Meteorology*, 65(1):1–16.

Boe, J., Hall, A., and Qu, X. (2009). September sea-ice cover in the Arctic Ocean projected to vanish by 2100. *Nature Geoscience*, 2(5):341–343.

Boggs, P. T., Donaldson, J. R., Byrd, R. h., and Schnabel, R. B. (1989). Algorithm 676: ODRPACK: Software for Weighted Orthogonal Distance Regression. *ACM Transactions on Mathematical Software*, 15(4):348–364.

Bracegirdle, T. J. and Stephenson, D. B. (2013). On the robustness of emergent constraints used in multimodel climate change projections of arctic warming. *Journal of Climate*, 26(2):669–678.

Bretherton, C. and Caldwell, P. (2020). Combining Emergent Constraints for Climate Sensitivity. *Journal of Climate*, 33:7413–7430.

Brient, F. (2020). Reducing uncertainties in climate projections with emergent constraints: Concepts, examples and prospects. *Advances in Atmospheric Sciences*, 37:1–15.

Brient, F. and Schneider, T. (2016). Constraints on Climate Sensitivity from Space-Based Measurements of Low-Cloud Reflection. *Journal of Climate*, 29(16):5821–5835.

Brown, P. T. and Caldeira, K. (2017). Greater future global warming inferred from Earth's recent energy budget. *Nature*, 552(7683):45–50.

Brown, P. T., Stolpe, M. B., and Caldeira, K. (2018). Assumptions for emergent constraints. *Nature*, 563(7729):E1–E3.

Budyko, M. I. (1969). The effect of solar variations on the climate of the Earth. *Tellus*, 7(5):611–619.

Burke, E. J., Chadburn, S. E., Huntingford, C., and Jones, C. D. (2018). CO_2 loss by permafrost thawing implies additional emissions reductions to limit warming to 1.5 or 2 ℃. *Environmental Research Letters*, 13(2):024024.

Caldeira, K. and Myhrvold, N. P. (2013). Projections of the pace of warming following an abrupt increase in atmospheric carbon dioxide concentration. *Environmental Research Letters*, 8(3):34039–10.

Caldeira, K. and Wickett, M. E. (2003). Anthropogenic carbon and ocean pH. *Nature*, 425(6956):365–365.

Caldwell, P. M., Bretherton, C. S., Zelinka, M. D., Klein, S. A., Santer, B. D., and Sanderson, B. M. (2014). Statistical significance of climate sensitivity predictors obtained by data mining. *Geophysical Research Letters*, 41(5):1803–1808.

Caldwell, P. M., Zelinka, M. D., and Klein, S. A. (2018). Evaluating Emergent Constraints on Equilibrium Climate Sensitivity. *Journal of Climate*, 31:3921–3942.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, Articles*, 76(1):1–32.

Carroll, R. J. and Ruppert, D. (1996). The use and misuse of orthogonal regression in linear errors-in-variables models. *The American Statistician*, 50(1):1–6.

Chancel, L. and Piketty, T. (2015). Carbon and inequality: From Kyoto to Paris Trends in the global inequality of carbon emissions (1998-2013) & prospects for an equitable adaptation fund World Inequality Lab. PSE Working Papers halshs-02655266, HAL.

Chang, J. and Cooper, G. (1970). A practical difference scheme for Fokker-Planck equations. *Journal of Computational Physics*, 6(1):1 – 16.

Chen, X. L., Zhou, T. J., and Guo, Z. (2014). Climate sensitivities of two versions of FGOALS model to idealized radiative forcing. *Science China Earth Sciences*, 57(6):1363–1373.

Chikamoto, Y., Timmermann, A., Widlansky, M. J., Zhang, S., and Balmaseda, M. A. (2019). A Drift-Free Decadal Climate Prediction System for the Community Earth System Model. *Journal of Climate*, 32(18):5967–5995.

Clement, A. C., Burgman, R., and Norris, J. R. (2009). Observational and Model Evidence for Positive Low-Level Cloud Feedback. *Science*, 325(5939):460–464.

Colman, R. and Power, S. B. (2018). What can decadal variability tell us about climate feedbacks and sensitivity? *Climate Dynamics*, 51:3815–3828.

Cox, P. M., Huntingford, C., and Williamson, M. S. (2018a). Emergent constraint on equilibrium climate sensitivity from global temperature variability. *Nature*, 553(7688):319–322.

Cox, P. M., Pearson, D., Booth, B. B., Friedlingstein, P., Huntingford, C., Jones, C. D., and Luke, C. M. (2013). Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability. *Nature*, 494(7437):341–344.

Cox, P. M., Williamson, M. S., Nijsse, F. J. M. M., and Huntingford, C. (2018b). Cox et al. reply. *Nature*, 563(7729):E10–E15.

Crowley, T. J., Baum, S. K., Kim, K. Y., Hegerl, G. C., and Hyde, W. T. (2003). Modeling ocean heat content changes during the last millennium. *Geophysical Research Letters*, 30(18).

Crowley, T. J., Zielinski, G., Vinther, B., Udisti, R., Kreutz, K., Cole-Dai, J., and Castellano, E. (2008). Volcanism and the Little Ice Age. *PAGES news*, 16(2):22–23.

DeAngelis, A. M., Qu, X., Zelinka, M. D., and Hall, A. (2015). An observational radiative constraint on hydrologic cycle intensification. *Nature*, 528(7581):249–253.

