A new framework for climate sensitivity and prediction

Francesco Ragone[1,2], Valerio Lucarini[2,3,4], Frank Lunkeit[2]

1) Klimacampus, Institut für Meereskunde, University of Hamburg, Hamburg, Germany

2) Klimacampus, Meteorologisches Institut, University of Hamburg, Hamburg, Germany.

3) Department of Mathematics and Statistics, University of Reading, Reading, UK

4) Walker Institute for Climate System Research, University of Reading, Reading, UK

The sensitivity of the climate system to increasing CO_2 concentration and the response at decadal time-scales are still major factors of uncertainty for the assessment of the long and short term effects of anthropogenic climate change. Here we demostrate that it is possible to use Ruelle's response theory to predict the impact of an arbitrary CO_2 forcing scenario on the global surface temperature of a general circulation model. Response theory puts the concept of climate sensitivity on firm theoretical grounds, and addresses rigorously the problem of predictability at different time-scales. Conceptually, our results show that climate change assessment is a well-defined problem from a physical and mathematical point of view. Practically, our results show that considering one single CO_2 forcing scenario is enough to construct operators able to predict the response of climatic observables to any other CO_2 forcing scenario, without the need to perform additional numerical simulations, thus paving the way for redesigning climate change experiments from a radically new perspective.

One of the main goals of climate science is to predict how modulations on different time scales of internal or external parameters (such as the greenhouse gases - GHGs - concentration or the solar constant) impact the statistical properties of the system. Due to the observational evidence of the ongoing anthropogenic climate change, the problem of the response of the system to increasing greenhouse gases (GHGs) and in particular CO_2 concentration ($[CO_2]$) is of particular social and environmental relevance, as stressed by the Intergovernmental Panel for Climate Change (IPCC)[1]. The assessment of the future impacts of climate change under a variety of CO2 forcing scenarios[2,3]

and the evaluation of the climate sensitivity[4,5,6] to the increase of [CO_2] mostly rely on the use of general circulation models (GCMs). Despite the intense efforts put by the scientific community having led to impressive improvements in the GCMs complexity and computational performances, after several decades large uncertainties[3,6,7] are still present even in the evaluation of the Equilibrium Climate Sensitivity[3,6] (ECS), the change of the globally averaged surface temperature for doubling [CO_2], the most basic measure of climate sensitivity, as well as of the response of the surface temperature at different time scales for CO_2 forcings with non-trivial temporal evolution[3]. The importance of addressing the properties of the climate system (CS) on various temporal horizons is also emphasized by the presence of substantial uncertainties in the predictive skills of the models at decadal time scales[3]. The need of substantial advances in the scientific ideas at the basis of climate modelling strategies is more and more clear in the climate community[8]. An illuminating perspective on mathematical frameworks suited for a theory of climate sensitivity has been recently presented[9]. Here, we follow a complementary approach.

In the past, several attempts have been aimed at constructing some sort of response operator for the CS. Early attempts[10] tried to address empirically the problem of the cold start of climate simulations. More recently, the fluctuation-dissipation theorem[11] (FDT) has inspired various authors to reconstructing the climatic response to perturbations from some statistical properties of the unperturbed system, typically using severely simplified, quasi-Gaussian approximations[12,13,14,15], or constructing blended response algorithms[16]. A basic obstacle is that, as already anticipated by Lorenz[17], forced and free fluctuations are not equivalent. Ruelle[18,19,20] introduced general methods for studying how nonequilibrium systems respond to external perturbations. Ruelle's response theory (RRT) clarifies that Lorenz's intuition[17] applies to general nonequilibrium systems, so that applying FDT can lead in principle to large errors. Nonetheless, it is indeed possible to compute deviations from a nonequilibrium steady state (NESS) due to weak forcings through response formulas formally similar to those of the equilibrium case. Recently the RRT has been proposed as a rigorous framework for computing climate response and its applicability has been tested on the

Lorenz 96 model[21].

Going in the direction of bridging the gap between statistical mechanical theories and climate science, we show how the formalism of nonequilibrium statistical mechanics can be used for predicting the response of the globally averaged temperature of a GCM to an arbitrary temporal evolution of the CO_2 forcing, once the Green function of the system is computed starting from an ensemble of runs forced by instantaneous CO_2 doubling. While the CS is extremely nonlinear, the statistical properties of the response are remarkably well captured by linear version of RRT even when finite forcings of practical interest are applied. Our results show that only one single CO_2 forcing scenario has effectively to be taken into account in order to predict the response of the system to any other concentration scenario at infinite and lead-time. This suggest that one could reduce the vast range of forcing scenarios considered in the IPCC protocol with few selected scenarios, and extrapolate more information from currently available data. While what we report here refers to linear response, the theory behind these findings allows for treating nonlinear effects as well[22,23].

Linear response theory

RRT strictly applies to Axiom A dynamical systems[18,19,20]. However, the use of response formulas in most cases of physical interest is justified thanks to the Chaotic Hypothesis[24], which states that systems with many degrees of freedom effectively behave as Axiom A systems, when macroscopic observables are considered. In general, when we compute the expectation value of an observable in a numerical model as the long-term average once a stationary state has been reached, we are in fact implictly assuming that the system is Axiom A-like[20]. If we consider low dimensional models, or dynamical systems living at the edge of chaos, the basic tenets of response theory may fail[25].

We apply a weak forcing to the system so that the evolution equation can be written as $\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}) + \boldsymbol{X}(\boldsymbol{x})f(t)$, where $\boldsymbol{F}(\boldsymbol{x})$ represents the unperturbed dynamics with set boundary conditions, $\boldsymbol{X}(\boldsymbol{x})$ is a vector field defining the pattern of the forcing in the phase space, and $f(t)$ is

the time modulation of the forcing. We consider the case where the base dynamics is autonomous, but one could accommodate more general time-dependent flows by resorting to the concept of pullback attractor[9]. Ruelle showed that the expectation value of an observable Φ in the forced system can be computed as a perturbative expansion $\langle\Phi\rangle_f(t) = \langle\Phi\rangle_0 + \sum_{n=1}^{+\infty}\langle\Phi\rangle_f^{(n)}(t)$, where $\langle\Phi\rangle_0$ is the expectation value in the unperturbed state and the perturbative terms $\langle\Phi\rangle_f^{(n)}(t)$ can be computed by performing an n-convolution of the n^{th} order Green function with $f(t)$. In particular, the first term in the series gives the linear response:

$$\langle\Phi\rangle_f^{(1)}(t) = \int_{-\infty}^{+\infty} d\sigma_1 G_{\mathbf{\Phi}}^{(1)}(\sigma_1)\, f(t-\sigma_1) \tag{1}$$

where $G_{\mathbf{\Phi}}^{(1)}(t)$ is the first order Green function, given by the expectation value of a (complicated) observable defined on the unperturbed flow. In general $G_{\mathbf{\Phi}}^{(1)}(t)$ is a causal function ($G_{\mathbf{\Phi}}^{(1)}(t) = 0$ if $t < 0$). By taking the Fourier transform of Equation (1) we derive

$$\widetilde{\langle\Phi\rangle}_f^{(1)}(\omega) = \chi_{\Phi}^{(1)}(\omega)\tilde{f}(\omega) \tag{2}$$

