Beyond forcing scenarios: predicting climate change through response operators in a coupled general circulation model

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. ISSN 2045-2322 doi: https://doi.org/10.1038/s41598-020-65297-2 Available at http://centaur.reading.ac.uk/90344/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1038/s41598-020-65297-2

Publisher: Nature Publishing Group
including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading’s research outputs online
Beyond Forcing Scenarios: Predicting Climate Change through Response Operators in a Coupled General Circulation Model

Valerio Lembo1, Valerio Lucarini1,2,3✉ & Francesco Ragone4

Global Climate Models are key tools for predicting the future response of the climate system to a variety of natural and anthropogenic forcings. Here we show how to use statistical mechanics to construct operators able to flexibly predict climate change. We perform our study using a fully coupled model - MPI-ESM v.1.2 - and for the first time we prove the effectiveness of response theory in predicting future climate response to CO2 increase on a vast range of temporal scales, from inter-annual to centennial, and for very diverse climatic variables. We investigate within a unified perspective the transient climate response and the equilibrium climate sensitivity, and assess the role of fast and slow processes. The prediction of the ocean heat uptake highlights the very slow relaxation to a newly established steady state. The change in the Atlantic Meridional Overturning Circulation (AMOC) and of the Antarctic Circumpolar Current (ACC) is accurately predicted. The AMOC strength is initially reduced and then undergoes a slow and partial recovery. The ACC strength initially increases due to changes in the wind stress, then undergoes a slowdown, followed by a recovery leading to a overshoot with respect to the initial value. Finally, we are able to predict accurately the temperature change in the North Atlantic.

Climate change is arguably one of the greatest contemporary societal challenges1 and one of the grand contemporary scientific endeavours2. The provision of new and efficient ways to understand its mechanisms and predict its future development is one of the key goals of climate science. Global climate models (GCMs) are currently the most advanced tools for studying future climate change; their future projections are key ingredients of the reports of the Intergovernmental Panel on Climate Change (IPCC) and are key for climate negotiations3. For IPCC-class GCMs, future climate projections are usually constructed by defining a few climate forcing scenarios, given by changes in the composition of the atmosphere and in the land use, each corresponding to a different intensity and time modulation of the equivalent anthropogenic forcing. Typically, for each scenario an ensemble of simulations is performed, with each member differing in terms of initial conditions, applied forcing or chosen physical parameterizations. Subsequent phases of the Coupled Model Intercomparison Project (CMIP, currently the sixth phase CMIP6 is active4), which is part of the Program for Climate Model Diagnosis and Intercomparison (PCMDI), allowed the definition of standardized experimental protocols for numerical simulations performed with GCMs and for the evaluation of the GCMs runs5,6. A bottleneck of this approach is that the consideration of an additional forcing scenario requires running a new ensemble of simulations. Additionally, for each forcing scenario, it is hard to disentangle the impact of each component of the forcing, e.g. different greenhouse gases with their concentration pathways and land surface alterations in geographically distinct regions. Finally, no rigorous prescription exists for translating the climate change projections if one wants to consider different time modulations of a given forcing, e.g. a faster or slower CO2 increase.

Response Theory and Climate Change
A possible strategy for delivering flexible and accurate climate change projections is the construction of response operators able to transform inputs given by forcing scenarios into outputs in the form of climate change signal. At this regard, it is tempting to use the fluctuation-dissipation theorem (FDT)7,8, which provides a dictionary for
translating the statistics of free fluctuations of a system into its response to external forcings. The idea of using the FDT to predict climate change from climate variability has been first proposed by Leith and used by several authors thereafter. The FDT has recently been key to inspiring the theory of emergent constraints, which are tools for reducing the uncertainties on climate change by looking at empirical relations between climate response and variability of some given observables.

In the case of nonequilibrium systems, forced fluctuations contain features that are absent from the free fluctuations of the unperturbed system. Therefore, the applicability of the FDT for such systems faces some serious theoretical and practical challenges. The climate is a nonequilibrium system whose dynamics is primarily driven by the heterogeneous absorption of solar radiation. The motions of the geophysical fluids with the associated transports of mass and energy, as well as the exchanges of infrared radiation, tend to re-equilibrate the system and allow it to reach a steady state. Since the climate is not in equilibrium, climate change projects only partially on the modes of climate variability, whilst climate surprises - unprecedented events - are indeed possible when forcings are applied. See Ghil and Lucarini for a comprehensive mathematical and physical discussion of the relationship between climate variability and climate response to forcings.

Response theory is a generalisation of the FDT that allows one to predict how the statistical properties of general - near or far from equilibrium, deterministic or stochastic - systems change as a result of forcings. After the pioneering work by Kubo, response theory has been firmly grounded in mathematical terms for stochastic and deterministic systems; see Ref. for a link between the two perspectives. The use of the response theory introduced by Ruelle for predicting climate change has been successful in various numerical investigations performed on models of various degrees of complexity, ranging from rather simple ones, to intermediate complexity ones, up to simple Earth-like climate models. The key step is the computation of the Green function for each observable of interest. Then, the corresponding climate change signal is predicted by convolving the Green function with the temporal pattern of forcing. Once the Green function is known, response theory allows one to treat in a unified and comprehensive way forcings with any temporal modulation, ranging from instantaneous to adiabatic changes.