Deser, C., Lehner, F., Rodgers, K. B., Ault, T., Delworth, T. L., DiNezio, P. N., Fiore, A., Frankignoul, C., Fyfe, J. C., Horton, D. E., Kay, J. E., Knutti, R., Lovenduski, N. S., Marotzke, J., McKinnon, K. A., Minobe, S., Randerson, J., Screen, J. A., Simpson, I. R., and Ting, M. (2020). Insights from Earth system model initial-condition large ensembles and future prospects. *Nature Climate Change*, 10:277–286.

Dlugokenchy, E. and Tans, P. (2019). Trends in atmospheric carbon dioxide. https://www.esrl.noaa.gov/gmd/ccgg/trends/gl_gr.html.

Donat, M. G., Pitman, A. J., and Angélil, O. (2018). Understanding and Reducing Future Uncertainty in Midlatitude Daily Heat Extremes Via Land Surface Feedback Constraints. *Geophysical Research Letters*, 45(19):10,610–627,636.

Dong, B., Gregory, J. M., and Sutton, R. T. (2009). Undestanding land-sea warming contrast in response to increasing greenhouse gases. Part I: Transient adjustment. *Journal of Climate*, 22(11):3079–3097.

Emile-Geay, J., McKay, N. P., Kaufman, D. S., von Gunten, L., Wang, J., Anchukaitis, K. J., Abram, N. J., Addison, J. A., Curran, M. A., Evans, M. N., Henley, B. J., Hao, Z., Martrat, B., McGregor, H. V., Neukom, R., Pederson, G. T., Stenni, B., Thirumalai,

K., Werner, J. P., Xu, C., Divine, D. V., Dixon, B. C., Gergis, J., Mundo, I. A., Nakatsuka, T., Phipps, S. J., Routson, C. C., Steig, E. J., Tierney, J. E., Tyler, J. J., Allen, K. J., Bertler, N. A., Björklund, J., Chase, B. M., Chen, M.-T., Cook, E., de Jong, R., DeLong, K. L., Dixon, D. A., Ekaykin, A. A., Ersek, V., Filipsson, H. L., Francus, P., Freund, M. B., Frezzotti, M., Gaire, N. P., Gajewski, K., Ge, Q., Goosse, H., Gornostaeva, A., Grosjean, M., Horiuchi, K., Hormes, A., Husum, K., Isaksson, E., Kandasamy, S., Kawamura, K., Kilbourne, K. H., Koç, N., Leduc, G., Linderholm, H. W., Lorrey, A. M., Mikhalenko, V., Mortyn, P. G., Motoyama, H., Moy, A. D., Mulvaney, R., Munz, P. M., Nash, D. J., Oerter, H., Opel, T., Orsi, A. J., Ovchinnikov, D. V., Porter, T. J., Roop, H. A., Saenger, C., Sano, M., Sauchyn, D., Saunders, K. M., Seidenkrantz, M.-S., Severi, M., Shao, X., Sicre, M.-A., Sigl, M., Sinclair, K., St. George, S., St. Jacques, J.-M., Thamban, M., Kuwar Thapa, U., Thomas, E. R., Turney, C., Uemura, R., Viau, A. E., Vladimirova, D. O., Wahl, E. R., White, J. W., Yu, Z., Zinke, J., and Consortium, P. (2017). A global multiproxy database for temperature reconstructions of the common era. *Scientific Data*, 4(1):170088.

Fahey, D. W., Doherty, S. J., Hibbard, K. A., Romanou, A., and Taylor, P. C. (2017). Physical drivers of climate change. In Wuebbles, D. J., Fahey, D. W., Hibbard, K. A., Dokken, D. J., Stewart, B. C., and Maycock, T. K., editors, *Climate Science Special Report: Fourth National Climate Assessment, Volume I*, pages 73–113. U.S. Global Change Research Program.

Fasullo, J. T. and Trenberth, K. E. (2012). A Less Cloudy Future: The Role of Subtropical Subsidence in Climate Sensitivity. *Science*, 338(6108):792–794.

Flato, G., Marotzke, J., B. Abiodun, P. B., Chou, S., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., and Rummukainen, M. (2013). Evaluation of climate models. In Stocker, T., D. Qin, G.-K. P., Tignor, M., Allen, S., J. Boschung, A. N., Xia, Y., Bex, V., and Midgley, P., editors, *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, United Kingdom and New York, NY, USA.

Forster, P. M., Andrews, T., Good, P., Gregory, J. M., Jackson, L. S., and Zelinka, M. (2013). Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models. *Journal of Geophysical Research Atmospheres*, 118(3):1139–1150.

Forster, P. M., Maycock, A. C., McKenna, C. M., and Smith, C. J. (2020). Latest climate models confirm need for urgent mitigation. *Nature Climate Change*, 10:7–10.

Foster, G., Annan, J. D., Schmidt, G. A., and Mann, M. E. (2008). Comment on "Heat capacity, time constant, and sensitivity of Earth's climate system" by S. E. Schwartz. *Journal of Geophysical Research Atmospheres*, 113(15):D15102.

Fraedrich, K. (1979). Catastrophes and resilience of a zero-dimensional climate system with ice-albedo and greenhouse feedback. *Quarterly Journal of the Royal Meteorological Society*, 105(443):147–167.

Fraedrich, K., Jansen, H., Kusch, U., and Lunkeit, F. (2005). The Planet Simulator: Towards a user friendly model. *Meteorologische Zeitschrift*, 14(3):299–304.