where the linear susceptibility $\chi_{\Phi}^{(1)}(\omega)$ is the Fourier transform of $G_{\mathbf{\Phi}}^{(1)}(t)$, and $\tilde{f}(\omega)$ is the Fourier transform of $f(t)$. The susceptibility gives the structure of the response of the system to forcings at different frequencies (time-scales). The real and imaginary parts of the function represent the in- and out-of-phase response of the system respectively to a sinusoidal forcing at frequency ω. Maxima in the absolute value of the susceptibility correspond to resonances of the system, where the response is enhanced due to positive feedbacks at the corresponding time-scales. Equations (1-2) provide the basis for deriving $G_{\mathbf{\Phi}}^{(1)}(t)$ (or equivalently $\chi_{\Phi}^{(1)}(\omega)$) from one set of experiments, when $f(t)$ is known and $\langle\Phi\rangle_f^{(1)}(t)$ is measured. We can then use Equation (1) to perform projections at any lead-time t when a different temporal evolution $g(t)$ of the forcing is used. Equation (2) implies that for reconstructing $\chi_{\Phi}^{(1)}(\omega)$ one must consider a broadband forcing, since the response $\widetilde{\langle\Phi\rangle}_f^{(1)}(\omega)$ does not contain frequencies that are not present in the forcing $\tilde{f}(\omega)$. Being $G_{\mathbf{\Phi}}^{(1)}(t)$ a causal function, its Fourier transform $\chi_{\Phi}^{(1)}(\omega)$ obeys the following identity[26]:

$$\chi_{\Phi}^{(1)}(\omega) = \frac{i}{\pi} P \int_{-\infty}^{\infty} d\omega' \frac{\chi_{\Phi}^{(1)}(\omega')}{\omega' - \omega} \tag{3}$$

where P indicates integration in principal part and the susceptibility is related to its complex conjugate through $\chi_{\Phi}^{(1)}(\omega) = [\chi_{\Phi}^{(1)}(-\omega)]^*$. Equation (3) can be recast in terms of conventional Kramers-Kronig relations (KK), linking the real and imaginary parts of $\chi_{\Phi}^{(1)}(\omega)$[26,27].

Climate response and prediction

Demonstrations of the validity of RRT for simple systems of geophysical interest have been proposed in the past[16,21]; here we assess for the first time the predictive power of the theory with a real GCM. The numerical model used in this study is the open-source model PLASIM[28]. The atmospheric model has a full set of physical parameterizations for radiative processes, convection, phase transitions of water, surface exchanges and sea-ice (see the Methods Summary). Even if indeed not competitive with state-of-the-art GCMs, PLASIM produces a fairly realistic present climate and is representative of the class of complex numerical models used for operational climate prediction. The atmospheric model is set up in T21 horizontal resolution with 10 vertical levels, and is coupled to a 1-layer slab model of the oceanic mixed layer for a total of O(10^5) degrees of freedom The climate response and sensitivities evaluated with PLASIM are due therefore only to the fast feedbacks[29,30], missing contributions from ocean, continental ice-sheets, vegetation, and dynamic GHG-climate interactions. In order to simplify the analysis, we have removed daily and seasonal cycles, so that the evolution equations in the reference state do not explicitly depend on time.

We consider as our observable the globally averaged surface temperature T_s and as forcing the convergence of radiative fluxes due to the increase in the logarithm of [CO_2] (the radiative forcing scales approximately logarithmically with [CO_2] within a reasonable range of concentrations[2,3]). Therefore from now on $\langle \Phi \rangle_f^{(1)} = \langle T_s \rangle_f^{(1)}$ represents the expectation value of the increase the globally averaged surface temperature T_s in the linear regime and $f(t)$ represents the temporal

evolution of the radiative forcing correspondent to a chosen CO_2 forcing scenario. Linear RRT in principle applies only in the limit of infinitesimal forcings. Nonetheless, we show that its range of validity extends to rather intense finite forcings. In order to compute the Green function $G_{T_s}^{(1)}(t)$ and the susceptibility $\chi_{T_s}^{(1)}(\omega)$, we consider a scenario where [CO_2] is doubled instantaneously from the present-day value of 360 ppm to 720 ppm. In this case the forcing is given by $f(t) = f_{CO_2}^{2x} H(t)$, where $H(t)$ is the Heaviside function and $f_{CO_2}^{2x}$ is a constant, so that its Fourier transform is $\tilde{f}(\omega) = f_{CO_2}^{2x} P(\pi\delta(\omega) + i/\omega)$, where P stands for principal part and $\delta(\omega)$ is the Dirac delta.

We consider an ensemble of 200 runs, each of 200 years, starting from different initial conditions. (see the Methods Summary). Figure 1 shows the time serie of the ensemble average of the increase of globally averaged surface temperature for the instantaneous doubling scenario (blue). The shaded area represents the 95% (two standard deviations) of the ensemble variability. The long-term increase of the surface temperature for the doubling scenario (the equilibrium climate sensitivity) is rather high if compared with what is typically obtained with standard IPCC models, being 8.1 K against typical estimates between 1.5 and 4.5 K[2,3]. This is due to the fact we have chosen a simplified setup without the daily and in particular the seasonal cycle, which greatly enhances the system response to the radiative forcing. While this is mathematically more convenient at the price of decreasing the realism of our experiments, it also pushes the RRT more to its limits, thus providing a more stringent test for our approach. We have performed additional experiments increasing istantaneously the [CO2] by a factor $\sqrt{2}$ (about 510 ppm), checking that the observable indeed scales reasonably linearly with the logarithm of the [CO2], the actual behavior being only slightly sublinear (not shown). In the insert of Figure 1 we show $G_{T_s}^{(1)}(t)$, derived from the doubling experiment output (see the Methods Summary). The Green function $G_{T_s}^{(1)}(t)$ is to a first impression consistent with a relaxation process with a characteristic time scale of about 10 years, even if relevant differences emerge, as explained later. In earlier works[10] the response of the GCM was fitted with a prescribed functional form to then compute $G_{T_s}^{(1)}(t)$ analitically; here we are

computing it numerically directly from the output of the GCM without loss of information.

In order to test the predictive power of the theory, we consider another scenario where [CO_2] is increased by 1% per year starting from the present-day value of 360 ppm until the value of 720 ppm is reached, and kept fixed afterwards. The corresponding radiative forcing is a ramp function $g_\tau(t) = f_{CO_2}^{2x} t/\tau$ for $0 \leq t \leq \tau$ and $g_\tau(t) = f_{CO_2}^{2x}$ for $t > \tau$ with $\tau \approx 70\ years$. In Figure 3 we compare the ensemble average of the simulations and the prediction for $\langle T_s \rangle_{g_\tau}^{(1)}$ obtained using the estimate of the Green function shown in Figure 2. The agreement is excellent both on the short and long term. Discrepancies of less that 10% are present during the transient in the window between 25 and 100 years, because of of the strong nonlinearities due to the activation of the ice-albedo feedback. The degree of precision of the prediction obtained with the linear RRT is remarkable, considering the complexity of the climate model in use and the presence of strong nonlinearities in the underlying equations.

Climate sensitivity in response theory framework

Figure 3 shows the real $\Re\left\{\chi_{T_s}^{(1)}(\omega)\right\}$ and imaginary $\Im\left\{\chi_{T_s}^{(1)}(\omega)\right\}$ part of the susceptibility, computed as the Fourier transform of $G_{\boldsymbol{T_s}}^{(1)}(t)$. The function is plotted against the frequency $\xi = \omega/2\pi$ measured in $years^{-1}$. The susceptibility to CO_2 forcing at a frequency ξ' gives a precise quantification of the climate response at the corresponding time scale $t' = 1/\xi'$. The estimate of the susceptibility becomes rather noisy for frequencies higher than 0.25 $years^{-1}$ (time-scales $\leq 4\ years$,), where we are not able to compute reliable climate projections. The overall quality of the experimentally obtained $\chi_{T_s}^{(1)}(\omega)$ is confirmed by the good agreement between the measured and KK-reconstructed real and imaginary parts (see Methods Summary). We have truncated the integral in Equation (3) at the highest measured frequency $\xi_h = 0.5/\Delta t$, where $\Delta t = 1\ year$, and have used a subtraction technique known as singly subtractive KK relations[26]). One must emphasize that all the bumps around $\xi \approx 0.15\ years^{-1}$ and $\xi \approx 0.20\ years^{-1}$ found in the $\chi_{T_s}^{(1)}(\omega)$ are not

due (only) to the presence of the noise, but correspond indeed to physical processes of the system (multiannual variability), as they are consistently captured by the KK relations. These spectral features are not consistent with a simple relaxation model of the climate response.