Note that, despite non-equilibrium conditions, a subtle relation exists between climate response and climate variability. Indeed, natural modes of variability of the unperturbed system that can be identified as prominent features in the autocorrelation or power spectra of climatic fields are described by the Ruelle-Pollicott poles. Such poles are responsible for the amplified response of the system to a resonating forcing with suitably defined spatial and temporal pattern.

A similar heuristic approach, although not rooted in formal Ruelle response theory, was also proposed in the seminal work of Hasselmann et al. While some doubts were raised in general terms regarding its applicability, the method showed excellent skills for a wide range of observables in GCM experiments. The crucial point is that in these works linear response formulas were applied to the output of individual experiments with GCMs. Ruelle’s response theory clarifies that the heuristic idea of Hasselmann et al. is instead mathematically grounded if one considers ensemble averages rather than individual experiments.

The climate models used so far to test response theory as formulated above lacked an active and dynamic ocean, so that the multiscale nature of climate processes was only partially represented. Capturing the slow oceanic processes is essential for a correct representation of the multidecadal and long-time climatic response. A proper treatment within the boundaries of the theory, based on ensemble of simulations with the same model featuring an active ocean component, is still lacking.

The response of a slow (oceanic) climatic observable of interest has been investigated so far in relation to the change in the dynamical properties of some other climatic observable, by constructing a linear regression between the predictand and predictor using the properties of the natural variability of the system. While indeed attractive and promising (and showing a good degree of success), this point of view cannot be rigorously ported to climate change studies. Using the resulting transfer function to predict the change of the observable of interest in a forced experiment would implicitly assume the validity of the FDT; see discussion above.

We remark that, in some cases (but not always), it is theoretically possible to use a climatic observable for predicting the response of another climatic observable of interest in the presence of an external forcing acting on the system. The possibility of using an observable as a surrogate forcing able to retain predictive power depends on non-trivial properties of its frequency-dependent response (its susceptibility, see below for a definition) and, in more practical terms, on the possibility of separating two or more dynamically relevant time scales in the system. See recent applications of this idea in Zappa et al. and, using a methodology based on adjoint modelling, in Smith et al.

Predicting Climate Change using the Ruelle Response Theory

In this paper we show for the first time how Ruelle’s response theory can be used to perform successful centennial climate predictions in the fully coupled climate model MPI-ESM v1.2. We consider two ensemble experiments. One experiment features an instantaneous CO2 doubling (2xCO2) taking place at year 1850, and the outputs of its ensemble members are used to compute the Green functions of several observables of interest, as described in the Methods section. We then perform an ensemble of runs where the CO2 concentration is increased - starting also at year 1850 - at the rate of 1% per year until doubling (1pctCO2), which takes place after about 70 years (y). We use the Green function computed with 2xCO2 to reconstruct the response of 1pctCO2 and compare the prediction with direct numerical simulations. As discussed in the Methods section, the Green function is defined for a period of 2000 y. Therefore, in what follows we are able to perform predictions beyond the time frame simulated in the 1pctCO2 scenario (which is only 1000 y long), thus showing very useful predictive power.
First, we analyse the response of the globally averaged near-surface temperature ($T_{2m}$) on short and long time scales. We then focus on two key aspects of the large-scale ocean circulation, namely the Atlantic Meridional Overturning Circulation (AMOC)\(^{41,42}\) and the Antarctic Circumpolar Current (ACC)\(^{43}\), and show that we can achieve excellent skill in predicting the the slow modes of the climate response. We also look into the global ocean heat uptake (OHU)\(^{44}\). A non-vanishing OHU indicates the presence of a global net energy imbalance. In current conditions, the ocean is well-known to absorb a large fraction of the Earth's energy imbalance due to global warming and to store it through its large thermal inertia, up to time scales defined by the deep ocean circulation\(^{45}\). The warming goes on for multicentennial scales, in a way that is not captured at all by models featuring a non-dynamic ocean\(^{15,27}\). The importance of the slow modes of climate response, associated with the oceanic thermal inertia, can be quantified considering the ratio between the transient climate response (TCR) and the equilibrium climate sensitivity (ECS), sometimes referred to as the realised warming fraction\(^{1}\); see the Methods section for the precise definition of these quantities. Here we have $\text{TCR/ECS} \approx 0.5 \ (\text{ECS} \approx 3.5 \ K)$, which is much smaller than what found ($\approx 0.85$) by Ragone et al.\(^{17}\), indicating a more prominent role of the slow modes of variability in the model investigated here. The prediction obtained via response theory shows the establishment of steady state conditions for times larger than 1000 y.

The Green function - see Fig. 5 - provides information on the time scales of the response. As a result of the presence of slow oceanic time scales, the Green function significantly departs from a simple exponential relaxation behavior, which is sometimes adopted to describe the relaxation of the climate system to forcings\(^{50,57}\). The idea of defining a general Green function as a sum of multiple or infinite\(^{29}\) exponential functions with different timescales, generalising Hasselmann’s ideas - the so- called pulse-response method - and along the lines of previous studies\(^{28,49,63}\) is beyond the scope of our analysis and will be investigated further in a future work. In our case, after a fast decrease for short time scales, the Green function tends to zero at a much slower pace for times longer than 100 y, in agreement with what reported by Held et al.\(^{37}\).

**Atlantic meridional overturning circulation.** The AMOC is strongly influenced by buoyancy perturbations in the Atlantic basin\(^{41}\). It is relevant at climatic level because it encompasses about 25% of the total (atmospheric and oceanic) meridional heat transport\(^{42}\). The time series of annual mean AMOC strength in the 1pctCO2 scenario is shown in Fig. 2a. The AMOC strength undergoes a decrease by about 30%, reaching its minimum in about 150 y. Successively, the AMOC slowly recovers.

