A mathematical approach to understanding emergent constraints

Femke J.M.M. Nijsse[1]†

[1]Institute for Marine and Atmospheric Research Utrecht, Utrecht University, Utrecht, the Netherlands

Emergent constraints are one of the tools to reduce uncertainty in climate model projections. These are physically explainable empirical relationships between characteristics of the current climate and long-term behavior that emerge in ensembles of climate models, where the long-term behavior is constrained using observations. So far, no general mathematical framework describing emergent constraints has been proposed. In this work, we introduce a classification for emergent constraints, depending on the process under consideration: we distinguish between emergent constraints that model variability, mean state or feedback strength. In addition we present a mathematical framework making use of linear response theory and apply this description to a set of conceptual (climate) models. We specifically focus on an emergent constraint that was previously found for the snow-albedo feedback. We conclude by discussing if and how this framework can be applied to GCMs.

1. Introduction

Decreasing the uncertainty in climate projections is one of the most important challenges in climate modeling. On the one hand, uncertainty can be reduced by producing more and more sophisticated numerical climate models. The slow progress in this regard, as illustrated by climate sensitivity which predicted range hasn't shrunk substantially over the last couple of decades, hints at the need for stronger methods to determine the accuracy of existing models. One of the proposed methods to accomplish this is the use of emergent constraints.

Emergent constraints are used to constrain future projections using current observations (Collins *et al.* 2012). In multimodel ensembles of complex climate models, a linear equilibrium relation can be found between a current aspect of the climate and some future aspect. More credibility is attached to models that match this aspect, be it the observed variability, feedback or mean state, well over the recent period. In this way, current observations provide a constraint to long term trends if the spread in observations in sufficiently small (Klein & Hall 2015). Some emergent constraints, however, may be spurious and could arise because of shared errors in a particular multimodel ensemble. In either way, the additional credibility in some models should not naively be interpreted as formal probabilities (Stephenson *et al.* 2012).

In recent years, emergent constraints have been found for a large set of climate variables: Arctic warming, snow-albedo feedback, tropical carbon uptake and global precipitation among others (Bracegirdle & Stephenson 2013; Hall & Qu 2006; Wenzel *et al.* 2014; Allen & Ingram 2002).

A dynamical mechanism for emergent constraints is still lacking. Under which circum-

† Email address for correspondence: f.j.m.m.nijsse@uu.nl

stances are they expected to arise? A mathematical framework might be used to identify the latter and give an indication as to where new emergent constraints might arise.

Ruelle's response theory (RRT), a generalization of linear response to high-dimensional chaotic dissipative dynamical systems, can be used to address the problem of predictability on different timescales (Ragone *et al.* 2016). With RRT, the response of nonequilibrium systems to external perturbations can be studied. Like the fluctuation-dissipation theorem (FDT), it uses the statistical properties of the unperturbed state only. In contrast with FDT, it does not assume that the unperturbed statistical steady-state is smooth. Recently, RRT has been proposed as a rigorous framework for computing the response of the climate system and its applicability has been tested on the Lorenz 96 model and on a simplified global climate model (Ragone *et al.* 2016; Lucarini & Sarno 2011).

In this work, we will investigate how and under what conditions emergent constraints appear and what this tells us about the physics of the climate system. As a starting point, we look at the emergent constraint found in Hall & Qu (2006) between the snow-albedo feedback on a seasonal timescale and the snow-albedo feedback over the course of the 22nd century on a warming earth in a CMIP4 ensemble. They proposed that the mechanism behind this emergent constraint is the maximum albedo snow can attain. With a high maximum albedo the contrast between snow-covered and snow-free areas is high, so that the snow-albedo feedback is higher both on a seasonal scale and on a century scale (Qu & Hall 2007).

To get an understanding of this emergent constraint we start by reproducing it in *PlaSim*, an intermediate complexity climate model. We proceed by formulating a dynamical framework in terms of susceptibilities, making use of linear response theory. Then, these are tested on an Ornstein-Uhlenbeck process in one and two dimensions. Subsequently, the theory is applied to two formulations of an energy balance model.

This paper is organized as follows: section 2 describes the snow-albedo feedback emergent constraint in *PlaSim*. Section 3 provides the basis for linear response theory. In section 4 a classification scheme for emergent constraints is proposed. Section 5 describes the applications of linear response theory to a set of toy (climate) models. Section 6 concludes with a summary and a discussion on how to generalize the method to global circulation models.

2. Snow-albedo feedback in *The Planet Simulator*

Hall & Qu (2006) found a correlation between SAF on a seasonal scale and SAF as a result of climate change. They define the SAF as:

$$\left(\frac{\partial Q}{\partial T_s}\right)_{SAF} = -Q_t \cdot \frac{\partial \alpha_p}{\partial \alpha_s} \cdot \frac{\Delta \alpha_s}{\Delta T_s}, \tag{2.1}$$

where Q_t is the incoming solar radiation as a function of time, α_p is the planetary albedo, α_s is the surface albedo and T_s the surface temperature. They show the last factor $\frac{\Delta \alpha_s}{\Delta T_s}$ shows the highest variability between models and use this for the emergent constraint. Qu & Hall (2007) provide a physical basis: they find that in models with a high snow albedo, the contrast between snow-covered and bare surfaces was largest and consequently the sensitivity to changes in temperature was largest.

The snow-albedo feedback emergent constraint can be reproduced in a multi-parameter ensemble in *PlaSim*. *PlaSim* is a numerical model of intermediate complexity, developed at the University of Hamburg to provide a fairly realistic present climate which can still be

simulated on a personal computer (Fraedrich *et al.* 2005). The atmospheric dynamics are modelled using the primitive equations formulated for temperature, vorticity, divergence and surface pressure. Moisture is included by transport of water vapor. The equations are solved using the spectral method. A full set of parameterizations is used for unresolved processed such as long and shortwave radiation with interactive clouds, boundary layer fluxes of latent and sensible heat and horizontal and vertical diffusion.

The atmospheric dynamics are coupled to a one-layer slab model of the ocean, which in our case has a parameterized horizontal diffusion for heat transport. The slab ocean model includes a thermodynamic sea ice module. For the present study uses a T21 horizontal resolution and 10 levels of vertical resolution, with a time step of 45 minutes. Daily and seasonal cycles are included in the simulation.

To model climate change, the historical forcing in *Plasim* was approximated by a CO_2 increase from 295 ppm at a rate of 0.3% per year in the 20th century and 1% per year in the 21st century before it stabilised at 720 ppm. A 50-year run up was used.

Snow albedo in this model is a function of surface temperature, snow depth and vegetation cover. The bare soil snow albedo A_{snow} in *PlaSim* is described by:

$$A_{snow} = \begin{cases} A_{max}, & \text{if } T_s \geqslant 10\,^{\circ}\text{C}. \\ A_{min} + (A_{max} - A_{min})\frac{T_s}{-10^{\circ}C} & \text{if } 0\,^{\circ}\text{C} < T_s < 10\,^{\circ}\text{C} \\ A_{min}, & \text{if } T_s > 0\,^{\circ}\text{C}. \end{cases} \tag{2.2}$$

This equation is modified in the presence of vegetation and in the case of shallow snow depth. See Lunkeit *et al.* (2011) for more details.

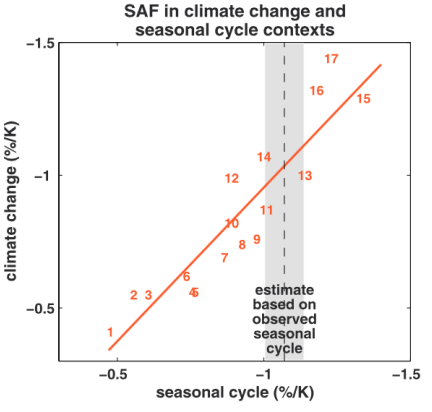

Figure 1: The emergent constraint on snow-albedo feedback $\frac{\Delta\alpha_s}{\Delta T_s}$ (from Hall & Qu (2006), α_s given in units of %)

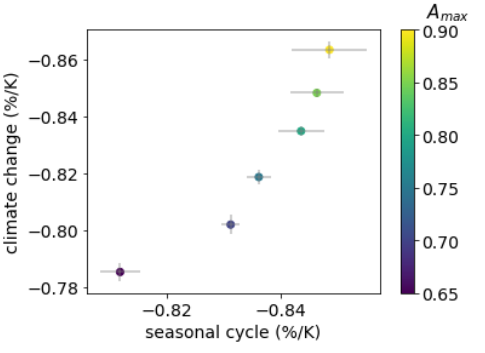

Figure 2: Same as figure 1, but now results from *PlaSim*

A set of simulations was performed with A_{max} varying between 0.650 and 0.900. In figure 1 and 2 the results from Hall & Qu (2006) and *PlaSim* are compared. Note that the variation of SAF in CMIP4 is significantly larger than the variation found in *PlaSim*, but that the *PlaSim* results do fit within the other models found by Hall & Qu (2006). Variations in other parameterizations, such as the maximum snow albedo over forested regions, increase the spread in *PlaSim* SAF further (not shown).