While the importance of $G_{T_s}^{(1)}(t)$ is due to its predictive power, $\chi_{T_s}^{(1)}(\omega)$ is a powerful tool for defining accurately climate sensitivity at different time scales. The equilibrium climate sensitivity is defined as $ECS = \lim_{t\to+\infty}\langle T_s\rangle^{(1)}(t)$ after an istantaneous doubling of [CO_2]. Using Equation (2) and the definition of $\chi_{T_s}^{(1)}(\omega)$ one can show that, in linear approximation,

$$ECS = f_{CO_2}^{2x}\chi_{T_s}^{(1)}(0) \tag{4}$$

Therefore *ECS* is proportional to the zero frequency value of the susceptibility. From Figure 3 (and Figure 1) we see that $f_{CO_2}^{2x}\chi_{T_s}^{(1)}(0)$ ≈8.1 K. Therefore, when computing the *ECS* as long-term average of the T_s increase from a GCM run, we actually compute $\chi_{T_s}^{(1)}(\omega)$ for a specific frequency $\omega = 0$. By using Equations (2-3) and the expression of $\tilde{f}(\omega)$ we obtain:

$$ECS = \frac{2}{\pi}P\int_0^{+\infty} d\omega\,\frac{Im\left[f_{CO_2}^{2x}\chi_{T_s}^{(1)}(\omega)\right]}{\omega} = \frac{2}{\pi}P\int_0^{+\infty} d\omega\, Re\left[\widetilde{\langle T_s\rangle}_f^{(1)}(\omega)\right] \tag{5}$$

The RRT formalism permits to link the *ECS* to the imaginary part of the susceptibility at all frequencies or, equivalently, to the real part of the response at all frequencies for the imposed step function forcing $f(t)$. Equation (5) clarifies that the values of the response at all frequencies are relevant for determining the long-term response. One may compare the integrand of Equation (5) obtained for two different models in order to test their agreement. In so doing, one would find out which time-scales (and therefore which physical processes) are mostly responsible for possible discrepancies in their *ECSs*. Alternatively, one may find that two models with similar *ECS*s differ substantially regarding their response at different time scales.

The Transient Climate Response (TCR) is defined as the T_s increase at the moment (after about 70 years) [CO_2] has doubled following a 1% per year increase[6]. In our case $TCR(\tau) = \langle T_s\rangle_{g_\tau}^{(1)}(\tau)$ when the forcing is given by the ramp function $g_\tau(t)$. The forcing is the same used for the scenario

of Figure 3, and we obtain $TCR(\tau) = 7.2\,K$. In frequency domain, the forcing is $\tilde{g}_\tau(\omega) = f_{CO_2}^{2x} P\left(\pi\delta(\omega)e^{2i\omega\tau} + i\,\mathrm{sinc}\,(\omega\tau)e^{i\omega\tau}/\omega\right)$, where $sinc(x) = \sin(x)/x$. Using Equations (2-5) one gets (see the Methods Summary):

$$ECS - TCR(\tau) = INR(\tau)$$
$$= f_{CO_2}^{2x} P \int_{-\infty}^{+\infty} d\omega\, \chi_{T_s}^{(1)}(\omega) \frac{1 + \mathrm{sinc}(\omega\tau/2)e^{-i\omega\tau/2}}{2\pi i\omega} \quad (6)$$

The difference between *ECS* and *TCR* is given by a weighted integral of the susceptibility, accounting for the contribution of processes and feedbacks occurring at different time scales. The integral in Equation (6) by KK relations gives $\chi_{T_s}^{(1)}(0)$ in the limit $\tau \to 0$, decreases monotonically with τ, and vanishes in the limit $\tau \to \infty$. $INR(\tau)$ provides a measure of the *inertia* of the system at the timescale τ, due to the overall contribution of the internal physical processes and characteristic time-scales of the relevant climatic sub-systems[31,32]. By changing τ, $INR(\tau)$ allows one to deal with different rates of increase of [CO_2] (Figure 4). In Figure 4 we also plot $\langle T_s\rangle_f^{(1)}(\tau) - \langle T_s\rangle_{g_\tau}^{(1)}(\tau)$, which instead measures the difference in the transient response at time τ between the case where the forcing is modulated by $f(t)$ and by $g_\tau(t)$, respectively. This quantity approximately coincides with $INR(\tau)$ for $\tau > 50$ years but has a completely different behavior for small values of τ.

Horizon of predictability

We can quantify the limits to the predictive skills of the theory with a scale analyis. We focus on the high-frequency range. Given an ensemble of N realizations, the error in the estimate of $\langle\widetilde{\Phi}\rangle_f^{(1)}(\omega)$ with respect to the true expectation value can be represented as a random signal $\sigma(\omega)$ such that $|\sigma(\omega)| \approx \alpha N^{-1/2}$, where α is a suitable constant. If the forcing spectrum decays for high frequencies as $|f(\omega)| \sim \beta\omega^{-\nu}$, from Equation (2) we derive that the error on the estimate of the susceptibility is $\left|\delta\left[\chi_\Phi^{(1)}(\omega)\right]\right| \approx \alpha/\beta N^{-1/2}\omega^\nu$. Assuming that the true susceptibility asymptotically decays as $\left|\chi_\Phi^{(1)}(\omega)\right| \sim \gamma\omega^{-\kappa}$, the signal to noise ration approaches unity at

$$\omega_c = \left(\frac{\beta\gamma N^{1/2}}{\alpha}\right)^{1/(\nu+\kappa)} \quad (7)$$

For $\omega \gtrsim \omega_c$ the estimate of the susceptibility will be seriously deteriorated by the presence of noise so that we will have no skills in prediction at time scales smaller than $\tau_c = 2\pi/\omega_c$. From our data N=200, we estimate $\alpha \approx 0.2$, $\gamma \approx 0.63$, κ =1, while we have by consruction $\beta = 1$ and ν =1. Using Equation (7), we obtain an estimate of $\tau_c \approx 1$ year, which fits with the qualitative information provided by Figure 3. The design of the experiment and the processing of the output involve 3 parameters: the temporal resolution of the response signal Δt (tipically a coarse graining of the raw output of the numerical model), the length of the simulations T, and the size of the ensemble N. They define respectively the high frequency cutoff $\omega_h = \pi/\Delta t$, the low frequency cutoff $\omega_l = 2\pi/\mathrm{T}$ (that is also the spectral resolution $\Delta\omega = \omega_l$) and the critical value for predictability ω_c. Given the time scales of interest, the appropriate temporal resolution, integration length and ensemble size to have good predictive skills can thus be determined. The quality of prediction differs for different observables and forcings, with *red* observables featuring a slowly decaying susceptibility and forced by *red* modulations being better candidates.