The prediction of the AMOC change obtained via response theory captures very well the ensemble mean of the time evolution for the 1pctCO2 in the first 1000 y. The corresponding Green function is shown in the inset of Fig. 6a. On short time scales, we have a reduction of AMOC, as a result of the negative value of the Green function. On longer time scales ($>100 y$), a negative feedback acts as a restoring mechanism, associated with a positive sign in the Green function. The presence of fast (meaning here decadal) response associated with the GHG forcing has already been found in other models\(^{50,54}\), and is most likely related to the timescales of the sea-ice melting, consistently with paleoclimate simulations of the last interglacial climate with prescribed freshwater influx from reconstructed sea-ice melting\(^{52}\). The slow recovery of the AMOC might be understood as a heat\(^{53-55}\) and freshwater\(^{56}\) advection feedback.

**Figure 1.** Comparison between the simulated globally averaged near-surface temperature $T_{2m}$ (in K) in the 1pctCO2 scenario (thick red line with ensemble mean uncertainty) and the prediction performed using the linear Green function in Fig. 5 (thick blue).
In the 1001–2000 y period, response theory shows that a steady state is progressively reached over multi-centennial scales. The newly established AMOC is significantly weaker than the unperturbed AMOC, although a large ensemble spread is found. This is consistent with simulations obtained from higher resolution versions of the same model, intermediate-complexity models and other fully coupled models inclusive of an interactive carbon cycle.

The Antarctic Circumpolar Current. The ACC is by far the strongest large-scale oceanic current and its role in the general circulation is two-fold. On one hand, it isolates Antarctica from the rest of the system, being associated with a very strong zonal circulation in the Southern Ocean. On the other hand, although eminently wind-driven, it marks the area of outcropping of deep water occurring at the southern flank of the subtropical gyre, as part of the global-scale overturning circulation.

The Green function is shown in the inset of Fig. 6b. We find that the initial strengthening of the ACC can be associated with an increase in surface zonal wind stress (not shown here). Such a surface forcing determines...
an enhanced Eulerian mean ACC transport, consistently with previous low resolution simulations\(^\text{59}\). On decadal scales, we have a loss in the correlation between wind stress and ACC, corresponding to the Green function turning negative after about 30\,y. Beyond these time scales, we have time-wise coherent response of the AMOC and ACC, underlying the response of the global ocean circulation. Other models\(^\text{60,61}\) also feature such a behavior on intermediate time scales, consistently with the idea that the two circulations are related via the thermal wind balance\(^\text{62}\).

Figure 2b shows that the prediction of the ACC strength evolution in the \(\text{pctCO}_2\) scenario is rather accurate for the first 1000\,y, except for an underestimation of the positive short-term response, which is smoothed out. This points to an insufficient ability of response theory in representing the complex coupling between surface wind stress and downward momentum transfer. Furthermore, we observe the presence of a strong variability (on decadal time scales) of the predicted signal. This might result from either the small ensemble size or, more interestingly, could be the signature of the natural variability, encoded by a Ruelle-Pollicott pole\(^\text{2,28,29}\); see discussion in Sect. IA.

Note that in the 1001–2000\,yrs period the ACC reaches an approximate steady state a bit later than the AMOC, possibly as a result of having a larger inertia, consistently with the different depth scales of the two currents\(^\text{62,63}\). The AMOC maximum overturning depth scale is indeed located at about 1000 m, whereas the outcropping in the Southern Ocean is related to isopycnal surfaces reaching much deeper. This has profound implications for setting the time scales of the ACC and AMOC response. The propagation of deep water formation anomalies in the Northern Hemisphere is in fact mediated by Kelvin waves in the Northern Atlantic, whereas much slower interior adjustment through Rossby waves communicates the anomaly to the Southern Ocean\(^\text{64}\).

**Ocean Heat Uptake.** Looking at the 2000\,y of prediction in Fig. 3, we notice that response theory accurately predicts the response at all time scales. The linearity of the OHU increase in the 70\,y of integration comes from the convolution of the singular component of the corresponding Green function with the ramp, see the Methods section. After the CO\(_2\) concentration stabilizes, the OHU decreases towards vanishing values. In the last 1000\,y, response theory predicts a further decrease in the OHU down to a value of the order of \(\approx 5 \times 10^{13}\) W. As the cli-
The climate system as a whole relaxes towards the newly established energetic steady state through the negative Planck feedback, each of its subcomponents go through a process of relaxation. What we portray in Fig. 3 after year 1920 is the relaxation of the slowest climatic component, namely the global ocean. The remaining imbalance at the end of the prediction can also be interpreted as either resulting from the ultra-long time scales required for reaching rigorous steady state conditions, or as the signature of a model energy bias, associated with non-vanishing energy budget at steady state; on this aspect, see discussion and Fig. 1 in Lucarini and Ragone.

