QUARTERLY JOURNAL OF THE ROYAL METEOROLOGICAL SOCIETY
Q. J. R. Meteorol. Soc. **133**: 545–560 (2007)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/qj.62

Some considerations of the concept of climate feedback

J. R. Bates*
University College Dublin, Dublin, Ireland

ABSTRACT: A conceptual study of climate feedbacks is carried out using two simple linear two-zone models and the commonly-used zero-dimensional model to which they reduce under simplifying assumptions. The term 'feedback' is used in many different senses in the climate literature. Two prototype usages, stability-altering feedback (defined in terms of a system's asymptotic response to an impulsive forcing, negative when stability-enhancing) and sensitivity-altering feedback (defined in terms of a system's steady-state response to a step-function forcing, negative when sensitivity-diminishing) have been isolated for study. These two climate feedback concepts are viewed against the background of control theory, which provides a generalized feedback perspective embracing all forms of forcing and which is often seen as providing the paradigm for the concept of feedback as used in climate studies.

The relationship between the prototype climate feedbacks is simple in the context of the zero-dimensional model. Here, the stability-altering and sensitivity-altering feedbacks provided by a given interaction are of the same sign, and the sign of the stability-altering feedback as measured by initial tendencies always coincides with its sign as measured by the defining asymptotic tendencies. Even in this simple model, however, the sign of the prototype climate feedbacks can be opposite to the sign of the system's feedback as defined in control theory.

In the two-zone models, the relationship between the prototype climate feedbacks is not so simple. It is shown that, contrary to the common assumption, these feedbacks can be of opposite signs. Moreover, the sign of the stability-altering feedback as measured by initial tendencies can be opposite to its sign as measured by asymptotic tendencies. It is further shown that there is no simple relationship between the sign of either of the prototype climate feedbacks in the two-zone models and the sign of these models' feedback as defined in control theory.

These results point to the need for greater precision and explicitness in the definition and use of the term 'climate feedback', both to facilitate interdisciplinary dialogue in relation to feedback and to guard against erroneous inferences within the climate field. Explicit definitions of the two prototype categories of climate feedback studied here are proposed. Copyright © 2007 Royal Meteorological Society

KEY WORDS climate stability; climate sensitivity

Received 18 January 2006; Revised 12 October 2006; Accepted 29 December 2006

1. Introduction

Feedbacks in the climate system are currently receiving much attention, particularly as the future extent of anthropogenic global warming is believed to depend critically on their influence (e.g. NRC, 2003; Bony *et al.*, 2006). Considerable difficulty has been encountered in understanding the various climate feedbacks in operation. In a recent review of cloud feedbacks, Stephens (2005) has identified conceptual difficulties in the definition of feedback as one of the impediments to progress, pointing out that different assumptions about the nature of the climate system can lead to feedback measures that differ not only in magnitude but even in sign. The present paper expands on this theme, examining some further conceptual aspects of climate feedback. Attention is focused on two of the different senses in which the term 'climate feedback' is commonly used, and it is shown that, except in a zero-dimensional model, they can be in conflict.

We argue that careful attention to how climate feedback is defined is important for two reasons. First, interdisciplinary dialogue with scientists working outside the climate field, or discussion between scientists working in different areas of the climate field, can be greatly impeded if those involved have different understandings of what the term 'climate feedback' means. Secondly, an awareness of the different ways in which feedback is defined is necessary to guard against erroneous transfer of results from other disciplines into the climate field, or faulty reasoning within the climate field.

We begin by reviewing the concept of feedback as used in electronics and in the theory of automatic control, since it is from these areas that the concept of feedback as used in climate studies is frequently claimed to be derived.

Feedback as a technical concept originated in the field of electronics in the early part of the twentieth century, where it was used to describe the return of a portion of the output of an amplifier to the input, modifying the properties of the amplifier. In 1927, H. S. Black invented

* Correspondence to: J. R. Bates, School of Mathematical Sciences, University College Dublin, Belfield, Dublin 4, Ireland.
E-mail: Ray.Bates@ucd.ie

the negative-feedback audio amplifier and derived the now-classical formula for the amplification with feedback:

$$G_{\mathrm{F}} = \frac{G}{1 - GH},$$

where G_{F} is the gain (ratio of output to input signals) with feedback, G is the forward gain (or gain without feedback), and H is the feedback factor, defined as the ratio of the fed-back signal to the output signal. (For a historical view, see the retrospective article by Black (1977); for a general treatment of feedback in electronics, see (Smith, 1987)). Black's formula, which applies to both positive and negative feedback, describes the steady-state response of a stable single-loop amplifier to a sinusoidal input signal. (At the audio frequencies of relevance in amplifier theory, adjustment to the final steady state in an electronic circuit is effectively instantaneous). The output is a sinusoidal signal of the same frequency. In general, changes in phase and magnitude (both frequency-dependent) occur at both the amplification and the feedback stages within the feedback loop; thus, G and H are complex quantities. Negative feedback is said to occur when the amplitude and phase of the fed-back signal are such that the magnitude of G_{F} is less then the magnitude of G, and vice versa for positive feedback. A special case of negative feedback, corresponding to a signal fed back with a 180° phase change, occurs when the loop gain GH is real and negative.