This simulation shows that the constraint that emerges in a multi-model ensemble with structurally different formulations of the snow response can to some extent also

be reproduced using variations in one parameter. This provides the justification for simplifying further to energy balance models to examine the snow-albedo feedback emergent constraint.

3. Linear response and spectral theory

This section summarizes the mathematical tools we will use to describe emergent constraints. First, the basic elements of linear response theory are recapitulated and rewritten with the help of a spectral decomposition of the operators involved in the description of stochastic differential equations describing gradient systems. This is ultimately expanded to a more general set of hypoelliptic and hypocoersive systems.

Let us start with an one-dimensional forced SDE whose drift term is described by a potential:

$$dX_t = (-V'(X_t) + F(t))dt + \sqrt{2\beta^{-1}}dW_t. \tag{3.1}$$

Here $V(x)$ is a confining potential, meaning that a equilibrium solution exists for the unforced system, and $F(t)$ is a prescribed forcing. Furthermore, β a diffusion term, often referred to as the inverse temperature. The associated Wiener process is indicated by W_t. When $V'(x) = -\gamma x$, the solution of the unforced problem is the well-known Ornstein-Uhlenbeck process.

3.1. *Linear response theory*

The probability density function, say $\bar{p}$, of the unforced system described by

$$dX_t = -V'(X_t)dt + \sqrt{2\beta^{-1}}dW_t \tag{3.2}$$

satisfies the Fokker-Planck equation:

$$\frac{\partial \bar{p}}{\partial t} = \frac{\partial (V'(x)\bar{p})}{\partial x} + \frac{\sigma^2}{2}\frac{\partial^2 \bar{p}}{\partial x^2} = \mathscr{L}^*(\bar{p}), \tag{3.3}$$

also defining the operator $\mathscr{L}^*$.

The first step in linear response theory is to determine the equilibrium distribution, here indicated by $\bar{p}_e$ of the unforced system. This is a distribution given by

$$\bar{p}_e(x) = \frac{1}{Z}e^{-\beta V(x)}, \qquad Z = \int_{\mathbb{R}} e^{-\beta V(x)}\ dx, \tag{3.4}$$

where Z is the partition function. Linear response theory provides an expression for the change in the expectation value of the change in an observable A, say $\Delta A(t)$ when the system is forced, compared to the unforced case, i.e.

$$\Delta A(t) = E[A(X_t)] - E[A_e(X_t)]. \tag{3.5}$$

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

$$\Delta A(t) = \int_0^t R_A(t-s)F(s)ds, \tag{3.6}$$

where $R_A(t)$ is the response function

$$R_A(t) = \int_{-\infty}^{\infty} A(x)\ e^{\mathscr{L}^* t}\left(-\frac{\partial \bar{p}_e}{\partial x}\right)dx. \tag{3.7}$$

When (3.6) is Fourier transformed to eliminate the convolution, we find

$$\mathcal{F}(\Delta A(t)) = \chi(\omega)\hat{F}(\omega), \tag{3.8}$$

where $\chi(\omega)$ is the susceptibility. If we take a cosine forcing, i.e.,

$$F(t) = F_0 \cos \omega t, \tag{3.9}$$

then

$$\hat{F}(\omega) = F_0\pi(\delta(\omega - \omega_0) + \delta(\omega + \omega_0)), \tag{3.10}$$

so once we know $\chi(\omega)$, we can determine the response $\Delta A(t)$ with equation 3.6.

3.2. *Explicit expression for $\chi(\omega)$ for $A(x) = x$.*

For $A(x) = x$, we find from (3.7) that

$$R_A(t) = \int_{-\infty}^{\infty} x\ e^{\mathscr{L}^* t} \left(-\frac{\partial \bar{p}_e}{\partial x} \right) dx. \tag{3.11}$$

By differentiating the expression for $\bar{p}_e$ in e, we find

$$-\frac{\partial \bar{p}_e}{\partial x} = \beta V'(x)\bar{p}_e \tag{3.12}$$

and hence (3.11) becomes

$$R_A(t) = \int_{-\infty}^{\infty} x\ e^{\mathscr{L}^* t} \left(\beta V'(x)\bar{p}_e \right)\ dx. \tag{3.13}$$

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

$$\mathscr{L} u = V'(x)\frac{\partial u}{\partial x} + \beta^{-1}\frac{\partial^2 u}{\partial x^2}, \tag{3.14}$$

where u is a test function. It is the operator appearing in the backward Kolmogorov equation. Using this property of adjointness in (3.13), we find

$$\langle x, e^{\mathscr{L}^* t}(V'(x)\bar{p}_e)\rangle = \langle e^{\mathscr{L} t} x, V'(x)\bar{p}_e\rangle \tag{3.15}$$

and hence

$$R_A(t) = \beta \int_{-\infty}^{\infty} e^{\mathscr{L} t}(x)\ V'(x)\bar{p}_e\ dx. \tag{3.16}$$

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

$$\langle g, h\rangle_{\bar{p}_e} = \int_{-\infty}^{\infty} gh\bar{p}_e\ dx. \tag{3.17}$$

As a subsequent step, let λ_l and ϕ_l be the eigenvalues and eigenfunctions of the generator $\mathscr{L}$ respectively, i.e. solutions of

$$\mathscr{L}\phi = -\lambda\phi. \tag{3.18}$$

For the OU process, these eigenvalues and eigenfunctions are given by

$$\lambda_l = \gamma l\ ;\ \phi_l(x) = \frac{1}{\sqrt{l!}} H_n(\sqrt{\gamma\beta}x), \tag{3.19}$$

where H_n are the Hermite polynomials.

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

$$\frac{\partial u}{\partial t} = \mathscr{L} u \tag{3.20}$$

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

$$u(x,t) = \sum_{l=1}^{\infty} \alpha_l \phi_l(x) e^{-\lambda_l t} \tag{3.21}$$

From the initial condition, we find

$$\sum_{l=1}^{\infty} \alpha_l \phi_l(x) = x \tag{3.22}$$

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

$$\alpha_l = \langle x, \phi_l \rangle_{\bar{p}_e}. \tag{3.23}$$

On the other hand, substituting the expression for u into (3.16) gives

$$\int_{-\infty}^{\infty} \sum_{l=1}^{\infty} \alpha_l \phi_l(x) e^{-\lambda_l t} V'(x) \bar{p}_e \, dx = \sum_{l=1}^{\infty} \beta_l e^{-\lambda_l t}, \tag{3.24}$$

where

$$\beta_l = \alpha_l \langle V'(x), \phi_l \rangle_{\bar{p}_e} = \langle x, \phi_l \rangle_{\bar{p}_e} \langle V'(x), \phi_l \rangle_{\bar{p}_e}. \tag{3.25}$$

Repeating the derivation with a general observable $A(x)$ gives $\langle A(x), \phi_l \rangle_{\bar{p}_e} \langle V'(x), \phi_l \rangle_{\bar{p}_e}$. The first term in β_l denotes the projection of the observable on the eigenfunctions and could intuitively be interpreted (for $l > 0$) as the amenability of the observable to change. The second projection term in β_l can be understood to be the amenability of the whole system to change under the influence of the forcing field. Finally, the response function can be written as

$$R_A(t) = \beta \sum_{l=1}^{\infty} \beta_l e^{-\lambda_l t} \tag{3.26}$$

and the susceptibility is given by

$$\chi(\omega) = \beta \sum_{l=1}^{\infty} \frac{\beta_l}{\lambda_l + i\omega}. \tag{3.27}$$

The total response to the forcing $F(t) = F_0 \cos \omega_0$ is now determined from (3.6) by

$$\Delta X(t)_{\omega_0} = 2F_0 \beta \int_{-\infty}^{t} \sum_{l=1}^{\infty} \beta_l e^{-\lambda_l (t-s)} \cos(\omega_0 s) \, ds \tag{3.28}$$

or in Fourier space and dropping the subscript

$$\mathcal{F}\left(\Delta X_t\right)(\omega_0) = 4F_0 \beta \sum_{l=1}^{\infty} \frac{\beta_l \lambda_l}{\lambda_l^2 + \omega_0^2}. \tag{3.29}$$

Somewhat sloppy, we will often refer to the last quantity as the susceptibility, but do keep in mind that it is actually the susceptibility times the Fourier transform of the forcing function given by (3.10).