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., Bakker, D. C. E., Canadell, J. G., Ciais, P., Jackson, R. B., Anthoni, P., Barbero, L., Bastos, A., Bastrikov, V., Becker, M., Bopp, L., Buitenhuis, E., Chandra, N., Chevallier, F., Chini, L. P., Currie, K. I., Feely, R. A., Gehlen, M., Gilfillan, D., Gkritzalis, T., Goll, D. S., Gruber, N., Gutekunst, S., Harris, I., Haverd, V., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Joetzjer, E., Kaplan, J. O., Kato, E., Klein Goldewijk, K., Korsbakken, J. I., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lienert, S., Lombardozzi, D., Marland, G., McGuire, P. C., Melton, J. R., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Neill, C., Omar, A. M., Ono, T., Peregon, A., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Séférian, R., Schwinger, J., Smith, N., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., van der Werf, G. R., Wiltshire, A. J., and Zaehle, S. (2019). Global carbon budget 2019. *Earth System Science Data*, 11(4):1783–1838.

Frost, C. and Thompson, S. G. (2000). Correcting for regression dilution bias: Comparison of methods for a single predictor variable. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 163(2):173–189.

Fuglestvedt, J., Rogelj, J., Millar, R. J., Allen M., B. O., Cain, M., Forster, P. M., Kriegler, E., and D., S. (2018). Implications of possible interpretations of 'greenhouse gas balance' in the Paris Agreement. *Philosophical Transactions of the Royal Society A*, 376.

Gao, C., Robock, A., and Ammann, C. (2008). Volcanic forcing of climate over the past 1500 years: An improved ice core-based index for climate models. *Journal of Geophysical Research Atmospheres*, 113(23):D23111.

Geoffroy, O., Saint-Martin, D., Bellon, G., Voldoire, A., Olivié, D. J., and Tytéca, S. (2013a). Transient climate response in a two-layer energy-balance model. Part II: Representation of the efficacy of deep-ocean heat uptake and validation for CMIP5 AOGCMs. *Journal of Climate*, 26(6):1859–1876.

Geoffroy, O., Saint-Martin, D., Olivié, D. J., Voldoire, A., Bellon, G., and Tytéca, S. (2013b). Transient climate response in a two-layer energy-balance model. Part I: Analytical solution and parameter calibration using CMIP5 AOGCM experiments. *Journal of Climate*, 26(6):1841–1857.

Ghil, M. and Lucarini, V. (2020). The physics of climate variability and climate change. *Reviews of Modern Physics*, 92:035002.

Goodwin, P., Katavouta, A., Roussenov, V. M., Foster, G. L., Rohling, E. J., and Williams, R. G. (2018). Pathways to 1.5 °C and 2 °C warming based on observational and geological constraints. *Nature Geoscience*, 11(2):102–107.

Gordon, N. D. and Klein, S. A. (2014). Low-cloud optical depth feedback in climate models. *Journal of Geophysical Research: Atmospheres*, 119:6052–6065.

Gottwald, G. A., Wormell, J. P., and Wouters, J. (2016). On spurious detection of linear response and misuse of the fluctuation-dissipation theorem in finite time series. *Physica D: Nonlinear Phenomena*, 331:89–101.

Gregory, J. M. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters*, 31(3):L03205.

Gregory, J. M., Andrews, T., Ceppi, P., Mauritsen, T., and Webb, M. J. (2019). How accurately can the climate sensitivity to CO_2 be estimated from historical climate change? *Climate Dynamics*, 54(1-2):129–157.

Gregory, J. M. and Forster, P. M. (2008). Transient climate response estimated from radiative forcing and observed temperature change. *Journal of Geophysical Research Atmospheres*, 113(23):1–15.

Grise, K. M., Polvani, L. M., and Fasullo, J. T. (2015). Reexamining the Relationship between Climate Sensitivity and the Southern Hemisphere Radiation Budget in CMIP Models. *Journal of Climate*, 28(23):9298–9312.

Gritsun, A. and Lucarini, V. (2017). Fluctuations, response, and resonances in a simple atmospheric model. *Physica D*, 349:62–76.

Hahn, G. J. (1977). Fitting Regression Models With No Intercept Term. *Journal of Quality Technology*, 9(2):56–61.

Hairer, M. and Majda, A. J. (2010). A simple framework to justify linear response theory. *Nonlinearity*, 23(4):909–922.

Hall, A., Cox, P., Huntingford, C., and Klein, S. (2019). Progressing emergent constraints on future climate change. *Nature Climate Change*, 9(4):269–278.

Hall, A. and Qu, X. (2006). Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophysical Research Letters*, 33(3). L03502.

Hannachi, A., Jolliffe, I. T., and Stephenson, D. B. (2007). Empirical orthogonal functions and related techniques in atmospheric science: A review. *International Journal of Climatology*, 27(9):1119–1152.

Hansen, J., Russell, G., Lacis, A., Fung, I., Rind, D., and Stone, P. (1985). Climate response times: Dependence on climate sensitivity and ocean mixing. *Science*, 229:857–859.

Hargreaves, J. C., Annan, J. D., Yoshimori, M., and Abe-Ouchi, A. (2012). Can the Last Glacial Maximum constrain climate sensitivity? *Geophysical Research Letters*, 39(24).

Hasselmann, K. (1976). Stochastic Models of Climate Extremes: Theory and Observations. *Tellus*, 28(6):473–485.