The North Atlantic Cold Blob. Finally, we study the surface temperature response for a domain covering the North Atlantic (between lon 26° W and lon 53° W and between lat 53° N and lat 69° N), thus including the areas where the deep water formation occurs. The region is identified as a peculiar spot for the effects of the GHG forcing, since the sea-ice melting has been hypothesized to delay the surface warming of this area compared to surrounding regions, through a weakening of the overturning circulation. Indeed - see Fig. 4 - the surface warming over the North Atlantic region is remarkably different from the behavior of the extratropics, which features a time dependent response (not shown here) similar in shape but somewhat amplified with respect to the global mean depicted in Fig. 1. Indeed, a long-lasting plateau - a hiatus in the temperature increase of more than 100 y - is observed around the end of the CO₂ increase ramp in the North Atlantic. The plateau is well captured by response theory, and comes in agreement with the AMOC weakening predicted in Fig. 2a. This result is non-trivial, given that such local response results from an interplay of local factors and, as mentioned, the response of the large-scale circulation. This hints at the potential of using response theory to identify global quantities that can be used as predictors for the response of local observables.

Discussion and Conclusions
We have shown here that response theory is a valuable tool for flexibly predicting crucial features of the climate change signal. All relies on obtaining the observable-dependent Green function from model simulations. The Green function allows one to deal with a continuum of time-dependent forcings, beyond the standard use of reference scenarios. Our findings provide guidance to the climate modellers’ community on how to set up climate
change experimental protocols, minimising the need for computational resources. This is especially relevant when one wants to study the overall impact and individual effects of multiple climatic forcings or investigate geoengineering options.

We have made here use of a fully coupled climate model, highlighting the slow components of the response associated with the oceanic modes of variability. The presence of a vast range of active time scales in the system makes the prediction of the response theoretically challenging and practically extremely relevant. Compared with previous contributions to the literature that have attempted with a good degree of success the prediction of changes in slow climatic fields using some form of linear regression, our approach is mathematically more robust as it derives directly from basic results in non-equilibrium statistical mechanics.

We stress that response formulas based on Ruelle’s theory are rigorously valid, but only when considering ensemble averages. On one side, this clarifies the limits of validity of previous attempts to apply similar formulas to individual model runs. On the other side, it indicates the importance of considering large ensemble strategies when planning major computational efforts for climate change projections.

The predicted changes in the AMOC and ACC feature clearly distinguishable fast and slow regimes of response. The former is essentially different in the two branches of the global ocean circulation, being the ACC subject to the effect of surface wind stress anomalies (substantially underestimated, compared to actual simulations). The latter is found to be well correlated between AMOC and ACC, as a signature of the forced response of the global ocean circulation circulation, which they are both part of. Coherently with previous findings, the ACC reaches a steady state much later than the AMOC. The plateau in near-surface warming over the North Atlantic is also related to the initial slowdown of the AMOC, and contrasts with the regular increase of the surface temperature we find globally (and predict accurately).

Figure 7. Same as in Fig. 5, for OHU (in W). The linear Green function is in $W \cdot y^{-1}$.

Figure 8. Same as in Fig. 5, for the near-surface temperature in the North Atlantic (in K). The linear Green function is in $K \cdot y^{-1}$.
We remark that the time-dependent transient and long-term evolutions resulting from the CO₂ increase are qualitatively different for the various climatic observables investigated here. Nonetheless, in all considered cases response theory successfully predicts the time-dependent change.

Ruelle’s response theory provides a relatively simple yet robust and powerful set of diagnostic and prognostic tools to study the response of climatic observables to external forcings. The availability of a large number of ensemble members allows for constructing more accurate Green functions and for studying effectively the response of a broader class of climatic observables. The approach proposed here could be extremely useful to inform the planning of major computational efforts for climate projections, putting such endeavours on firm theoretical grounds and optimizing the use of computational resources.

We have dealt here with a forcing due only to changes in the CO₂ concentration. This means that the pattern of the forcing is determined by heating rate associated with a spatially homogeneous CO₂ mixing ratio. Within the linear regime, different sources of forcings can be treated independently. The next step is to investigate other nontrivial forcings (e.g. aerosol forcings, land-use change, land glaciers location and extension). As an example, response theory has been proposed as a tool for framing geoengineering strategies and understanding its limitations.

Response theory provides a powerful formalism to tackle different problems related to the concepts of feedback and sensitivity. A promising application is the definition of functional relations between the response of different observables of a system to forcings, in the spirit of some recent investigations (see, e.g. Zappa et al.66). This would allow to treat comprehensively the concept of feedback across different time scales and define causal links between variables67–69. On a different line, one of the most promising future applications will be given by the synergy between the formalism of response theory and the recently proposed theory of emergent constraints67,70. The combination of the two approaches could lead to much needed insights on the climate response to forcings.

Additionally, along the lines of the pulse-response approach, it is in principle possible to try to extract the characteristic time scales of the response of the system by fitting the Green functions of the considered observables as a weighted sum of (in principle, infinitely many) exponential functions. As explained in Refs. 79,80, the time scales of the exponential functions are the same for all observables. Instead, the weight of each exponential contribution does depend on the observable of interest, with the response of rapidly equilibrating observables being dominated by the fast time scales, while the opposite holds for observables associated with the slow components of the system. The optimal fit can be obtained through a global optimisation procedure, where the response of various different observables is simultaneously fitted to a sum of exponentials. We will delve into this interesting problem of inverse modelling in a separate publication.