3.3. *Higher-dimensional systems*

The previous analysis can be generalised to more dimensions. Take a vector $X_t = (x_t, y_t, ..)^T$. Then

$$dX_t = (-\nabla V(X_t) + F(t)\hat{x})dt + \sqrt{2\beta^{-1}}dW_t. \tag{3.30}$$

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

$$R_x(t) = \beta \sum_{l=1}^{\infty} g_l e^{-\lambda_l t}; \qquad R_y(t) = \beta \sum_{l=1}^{\infty} h_l e^{-\lambda_l t} \tag{3.31}$$

where $g_l = \langle x, \phi_l \rangle_{\bar{p}_e} \langle \frac{dV}{dx}, \phi_l \rangle_{\bar{p}_e}$ and h_l only differ in the first argument: $h_l = \langle y, \phi_l \rangle_{\bar{p}_e} \langle \frac{dV}{dx}, \phi_l \rangle_{\bar{p}_e}$. Calculating the response is analogous to the one-dimensional case, so that the Fourier transform of the response functions are given by

$$\mathcal{F}(\Delta x_t) = 2\beta \sum_{l=1}^{\infty} \frac{g_l \lambda_l}{\lambda_l^2 + \omega_0^2}; \qquad \mathcal{F}(\Delta y_t) = 2\beta \sum_{l=1}^{\infty} \frac{h_l \lambda_l}{\lambda_l^2 + \omega_0^2} \tag{3.32}$$

Note that for uncorrelated and equally large noise terms, the eigenfunctions are the tensor products of the eigenfunctions in the one-dimensional case, while the corresponding eigenvalues are the sum of the eigenvalues in the one-dimensional case. For Ornstein-Uhlenbeck the eigenfunctions $\phi_{l,m}$ are given by

$$\phi_{l,m} = \phi_l(x)\phi_m(y)$$

for ϕ_l as defined in Equation 3.19. These form an orthonormal complete basis.

3.4. *Irreversible and hypoelliptic systems*

Not all climate processes for which emergent constraints have been found can be simplified to differential equations with a gradient structure. In this subsection, we provide an generalization to irreversible processes that are not necessarily elliptic, but instead fall into the category of hypoelliptic systems. Irreversible processes are ubiquitous in climate: entropy is produced continuously in the presence of dissipation (Lucarini *et al.* 2010). For this group of systems the susceptibility can also be expressed in the form of equation 3.29. An example is the Langevin equation. In the Langevin equation the noise only works on the momentum variables (p).

$$\begin{aligned} dq_t &= p_t dt, \\ dp_t &= -\nabla_q V(q_t)dt - \gamma p_t dt + \sqrt{2\gamma\beta^{-1}}dW_t. \end{aligned} \tag{3.33}$$

To define a unique invariant distribution $\bar{\rho}_e$ (where ρ is used to avoid confusion with the momentum) there should be sufficient interaction between the variables on which noise works and variables on which it does not work. This condition is called hypoellipticity (see Pavliotis (2014) for a formal definition). Observe that for systems without noise, the Fokker-Planck equation reduces to the Liouville equation for which no unique invariant measure exists. Furthermore, systems should converge exponentially fast to equilibrium, for instance by meeting the property of hypocoercivity (Pavliotis 2014).

In the case of the Langevin equation the invariant distribution can be expressed again as an exponential:

$$\bar{\rho}_{e(p,q)} = \frac{1}{Z} e^{-\beta \mathcal{H}(p,q)}, \tag{3.34}$$

where Z is the partition function as in (3.4) and the Hamiltonian $\mathcal{H}(p,q) = \frac{1}{2}|p|^2 + V(q)$.

Due to the loss of reversibility, the generator is not self-adjoint in $L^2(\bar{\rho}_e)$ and we call its adjoint L_{kin}. The adjoint L_{kin} can be seen as the generator for the time-reversed process. As a consequence the eigenfunctions ϕ_n do not form a complete orthogonal set. Instead, the eigenfunctions satisfy the biorthonormality relation with the eigenfunctions ψ_n of L_{kin}:

$$\langle \phi_n, \psi_m \rangle_{\bar{\rho}_e} = \delta_{nm}. \tag{3.35}$$

Just as in equation 3.21, the solution u of $\frac{\partial u}{\partial t} = \mathscr{L}u$ can be expanded in eigenfunctions. The eigenvalues do not have to be real, however, since the generator is not self-adjoint or equivalently since the process is irreversible. From the initial condition and using the biorthonormality under the inner product $\langle , \rangle_{\bar{\rho}_e}$, we now find

$$\alpha_l = \langle x, \psi_l \rangle_{\bar{\rho}_e}. \tag{3.36}$$

The second factor in the definition of β_l, the inner product between the eigenvalues and the derivative of the potential in the direction of the forcing, in (3.25) should be reconsidered as well. The derivative of the potential was obtained from the derivation of the equilibrium distribution and should thus be replaced by the derivative of the Hamiltonian for irreversible systems.

$$\beta_l = \langle x, \psi_l \rangle_{\bar{\rho}_e} \langle \phi_l, \mathcal{H}'(x) \rangle_{\bar{\rho}_e}. \tag{3.37}$$

In general the Hamiltonian can be derived from the invariant measure: $\mathcal{H} = -\ln(\bar{\rho}_e)$, where the noise term is now included in the Hamiltonian. Redoing (3.26—3.29) for complex eigenvalues, gives us the following expression for the 'susceptibility':

$$\mathcal{F}\left(\Delta X_t\right)(\omega_0) = 4F_0\beta \sum_{l=1}^{\infty} \frac{\beta_l \Re(\lambda_l)}{\Re(\lambda_l)^2 + (\omega_0^2 - \Im(\lambda_l))}. \tag{3.38}$$

When considering this equation as a function of ω_0, the real part of the eigenvalue $\Re(\lambda_l)$ gives the width of the peak, while the imaginary part of the eigenvalue $\Im(\lambda_l)$ indicates the peak location.

4. Classification

Although a wide set of different emergent constraints have been found, we are not aware of any attempts to classify them. Here, a classification is proposed based on a characterization of the predictor and of the relationship between the predictor and the predictand. Using this classification, assessment of their validity and applicability should become easier. Furthermore, a classification is a prerequisite for a dynamical description of emergent constraints.

In table 1, three types of emergent constraints are described. Type I encompasses emergent constraints where the (short-term) variability in one parameter is linked to the future predictand. This variability can be daily, seasonally or time scales of ENSO. Type II refers to emergent constraints where the mean state of a certain variable is linked to the predictand. In other words, not the accuracy of a response to a forcing is used as a constraint, but the error in the mean state. The final category III contains relationships between the strength of a feedback on different time scales. The feedback can only be measured as a consequence of the forcing, but is not a direct response to the forcing. These three categories are further subdivided depending on whether the predictor and predictand describe the same (-s) observable or process or different ones (-d).

	Predictor, predictand the same	**Predictor, predictand different**
Variability	I-s	I-d
Mean state	II-s	II-d
Feedback	III-s	III-d

Table 1: Different types of emergent constraints can be classified. type I constraints link the short-time variability in a certain variable to a future predictand, while type II constraints link (the error) in the mean state of a variable to a change over a longer period. The final category links a short-time feedback mechanism to future feedbacks.

Emergent constraints of type I-s and III-s are the most intuitive. As long as the variations in the predictor are of a sufficient amplitude compared to the size of the predictand, a correlation between the predictor and predictand automatically points towards a common physical basis, namely a common dynamical response to an external forcing with sufficiently short time scales.

Emergent constraints of type I-d or III-d have not been described in CMIP5 models so far. We see no particular reason for these types of emergent constraints not to exist, so it might be interesting to also start looking for these kind of emergent constraints. In table 2 we apply our classification to emergent constraints found in literature.

4.1. *Conditions for emergent constraints*

For simplicity, the conditions are all given for elliptic gradient systems. The following equations can easily be extended for systems that are both hypoelliptic and irreversible.

4.1.1. *Type I-s*

Referring back to the theory described in section 3, conditions for emergent constraints can be formulated. For a type I emergent constraint in a system with varying parameters, the ratio of the responses to the frequencies ω_1 and ω_2 should be constant over the ensemble members e_i (recall that emergent constraints are here defined to be *linear* relationships), hence we find the condition from the ratio of the susceptibilities SR

$$\mathrm{SR}(e) = \frac{\mathcal{F}(\Delta A(t))(\omega_2)}{\mathcal{F}(\Delta A(t))(\omega_1)} = \frac{\sum_{l=1}^{\infty} \frac{\beta_l \lambda_l}{\lambda_l^2+\omega_2^2}}{\sum_{l=1}^{\infty} \frac{\beta_l \lambda_l}{\lambda_l^2+\omega_1^2}} = C, \tag{4.1}$$

where C is independent of e or the parameter(s) generating the ensemble.