Hegerl, G. C., Zwiers, F. W., Braconnot, P., Gillett, N. P., Luo, Y., Marengo Orsini, J., Nicholls, N., Penner, J. E., and Stott, P. A. (2007). Understanding and Attributing Climate Change. 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*. Cambridge, UK, Cambridge University Press.

Held, I. M., Winton, M., Takahashi, K., Delworth, T., Zeng, F., and Vallis, G. K. (2010). Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *Journal of Climate*, 23(9):2418–2427.

Hunter, S. J., Haywood, A. M., Dolan, A. M., and Tindall, J. C. (2019). The HadCM3 contribution to PlioMIP phase 2. *Climate of the Past*, 15(5):1691–1713.

Huntingford, C., Williamson, M. S., and Nijsse, F. J. M. M. (2020). CMIP6 climate models imply high committed warming. *Climatic Change*.

IPCC (2013a). *Annex III: Glossary*, book section AIII, page 1447–1466. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

IPCC (2013b). *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, United Kingdom and New York, NY, USA.

IPCC (2013c). Summary for Policymakers. In Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., editors, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, page 21. Cambridge University Press.

IPCC (2018). Summary for Policymakers. In Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R., Connors, S., Matthews, J., Chen, Y., Zhou, X., Gomis, M., Lonnoy, E., Maycock, T., Tignor, M., and Waterfield, T., editors, *Global Warming of 1.5℃. An IPCC Special Report on the impacts of global warming of 1.5℃ above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*, page 11.

Jebeile, J. and Crucifix, M. (2020). Multi-model ensembles in climate science: Mathematical structures and expert judgements. *Studies in History and Philosophy of Science Part A*, 83:44 – 52.

Jiménez-de-la Cuesta, D. and Mauritsen, T. (2019). Emergent constraints on Earth's transient and equilibrium response to doubled CO_2 from post-1970s global warming. *Nature Geoscience*, 12(11):902–905.

Kalman, R. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45.

Karpechko, A. Y., Maraun, D., and Eyring, V. (2013). Improving Antarctic Total Ozone Projections by a Process-Oriented Multiple Diagnostic Ensemble Regression. *Journal of the Atmospheric Sciences*, 70(12):3959–3976.

Kiehl, J. T. (2007). Twentieth century climate model response and climate sensitivity. *Geophysical Research Letters*, 34(22).

Klein, S. A. and Hall, A. (2015). Emergent Constraints for Cloud Feedbacks. *Current Climate Change Reports*, 1(4):276–287.

Kloeden, P. and Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg.

Knutti, R. (2010). The end of model democracy? *Climatic Change*, 102(3):395–404.

Knutti, R., Meehl, G. A., Allen, M. R., and Stainforth, D. A. (2006). Constraining Climate Sensitivity from the Seasonal Cycle in Surface Temperature. *Journal of Climate*, 19(17):4224–4233.

Knutti, R., Rugenstein, M. A., and Hegerl, G. C. (2017). Beyond equilibrium climate sensitivity. *Nature Geoscience*, 10(10):727–736.

Kubo, R. (1966). The fluctuation-dissipation theorem. *Reports on Progress in Physics*, 29(1):255–284.

Kwiatkowski, L., Bopp, L., Aumont, O., Ciais, P., Cox, P. M., Laufkötter, C., Li, Y., and Séférian, R. (2017). Emergent constraints on projections of declining primary production in the tropical oceans. *Nature Climate Change*, 7(5):355–358.

Lehoucq, R. B., Sorensen, D. C., and Yang, C. (1998). *ARPACK USERS GUIDE: Solution of Large Scale Eigenvalue Problems by Implicitly Restarted Arnoldi Methods*. SIAM.

Lei, L. and Hacker, J. P. (2015). Nudging, Ensemble, and Nudging-Ensembles for Data Assimilation in the Presence of Model Error. *Monthly Weather Review*, 143(7):2600–2610.

Leith, C. E. (1975). Climate Response and Fluctuation Dissipation. *Journal of the Atmospheric Sciences*, 32(10):2022–2026.

Lembo, V., Lucarini, V., and Ragone, F. (2020). Beyond forcing scenarios: Predicting climate change through response operators in a coupled general circulation model. *Scientific Reports*, 10(1):8668.

Lenssen, N., Schmidt, G., Hansen, J., Menne, M., Persin, A., Ruedy, R., and Zyss, D. (2019). Improvements in the GISTEMP Uncertainty Model. *Journal of Geophysical Research: Atmospheres*, 124(12):6307–6326.

Li, W., Duan, Q., Miao, C., Ye, A., Gong, W., and Di, Z. (2017). A review on statistical postprocessing methods for hydrometeorological ensemble forecasting. *Wiley Interdisciplinary Reviews: Water*, 4:1–24.

Liu, Y., Goodrick, S., and Heilman, W. (2014). Wildland fire emissions, carbon, and climate: Wildfire–climate interactions. *Forest Ecology and Management*, 317:80 – 96. Wildland fire emissions, carbon, and climate: Science overview and knowledge needs.

Lucarini, V. and Sarno, S. (2011). A statistical mechanical approach for the computation of the climatic response to general forcings. *Nonlinear Processes in Geophysics*, 18(1):7–28.

Lunkeit, F., Borth, H., Böttinger, M., and Fraedlich, F. (2011). *PLASIM Reference Guide*.

Lutsko, N. J. and Takahashi, K. (2018). What Can the Internal Variability of CMIP5 Models Tell Us About Their Climate Sensitivity? *Journal of Climate*, 31(13):JCLI–D–17–0736.1.