Discussion

The applicability of RRT has several important implications regarding climate change science. At a fundamental level, it demonstrates that the problem of climate change is indeed well-defined from a mathematical and a physical point of view, as we can construct a statistical mechanical framework. Applying the FDT to a non-equilibrum system like the climate can introduce mathematically uncontrollable errors in the estimates of the response or, anyway, requires a very accurate representation of the attractor of the system. While there might be ways to circumvent this problem[33], we expect to be able to reconstruct the climate response from its natural variability only in special cases. The approach proposed here bypasses some of these mathematical issues by exploiting formal properties of the response and allows for constructing rigorous definitions of

climate sensitivity at different time scales through the susceptibility function. We have provided a framework for relating the difference between transient and equilibrium climate sensitivity to the inertia of the CS, and have shown how these properties depend of the response of the system on all time scales. This partially addresses some issues debated in the literature regarding the specific relevance of the *ECS*[34] and provide viable ways to intercompare GCMs. Deviations from a simple relaxation behaviour have been detected, and point to the complexity of the climate response at multi-annual time-scales. Considering the corresponding spectral features is necessary for having a consistent and integrated picture of the overall climate response at all frequencies. Our approach clarifies the limits of simple linear feedbacks approximations.

On the practical side, our results provide a way to perform climate prediction (in an ensemble sense) at all lead-times. We have also shown how to estimate the predictability horizon and assess how it scales with different sizes of the ensemble of simulations. Inaccurancies in representing specific spectral features have serious impacts on our ability to predict climate response on the corresponding time scales, and our findings help understanding why, e.g., climate response at decadal time scales may be hard to capture. We have highlighted that our ability to predict the response may vary massively when different observables are considered. It appears that considering only one single CO_2 forcing scenario allows for reconstructing accurately the response of the system to other temporal patters of changes in [CO2]. Exploiting the linearity of the response it is possible to compose the effect of multiple forcings (like changes in the solar constant or in other GHGs) by simply adding the linear corresponding Green functions and susceptibilites; the extension to nonlinear cases is more cumbersome but theoretically doable[22,23], and basically points to a generalisation of the factor separation technique[35]. The vast number of diverse scenarios considered nowdays in the IPCC protocol could then be substituted by a more focused effort on few selected forcing experiments making use of ensemble methods. Or, conversely, given the present palette of forcing scenarios, one could fill in the gaps and create projections for combinations of scenarios without resorting to additional simulations. This is promising for applications,

considering that simulations for climate change assessment are extremely expensive and, at the same time, it is necessary to inform policymakers on wide ranges of climate change scenarios.

We have limited our analysis to one single observable of primary climatic interest. The analysis of other observables will shed light on the mechanisms determining the response of the CS to CO_2 forcing. As an example, the analysis of the response of large-scale meridional gradients of temperature at surface and in the middle troposphere will provide information on changes in the midla circulation. We expect the analysis to work better for globally or regionally integrated, rather than for local, quantities, in terms of signal to noise ratio in the estimate of the susceptibility at high frequencies. In general different observables will show different ranges of linearity. The response of water vapor related observables, in particular, is expected not to scale linearly in the same range as the surface temperature, because of the Clausis-Clapeyron equation[36]. In this case, if one wants to study large climate shifts the role of the nonlinear terms will likely be relevant. The existence of approximate functional relationship between the susceptibilities of different observables[22] would provide the key for defining rigorously the so-called emergent constraints[37,38].

A first tentative intercomparison analysis on several GCMs could be performed using CMIP5 data for the 4xCO2 scenario. The present analysis takes advantage of a rather large ensemble of simulations, while CMIP5 ensemble runs host only few ensemble members. Also, the 4xCO2 scenario realizes a stronger forcing than the 2xCO2 we have considered here; however, the GCMs hosted in the CMIP5 database feature a much lower sensitivity than our case, mostly because of the presence of the annual cycle, so that the linearity assumption should still be valid. The analysis of fully coupled atmosphere-ocean GCMs is crucial in order to understand the important role of ocean and interactive sea-ice on climate response. Obviously, considering models with interactive ocean introduces longer time scales and makes the analysis technically more difficult. RRT allows to diagnose rigorously discrepancies in the properties of the frequency dependent response of different models, supporting strategies for future GCM development aimed at improving the critical and persisting deficiencies in the model performances highlighted in the last IPCC report[3].

Methods Summary

The numerical model used in this study is PLASIM[28], a simplified GCM developed at the University of Hamburg and freely available at http://www.mi.uni-hamburg.de/plasim. The ensemble method to compute unbiased expectation values of the observables follows [21]. The data have been processed in MATLAB© V. 7.9 environment; Fourier transforms have been computed making use of the standard fft algorithm following [40]. The Kramers-Kronig analysis has been performed by means of the singly subtractive Kramers-Kronig method as described in [26].

Acknowledgments

The authors wish to thank C. Franzke and G. Gallavotti for commenting on an earlier version of the manuscript. F.R. wish to thank T. Bodai and S. Schubert for useful discussions. F.R. and V.L. acknowledge fundings from the Cluster of Excellence for Integrated Climate Science (CLISAP) and from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreement No. 257106. The authors acknowledge the Newton Institute for Mathematical Sciences (Cambridge, UK), hosting the 2013 programme "Mathematics for the Fluid Earth" during which part of this work was discussed.

Author contributions

F.R., V.L. and F.L. conceived and designed the experiments. F.L. performed the numerical simulations. F.R. analysed the data. V.L. performed the Kramers-Kronig analysis. F.R and V.L. wrote the paper. All the authors discussed the results and commented on the manuscript.

Author informations

Reprints and permission information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to F.R. (francesco.ragone@zmaw.de).

Figures legend

Figure 1 Ensemble average of increase of global surface temperature after an instantaneous increase of the [CO2] by a factor 2. The upper and lower limit of the bands are computed as two standard deviations of the ensemble distribution. Insert: Green function of the global surface temperature.

Figure 2 Comparison between GCM simulation (blue) and response theory prediction (red) for 1% per year increase of the CO2 concentratrion. The upper and lower limit of the bands are computed as two standard deviations of the ensemble distribution. The agreement between the GCM run and the prediction via linear response theory is remarkable, with only a slight discrepancy during the transient, most probably connected to the activation of the ice-albedo feedback.

Figure 3 Real (black) and imaginary (red) parts of the susceptibility computed as the Fourier transform of the Green function of Figure 1. Real (blue) and imaginary (pink) parts of the susceptibilty obtained via singly subtractive KK relations.

Figure 4 Measuring the relaxation of the system. Inertia of the system at time scale τ (blue) and difference at lead time τ between the response at the system to a step-like and ramp (of time scale τ) forcings (black).