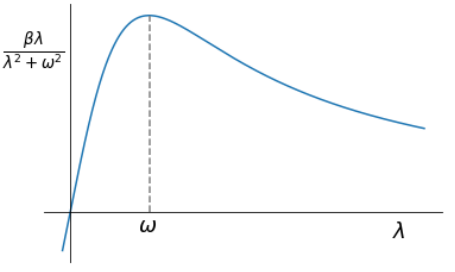

Figure 3: A single term in the infinite sum of Equation 4.1, as function of lambda

Physically, we expect that the same mechanisms to be responsible for the response at a short and fast time scale to obtain an emergent constraint of type I-s. Mathematically, this translates to the expectation that the system should have response times smaller than the timescale of the forcing or equivalently: the generator should have eigenvalues λ larger than the frequency of the forcing. Naturally, the response times $1/\lambda$ of the

Reference	Climate predictor	Future climate predictand	Type
Knutti *et al.* (2006)	Seasonal cycle land temperature amplitude	Climate sensitivity	II-s
Boe *et al.* (2009)	Arctic sea ice extent trend 1979-2007	Arctic sea ice extent	I-s
Clement *et al.* (2009)	Sensitivity LLC to pacific decal variability	Sign LLC feedback	I-s
Fasullo & Trenberth (2012)	Mid-tropospheric RH over ocean in subsidence region	ECS	II-d
Bracegirdle & Stephenson (2013)	Arctic SAT	Arctic SAT under climate warming	II-s
Gordon & Klein (2014)	Sensitivity of extra-tropical LLC optical depth to temperature	Extra-tropical LLC optical depth response to climate warming.	I-s
Qu *et al.* (2014)	Sensitivity of LLC cover to SST	LCC cover changes under climate warming	I-s
Hall & Qu (2006); Qu & Hall (2014)	Springtime SAF	SAF under climate warming	III-s
Sherwood *et al.* (2014)	Strength cloud-scale and large-scale lower tropospheric mixing over oceans	ECS	II-d
Su *et al.* (2014)	RH & cloud fraction tropics	ECS	II-d
Wenzel *et al.* (2014)	Short-term sensitivity of atmospheric carbon dioxide	Sensitivity tropical land carbon storage to climate warming	I-s
Tian (2015)	Precipitation & mid-tropospheric RH asymmetry bias (for ITCZ)	ECS	II-d
Trenberth & Fasullo (2010)	SH net radiation TOA	ECS	II-d

Table 2: A classification of emergent constraints found in literature. Abbreviations stand for RH: relative humidity, ITCZ: inter-tropical convergence zone, TOA: top of atmosphere, SH: southern hemisphere, ECS: equilibrium climate sensitivity, LLC: low-level cloud, SAF: snow-albedo feedback. SAT: surface air temperature. The last entry is an example of a relationship that appears to have been an coincidence: no physical mechanism was proposed and it did not appear in different ensembles, such as CMIP5 Grise *et al.* (2015))

dominant processes are expected to be at least smaller than the timescale of the slow forcing $1/\omega_1$.

This reasoning can be found back in the properties of equation 4.1. To better understand it, we examine a single term in the sum that makes up the susceptibility. Figure 3 shows such a term as a function of an eigenvalue and for an arbitrary β_l. The maximum is reached when $\lambda = \omega$. If the projection terms β_l stay constant over an ensemble, and indeed $\omega_1 < \lambda_l$ for all l, the susceptibility to low frequencies decreases over an ensemble when eigenvalues increase. The ratio can then only stay constant over the ensemble if the susceptibility at high frequency also decreases; so to fulfil the condition in the case that β_l stays (relatively) constant over the ensemble, the dominant projection terms should correspond to eigenvalues that are larger than ω_2.

For the Ornstein-Uhlenbeck process, using equation 3.19, the ratio of susceptibilities

reduces to

$$SR(\gamma) = \frac{\beta_1\lambda_1}{\lambda_1^2+\omega_2^2} \Big/ \frac{\beta_1\lambda_1}{\lambda_1^2+\omega_1^2} = \frac{\gamma^2+\omega_1^2}{\gamma^2+\omega_2^2} \neq C, \tag{4.2}$$

since both the observable x and the derivative of the potential are orthogonal to all eigenfunctions other than ϕ_1. This ratio is dependent on γ. Although linear emergent relationships are only found in a few uninteresting limiting cases — either $\gamma \gg \omega_i$ or $\gamma \ll \omega_i$ for $i \in \{1, 2\}$ — the two susceptibilities correlate positively which we call a nonlinear or weak emergent relationship.

4.1.2. *Type I-d*

In this subsection, the analysis above is repeated in the case the observable reacting to a high-frequency forcing is unequal to the observable reacting to a low-frequency forcing. Let us add a forcing in the x-direction and take as observables x and y. Mutatis mutandis, a condition very similar to 4.1 is found.

$$\frac{\mathcal{F}(\Delta A_1(t))(\omega_2)}{\mathcal{F}(\Delta A_2(t))(\omega_1)} = \frac{\sum_{l=1}^{\infty} \frac{g_l\lambda_l}{\lambda_l^2+\omega_2^2}}{\sum_{l=1}^{\infty} \frac{h_l\lambda_l}{\lambda_l^2+\omega_1^2}} = C, \tag{4.3}$$

where g_l and h_l are defined as in equation 3.31.

It is not possible to find an analytic solution in terms a coupled Ornstein-Uhlenbeck system. When coupling is introduced in the drift, but not in the noise, it is not possible to write the 2D-eigenfunctions as tensor products of the lower-dimension eigenfunctions; the problem is not separable.

For type I-s we found a weak restriction on the values of the eigenvalues: the eigenvalues of the dominant terms in the susceptibility should be larger than ω_2, the frequency of the fast forcing term. In the case of a type I-d emergent constraint the projection terms corresponding to the same eigenvalue in the numerator and denominator have different values, so it might be more difficult to identify dominant terms. However, the same logic is applicable and also to meet the condition of type I-d, the eigenvalues of dominant β_l are expected to be larger than ω_2 if the projection terms are relatively constant over the ensemble.

The response in one direction might have a different sign compared to the response in the other direction. Therefore, C can either be positive or negative, in contrast to the constant in the condition for type I-s. The sign of the constant should generally be determinable from physical arguments.

4.1.3. *Type II*

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

For emergent constraints of type II-s, the linear relationship between the response and the mean state is not expected to pass through the origin, since the mean will in general be nonzero. Therefore, an additional term I is added to the ratio, denoting the intercept of the line between the mean state and the response with varying parameters. Instead, the susceptibility is compared to the mean state and the following condition is derived, where C should again be a constant independent of parameter(s) that is used to generate

the ensemble:

$$\frac{E[A_t] - I}{\mathcal{F}(\Delta A(t))} = \frac{\int_{-\infty}^{\infty} \bar{p}_e A(x)\ dx - I}{\sum_{l=1}^{\infty} \frac{h_l \lambda_l}{\lambda_l^2 + \omega_1^2}} = C. \tag{4.4}$$

Again C can either be positive or negative, depending on the physics under consideration.

4.1.4. *Type III*

In general, a feedback is the process in which changing one quantity changes a second quantity and the change in the second quantity in turn changes the first. These changes can be represented by derivatives, so that feedback strength generally scales with a derivative of the secondary quantity to the first quantity in the linear case. See for instance the snow-albedo feedback (equation 2.1), which scales with $\frac{d\alpha}{dT}$. At first glance, one would gather that you can proceed as with Type I-s and use the derivative $\frac{d\alpha}{dT}$ as an observable. Note however that linear response theory does not give the expectation value of the observable, but the expectation value of the deviation due to the forcing. In the case of a feedback, the susceptibility would not give the strength of the feedback, but the sensitivity of that feedback to the forcing.

So instead, the feedback strength can be described by a ratio of susceptibilities of the two observables under consideration. It is possible that the susceptibilities of the two observables do not have the same phase; this phase difference can be used as a physical check. In the case of the snow-albedo feedback the temperature is expected to start responding first to a radiative forcing. As with the Type I emergent constraints, the ratio of feedback strength RFS should be constant over the ensemble. Let us take the derivative $\frac{dx}{dy}$ to scale with the feedback strength, then we find the following condition:

$$\mathrm{RFS} = \frac{\mathcal{F}(\Delta x(t))(\omega_2)/\mathcal{F}(\Delta y(t))(\omega_2)}{\mathcal{F}(\Delta x(t))(\omega_1)/\mathcal{F}(\Delta y(t))(\omega_1)} = C \tag{4.5}$$

where the C should be independent of the ensemble member.

For type I and type III there is one assumption each that should still be made explicit. In posing the conditions, it is assumed for type I that *all* variability on a seasonal scale has the same physical mechanism as the change on the long term and for type III that the complete feedback is present on both scales. Consequently the graph of the susceptibility as a response to a high-frequency forcing and the susceptibility to a low-frequency forcing passes through the origin. An additional intercept term I just as in equation 4.4 can be added to the condition whenever this assumption is not valid.

5. Applications

In this section we test the conditions put forward in section 4 on a set of simple dynamical models. Firstly, the OU process is examined in 1D and 2D. This provides information on type I-s and I-d emergent constraints respectively. We continue by examining the snow-albedo feedback emergent constraint in two different formulations of the energy balance model, as a model of type III emergent constraints.

In terms of numerics, the eigenvalues and eigenfunctions of the generator were indirectly determined using the fact that the eigenvalues of the Fokker-Planck operator $\mathscr{L}^*$ are equal to the eigenvalues of the generator and that the eigenfunctions are computed from the transformation:

$$\phi_{\mathscr{L}} = \bar{p}_e^{-1} \phi_{\mathscr{L}^*} \tag{5.1}$$

The Fokker-Planck operator was discretized with use of Chang-Cooper algorithm (Chang & Cooper 1970). Eigenvalues and eigenvectors were determined using an Implicitly Restarted Arnoldi Method (Lehoucq *et al.* 1998). Explicit calculations of the trajectories of the SDEs were performed the Runge-Kutta method for SDEs.

5.1. *Ornstein-Uhlenbeck*

First, the framework and code are tested with a simulation of the one-dimensional Ornstein-Uhlenbeck (OU) process for which an analytic solution of its response exists. Forcing $F(t) = \sin 2\pi t\omega_i$ with two different frequencies $\omega_1 = 0.001$ and $\omega_2 = 0.1$ are added in two simulations, which both end after $1/4$ of the phase of the slow forcing. The computed eigenvalues are in agreement with equation 3.19: the system responds faster for higher contraction rates γ. The projection terms β_l are invariant under a change in γ.