The term 'feedback' was subsequently adopted in the field of automatic control, where it is used to describe an operation in which the difference between the desired value of a variable (the input) and its measured value (the output) is returned to a controller that drives the variable in such a way as to reduce the difference (e.g. Dorf and Bishop, 2005). General feedback control theory embraces systems that may involve electrical, mechanical, hydraulic and chemical components. Unlike in the case of the audio amplifier, instantaneous adjustment of such general systems to a final steady state cannot usefully be assumed: the transient response and the eventual steady-state response to a given input are both of concern, and the transient features (such as the stability properties and the swiftness of response) are generally the ones of primary interest.

In the design of a feedback control system, a mathematical model of the system is usually constructed, and the system's performance under normal operating conditions assessed using test input signals of various forms – for example, an impulse function, a step function, or a sinusoidal function. (Maxwell (1868) initiated the mathematical modelling of what are now called feedback control systems before the term 'feedback' was coined. He was mainly concerned with the stability of such systems.) It is the usual practice in control theory to linearize the governing equations of the model and to solve them for the various test inputs using the method of Laplace transforms. For a single-loop system, a gain formula analogous to Black's amplifier formula then arises, but with the complex numbers G_{F}, G and H replaced by the corresponding transforms $G_{\mathrm{F}}(s)$, $G(s)$ and $H(s)$, which are explicit functions of the complex Laplace variable s. For a negative-feedback control system (defined as one in which the fed-back signal is subtracted from the input signal), the gain with feedback $G_{\mathrm{F}}(s)$ can be expressed in terms of the forward gain $G(s)$ and the loop gain $G(s)H(s)$ as:

$$G_{\mathrm{F}}(s) = \frac{G(s)}{1 + G(s)H(s)}.$$

For a positive-feedback control system (defined as one in which the fed-back signal is added to the input signal),

$$G_{\mathrm{F}}(s) = \frac{G(s)}{1 - G(s)H(s)}.$$

The gain formula (or closed-loop transfer function) thus expressed explicitly shows the sign of the feedback and provides a means of determining both the transient and the steady-state response of the system to any form of input.

It might be expected that, for a sinusoidal input, in the limit as the transients decay to zero in a stable system, the sign of the feedback as defined in control theory would coincide with that of the feedback as defined in electronics. We shall show here, however, that it is possible in this limit for a system that has negative feedback according to the control-theory definition to have positive feedback according to the electronics definition. This gives an indication from outside the climate field of the need for caution in relation to how feedback is defined.

The term 'feedback' entered the climate literature in the 1960s, and its use is now widespread within this field. Even within this restricted area, there are many different usages. From these, two prototype senses can be isolated. The first, which we shall call *stability-altering* feedback, describes the effect of an interaction between physical processes that influences the asymptotic stability of impulsively-forced perturbations in the global climate system or its components. The second, which we shall call *sensitivity-altering* feedback, describes the effect of an interaction between physical processes that influences the equilibrium change in global mean surface temperature resulting from a constant increment in external forcing.

Climate feedback in the stability-altering sense was introduced by Stommel (1961) in a theoretical study of the thermohaline circulation. The interaction providing the feedback in his case was that between the local density transforming processes (heating and cooling, evaporation and precipitation) in the warm and cold reservoirs of his ocean model and the flow between the reservoirs, which is itself driven by the density difference. This interaction implies the existence of three possible equilibrium states of the model for the same basic forcing, and determines whether small perturbations about these states are asymptotically stable or unstable. Feedback in