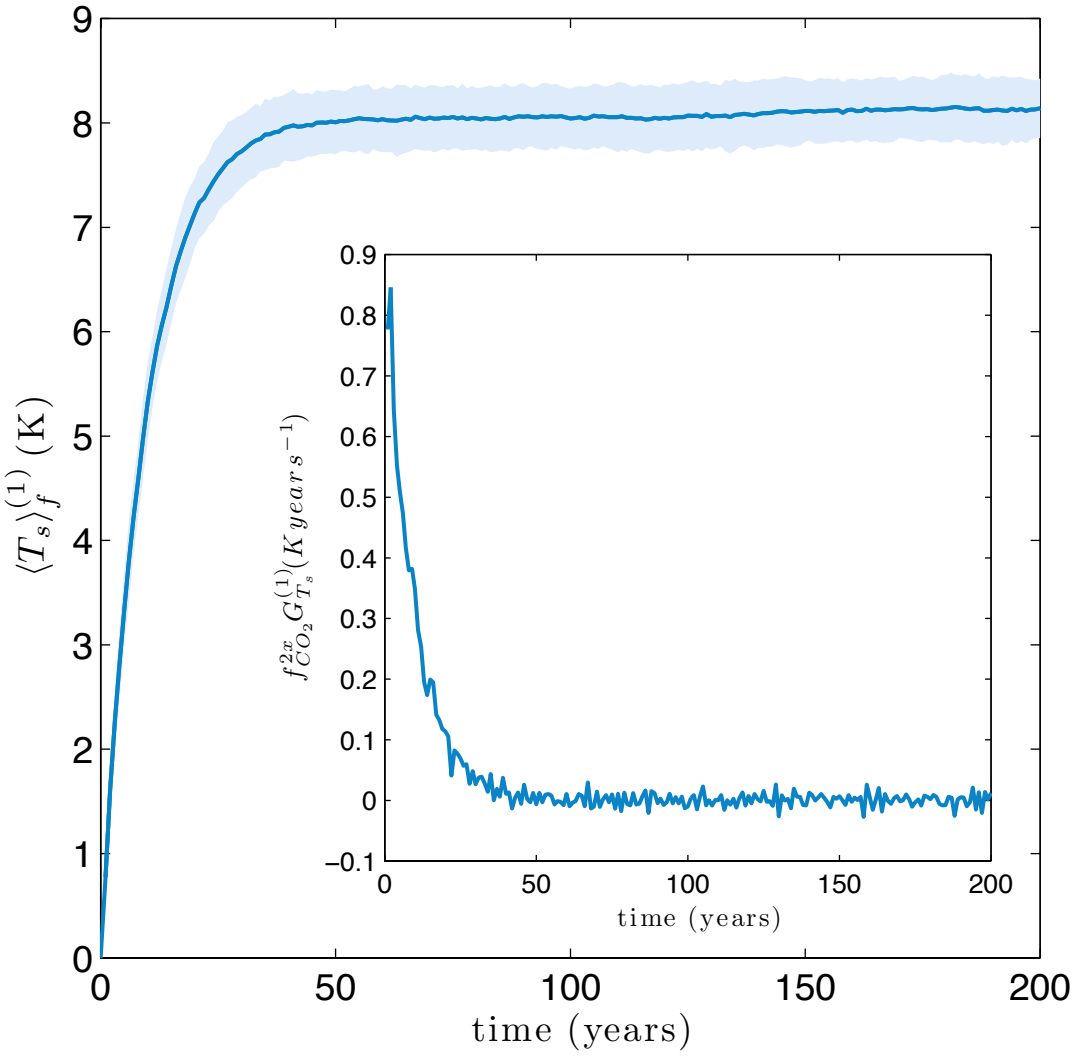
$\langle T_s \rangle_f^{(1)}$ (K)
time (years)
$f_{CO_2}^{2x} G_{T_s}^{(1)} (K\, years^{-1})$
time (years)

Figure 1

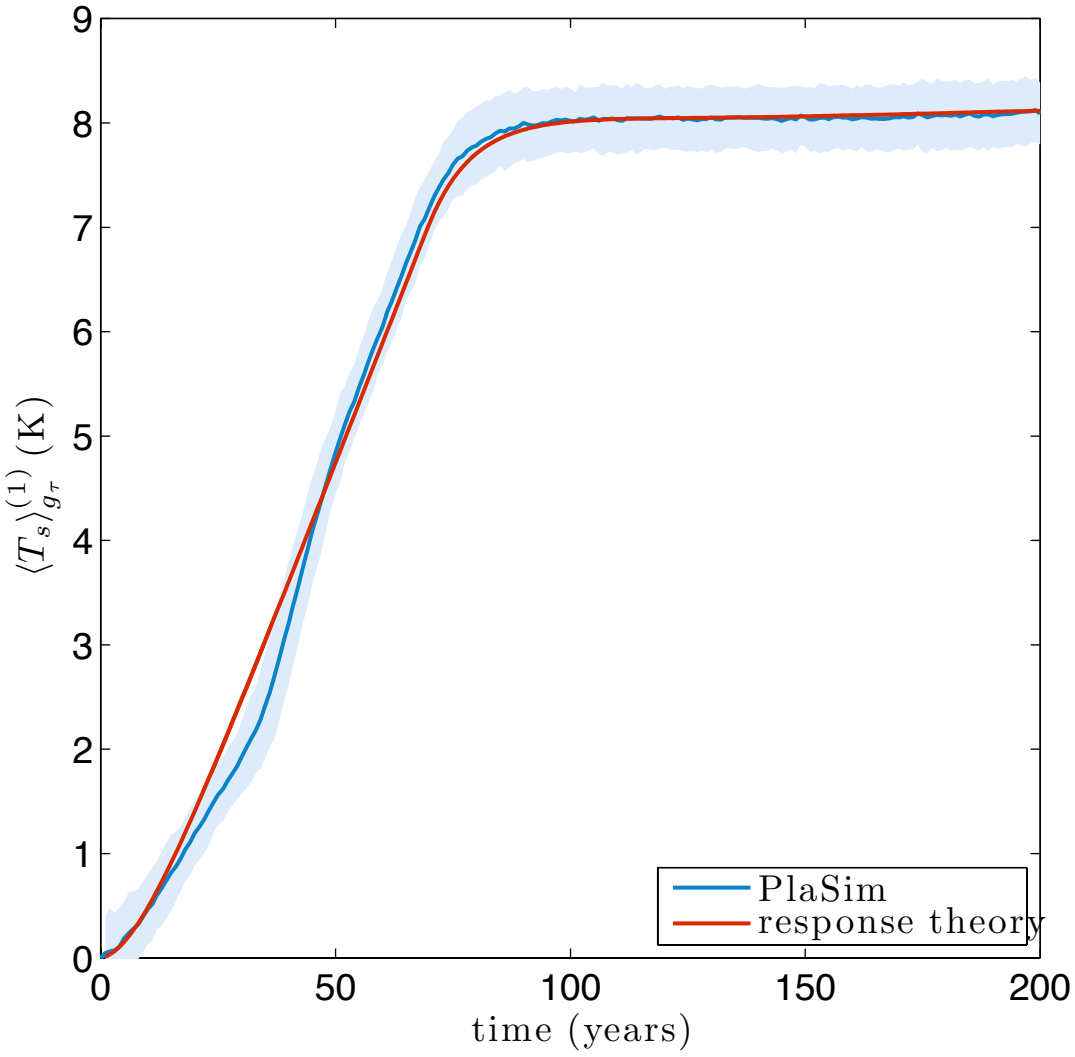
9
8
7
6
5
4
3
2
1
0
0
50
100
150
200
time (years)
$\langle T_s \rangle^{(1)}_{g_\tau}$ (K)
PlaSim
response theory