Clearly, in some applications one may want to test accurately to what extent nonlinear effects are relevant, as the theory is also applicable to higher orders66. Some insights into the non linear component of the response could also be obtained by appropriately combining forcings differing in sign and magnitude66,67,69. Nonetheless, being based on a perturbative approach, response theory (linear and nonlinear) has, by definition, only a limited range of applicability (e.g. one cannot use it to treat arbitrarily strong forcings). Still, the non-applicability of response theory has itself fundamental implications for the knowledge of the dynamics of the system one is studying. At a tipping point69–72 (or critical transition) the negative feedbacks of a system are overcome by the positive ones and any linear Green function diverges as a result of the increase in the time correlations of the system due to a critical Ruelle-Pollicott pole38, signalling the crisis of the chaotic attractor66. Instead, near a critical transition, response operators do not converge unless one considers very weak forcings8,37. The experimental design provided here is thus also a clear and mathematically sound strategy for the study of conditions leading to tipping points and their role for the climate response in state-of-the-art climate models.

Appendix A: Methods

Simulations. The analysis is based on two ensembles of simulations with Max Planck Institute Earth System Model (MPI-ESM) v.1.2.40, using its coarse resolution (CR) version. It features, for the atmospheric module ECHAM6, a T31 spectral resolution (amounting to 96 gridpoints in longitude and 48 in latitude) and 31 vertical levels, for the oceanic module MPI-OM, a curvilinear orthogonal bipolar grid (GR30) (122 longitudinal and 101 latitudinal gridpoints) with 40 vertical levels. The two ensembles, each including 20 runs, are based on two different scenarios. The first one features an instantaneous doubling in CO₂ concentrations (from a reference value of 280 ppm, characteristic of pre-industrial conditions) at the beginning of the simulations (2xCO₂), the other one an increase in the CO₂ concentration at the constant rate of by 1% per year, until the 2xCO₂ level is reached after about 70 years; afterwards, the CO₂ concentration is kept constant (1pctCO₂). The procedure for the construction of the ensemble is analogous to the protocol for CMIP575 and Grand Ensemble76 experiments. A control run is performed for 2000 y with pre-industrial conditions; the initial conditions of each ensemble member is initialized from a state of the control run. The initial conditions are sampled from the control run every 100 y, in order to ensure sufficient decorrelation among the respective oceanic states (at least in the mixed layer77,78).

Retrieval of AMOC and ACC. Typically, the large-scale circulation in the ocean is measured in terms of the mass transport across a suitably chosen section of a basin. The strength of the AMOC is computed as the vertically integrated mass weighted meridional mass streamfunction across lat 26.5° N80. This is a standard diagnostics of the MPI-ESM model. The ensemble average of the AMOC volume transport converges to 17.7 Sv (1 Sv = 10⁶ m³ s⁻¹), which is consistent with recent available measurements from the RAPID monitoring array81. The ACC is roughly zonally symmetric, and its location is closely related to the isopycnal slopes in the Southern Ocean.
Traditionally, it has been measured in terms of the strength of the mass transport across the Drake passage. Similarly, we take the vertically integrated barotropic streamfunction difference between the $2^\circ \times 2^\circ$ boxes centered around the lon $68^\circ W$, lat $54^\circ S$ and lon $60^\circ W$, lat $65^\circ S$ locations. The ensemble average of the ACC is 138 Sv, which is consistent with the multi-model mean estimate found in Mejers et al., amounting to $155 \pm 51$ Sv. It is also not far from the value commonly used as benchmark for the assessment of climate models ($173$ Sv).

**Linear Response Theory.** Response theories allow one to predict how the statistical properties of a system changes as a result of acting modulations in its external or internal parameters. The validity of the corresponding response formulas is heavily dependent on the hypothesis that the unperturbed system is structurally stable, i.e., roughly speaking, far from bifurcations, or, in terms of geophysical systems, from tipping points (see related discussions in a climate context). Rigorous derivations of response theories have been provided for the case of deterministic and stochastic dynamics. We only remark here that statistical mechanical arguments encoded by the chaotic hypothesis (a non-equilibrium analogue of the ergodic hypothesis) indicate the feasibility of the methodology proposed here.

In this paper we follow to a large extent the approach presented by Lucarini et al. and by Ragone et al. (see also Refs. 16,26) for the study of a large ensemble of intermediate-complexity atmospheric model runs and follow the deterministic route for response theory. Let us consider a dynamical system described by the state vector $x \in \mathbb{R}^3$, whose dynamics is described by the set of differential equations $\dot{x} = F(x)$, where $F \in \mathbb{R}^3$ is a smooth vector field. We add a perturbation to the dynamics in the form $\Psi(x, t) = X(x)f(t)$, where $X \in \mathbb{R}^3$ is a smooth vector field that gives the structure of the forcing in the phase space, whilst $f$ is its time modulation. The expectation value of any observable $\Phi = \Phi(x)$ can be written as:

$$\langle \Phi \rangle (t) = \langle \Phi \rangle_0 + \sum_{n=1}^{\infty} \langle \Phi^{(n)} \rangle f(t) \tag{A1}$$

where $\langle \Phi \rangle_0$ is the expectation value in the unperturbed state, and the term $\langle \Phi^{(n)} \rangle f(t)$ gives the $n^{th}$ order perturbative contribution. We consider here only the first order correction $\langle \Phi^{(1)} \rangle f(t)$. The linear correction is given by the convolution of the linear Green function with the time modulation of the perturbation:

$$\langle \Phi \rangle^{(1)} (t) = \int dx \langle \Phi(x) \rangle_0 f(t) \tag{A2}$$