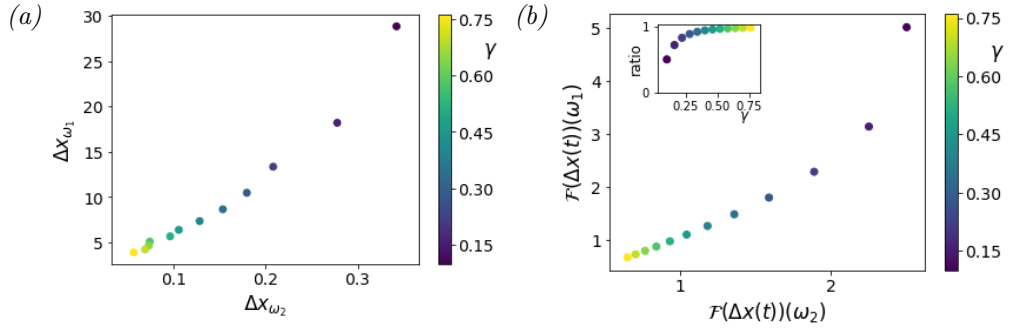

Figure 4: *(a)* Response to forcings at two different frequencies of the 1D Ornstein-Uhlenbeck process. Shown is the average of a 3000-member simulation of trajectories *(b)* The susceptibility at these frequencies, whose ratio is given in the inset figure.

To examine the correspondence between the mean of the trajectories of the SDE, averaged over 3000 trajectories, and the separately calculated susceptibility, both are shown in figure 4. The calculated susceptibilities give a good indication of the actual response. No linear emergent relationship is found for OU. For decreasing γ the susceptibility at low frequency increases faster than the susceptibility at high frequency. This is in agreement the analytic solution in equation 4.2.

In the 2D system the forcing $F_i(t) = \sin 2\pi\omega_i t$ with the same frequencies as in the 1D case is added in the first dimension only. The sensitivity to this forcing in the second dimension is examined. The process is given by:

$$dX_t = \left[\begin{pmatrix} -\gamma_1 & \delta \\ \delta & -\gamma_2 \end{pmatrix} X_t + \begin{pmatrix} F_i(t) \\ 0 \end{pmatrix} \right] dt + \sqrt{2\beta^{-1}} dW_t. \tag{5.2}$$

An ensemble is generated by changing the contraction rate γ_1. Two ensembles are compared: first with the cross term $\delta = 0.2$, while in the second case a stronger coupling is used: $\delta = 0.5$. The contraction term in the second dimension is held constant at $\gamma_2 = 0.7$.

Figure 5 shows the eigenvalues and susceptibility ratios for the 2D-simulations. In the case of a weak coupling, $\delta = 0.2$, all nonzero eigenvalues are larger than the forcing frequency ω_2, and therefore naturally larger than ω_1. On the other hand, the stronger coupling $\delta = 0.5$ leads to a slowing down of the system, so that some eigenvalues are now smaller than ω_2. In these cases ($\gamma_1 < 0.5$) the system does not have time to portray the full response to a forcing, while for others ($\gamma_1 > 0.5$) it does. Consequently, the strength

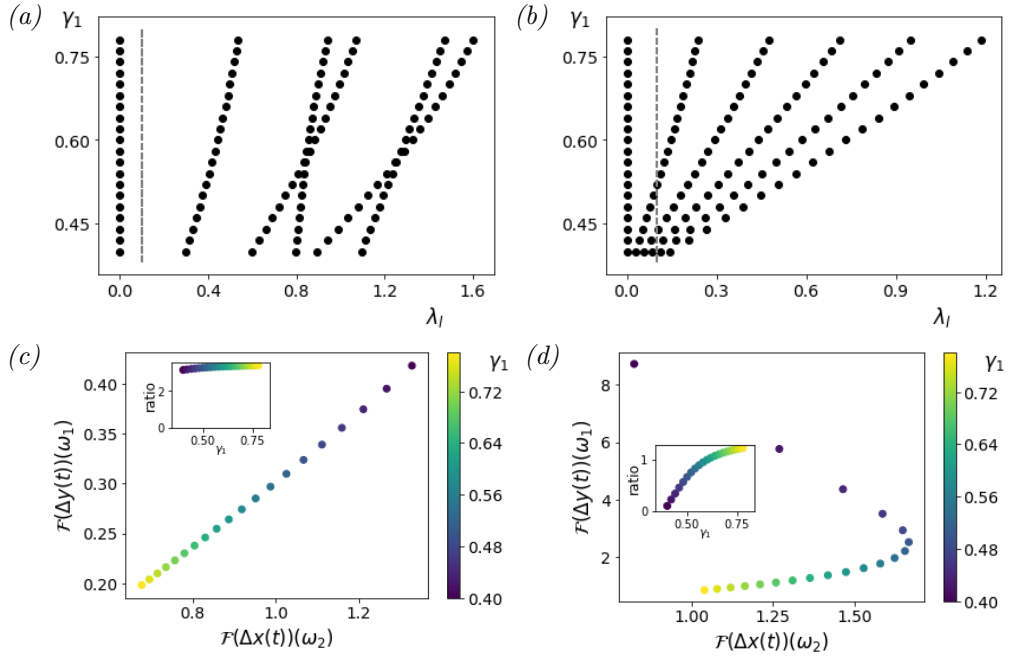

Figure 5: Eigenvalue spectrum for *(a)* $\delta = 0.2$ and *(b)* $\delta = 0.5$. The dashed line corresponds to the frequency ω_2 of the fast forcing *(c,d)* Corresponding susceptibilities, with their ratio in the inset figures.

of the response actually decreases for $\gamma_1 < 0.5$. Directly calculating the expectation value as the mean of 2000 stochastic trajectories confirms this image.

The projections are shown in figure 6. Although the variations in the projection terms are significant, they appear to compensate each other and the analysis based on the eigenvalues seems to be sufficient to explain the responses. In the case of strong coupling, the projections and thus the eigenfunctions, vary less.

In the low-coupling system, the susceptibility ratio is almost constant and an emergent linear relationship is found. This example stresses that the size of the the response, which is substantially larger in the high-coupling case, does not indicate whether an emergent relationship exists. Of course, the size of the spread of the response should be large enough so that it can be constrained by observations.

5.2. *Energy Balance model*

In this subsection we turn to the snow-albedo feedback again. one emergent constraint is examined in more detail, namely the one pertaining to the snow-albedo feedback (SAF) first described by Hall & Qu (2006). This emergent constraint falls in category III-s.

To study this emergent constraint we modify a simple energy balance model and make the albedo temperature-dependent. We change a parameter in the albedo function to obtain an ensemble of runs.

With constant albedo, the energy balance model reads:

$$\frac{dT}{dt} = \frac{1}{c_T}\left(Q(1-\alpha) + A\ln\frac{C}{C_{ref}} + G - \epsilon\sigma T^4\right), \tag{5.3}$$

where dT is the temperature change, c_T the heat capacity of the earth, Q the solar

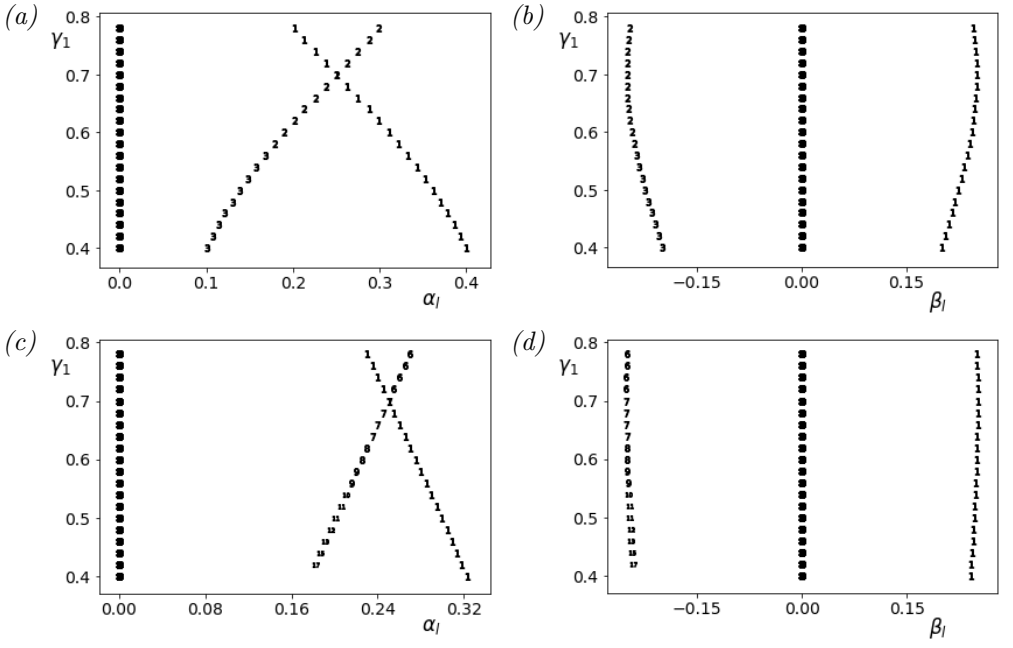

Figure 6: *(a,b)* Projections α_l and β_l respectively for a weakly coupled OU 2D system with $\delta = 0.2$, *(c,d)* same for $\delta = 0.5$.

insolation, α the albedo, C the concentration of greenhouse gases, C_{ref} a reference concentration, G a constant for the background greenhouse gas forcing, σ the Stephan-Boltzmann constant, ϵ the earth emissivity. Their numerical values are found in table 3. Temperature T can be calculated analytically for an equilibrium state, by setting $dT = 0$. To include a seasonal variation in the model, we restrict the attention to a single hemisphere and assume no transport of heat between the two hemispheres.