Figure 2

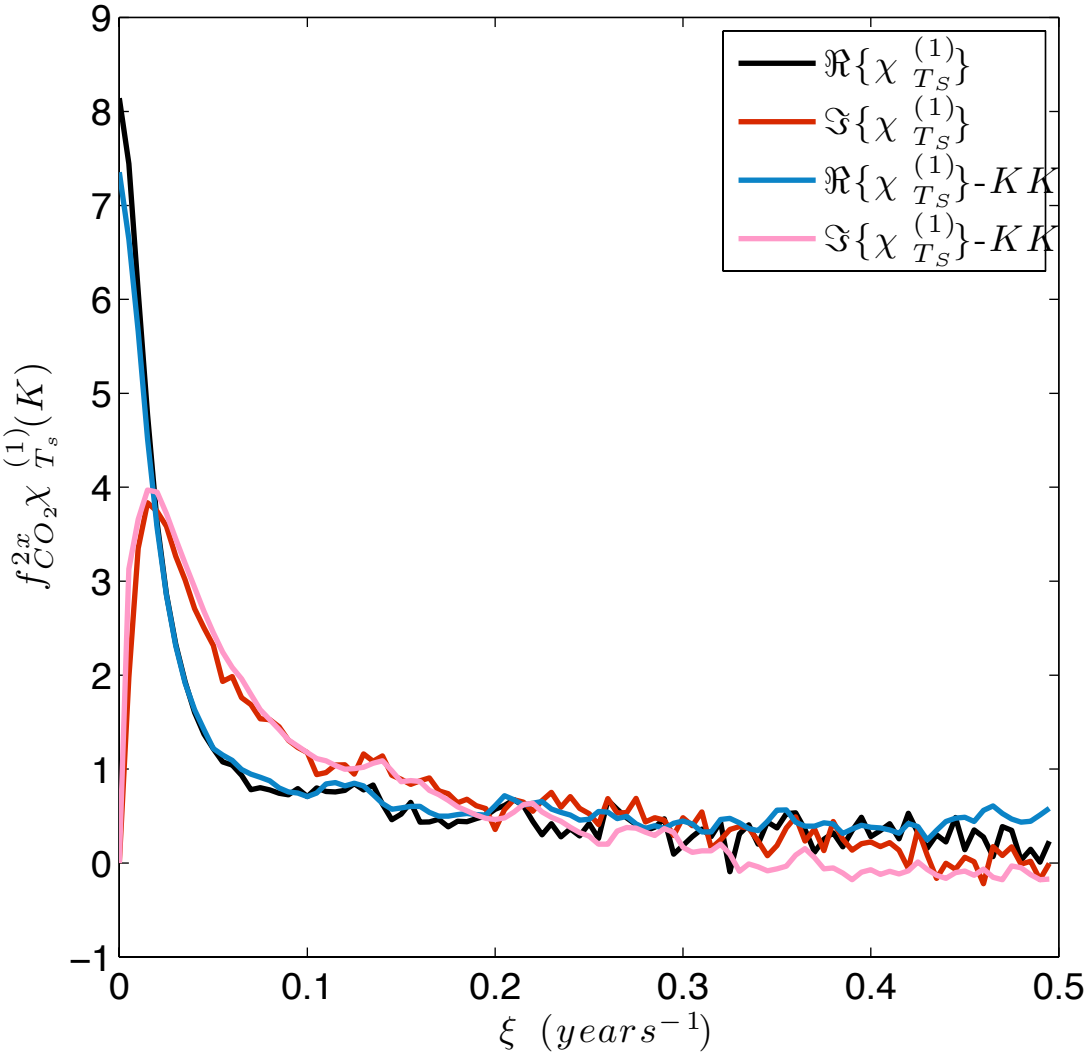
9
8
7
6
5
4
3
2
1
0
−1
0
0.1
0.2
0.3
0.4
0.5
$\xi \ \ (years^{-1})$
$f^{2x}_{CO_2}\chi^{\ (1)}_{T_s}(K)$
$\Re\{\chi^{\ (1)}_{T_S}\}$
$\Im\{\chi^{\ (1)}_{T_S}\}$
$\Re\{\chi^{\ (1)}_{T_S}\}$-$KK$
$\Im\{\chi^{\ (1)}_{T_S}\}$-$KK$

Figure 3

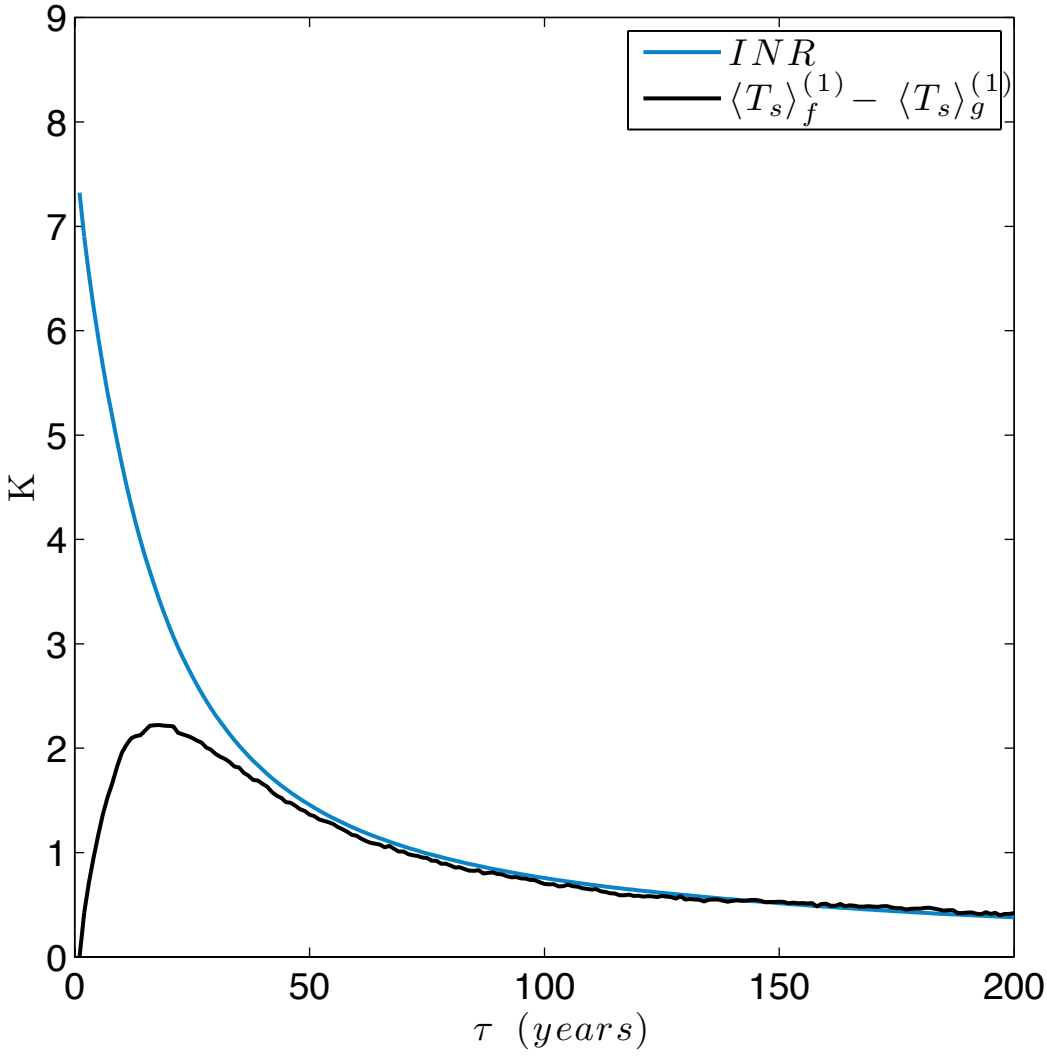
INR
$\langle T_s \rangle_f^{(1)} - \langle T_s \rangle_g^{(1)}$
K
τ $(years)$
0
1
2
3
4
5
6
7
8
9
0
50
100
150
200

Figure 4

Methods

Description of the model

The numerical model used in this study is PLASIM[28], a simplified GCM developed at the University of Hamburg. The dynamical core is based on the Portable University Model of the Atmosphere PUMA[39]. The primitive equations are solved by the spectral transform method. The model has a full set of physical parameterizations for unresolved processes. Parameterizations include long and shortwave radiation with interactive clouds, horizontal and vertical diffusion, boundary layer fluxes of latent and sensible heat. Stratiform precipitation is generated in supersaturated states, and the Kuo scheme is used for deep moist convection, while shallow cumulus convection is parameterized by means of vertical diffusion. See the Reference Manual freely available toghether with the code at http://www.mi.uni-hamburg.de/plasim for a detailed and referenced description of the parameterizations included in the model. The atmospheric model is coupled to a 1-layer slab model of the oceanic mixed layer with a depth of 50 m. For the present study we set the model at T21 horizontal resolution and 10 levels vertical resolution, with a time-step of 45 minutes. The diurnal and annual cycles, as well as the 11-years cycle in the ozone parameter have been removed, so that no time-dependent forcings are present in the control run.