DOI: 10.1002/qj

the stability-altering sense is now a common concept in the oceanographic literature, with positive feedback used to characterize a destabilizing interaction and negative feedback a stabilizing one (e.g. Nakamura *et al*., 1994; Rahmstorf and Willebrand, 1995; Marotzke, 1996; Jayne and Marotzke, 1999). It is also common in the general climate literature. For example, Charney *et al*. (1977) used the concept in their study of the stability of desert margins using a general-circulation model (GCM); the interaction providing the feedback involved surface albedo, radiation, evaporation and rainfall. Ramanathan and Collins (1991) and Lindzen *et al*. (2001), using satellite observations, proposed stabilizing feedbacks on tropical sea surface temperature (SST) perturbations involving cloud–radiation–temperature interactions, based respectively on increased albedo and increased infrared emission through a dryer and more cloud-free upper troposphere as the tropical SST increases. Bates (1999) proposed a stabilizing feedback on large-scale SST perturbations based on the interaction between dynamically-induced changes in evaporation (caused by changes in eddy angular momentum transport and the consequent changes in surface wind) and the local physical processes influencing SST. Caballero (2001) further investigated this stabilizing feedback using a single-column model for the Tropics interacting with an Extratropics where the SST is held fixed, a tropical upper-tropospheric humidity profile being prescribed.

Climate feedback in the sensitivity-altering sense was introduced by Manabe and Wetherald (1967) in a study of the equilibrium temperature response of a one-dimensional radiative–convective model to a constant increment in external forcing, such as a change in the solar constant or a doubling of the atmospheric CO_2 concentration. The interaction providing the feedback in their case was that between the changes in the water-vapour field with a fixed relative-humidity profile and the radiative–convective processes determining the temperature field, the zero-interaction case corresponding to a fixed *absolute*-humidity profile. This type of feedback is usually referred to in the climate literature simply as the water-vapour feedback. Sensitivity-altering feedback was soon extended to the GCM context and to other types of interactions. Thus, Wetherald and Manabe (1988) studied the effect of cloud–radiation interactions on the equilibrium response of a GCM to a CO_2 doubling, comparing the results with interactive and prescribed cloud cover. Sensitivity-altering feedback is defined as positive if the equilibrium global-mean surface temperature increment due to a constant increment in global-mean forcing is enhanced by a specified interaction between physical processes, and negative if it is diminished. Feedback in the sensitivity-altering sense is now a common concept in the climate-change literature (e.g. IPCC, 2001; NRC, 2003).

An unstated assumption generally made in the climate literature is that the two usages of the term 'climate feedback' described above are interchangeable. In other words, it is implicitly assumed that if an interaction between physical processes provides a stability-altering feedback, then it provides a sensitivity-altering feedback of the same sign. (Exceptions where this assumption is not made are the studies of Lindberg (2002) and Alexeev (2003). In these studies, the complexity of the relationship between climate stability and sensitivity is recognized, but the authors' concerns are with matters other than those discussed here.) Here are some examples to illustrate this point:

- In (Peixoto and Oort, 1992), chapter 2, the mathematical treatment in sections 2.5.1 and 2.5.2 is concerned with sensitivity-altering feedback, whereas the subsequent verbal discussion in section 2.5.3 is concerned mainly with stability-altering feedback. There us no mention of any transition to a different concept of feedback.
- In (AMS, 2000), the following definition is given:

> Feedback – A sequence of interactions that determines the response of a system to an initial perturbation. Feedbacks may either amplify (positive feedback) or reduce (negative feedback) the ultimate state of the system.

  The first sentence, in referring to the response to an initial perturbation, is concerned with stability. The second, in referring to the ultimate state, is concerned with sensitivity. No distinction between the two senses of feedback is made.
- In (IPCC, 2001), chapter 7 'Physical Climate Processes and Feedbacks' is concerned mainly with sensitivity-altering feedback, but also deals with stability-altering feedback (for example, in relation to the stability of the thermohaline circulation). When discussing water-vapour and cloud feedbacks, the term is used in some places in a sensitivity-altering sense and in others in a stability-altering sense, with no discussion of the possible implications of the alternation in usage. The glossary of the report provides the following definition:

> Climate Feedback. An interaction mechanism between processes in the climate system is called a climate feedback when the result of an initial process triggers changes in a second process that in turn influence the initial one. A positive feedback intensifies the initial process and a negative feedback reduces it.

  Since this definition does not mention a constant increment in external forcing or the equilibrium response thereto, it is not a definition of sensitivity-altering feedback. It can be interpreted as a definition of stability-altering feedback, though it appears to define the sign of the feedback in terms of initial rather than asymptotic tendencies. (These tendencies can be of opposite sign, and the sign of stability-altering feedback as indicated by initial tendencies can be misleading: see Section 3.4 below.) Both in chapter 7 and in the glossary definition, there is clearly an assumption that the stability-altering and sensitivity-altering senses of climate feedback are interchangeable.