where $\langle \Phi(x) \rangle_0$ is the linear Green function of the generic observable $\Phi$. Because of causality, the linear Green function vanishes for negative times. For ease of notation we have not indicated in Eq. A1 the dependence of the response on $X$, as in the applications considered in this paper $X$ is fixed and only the time modulation $f$ is varied. Note that for a time modulation $f$ such that $\lim_{t \to 0} f(t) = f(0)$ finite, and $f(t) = 0$ if $t < 0$, as in the case of $f(t) = e^H(t)$, where $H$ is a nonvanishing constant and $H(t)$ is the Heaviside distribution ($H(t) = 0$ for $t < 0$ and $H(t) = 1$ for $t > 0$), one typically has that $\langle \Phi^{(1)} \rangle f(t) = 0$, as observed in this paper for all observables except the OHU. In this case, one has $\lim_{t \to 0} \langle \Phi \rangle f(t) = 0$ because the Green function has a singularity (in the form of a Dirac’s $\delta$ contribution) for $t = 0$.

We remark that in previous works, the linear prediction of the desired climate observable is instead obtained by convolving time pattern of another climatic observable - assumed to be the driver - rather than the actual external forcing - with an effective transfer function - rather than the true Green function. The conditions under which climate observables can be used as both predictands and predictors have been discussed by Lucarini.

By taking the Fourier transform of the Green function $\langle \Phi(x) \rangle_0$, one obtains the linear susceptibility of the observable $\chi^{(1)}_\Phi (\omega)$, where $\omega$ is the frequency. The susceptibility gives the frequency response to a forcing $f(t)$ as $\Psi^{(1)}_\Phi (\omega) = \chi^{(1)}_\Phi (\omega) f(\omega)$, where $f$ we indicate the Fourier transform. The susceptibility gives a spectroscopic description of the properties of the response of the observable, and its analysis can give interesting information on the most relevant time scales and related processes that determine the response of the observable.

**Procedure for the Retrieval of the Green Function.** The strategy for testing the prediction of the mentioned key variables with the coupled model ensembles is as follows. First, we compute the Green function from the 2xCO2 experiment. The time variable is defined in such a way that the instantaneous doubling occurs at $t = 0$. Hence, the time modulation of the forcing is given in this case by $f(t) = f_{\text{2xCO2}} H(t)$, where $f_{\text{2xCO2}}$ is a constant depending on the amplitude of the forcing. Since the radiative forcing is approximately proportional to the logarithm of the CO2 concentration, such a constant is given by $f_{\text{2xCO2}} = \log(2)$ (it would be $f_{\text{2xCO2}} = \log(p)$ if the final CO2 concentration were $p$ times as large as the initial one). Equation A2 can be thus rewritten as:

$$\frac{d}{dt} \Phi^{(1)}_{\text{2xCO2}} (t) = f_{\text{2xCO2}} G^{(1)}_b(t) \tag{A3}$$

The outputs of the 2xCO2 experiments and the corresponding Green functions for the observables described above are presented in Figs. 5, 6, 7 and 8.

In particular, the time evolution of OHU for this scenario is shown in Fig. 7. The positive forcing due to the instantaneous CO2 doubling leads to an instantaneous jump in the OHU, leading to an annual average value of more than 1 PW in the first year. Equation A3 then suggests that the Green function $G^{(1)}_{\text{OHU}}$ has a singular behav-
ior at $t = 0$ (cfr. Ref. 29), while a regular behaviour is found for $t > 0$, corresponding to the negative radiative Planck feedback.

For all observables, we then use the Green functions above to perform predictions for the 1pctCO2 scenario using Eq. A2. Thanks to the proportionality of the radiative forcing to the logarithm of the CO2 concentration, the time modulation of such forcing can be expressed as $f = f_{\text{pctCO2}} r(t)$, where $r(t)$ is a ramp function (cfr. Refs. 18 and 29):

$$r(t) = \begin{cases} 0 & t < 0 \\ \frac{t}{\tau} & 0 \leq t \leq \tau \\ 1 & t > \tau \end{cases}$$

where the time scale $\tau \approx 70$ y denotes the time needed to reach the doubling in the CO2 concentration and where $f_{\text{pctCO2}} = f_{\text{CO2}}$ because at the end of the ramp the CO2 concentration is doubled.

From the Green function, one could in principle compute the susceptibility and perform a spectral analysis of the properties of the response. However, the correct identification of spectral peaks in the susceptibility requires a much richer statistics than what we have available here. The reason is that, while the Green function is an integral kernel whose specific values at each $t$ are not of crucial importance per se, as it is their integrated contribution through convolution with the cosdered time pattern that determines the response, in the case of the susceptibility it is extremely important to make sure that the signal to noise ratio is very large at each individual value of $\omega$. This translates into the fact that, despite the Green function and the susceptibility being strictly connected, to obtain a satisfactory estimate for the latter require a statistics orders of magnitude larger than for the former, and possibly different and dedicated numerical estimation approaches13,17,28. An analysis of the susceptibility in experiments similar to what done in this work was attempted in Ragone et al.27, but using ten times more ensemble members. We therefore do not present an analysis of the susceptibility. While the analysis of the detailed frequency response of a climate model remains a very interesting and promising topic, it has to likely wait until experiments with at least several hundreds ensemble members will be available.