Description of the experiments

We consider a control run of 2400 years. The [CO_2] is set to the value of 360 ppm, representative of the present-day value. We then perform two set of forcing experiments, prototypical of the scenarios proposed by the IPCC protocol for estimating the climate sensitivity and the impact of CO2 forcings on the CS. In the first set of experiments we double istantaneously the CO2 concentration and we keep it fixed at 720 ppm afterwards. This corresponds to a radiative forcing represented by an Heaviside function times a constant. This is the scenario in which one computes the *ECS*. In the second set of experiments [CO_2] is increased by 1% per year until reaching 720 ppm (after about 70 years), and it is kept fixed afterwards. This corresponds to a radiative forcing linearly increasing for the first 70 years until it has doubled, and constant afterwards. This is the

scenario in which one computes the TCR. In both experiments the forcing is applied homogeneously at each point of the model. In both cases we have performed an ensemble of 200 experiments starting from different initial conditions[21,22]. The 200 initial conditions are taken from the control run at intervals of 10 years starting from year 200, in order to guarantee their statistical independence and performing in this way a reasonable sampling of the attractor of the unperturbed system. Each run is 200 years long. For each set of forcing experiments we consider yearly averaged values of T_s as output. We compute the expectation values of the temperature difference $\langle T_s \rangle_f^{(1)}$ and $\langle T_s \rangle_{g_\tau}^{(1)}$ by averaging over all the ensemble members and subtracting the average value computed from the control run. The results for the two forced experiments are presented in blue in Figures 1-2. The shaded areas represent two standard deviations of the ensemble distribution, accounting for 95% of the ensemble variability.

Data analysis

The susceptibility $\chi_{T_S}^{(1)}(\omega)$ can be computed from the response of the surface temperature $\langle \widetilde{T_s} \rangle_f^{(1)}$ to a general modulation $\tilde{f}(\omega)$ by using Equation (2), and then $G_{T_S}^{(1)}(t)$ can be derived by taking its inverse Fourier transform. If the forcing is modulated by a Heaviside function like in the of instantaneous [CO_2] doubling, we have

$$G_{T_S}^{(1)}(t) = \frac{d}{dt} \langle T_s \rangle_f^{(1)}(t) \tag{8}$$

The susceptibility can then be computed taking the Fourier transform of the Green function.

All the computations have been performed in MATLAB® V. 7.9 environment using the standard Fast Fourier Tranform (fft) algorithm. When applying fft, we implictly force the input function to be periodic outside the time domain where it is defined. Applying this method to the time series of $\langle T_s \rangle_f^{(1)}$ introduces a frequency-dependent bias in the estimate of the spectrum due to the long-term behaviour of $\langle T_s \rangle_f^{(1)}$. We need to keep the information on such long-term behaviour, in order to estimate properly the low-frequency variability. It is possible to cure this issue by subtracting a suitably defined function before applying fft[40].

The KK relations have been computed by means of the singly subtractive KK method[26] (SSKK). Conventional KK relations stems from considering the real and imaginary parts of Equation 3 and making use of the simmetries of the susceptibility, obtaining

$$\Re\left\{\chi_{\Phi}^{(1)}(\omega)\right\} = \frac{2}{\pi} P \int_0^{+\infty} d\omega' \frac{\omega' \Im\left\{\chi_{\Phi}^{(1)}(\omega)\right\}}{\omega'^2 - \omega^2}$$
$$\Im\left\{\chi_{\Phi}^{(1)}(\omega)\right\} = -\frac{2\omega}{\pi} P \int_0^{+\infty} d\omega' \frac{\Re\left\{\chi_{\Phi}^{(1)}(\omega)\right\}}{\omega'^2 - \omega^2} \tag{9}$$

KK relations provide a self-consistency test for measured or observed data in many fields of physics, notably in optics[26]. The accuracy of the KK inversion depends on the quality of the observed data. The SSKK method consists in referring to an anchor point $\omega_1$where it is supposed that the direct estimate of $\chi_{\Phi}^{(1)}(\omega)$ from the data is reliable[26]. The SSKK relations then read

$$\Re\left\{\chi_{\Phi}^{(1)}(\omega)\right\} - \Re\left\{\chi_{\Phi}^{(1)}(\omega_1)\right\} = \frac{2(\omega^2 - \omega_1^2)}{\pi} P \int_0^{+\infty} d\omega' \frac{\omega' \Im\left\{\chi_{\Phi}^{(1)}(\omega)\right\}}{(\omega'^2 - \omega^2)(\omega'^2 - \omega_1^2)}$$
$$\omega^{-1}\Im\left\{\chi_{\Phi}^{(1)}(\omega)\right\} - \omega_1^{-1}\Im\left\{\chi_{\Phi}^{(1)}(\omega_1)\right\} = -\frac{2(\omega^2 - \omega_1^2)}{\pi} P \int_0^{+\infty} d\omega' \frac{\Re\left\{\chi_{\Phi}^{(1)}(\omega)\right\}}{(\omega'^2 - \omega^2)(\omega'^2 - \omega_1^2)} \tag{10}$$

With a careful choice of the anchor point the SSKK relations improve the accuracy of the KK analysis[26]. In our case we have taken $\omega_1 = 2\pi\xi_1$ with $\xi_1 = 0.1\ years^{-1}$, obtaining a rather good quality of the inversion. We remark anyway that the results are rather robust with respect to different choices of the anchor point ω_1.

Derivation of Equation (6)

The transient climate response is defined as the temperature increase at the instant of CO_2 doubling after a 1% increase per year. With the doubling time $\tau \approx 70$ years, $TCR = \langle T_s \rangle_{g_\tau}^{(1)}(\tau)$ when the $g_\tau(t)$ forcing is applied. Considering the Fourier representation of $\langle T_s \rangle_{g_\tau}^{(1)}(t)$ evaluated for $t = \tau$ and using Equation (2)

$$\langle T_s \rangle_{g_\tau}^{(1)}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} d\omega\, \langle \widetilde{T_s} \rangle_{g_\tau}^{(1)}(\tau) e^{-i\omega\tau} = \frac{1}{2\pi}\int_{-\infty}^{+\infty} d\omega\, \chi_{\mathbf{T_S}}^{(1)}(\omega)\tilde{g}_\tau(\omega) e^{-i\omega\tau} \tag{11}$$

The forcing is the sum of the ramp function up to time τ and the Heaviside function traslated by τ, therefore its Fourier transform is given by $\widetilde{g_\tau}(\omega) = f_{CO_2}^{2x} P\left(\pi\delta(\omega)e^{i\omega\tau} + i\,\mathrm{sinc}\,(\omega\tau/2)e^{i\omega\tau/2}/\omega\right)$. Therefore

$$\langle T_s\rangle_{g_\tau}^{(1)}(\tau) = \frac{1}{2} f_{CO_2}^{2x}\chi_{\mathbf{\Phi}}^{(1)}(0) - P\int_{-\infty}^{+\infty} d\omega\, f_{CO_2}^{2x}\chi_{\mathbf{\Phi}}^{(1)}(\omega)\frac{\mathrm{sinc}\,(\omega\tau/2)e^{-i\omega\tau/2}}{2\pi i\omega} \tag{12}$$

Using the Cauchy formula

$$\frac{1}{2} f_{CO_2}^{2x}\chi_{\mathbf{T_S}}^{(1)}(0) = P\int_{-\infty}^{+\infty} d\omega \frac{f_{CO_2}^{2x}\chi_{\mathbf{T_S}}^{(1)}(\omega)}{2\pi i\omega} \tag{13}$$

we derive

$$\langle T_s\rangle_{g_\tau}^{(1)}(\tau) = f_{CO_2}^{2x}\chi_{\mathbf{T_S}}^{(1)}(0) - P\int_{-\infty}^{+\infty} d\omega\, f_{CO_2}^{2x}\chi_{\mathbf{T_S}}^{(1)}(\omega)\frac{1+\mathrm{sinc}\,(\omega\tau/2)e^{-i\omega\tau/2}}{2\pi i\omega} \tag{15}$$

which gives Equation (6).