Marsh, D. R., Lamarque, J. F., Conley, A. J., and Polvani, L. M. (2016). Stratospheric ozone chemistry feedbacks are not critical for the determination of climate sensitivity in CESM1(WACCM). *Geophysical Research Letters*, 43(8):3928–3934.

Massey Jr., F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, 46(253):68–78.

Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R., Connors, S., Matthews, J., Chen, Y., Zhou, X., Gomis, M., Lonnoy, E., Maycock, T., Tignor, M., and Waterfield, T., editors (2018). *Global Warming of 1.5℃. An IPCC Special Report on the impacts of global warming of 1.5℃ above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*. IPCC.

Medhaug, I., Stolpe, M. B., Fischer, E. M., and Knutti, R. (2017). Reconciling controversies about the 'global warming hiatus'. *Nature*, 545(7652):41–47.

Meehl, G. A., Washington, W. M., Arblaster, J. M., Hu, A., Teng, H., Kay, J. E., Gettelman, A., Lawrence, D. M., Sanderson, B. M., and Strand, W. G. (2013). Climate

change projections in CESM1(CAM5) compared to CCSM4. *Journal of Climate*, 26(17):6287–6308.

Mengis, N., Partanen, A.-I., Jalbert, J., and Matthews, H. D. (2018). 1.5 ℃ carbon budget dependent on carbon cycle uncertainty and future non-CO_2 forcing. *Scientific Reports*, 8(1):5831.

Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *Journal of Geophysical Research: Atmospheres*, 117(D8).

Myhre, G., Myhre, C. L., Forster, P. M., and Shine, K. P. (2017). Halfway to doubling of CO_2 radiative forcing. *Nature Geoscience*, 10(10):710–711.

Myhre, G., Shindell, D., Bréon, F.-M., W. Collins, J. F., Huang, J., Koch, D., J.-F. Lamarque, D. L., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T., and Zhang, H. (2013). Anthropogenic and natural radiative forcing. In Stocker, T., D. Qin, G.-K. P., Tignor, M., Allen, S., J. Boschung, A. N., Xia, Y., Bex, V., and Midgley, P., editors, *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, United Kingdom and New York, NY, USA.

Myung, J. I., Tang, Y., and Pitt, M. A. (2009). Evaluation and comparison of computational models. *Methods in enzymology*, 454:287–304.

Neukom, R., Barboza, L. A., Erb, M. P., Shi, F., Emile-Geay, J., Evans, M. N., Franke, J., Kaufman, D. S., Lücke, L., Rehfeld, K., Schurer, A., Zhu, F., Brönnimann, S., Hakim, G. J., Henley, B. J., Ljungqvist, F. C., McKay, N., Valler, V., von Gunten, L., and PAGES 2k Consortium (2019a). Consistent multidecadal variability in global temperature reconstructions and simulations over the common era. *Nature Geoscience*, 12(8):643–649.

Neukom, R., Barboza, L. A., Erb, M. P., Shi, F., Emile-Geay, J., Evans, M. N., Franke, J., Kaufman, D. S., Lücke, L., Rehfeld, K., and et al. (2019b). Global mean temperature reconstructions over the Common Era. https://figshare.com/collections/Global_mean_temperature_reconstructions_over_the_Common_Era/4507043/2 at figshare.

Neukom, R., Steiger, N., Gómez-Navarro, J. J., Wang, J., and Werner, J. P. (2019c). No evidence for globally coherent warm and cold periods over the preindustrial Common Era. *Nature*, 571(7766):550–554.

Newsom, E. R., Bitz, C. M., Bryan, F. O., Abernathey, R., and Gent, P. R. (2016). Southern Ocean Deep Circulation and Heat Uptake in a High-Resolution Climate Model. *Journal of Climate*, 29(7):2597–2619.

Nijsse, F. J., Cox, P. M., and Williamson, M. S. (2020a). Observed warming and climate sensitivity [source code]. https://codeocean.com/capsule/5713020/tree/v1.

Nijsse, F. J. M. M., Cox, P. M., Huntingford, C., and Williamson, M. S. (2019a). Decadal global temperature variability increases strongly with climate sensitivity. *Nature Climate Change*, 9(8):598–601.

Nijsse, F. J. M. M., Cox, P. M., Huntingford, C., and Williamson, M. S. (2019b). Decadal variability and climate sensitivity. https://doi.org/10.24433/CO.6969545.v1 at Code Ocean.

Nijsse, F. J. M. M., Cox, P. M., and Williamson, M. S. (2020b). Emergent constraints on transient climate response (TCR) and equilibrium climate sensitivity (ECS) from historical warming in CMIP5 and CMIP6 models. *Earth System Dynamics*, 11(3):737–750.

Nijsse, F. J. M. M. and Dijkstra, H. A. (2018). A mathematical approach to understanding emergent constraints. *Earth System Dynamics*, 9(3):999–1012.

O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J. F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., and Sanderson, B. M. (2016). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, 9(9):3461–3482.

Otto, A., Otto, F. E. L., Boucher, O., Church, J., Hegerl, G., Forster, P. M., Gillett, N. P., Gregory, J., Johnson, G. C., Knutti, R., Lewis, N., Lohmann, U., Marotzke, J., Myhre, G., Shindell, D., Stevens, B., and Allen, M. R. (2013). Energy budget constraints on climate response. *Nature Geoscience*, 6(6):415–416.