**Equilibrium Climate Response and Transient Climate Sensitivity.** Response theory allows to place on solid formal ground operational definitions of the sensitivity of the climate system14.18,27. One of the most important indicators of the global properties of the response of the system to climate change is the equilibrium climate sensitivity (ECS), which is the long term ($t \rightarrow \infty$) response of the observable $T_{\text{eq}}$ to an abrupt doubling of CO2 concentration2. Another common measure of the response is the transient climate response (TCR), which is the change in $T_{\text{eq}}$ realised in the 1pctCO2 scenario at the end of the ramp of CO2 increase4.

Using the formalism discussed in this paper, the ECS can be straightforwardly linked to the susceptibility, because $ECS = f_{\text{CO2}} \chi^{(1)}(0)2,18,27$. Additionally, the TCR can be computed as the result at time $t = \tau$ of the convolution of the Green function of $T_{\text{eq}}$ with the forcing given in Eq. A4. Indeed, more generally, the susceptibility can be interpreted as a generalised sensitivity function. In particular, as explained in Ragone et al.27, one can find an explicit functional relation for the realised warming fraction5, given by the ratio between TCR and ECS:

$$\frac{TCR}{ECS} = 1 - \int_{-\infty}^{+\infty} 1 + \frac{\text{sinc}(\omega \tau/2) e^{-\omega \tau/2} \chi^{(1)}(\omega)}{2 \pi k_\omega} \chi^{(1)}(0) d\omega$$

where $\text{sinc}(x) = \sin(x)/x$. The integrand in the second term on the right hand side of Eq. A5 gives the contribution of each time scale to the inertia of the system. Note that this approach allows for treating seamlessly - by changing the value of $\tau$ - the case of transient response to steeper or gentler increases of CO2. However, a detailed analysis of the relationship between TCR and ECS requires an accurate estimate of the susceptibility, that as explained above is beyond the scope of this work.

The theory of emergent constraint has been used to study the ECS85 and the TCR86 in climate models. The relationship between ECS and TCR proposed in Eq. A5 might be helpful elucidating and better understanding such results.