References

[1] IPCC, 2007: in Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden and C.E. Hanson, Eds., Cambridge University Press, Cambridge, UK, 976pp

[2] IPCC, 2007: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 996p.

[3] IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P. M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535pp.

[4] Knutti, R. & G. C. Hegerl, The equilibrium sensitivity of the Earth's temperature to radiation changes, *Nature Geoscience,* **1**, 735-743 (2008).

[5] Galloway, M. Heimann, C. Le Quéré, S. Levitus & V. Ramaswamy Climate sensitivity in the Anthropocene. *Q. J. R. Meteorol. Soc.*, **139**, 1121-1131 (2013).

[6] Otto, A. et al. Energy budget constraints on climate response. *Nature Geoscience* **6**, 415-416 (2013).

[7] Sherwood, S.C., S. Bony & J.-L. Dufresne. Spread in model climate sensitivity traced to atmospheric convective mixing. *Nature* **505**, 37-42, (2014).

[8] Shukla, J., Hagedorn, R., Hoskins, B., Kinter, J., Marotzke, J., Miller, M., Palmer, T., and Slingo J.: Revolution in climate prediction is both necessary and possible: A declaration at the world modelling summit for climate prediction, *B. Am. Meteor. Soc.*, **90**, 175–178 (2009).

[9] Chekroun, M. D., E. Simonnet, & M. Ghil, Stochastic climate dynamics: Random attractors and time-dependent invariant measures, *Physica D: Nonlinear Phenomena,* **240**, 1685–1700 (2011).

[10] Hasselmann, K., Sausen, R., Maier-Reimer, E. & Reinhard, V. On the cold start problem in transient simulations with coupled atmosphere ocean models. *Climate Dynamics* **9,** 53-61 (1993).

[11] Kubo, R.: The fluctuation-dissipation theorem, *Rep. Prog. Phys.*, **29**, 255-284, (1966).

[12] Langen, P. L., & V. A. Alexeev: Estimating 2 x CO2 warming in an aquaplanet

GCM using the fluctuation-dissipation theorem, *Geophysical Research Letters*, 32(23), L23,708 (2005).

[13] Gritsun, A., & G. Branstator, Climate response using a three-dimensional operator based on the FluctuationDissipation theorem, *Journal of the Atmospheric Sciences*, **64**, 2558–2575, (2007).

[14] Cooper, F. C., and P. H. Haynes, Climate sensitivity via a nonparametric fluctuation-dissipation theorem, *Journal of the Atmospheric Sciences*, **68**, 937–953 (2011).

[15] Cooper, F. C., J. G. Esler, and P. H. Haynes, Estimation of the local response to a forcing in a high dimensional system using the fluctuation-dissipation theorem, *Nonlinear Processes in Geophysics*, **20**, 239–248, (2013).

[16] Abramov, R. V., & A. Majda, New approximations and tests of linear fluctuation-response for chaotic nonlinear forced-dissipative dynamical systems, Journal of Nonlinear Science, **18**, 303–341, (2008)

[17] Lorenz, E., Forced and free variations of weather and climate, *J. Atmos. Sci*, **36**, 1367–1376 (1979).

[18] Ruelle, D., General linear response formula in statistical mechanics, and the fluctuation-dissipation theorem far from equilibrium, *Phys. Lett. A*, **245**, 220–224 (1998).

[19] Ruelle, D., Nonequilibrium statistical mechanics near equilibrium: computing higher-order terms, *Nonlinearity*, **11**, 5–18 (1998).

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

[21] Lucarini, V. and Sarno, S. A statistical mechanical approach for the computation of the climatic response to general forcings. *Nonlin. Processes Geophys.* **18,** 7–28 (2011)

[22] Lucarini, V.: Evidence of Dispersion Relations for the Nonlinear Response of Lorenz 63 System, *J. Stat. Phys.*, **134**, 381–400, (2009).

[23] Lucarini, V., & M. Colangeli, Beyond the linear fluctuation-dissipation theorem: the role of causality, *Journal of Statistical Mechanics: Theory and Experiment,* 2012(05), P05,013 (2012).

[24] Gallavotti, G.: Chaotic hypotesis: Onsanger reciprocity and fluctuation-dissipation theorem, *J. Stat. Phys.*, **84**, 899–926, (1996).

[25] Chekroun M.D., D. Kondrashov & M. Ghil: Predicting stochastic systems by noise sampling, and application to the El Nino-Southern Oscillation, *Proc. Natl. Acad. Sci USA,* 108,11766—11771 (2011).

[26] Lucarini, V., Saarinen, J. J., Peiponen, K.-E. & Vartiainen, E. M.: Kramers-Kronig relations in Optical Materials Research, Springer, Heidelberg, (2005).

[27] Lucarini, V.: Response theory for equilibrium and non-equilibium statistical mechanics: causality and generalized Kramers-Kronig relations, *J. Stat. Phys.*,**131**, 543–558, (2008).

[28] Fraedrich, K., Jansen, H., Luksch U. & Lunkeit, F. The planet simulator: Towards a user friendly model. *Meteorologische Zeitschrift* **14,** 299–304 (2005).

[29] Lunt, D.J, A.M. Haywood, G.A. Schmidt, U. Salzmann, P.J. Valdes & H.J. Dowset: Earth system sensitivity inferred from Pliocene modelling and data. *Nature Geoscience* **3**, 60 - 64 (2010)

[30] Previdi, M., B.G. Liepert, D. Peteet, J. Hansen, D.J. Beerling, A.J. Broccoli, S. Frolking, J.N. Galloway, M. Heimann, C. Le Quéré, S. Levitus, & V. Ramaswamy: Climate sensitivity in the Anthropocene. *Q. J. R. Meteorol. Soc.*, **139**, 1121-1131 (2013).

[31] Saltzman, B., Dynamical Paleoclimatology, Academic Press, New York, (2001).

[32] Winton, M., K. Takahashi & I.M. Held, Importance of Ocean Heat Uptake Efficacy to Transient Climate Change, *J. Climate*, **23**, 2333-2344, (2010).

[33] Colangeli M. & V. Lucarini, Elements of a unified framework for response formulae, J. *Stat. Mech.*, P01002, (2014).

[34] Allen, M. R. & Frame, D. J. Call off the quest. *Science,* **318**, 582–583 (2007).

[35] Stein U. & P. Alpert: Factor Separation in Numerical Simulations, *J. Atmos. Sci.* **50**, 2107-2115, (1993).

[36] Held I.M. & B.J. Soden: Robust Responses of the Hydrological Cycle to Global Warming, *J. Climate*, **19**, 5686–5699, (2006).

[37] Bracegirdle, Thomas J., & David B. Stephenson: On the Robustness of Emergent Constraints Used in Multimodel Climate Change Projections of Arctic Warming., *J. Climate*, **26**, 669–678, (2013).

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

[39] Fraedrich, K., Kirk, E. & Lunkeit, F. Puma: Portable university model of the atmosphere. *Technical Report 16, Deutsches Klimarechenzentrum* (1998)

[40] Nicolson, A. M. Forming the Fast Fourier Transform of a step response in time-domain metrology. *Electronics Letters* **9**, 317-318 (1973)