Before examining the snow-albedo feedback note that for some variables, notably the climate sensitivity, a simple EBM can show different sensitivities to forcing from solar insolation or greenhouse gases. For simplicity we take $H = G + A \ln C / \ln C_{ref}$ and $\epsilon = 1$.

$$\frac{\partial}{\partial \alpha}\frac{\partial T}{\partial Q} = (Q(1-\alpha)+H)^{-3/4}\,\sigma^{1/4}\left(-\frac{1}{4}+\frac{3}{4}\frac{Q(1-\alpha)}{Q(1-\alpha)+H}\right) \qquad < 0 \qquad (5.4a)$$

$$\frac{\partial}{\partial \alpha}\frac{\partial T}{\partial H} = \frac{3}{16}Q\sigma^{1/4}\left(Q(1-\alpha)+H\right)^{-7/4} \qquad > 0 \qquad (5.4b)$$

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

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

$$\alpha_r(T) = \alpha_{min} + \frac{\alpha_{amp}}{1+\exp k(T-T_h)} \qquad (5.5)$$

where α_{min} is the minimum the albedo takes, α_{amp} is the amplitude, k is a steepness factor and T_h is the temperature at which half of the amplitude is reached. The amplitude is the parameter that is varied to generate the ensemble. Of course, in this simple

Constant	Value	Constant	Value
c_T	5.0×10^8 J/m^2/K	ϵ	1.0
A	20.5 W/m^2	σ	5.67×10^{-8} W/m^2/K^4
Q_0	342 W/m^2	α_{min}	0.2
Q_s	115 W/m^2	α_{amp}	0.05–0.5
G	150 W/m^2	k	0.5
C_{ref}	280 ppmv	T_h	284 K
c_{snow}	4.0×10^6 s	β_T	2.0×10^7 s/K^2
β_α	2.0×10^5 s		

Table 3: Constants for the energy balance model.

formulation the ice-albedo feedback and snow-albedo feedback amount to the same. For simplicity the two feedbacks together are henceforth referred to as the snow-albedo feedback.

The seasonal snow-albedo feedback is computed in a simulation where Q_0 is modulated by adding a sine with amplitude Q_s and a period of one year. The snow-albedo feedback is then computed by dividing the amplitude of the albedo cycle by the amplitude of the temperature cycle. A second simulation is performed in which C is increased exponentially at a rate of 0.3% from 280 to 720 ppmv. Here the snow-albedo feedback is computed by dividing the total albedo response by the total temperature response.

The parameter values for this simulation can be found in table 3. The parameters of the albedo function are chosen to ensure that no bistability is present in the model, in which case linear response theory would break down.

In the framework of linear response theory, two different forcing terms are used. For the greenhouse gas forcing, the forcing field added to equation 5.3 is simply a constant, features as a constant, so that the forcing is described by $F_1(t) = F(t)$ for $F(t)$ as defined in equation **??**. In the case of a modulation in the insolation, the drift term is changed differently; namely by adding $F_2(t) = (1-a(T))F(t)$. Adding these terms to equation 5.3 gives:

$$dT = \frac{1}{c_T}\left(Q(1-\alpha_r(T)) + A\ln\frac{C}{C_{ref}} + F_1(t) + G - \epsilon\sigma T^4\right)dt + \sqrt{2\beta_T^{-1}}dW_t, \qquad (5.6a)$$

$$dT = \frac{1}{c_T}\left((Q+F_2(t))(1-\alpha_r(T)) + A\ln\frac{C}{C_{ref}} + G - \epsilon\sigma T^4\right)dt + \sqrt{2\beta_T^{-1}}dW_t. \qquad (5.6b)$$

The noise terms represent the effect of unmodelled variables, such as the heat transport between the two hemispheres.

5.2.1. *Formulation of condition for EBM*

As a type III emergent constraint, the SAF emergent constraint is described by two observables: the albedo and the temperature which feature in the equation for the snow-albedo feedback (equation 2.1) as $\frac{d\alpha}{dT}$.

$$\begin{aligned}\mathrm{RFS}(\alpha_{amp}) &= \frac{\mathcal{F}(\Delta a(t)|_Q)(\omega_2)/\mathcal{F}(\Delta T(t)|_Q)(\omega_2)}{\mathcal{F}(\Delta a(t)|_C)(\omega_1)/\mathcal{F}(\Delta T(t)|_C)(\omega_1)} \\ &= \frac{\sum_{l=1}^{\infty}\frac{\alpha_l\lambda_l}{\lambda_l^2+\omega_2^2}/\sum_{l=1}^{\infty}\frac{\beta_l\lambda_l}{\lambda_l^2+\omega_2^2}}{\sum_{l=1}^{\infty}\frac{\gamma_l\lambda_l}{\lambda_l^2+\omega_1^2}/\sum_{l=1}^{\infty}\frac{\delta_l\lambda_l}{\lambda_l^2+\omega_2^2}} = C,\end{aligned} \tag{5.7}$$

where

$$\begin{aligned}\alpha_l &= \langle\alpha,\phi_l\rangle_{\bar{p}_e}\langle(1-\alpha(T))V'(T),\phi_l\rangle_{\bar{p}_e}, \\ \beta_l &= \langle T,\phi_l\rangle_{\bar{p}_e}\langle(1-\alpha(T))V'(T),\phi_l\rangle_{\bar{p}_e}, \\ \gamma_l &= \langle\alpha,\phi_l\rangle_{\bar{p}_e}\langle V'(T),\phi_l\rangle_{\bar{p}_e}, \\ \delta_l &= \langle T,\phi_l\rangle_{\bar{p}_e}\langle V'(T),\phi_l\rangle_{\bar{p}_e}.\end{aligned} \tag{5.8}$$

With only one dominant term in all sums, this reduces to $C = \frac{\alpha_l\delta_l}{\beta_l\gamma_l} = 1$ for any l.

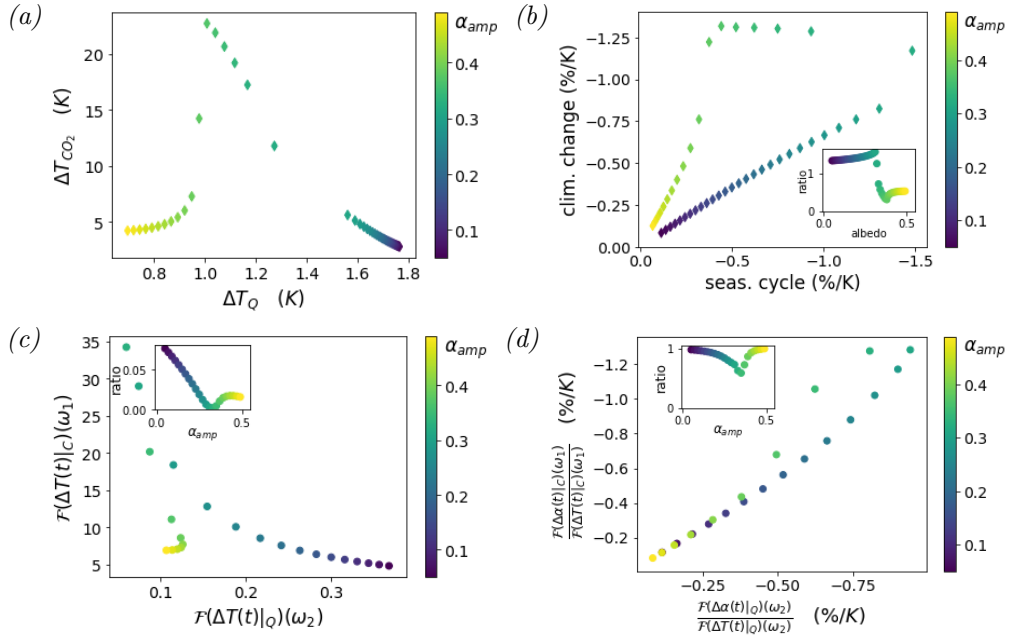

Figure 7: *(a)* The relation between temperature response to the seasonal cycle and the temperature response to greenhouse gas forcing *(b)* The strength of the snow-albedo feedback to solar and greenhouse gas forcing on different time scales. In the inset: their ratio as a function of α_{max}. For clarity, *(a,b)* are shown without noise. *(c)* The susceptibilities for temperature as the observable *(d)* The ratio of albedo and temperature susceptibilities and their ratio (RFS).