Received: 2 January 2020; Accepted: 23 April 2020;
Published online: 26 May 2020

**References**

1. Palmer, T. & Stevens, B. The scientific challenge of understanding and estimating climate change. *Proceedings of the National Academy of Sciences* **116**(49), 24390–24395 (2019).
2. Ghil, M. & Lucarini, V. The Physics of Climate Variability and Climate Change, *Reviews of Modern Physics*, In press; arXiv:1910.00583. (2020).
3. Intergovernmental Panel on Climate Change. *Climate Change 2013: The Physical Science Basis*. (Press, Cambridge University, Cambridge Mass., 2013).
4. Eyring, V. et al. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development* **9**(5), 1937–1958 (2016).
5. Eyring, V. et al. ESMValTool (v1.0) – a community diagnostic and performance metrics tool for routine evaluation of Earth system models in CMIP. *Geoscientific Model Development* **9**(5), 1747–1802 (2016).
6. Lembo, V., Lunkeit, F. & Lucarini, V. ThEiaTo (v1.0) – a new diagnostic tool for water, energy and entropy budgets in climate models. *Geoscientific Model Development* **12**(8), 3805–3834 (2019).
51. Zickfeld, K. et al. Long-Term Climate Change Commitment and Reversibility: An EMIC Intercomparison. Journal of Climate 26, 5782–5809 (2013).
52. Loutre, M. F. et al. Factors controlling the last interglacial climate as simulated by LOVECLIM1.3. Climate of the Past 10(4), 1541–1565 (2014).
53. Rahmstorf, S., Marotzke, J. & Willebrand, J. Stability of the thermohaline circulation. In The Warmwatersphere of the North Atlantic Ocean, pages 129–157 (Borntreager, 1996).
54. Scott, J. R., Marotzke, J. & Stone, P. H. Interhemispheric Thermohaline Circulation in a Coupled Box Model. Journal of Physical Oceanography 29(3), 351–365 (1999).
55. Lucarini, V. & Stone, P. H. Thermohaline Circulation Stability: A Box Model Study. Part I: Uncoupled Model. Journal of Climate 18(4), 501–513 (2005).
56. Jackson, L. C. Shutdown and recovery of the AMOC in a coupled global climate model: The role of the advective feedback. Geophysical Research Letters 40(6), 1182–1188 (2013).
57. Li, C., von Storch, J. S. & Marotzke, J. Deep-ocean heat uptake and equilibrium climate response. Climate Dynamics 40(5–6), 1071–1086 (2013).
58. Zickfeld, K., Arora, V. K. & Gillett, N. P. Is the climate response to CO2 emissions path dependent? Geophysical Research Letters 39(5), 105703 (2012).
59. Pye, J. C. & Saenko, O. A. Simulated changes in the extratropical Southern Hemisphere winds and currents. Geophysical Research Letters 33(6), L06701 (2006).
60. Meiers, A. J. S. et al. Representation of the Antarctic Circumpolar Current in the CMIP5 climate models and future changes under warming scenarios. Journal of Geophysical Research: Oceans, 117(C12208) (2012).
61. Heazé, C., Heywood, K. J., Stevens, D. P. & Ridley, J. K. Changes in Global Ocean Bottom Properties and Volume Transports in CMIP5 Models under Climate Change Scenarios. Journal of Climate 28(8), 2917–2944 (2015).
62. Marshall, D. P. & Johnson, H. L. Relative strength of the Antarctic Circumpolar Current and Atlantic Meridional Overturning Circulation. Tellus, Series A: Dynamic Meteorology and Oceanography 69(1), 133884 (2017).
63. Koltermann, K. P., Gouretski, V. & Jancke, K. Hydrographic Atlas of the World Ocean Circulation Experiment (WOCE): Volume 3: Atlantic Ocean, National Oceanography Centre (2011).
64. Johnson, H. L. & Marshall, D. P. A Theory for the Surface Atlantic Response to Thermohaline Variability. Journal of Physical Oceanography 32(4), 1121–1132 (2002).
65. Lucarini, V. & Ragone, F. Energetics of climate models: Net energy balance and meridional enthalpy transport. Reviews of Geophysics 49(1) (2011).
66. Bódai, T., Lucarini, V. & Lunkeit, F. Can we use linear response theory to assess geengineering strategies? Chaos: An Interdisciplinary Journal of Nonlinear Science 30(2), 023124 (2020).
67. Hall, A., Cox, P., Huntingford, C. & Klein, S. Progressing emergent constraints on future climate change. Nat. Clim. Chang 9, 269–278 (2019).
68. Tantet, A., Lucarini, V., Lunkeit, F. & Dijkstra, H. A. Crisis of the chaotic attractor of a climate model: a transfer operator approach. Nonlinearity 31(5), 2221–2251 (2018).
69. Lentz, T. M. et al. Tipping points in the Earth’s climate system. Proceedings of the National Academy of Sciences 105(6), 1786–1793 (2008).
70. Feudel, U., Pisarchik, A. N. & Showalter, K. Multistability and tipping: From mathematics and physics to climate and brain—A review. Chaos: An Interdisciplinary Journal of Nonlinear Science 28(3), 033501 (2018).
71. Ashwin, P. & von der Heydt, A. S. Extreme Sensitivity and Climate Tipping Points. Journal of Statistical Physics, https://doi.org/10.1007/s10955-019-02425-x (2019).
72. Lucarini, V. & Bódai, T. Transitions across melancholia states in a climate model: Reconciling the deterministic and stochastic points of view. Phys. Rev. Lett. 122, 158701 (2019).
73. Stevens, B. et al. Atmospheric component of the MPI-M Earth System Model: ECHAM6. Journal of Advances in Modeling Earth Systems 5(2), 146–172 (2013).
74. Jungclaus, J. H. et al. Characteristics of the ocean components in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model. Journal of Advances in Modeling Earth Systems 5(2), 422–446 (2013).
75. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. Bulletin of the American Meteorological Society 93(4), 485–498 (2012).
76. Maher, N., Matei, D., Milinski, S. & Marotzke, J. ENSO Change in Climate Projections: Forced Response or Internal Variability? Geophysical Research Letters 45(20), 11,390–11,398 (2018).
77. Ganopolski, A. & Rahmstorf, S. Abrupt Glacial Climate Changes Due to Stochastic Resonance. Physical Review Letters 88(3), 038501 (2002).
78. Wunsch, C. & Heimbach, P. How long to oceanic tracer and proxy equilibrium? Quaternary Science Reviews 27(7–8), 637–651 (2008).
79. Mauritzen, T. et al. Tuning the climate of a global model. Journal of Advances in Modeling Earth Systems 4(3), M00A01 (2012).
80. Bryden, H. L., Mujahid, A., Cunningham, S. A. & Kanwos, T. Adjustment of the basin-scale circulation at 26°N to variations in Gulf Stream, deep western boundary current and Ekmans transports as observed by the Rapid array. Ocean Science 5, 421–433 (2009).
81. McCarthy, G. D. et al. Measuring the Atlantic Meridional Overturning Circulation at 26°N. Progress in Oceanography 130, 91–111 (2015).
82. Donohue, K. A., Tracey, K. L., Watts, D. R., Chidichimo, M. P. & Chereskin, T. K. Mean Antarctic Circumpolar Current transport measured in Drake Passage. Geophysical Research Letters 43(22), 11,760–11,767 (2016).
83. Gallavotti, G. Chaotic hypothesis: Onsager reciprocity and fluctuation-dissipation theorem. Journal of Statistical Physics 84(5–6), 899–925 (1996).
84. Otto, A. et al. Energy budget constraints on climate response. Nature Geoscience 6(6), 415–416 (2013).
85. Caldwell, P. M., Zelinka, M. D. & Klein, S. A. Evaluating Emergent Constraints on Equilibrium Climate Sensitivity. Journal of Climate 31(10), 3921–3942 (2018).
86. Nisie, F. J. M. M., Cox, P. M. & Williamson, M. S. An emergent constraint on Transient Climate Response from simulated historical warming in CMIP6 models. Earth System Dynamics Discussions 2020, 1–14 (2020).

Acknowledgements
Valerio Lembo was supported by the Collaborative Research Centre TRR181 “Energy Transfers in Atmosphere and Ocean” funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), project No. 274762653. Valerio Lucarini was partially supported by the SFB/Transregio TRR181 project and by the Horizon2020 projects CRESCENDO (Grant no. 641816) Blue-Action (Grant No. 727852) and TiPES (Grant No. 820970). Simulations were performed with Mistral, the Deutsches Klimarechenzentrum (DKRZ) High Performance Computing system for Earth system research (HLRE-3) supercomputer, projects No. um0005 and bu1085.
Author contributions
V.Le. ran the ensemble of simulations, performed the retrieval of the Green functions that were used for the predictions and analysed the results. V.Lu. proposed and directed the work and commented on the results. F.R. supported the numerical implementation of the response operators and commented on the results. All authors equally contributed to the writing of the paper.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to V.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020