Parsons, L. A., Brennan, M. K., and Proistosescu, C. (2020). Magnitudes and spatial patterns of interdecadal temperature variability in CMIP6. *Geophysical Research Letters*, 47:1–11.

Pavliotis, G. A. (2014). *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*. Springer New York, New York, NY.

Previdi, M., Liepert, B. G., Peteet, D., Hansen, J., Beerling, D. J., Broccoli, A. J., Frolking, S., Galloway, J. N., Heimann, M., Le Quéré, C., Levitus, S., and Ramaswamy, V. (2013). Climate sensitivity in the anthropocene. *Quarterly Journal of the Royal Meteorological Society*, 139(674):1121–1131.

Qu, X. and Hall, A. (2007). What Controls the Strength of Snow-Albedo Feedback? *Journal of Climate*, 20(15):3971–3981.

Qu, X. and Hall, A. (2014). On the persistent spread in snow-albedo feedback. *Climate Dynamics*, 42(1):69–81.

Qu, X., Hall, A., Klein, S. A., and Caldwell, P. M. (2014). On the spread of changes in marine low cloud cover in climate model simulations of the 21st century. *Climate Dynamics*, 42(9):2603–2626.

Ragone, F., Lucarini, V., and Lunkeit, F. (2016). A new framework for climate sensitivity and prediction: a modelling perspective. *Climate Dynamics*, 46(5):1459–1471.

Raper, S. C. and Braithwaite, R. J. (2009). Glacier volume response time and its links to climate and topography based on a conceptual model of glacier hypsometry. *The Cryosphere*, 3(2):183.

Reichle, R. H. (2008). Data assimilation methods in the Earth sciences. *Advances in Water Resources*, 31(11):1411 – 1418. Hydrologic Remote Sensing.

Renoult, M., Annan, J. D., Hargreaves, J. C., Sagoo, N., Flynn, C., Kapsch, M.-L., Li, Q., Lohmann, G., Mikolajewicz, U., Ohgaito, R., Shi, X., Zhang, Q., and Mauritsen, T. (2020). A Bayesian framework for emergent constraints: case studies of climate sensitivity with PMIP. *Climate of the Past*, 16(5):1715–1735.

Riahi, K., Van Vuuren, D. P., Kriegler, E., Edmonds, J., O'neill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., et al. (2017). The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: an overview. *Global Environmental Change*, 42:153–168.

Risken, H. and Frank, T. (1996). *The Fokker-Planck Equation: Methods of Solution and Applications*, volume 5 of *Springer Series in Synergetics*. SpringerLink.

Roberts, C. D., Palmer, M. D., McNeall, D., and Collins, M. (2015). Quantifying the likelihood of a continued hiatus in global warming. *Nature Climate Change*, 5(4):337–342.

Roe, G. (2009). Feedbacks, Timescales, and Seeing Red. *Annual Review of Earth and Planetary Sciences*, 37(1):93–115.

Roemmich, D., Alford, M. H., Claustre, H., Johnson, K., King, B., Moum, J., Oke, P., Owens, W. B., Pouliquen, S., Purkey, S., Scanderbeg, M., Suga, T., Wijffels, S., Zilberman, N., Bakker, D., Baringer, et al. (2019). On the Future of Argo: A Global, Full-Depth, Multi-Disciplinary Array. *Frontiers in Marine Science*, 6:1–28.

Rogelj, J., Shindell, D., Jiang, K., Fifita, S., Forster, P., Ginzburg, V., Handa, C., Kheshgi, H., Kobayashi, S., Kriegler, E., Mundaca, L., Séférian, R., , and Vilariño, M. (2018). Mitigation Pathways Compatible with 1.5℃ in the Context of Sustainable Development. In Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R., Connors, S., Matthews, J., Chen, Y., Zhou, X., Gomis, M., Lonnoy, E., Maycock, T., Tignor, M., , and Waterfield, T., editors, *Global Warming of 1.5℃. An IPCC Special Report on the impacts of global warming of 1.5℃ above pre-industrial levels and related global greenhouse gas emission pathways and in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*.

Rohde, R., Muller, R., Jacobsen, R., Perlmutter, S., and Mosher, S. (2013). Berkeley Earth Temperature Averaging Process. *Geoinformatics & Geostatistics: An Overview*, 1(2).

Ruelle, D. (1998). General linear response formula in statistical mechanics, and the fluctuation-dissipation theorem far from equilibrium. *Physics Letters A*, 245(3-4):220–224.

Ruelle, D. (2009). A review of linear response theory for general differentiable dynamical systems. *Nonlinearity*, 22(4):855–870.

Rugenstein, M., Bloch-Johnson, J., Gregory, J., Mauritsen, T., Li, C., Frölicher, T., Danabasoglu, G., Yang, S., Dufresne, J.-L., Schmidt, G. A., Abe-Ouchi, A., and Geoffroy, O. (2020). Equilibrium climate sensitivity estimated by equilibrating climate models. *Geophysical Research Letters*, 47.

Rypdal, M., Fredriksen, H. B., Rypdal, K., and Steene, R. J. (2018). Emergent constraints on climate sensitivity. *Nature*, 563(7729):E4–E5.

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

Sanderson, B. M., Knutti, R., and Caldwell, P. (2015). Addressing interdependency in a multimodel ensemble by interpolation of model properties. *Journal of Climate*, 28(13):5150–5170.

Sansom, P. G. (2014). *Statistical methods for quantifying uncertainty in climate projections from ensembles of climate models*. PhD thesis, University of Exeter.