5.2.2. *Results*

In figure 7a the sensitivity of temperature to varying α_{amp} is shown. No emergent relationship is found for climate sensitivity, consistent with the prediction in the case of constant albedo. In figure 7c the susceptibilities for the temperature observable are shown. Although the match to the actual sensitivity in figure 7a is far from perfect, the major features are represented: the peak α_{amp} and the climate sensitivity. The seasonal

sensitivity is represented less well. This poor correspondence can possibly be explained by the fact that this system is not far from bistability, where linear response theory breaks down. This is supported by the fact that the response to an unrealistically small seasonal cycle and greenhouse gas forcing are represented well.

The emergent relationship of SAF is shown in figure 7b. In the warm regime (low albedo, lower 'line' in the figure), the SAF strength becomes larger for larger α_{amp}. The higher the maximum albedo, the steeper the logistic albedo function $\alpha_r(T)$. A second effect also takes place: with higher maximum albedo it gets warmer. Consequently, sensitivity of the albedo function is smaller. This decrease in sensitivity also takes place in the cold regime: the colder it gets, the less sensitive the albedo gets. Considering that an increasing α_{amp} causes the SAF strength to decrease, this second mechanism dominates in the cold regime.

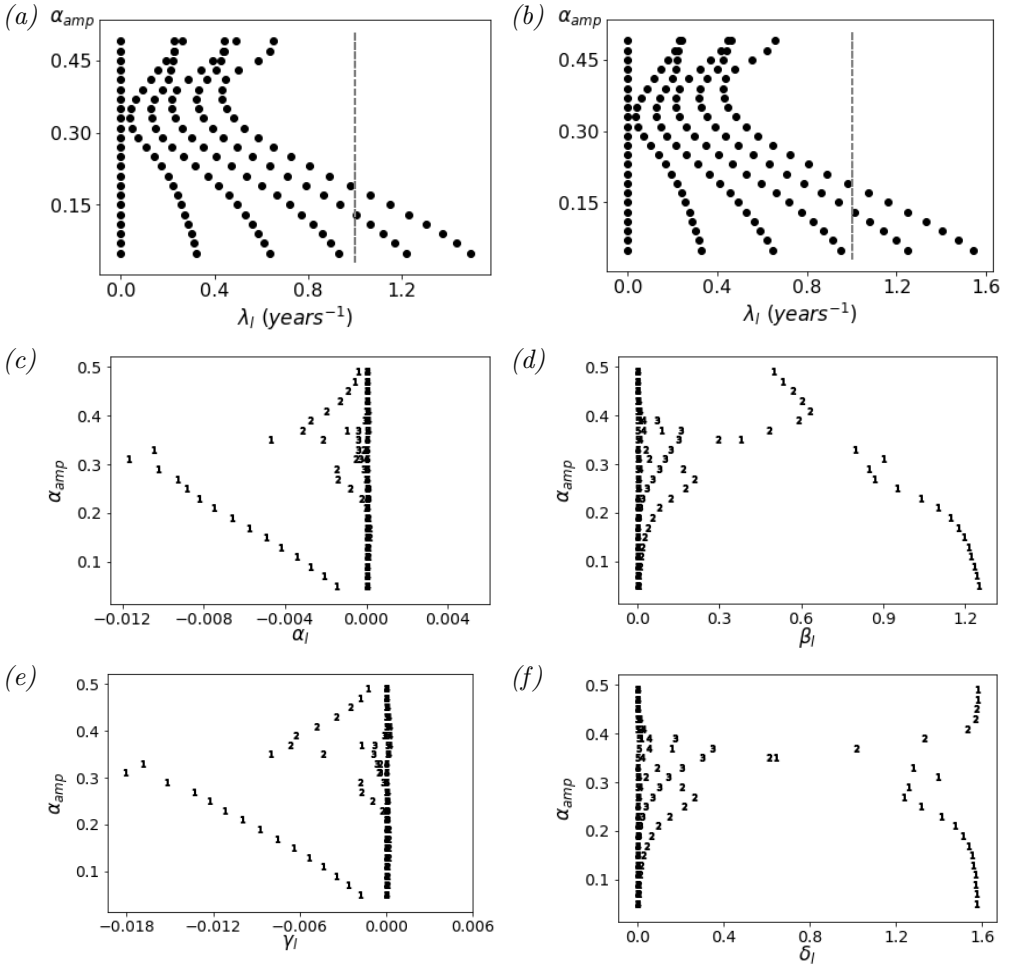

Figure 8: *(a)* Eigenvalues of the EBM depending on the amplitude of the albedo function for the simple EBM *(b)* and the extended EBM. *(c)* Albedo projection terms for solar forcing (α_l) as defined in Equation 5.8 where the markers denote l *(d)* Same for temperature β_l *(e,f)* projections terms for GHG forcing γ_l and δ_l respectively.

The eigenvalues and projections of this system are shown in figure 8. The first nonzero

eigenvalues are smaller than the forcing frequency, which means that the system has no time to fully respond to the seasonal forcing. Apparently, the SAF in the final part of the response scales with SAF on the seasonal scale, so that a linear relationship is still found. The projections reflect the features of the response reasonably: firstly, the albedo projection is negative while the temperature projections are all positive. This is to be expected from the fact that with increasing GHG concentration or solar forcing the albedo should decrease while the temperature should increase. Secondly the susceptibility to solar forcing is smaller than to greenhouse gas forcing, especially for large α_{amp}. In the latter case the average albedo is larger so the forcing term $(1 - \alpha(T))F(t)$ becomes smaller, which becomes apparent when comparing figure 8c and 8e or figure 8d and 8f.

The diffusion term β in the generator was chosen somewhat arbitrarily. Changing this parameter does not influence the eigenvalues of the simulation as expected from the theory (Pavliotis 2014). While the projections of the eigenfunctions did change slightly, the susceptibility ratio was not influenced significantly by a variation of the diffusion ($\beta \pm 20\%$, not shown).

5.2.3. *Extended EBM*

In reality and in *PlaSim*, snow does not react instantaneously to a temperature change. In this part we examine to what extent the spectrum and projections of the energy balance change when adding a separate equation for snow and ice, in which albedo relaxes towards the logistic albedo reference function $\alpha_r(T)$ given in equation 5.5 in a few weeks, a typical value for the reaction time of snow (Hall & Qu 2006).

$$\begin{aligned} dT &= \frac{1}{c_T}\left(Q(1-\alpha_r) + A\ln\frac{C}{C_{ref}} + G - \epsilon\sigma T^4\right)dt + \sqrt{2\beta_T^{-1}}dW_t, \\ d\alpha &= -\frac{1}{c_{snow}}\left(\alpha - \alpha_r(T)\right)dt + \sqrt{2\beta_\alpha^{-1}}dW_t. \end{aligned} \tag{5.9}$$

Here c_{snow} is a constant indicating the response time of the albedo. The drift term in the Fokker-Planck equation corresponding to equation 5.9 is not the gradient of a potential and therefore the system might not be reversible. Using the findings in subsection 3.4, we notice that the eigenvalues of the system can still be computed in the same fashion as for reversible systems.

Extending the model with an explicit albedo function does not change the dynamics of the system significantly, nor the eigenvalues and eigenvectors. Figure 8b shows the eigenvalues of the extended EBM to be almost exactly equal to the eigenvalues of the nonextended model, the complex part continuing to be zero. The projections, naively calculated without taking the adjoint into account, are very similar as well (not shown). Thus, the inclusion of a smaller time scale does not improve the prediction of the response.

6. Summary, discussion and conclusion

In this paper we have proposed a classification for the different kinds of emergent constraints. The primary dividing characteristic is the process under consideration, which can be a sensitivity to a forcing (type I), a mean state (type II) or a feedback strength (type III). A second criterion of our classification is whether the predictor and predictand are the same quantity. If not, there should be a sufficiently strong link between the two variables. For a weak link, an emergent relationship might still be found in simplified climate models, but it is a question whether this relationship is strong enough to overcome

other influences that invariably exist in GCMs. Any (linear) relationship should thus be eyed with suspicion if the link or physical basis is too weak.

With the help of linear response theory and the spectral characteristics of the Fokker-Planck operator and its corresponding generator, we derived expressions for conditions of the different types of emergent constraints in terms of eigenvalues and two projections; firstly the projection of the observable on the eigenfunctions of the generator and secondly the projection of the forcing field on the eigenfunctions. The expressions were derived for gradient systems and then extended to irreversible (hypoelliptic) systems, where the observables were projected on the eigenfunctions of the adjoint of the generator instead of on the eigenfunctions of the generator itself.

For a type I emergent constraint, the derived condition states that the ratio of susceptibilities at the two frequencies under consideration should be constant over the ensemble, specifically a positive constant for type I-s. Type II emergent constraint are encountered when a linear relationship is found between the expectation value of the observable and the susceptibility at the frequency of the change. Finally, a type III emergent constraint has a condition posed, not on the ratio of susceptibilities of one observable, but on the ratio of the ratio between the susceptibilities of the two observables that make up the feedback in the system.