Saunois, M., Stavert, A. R., Poulter, B., Bousquet, P., Canadell, J. G., Jackson, R. B., Raymond, P. A., Dlugokencky, E. J., Houweling, S., Patra, P. K., Ciais, P., Arora, V. K., Bastviken, D., Bergamaschi, P., Blake, D. R., Brailsford, G., Bruhwiler, L., Carlson, K. M., Carrol, M., Castaldi, S., Chandra, N., Crevoisier, C., Crill, P. M., Covey, K., Curry, C. L., Etiope, G., Frankenberg, C., Gedney, N., Hegglin, M. I., Höglund-Isaksson, L., Hugelius, G., Ishizawa, M., Ito, A., Janssens-Maenhout, G., Jensen, K. M., Joos, F., Kleinen, T., Krummel, P. B., Langenfelds, R. L., Laruelle, G. G., Liu, L., Machida, T., Maksyutov, S., McDonald, K. C., McNorton, J., Miller, P. A., Melton, J. R., Morino, I., Müller, J., Murguia-Flores, F., Naik, V., Niwa, Y., Noce, S., O'Doherty, S., Parker, R. J., Peng, C., Peng, S., Peters, G. P., Prigent, C., Prinn, R., Ramonet, M., Regnier, P., Riley, W. J., Rosentreter, J. A., Segers, A., Simpson, I. J., Shi, H., Smith, S. J., Steele, L. P., Thornton, B. F., Tian, H., Tohjima, Y., Tubiello, F. N., Tsuruta, A., Viovy, N., Voulgarakis, A., Weber, T. S., van Weele, M., van der Werf, G. R., Weiss, R. F., Worthy, D., Wunch, D., Yin, Y., Yoshida, Y., Zhang, W., Zhang, Z., Zhao, Y., Zheng, B., Zhu, Q., Zhu, Q., and Zhuang, Q. (2020). The global methane budget 2000–2017. *Earth System Science Data*, 12(3):1561–1623.

Schennach, S. M. (2016). Recent Advances in the Measurement Error Literature. *Annual Review of Economics*, 8(1):341–377.

Schmidt, G. A., Annan, J. D., Bartlein, P. J., Cook, B. I., Guilyardi, E., Hargreaves, J. C., Harrison, S. P., Kageyama, M., Legrande, A. N., Konecky, B., Lovejoy, S., Mann, M. E., Masson-Delmotte, V., Risi, C., Thompson, D., Timmermann, A., and Yiou, P. (2014). Using palaeo-climate comparisons to constrain future projections in CMIP5. *Climate of the Past*, 10(1):221–250.

Schwalm, C. R., Glendon, S., and Duffy, P. B. (2020). RCP8.5 tracks cumulative CO_2 emissions. *Proceedings of the National Academy of Sciences*, 117(33):19656–19657.

Schwartz, S. E. (2007). Heat capacity, time constant, and sensitivity of Earth's climate system. *Journal of Geophysical Research Atmospheres*, 112(24):1–12.

Sellers, W. D. (1969). A Global Climatic Model Based on the Energy Balance of the Earth-Atmosphere System. *Journal of Applied Meteorology*, 8(3):392–400.

Sherwood, S., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., Hegerl, G., Klein, S. A., Marvel, K. D., Rohling, E. J., Watanabe, M., Andrews, T., Braconnot, P., Bretherton, C. S., Foster, G. L., Hausfather, Z., Heydt, A. S. v. d., Knutti, R., Mauritsen, T., Norris, J. R., Proistosescu, C., Rugenstein, M., Schmidt, G. A., Tokarska, K. B., and Zelinka, M. D. (2020). An assessment of earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, 58(4):e2019RG000678.

Sherwood, S. C., Bony, S., and Dufresne, J. L. (2014). Spread in model climate sensitivity traced to atmospheric convective mixing. *Nature*, 505(7481):37–42.

Smith, D. M., Booth, B. B., Dunstone, N. J., Eade, R., Hermanson, L., Jones, G. S., Scaife, A. A., Sheen, K. L., and Thompson, V. (2016). Role of volcanic and anthropogenic aerosols in the recent global surface warming slowdown. *Nature Climate Change*, 6(10):936–940.

Smith, R. J. (2009). Use and misuse of the reduced major axis for line-fitting. *American Journal of Physical Anthropology*, 140(3):476–486.

Stan Development Team (2018). Stan Modeling Language Users Guide and Reference Manual, Version 2.18.0.

Steele, K. and Werndl, C. (2013). Climate Models, Calibration, and Confirmation. *The British Journal for the Philosophy of Science*, 64(3):609–635.

Stevens, B., Fiedler, S., Kinne, S., Peters, K., Rast, S., Müsse, J., Smith, S. J., and Mauritsen, T. (2017). MACv2-SP: A parameterization of anthropogenic aerosol optical properties and an associated Twomey effect for use in CMIP6. *Geoscientific Model Development*, 10(1):433–452.

Stocker, M., Ladstädter, F., Wilhelmsen, H., and Steiner, A. K. (2019). Quantifying Stratospheric Temperature Signals and Climate Imprints From Post-2000 Volcanic Eruptions. *Geophysical Research Letters*, 46:12486– 12494.

Stocker, T. F., Dahe, Q., Plattner, G.-K., Alexander, L. V., Allen, S. K., Bindoff, N. L., Bréon, F.-M., Church, J. A., Cubash, U., Emori, S., Forster, P., Friedlingstein, P., Talley, L. D., Vaughan, D. G., and Xie, S.-P. (2013). Technical Summary. 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*, pages 33–115. Cambridge University Press.

Stott, P. A. and Kettleborough, J. A. (2002). Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. *Nature*, 416(6882):723–726.

Strogatz, S. H. (2000). *Nonlinear dynamics and chaos : with applications to physics, biology, chemistry, and engineering*. Westview Press.

Su, H., Jiang, J. H., Zhai, C., Shen, T. J., Neelin, J. D., Stephens, G. L., and Yung, Y. L. (2014). Weakening and strengthening structures in the Hadley Circulation change under global warming and implications for cloud response and climate sensitivity. *Journal of Geophysical Research: Atmospheres*, 119(10):5787–5805.

Sutera, A. (1981). On stochastic perturbation and long-term climate behaviour. *Quarterly Journal of the Royal Meteorological Society*, 107(451):137–151.

Tanaka, K. and O'Neill, B. C. (2018). The Paris Agreement zero-emissions goal is not always consistent with the 1.5 ℃ and 2 ℃ temperature targets. *Nature Climate Change*, 8(4):319–324.

Tantet, A. (2016). *Ergodic theory of climate: variability, stability and response*. PhD thesis, Utrecht University.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A. (2012). An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society*, 93(4):485–498.

Terhaar, J., Kwiatkowski, L., and Bopp, L. (2020). Emergent constraint on Arctic Ocean acidification in the twenty-first century. *Nature*, 582(7812):379–383.

Thackeray, C. W., Qu, X., and Hall, A. (2018). Why do models produce spread in snow albedo feedback? *Geophysical Research Letters*, 45(12):6223–6231.

Tian, B. (2015). Spread of model climate sensitivity linked to double-Intertropical Convergence Zone bias. *Geophysical Research Letters*, 42(10):4133–4141.

Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F., and Knutti, R. (2020). Past warming trend constrains future warming in CMIP6 models. *Science Advances*, 6(12).

Trenberth, K. E. and Fasullo, J. T. (2010). Simulation of Present-Day and Twenty-First-Century Energy Budgets of the Southern Oceans. *Journal of Climate*, 23(2):440–454.

United Nations General Assembly (1992). United Nations Framework Convention on Climate Change. New York.

Van Schaeybroeck, B. and Vannitsem, S. (2012). Toward post-processing ensemble forecasts based on hindcasts. Koninklijk Meteorologisch Instituut van België.

Walsh, B., Ciais, P., Janssens, I. A., Peñuelas, J., Riahi, K., Rydzak, F., van Vuuren, D. P., and Obersteiner, M. (2017). Pathways for balancing CO_2 emissions and sinks. *Nature Communications*, 8(1):14856.

Wenzel, S., Cox, P. M., Eyring, V., and Friedlingstein, P. (2014). Emergent constraints on climate-carbon cycle feedbacks in the CMIP5 Earth system models. *Journal of Geophysical Research: Biogeosciences*, 119(5):794–807.

Wigley, T. M. and Raper, S. C. (1990). Natural variability of the climate system and detection of the greenhouse effect. *Nature*, 344(6264):324–327.

Williamson, D. B., Goldstein, M., Allison, L., Blaker, A., Challenor, P., Jackson, L., and Yamazaki, K. (2013). History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble. *Climate Dynamics*, 41(7-8):1703–1729.

Williamson, D. B. and Sansom, P. G. (2020). How Are Emergent Constraints Quantifying Uncertainty and What Do They Leave Behind? *Bulletin of the American Meteorological Society*, 100(12):2571–2588.

Williamson, M. S., Cox, P. M., and Nijsse, F. J. M. M. (2018). Theoretical foundations of emergent constraints: relationships between climate sensitivity and global temperature variability in conceptual models. *Dynamics and Statistics of the Climate System*, 3(1).

Winkler, A. J., Myneni, R. B., Alexandrov, G. A., and Brovkin, V. (2019). Earth system models underestimate carbon fixation by plants in the high latitudes. *Nature Communications*, 10(1):885.

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

World Meteorological Organization (2020). *WMO Statement on the State of the Global Climate in 2019*. WMO.

Yang, J., Gong, P., Fu, R., Zhang, M., Chen, J., Liang, S., Xu, B., Shi, J., and Dickinson, R. (2013). The role of satellite remote sensing in climate change studies. *Nature Climate Change*, 3(10):875–883.

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., Klein, S. A., and Taylor, K. E. (2020). Causes of Higher Climate Sensitivity in CMIP6 Models. *Geophysical Research Letters*, 47(1):e2019GL085782. e2019GL085782 10.1029/2019GL085782.

Zhang, H.-M., Huang, B., Lawrimore, J., Menne, M., and Smith, T. M. (2019). Global Surface Temperature Dataset (NOAAGlobalTemp), Version 5.0. NOAA National Centers for Environmental Information.

Zhang, J., Li, L., Zhou, T., and Xin, X. (2013). Variation of surface temperature during the last millennium in a simulation with the FGOALS-gl climate system model. *Advances in Atmospheric Sciences*, 30(3):699–712.

Zhao, M., Golaz, J. C., Held, I. M., Ramaswamy, V., Lin, S. J., Ming, Y., Ginoux, P., Wyman, B., Donner, L. J., Paynter, D., and Guo, H. (2016). Uncertainty in model climate sensitivity traced to representations of cumulus precipitation microphysics. *Journal of Climate*, 29(2):543–560.