Two applications of framework of conditions were examined. Firstly, it was tested on an Ornstein-Uhlenbeck system in one and two dimensions. No linear relationship was found in 1D, and indeed the condition in terms of the susceptibility ratio for type I-s was not met. In the 2D case a linear relationship was found, a type I-d emergent constraint. The relationship was only present for a weak coupling between the two dimensions. The condition for an emergent constraint type I-d was indeed met.

We continued by applying the framework to the snow-albedo feedback as found in a multi-model ensemble by Hall & Qu (2006), a type III emergent constraint. First, it was shown that this emergent constraint in the SAF could at least in part be reproduced by an ensemble generated with a varying parameter in the intermediate complexity climate model of *PlaSim*. From this, we continued with a modified Energy Balance Model, in which the albedo is a function of temperature. The ratio of the feedback strengths at two different time scales, as measured by a ratio of susceptibilities for the two observables, did indeed deviate from a constant in the domain for which no emergent relationship was found. Incorporating the snow response by a relaxation towards the albedo function of the first model did not change the outcome of the calculation of susceptibilities. In contrast to the OU process, the eigenvalues did turn out to be the determining factor, possibly due to the fact that a feedback can be measured almost instanteneously compared to the timescales involved. Instead, the projections varied substantially over the parameter domain.

Further generalizations of our framework to for instance GCMs prove difficult. Firstly, the system is not hypoelliptic anymore; the noise does not spread to all dimensions of the system. As a consequence, the system has a (strange) attractor with dimensionality lower than the full phase space. The invariant measure supported on this attractor is not continuous anymore with respect to the Lebesgue measure. By invoking the Chaotic Hypothesis we assume that the system possesses a Sinai-Ruelle-Bowen (SRB) invariant measure which guarantees (1) the asympotitc equivalence of time and ensemble averages and (2) the stability of the statistical properties to a weak stochastic forcing, for coarse-grained observables (e.g. regionally or globally integrated variables) (Ragone *et al.* 2016; Tantet *et al.* submitted for publication.). As in Ragone *et al.* (2016) we assume that the Chaotic Hypothesis allows using Ruelle's Response theory.

With the use of Ruelle's Response theory, a susceptibility can be defined as the Fourier

transform of the response function. Numerically, it is rather straightforward to compute the response function by exploiting the properties of the convolution in the equation for the response. Applying the force in the form of Dirac delta function or a Heaviside function allows to compute the response functions as the response in a certain observable and the derivative of the response respectively. Even so, a relatively large ensemble should be taken to guarantee reliability of the method (Ragone *et al.* 2016).

Analytically, the susceptibility cannot be expressed in terms of eigenfunctions and eigenvalues for systems with a SRB measure, as far as we know. Further examination of the link between the eigenfunctions of eigenvalues of the system could still be of interest in the analysis of emergent constraints though. To what extent are the same features found for simple models compared to high-dimensional models in terms of eigenvalues with a typical value larger than the frequency of the forcing and how do projections on eigenfunctions change between the different models?

In a complex high-dimensional dynamical system eigenfunctions and eigenvalues can be accessed with the help of transfer operators. The eigenfunctions that lie on the invariant measure are then computed by making use of the ergodic nature of the climate system. To overcome the burden of high-dimensionality, a reduced transfer operator can be computed on a very long simulation, from which the eigenfunctions on the attractor are approximated (Tantet 2016). However, computing the eigenvalues in the full phase space is prohibitively computationally expensive. A forcing on the system does not generally lie only on the attractor and should be split into a part parallel and perpendicular to the attractor. Consequently, the eigenvectors off the attractor cannot a priori be ignored (Lucarini & Sarno 2011). Gritsun & Lucarini (2017) showed that indeed for some geophysical systems, specifically quasi-geostrophic flow with orographic forcing, the fluctuation-dissipation theorem is violated and the response has no resemblance to the unforced variability in the same range of spatial and temporal scales.

The classification of emergent constraints given above gives a hint to which kind of emergent constraints one can look out for. Using the susceptibilities to find new emergent constraints however does not seem to have an advantage above directly looking for plausible correlations. An attempt to directly find an emergent constraint for climate sensitivity by data mining in a CMIP5 ensemble proved fruitless however (Caldwell *et al.* 2014). Susceptibilities might provide additional information on the emergent constraint. For example, when a susceptibility shows a resonance at a certain frequency over the ensemble, this could suggest that the same feedback is present in all simulations.

To conclude, in this paper a classification for emergent constraints was laid out with three different primary types. We successfully derived a set of conditions for the different types of emergent constraints. The eigenvalue spectrum can in some special cases indicate whether an emergent constraint is present, but often more information of the system is needed. By using transfer operators and Ruelle's Response theory, application to GCMs seems possible, but is far from straightforward.

Acknowledgements

Front and centre, I'd like to thank my supervisor Henk Dijkstra for introducing me into this beautiful topic. He gave me the freedom to tackle problems in my own way while always providing guidance and introduced me to the academic world. Many thanks also to Alexis Tantet who was always there to provide insight in the mathematical details of response theory and its applications to climate systems. Thanks to Peter Cox for the discussions we had on emergent constraints. Frank Lunkeit provided invaluable insight on the functioning of *PlaSim* and Micheal made sure that it got running. Furthermore I'd

like to thank the people at IMAU for creating a very pleasant atmosphere and Matthias for the lively discussions during tea breaks and the critical questions regarding my thesis. Finally, I'd like to thank my friends and family who were always there for me, notably Evelien, Sebastian, Noortje, Etienne, Tim, my parents and my brother Youri.

REFERENCES

Allen, Myles R. & Ingram, William J. 2002 Constraints on future changes in climate and the hydrologic cycle. *Nature* **419**, 224–232.

Boe, Julien, Hall, Alex & Qu, Xin 2009 September sea-ice cover in the Arctic Ocean projected to vanish by 2100. *Nature Geoscience* **2** (5), 341–343.

Bracegirdle, Thomas J. & Stephenson, David B. 2013 On the Robustness of Emergent Constraints Used in Multimodel Climate Change Projections of Arctic Warming. *Journal of Climate* **26** (2), 669–678.

Caldwell, Peter M., Bretherton, Christopher S., Zelinka, Mark D., Klein, Stephen A., Santer, Benjamin D. & Sanderson, Benjamin M. 2014 Statistical significance of climate sensitivity predictors obtained by data mining. *Geophysical Research Letters* **41** (5), 1803–1808.

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

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

Collins, Matthew, Chandler, Richard E., Cox, Peter M., Huthnance, John M., Rougier, Jonathan & Stephenson, David B. 2012 Quantifying future climate change. *Nature Climate Change* **2**, 403–409.

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

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

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

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

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

Hall, Alex & Qu, Xin 2006 Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophysical Research Letters* **33** (3), l03502.

Klein, Stephen A. & Hall, Alex 2015 Emergent Constraints for Cloud Feedbacks. *Current Climate Change Reports* **1** (4), 276–287.

Knutti, Reto, Meehl, Gerald A., Allen, Myles R. & Stainforth, David A. 2006 Constraining Climate Sensitivity from the Seasonal Cycle in Surface Temperature. *Journal of Climate* **19** (17), 4224–4233.

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

Lucarini, V., Fraedrich, K. & Lunkeit, F. 2010 Thermodynamics of climate change: generalized sensitivities. *Atmospheric Chemistry and Physics* **10** (20), 9729–9737.

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

Lunkeit, F., Borth, H., Böttinger, M. & Fraedlich, F. 2011 *PLASIM Reference Guide*.

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

Qu, Xin & Hall, Alex 2007 What Controls the Strength of Snow-Albedo Feedback? *Journal of Climate* **20** (15), 3971–3981.

Qu, Xin & Hall, Alex 2014 On the persistent spread in snow-albedo feedback. *Climate Dynamics* **42** (1), 69–81.

Qu, Xin, Hall, Alex, Klein, Stephen A. & Caldwell, Peter M. 2014 On the spread of changes in marine low cloud cover in climate model simulations of the 21st century. *Climate Dynamics* **42** (9), 2603–2626.

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

Sherwood, Steven C., Bony, Sandrine & Dufresne, Jean-Louis 2014 Spread in model climate sensitivity traced to atmospheric convective mixing. *Nature* **505** (7481), 37–42.

Stephenson, David B., Collins, Matthew, Rougier, Jonathan C. & Chandler, Richard E. 2012 Statistical problems in the probabilistic prediction of climate change. *Environmetrics* **23** (5), 364–372.

Su, Hui, Jiang, Jonathan H., Zhai, Chengxing, Shen, Tsaepyng J., Neelin, J. David, Stephens, Graeme L. & Yung, Yuk L. 2014 Weakening and strengthening structures in the Hadley Circulation change under global warming and implications for cloud response and climate sensitivity. *Journal of Geophysical Research: Atmospheres* **119** (10), 5787–5805.

Tantet, Alexis 2016 Ergodic theory of climate: variability, stability and response. PhD thesis, Utrecht University.

Tantet, A., Lucarini, V., Lunkeit, F. & Dijkstra, H. submitted for publication. Crisis of the Chaotic Attractor of a Climate Model: A Transfer Operator Approach .

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

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

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