DOI: 10.1002/qj

The main purpose of the present paper is to show that stability-altering feedback and sensitivity-altering feedback as used in climate are two separate concepts, and that neither of them coincides with the concept of feedback as defined in control theory or in electronics. The signs of stability-altering and sensitivity-altering feedback coincide for the simple case of a zero-dimensional climate model, but not necessarily if the model is extended to include two zones with dynamical interaction between them. To demonstrate this, results are presented from two two-zone models of a hemisphere that include dynamical interaction between the Tropics and the Extratropics through baroclinic eddy transports. If interhemispheric symmetry is assumed, the models are global.

An additional purpose of the paper is to investigate the relationship between the initial and asymptotic tendencies of impulsively-forced perturbations. Here again, it is shown that the simple relationship between these tendencies and stability-altering feedback that holds for a zero-dimensional climate model can break down when the model is generalized to include two zones.

The models used are based on linearized energy equations for an idealized climate system in which the surface fluxes determining the tropical and extratropical SSTs are calculated using, respectively, top-of-the-atmosphere and surface parametrizations. These two models will be referred to as the TOA and SFC models, respectively. More complex versions of them have been used in (Bates, 1999) to examine climate stability and in (Bates, 2004) to examine climate sensitivity. The results presented here, however, are new, involving parameter values chosen specifically to demonstrate that the stability-altering and sensitivity-altering feedbacks provided by the interaction between given physical processes can be of opposite signs. The required parameter values differ, though not drastically, from the observational estimates used previously.

Brief descriptions of both models are given. These are intended to demonstrate that both are internally consistent and physically possible low-order models of an idealized climate system. The models differ in assigning the determining roles to different physical processes. The question of which model may come closest to providing a realistic first-order description of the actual climate system has been discussed elsewhere (Bates, 1999, 2004), and will not be taken up here; it is not necessary for present purposes that either model provide an accurate description of the climate system. For the most part, the same results regarding conceptual aspects of feedbacks emerge from both models, but the overall results of our investigation are more clearly seen using the SFC model.

2. The models

The models used are two-zone box models of a hemisphere on an aquaplanet, zone 1 representing the Tropics (taken as 0–30°) and zone 2 the Extratropics (taken as 30°–90°), with a wall at the Equator. The models do not resolve the seasonal cycle. A mixed-layer ocean of depth D (taken here as 100 m) is assumed in both zones. The basic equations for both models are derived from the ocean energy equations:

$$c_0 \frac{\mathrm{d}T_1}{\mathrm{d}t} = (F^{\downarrow}_{\mathrm{SFC}})_1 - F_{\mathrm{OH}}, \tag{1}$$

$$c_0 \frac{\mathrm{d}T_2}{\mathrm{d}t} = (F^{\downarrow}_{\mathrm{SFC}})_2 + F_{\mathrm{OH}}, \tag{2}$$

where T_1 and T_2 are the SSTs in zones 1 and 2 respectively, c_0 is the mixed-layer heat capacity in each zone, $F^{\downarrow}_{\mathrm{SFC}}$ is the net downward energy flux at the surface in a zone, and F_{OH} is the poleward ocean energy transport at 30°. We calculate c_0 as

$$\pi a^2 c_{\mathrm{pw}} \rho_{\mathrm{w}} D,$$

where a is the Earth's radius (6.37×10^6 m), c_{pw} is the specific heat of seawater (4187 J kg^{-1}K^{-1}) and ρ_{w} is the density of seawater (10^3 kg m^{-3}); thus, $c_0 = 5.337 \times 10^{22}$ JK^{-1}. The atmosphere, having a much lower heat capacity than the mixed-layer ocean, is assumed to be in energetic balance on the long time-scales of interest; thus,

$$(F^{\downarrow}_{\mathrm{SFC}})_1 = (F^{\downarrow}_{\mathrm{TOA}})_1 - F_{\mathrm{E}}, \tag{3}$$

$$(F^{\downarrow}_{\mathrm{SFC}})_2 = (F^{\downarrow}_{\mathrm{TOA}})_2 + F_{\mathrm{E}}, \tag{4}$$

where $F^{\downarrow}_{\mathrm{TOA}}$ is the net downward radiative energy flux at the top of the atmosphere in a zone and F_{E} is the poleward transport of moist static energy by atmospheric motions at 30°. Eliminating the surface fluxes from Equations (1) and (2) using Equations (3) and (4) gives

$$c_0 \frac{\mathrm{d}T_1}{\mathrm{d}t} = (F^{\downarrow}_{\mathrm{TOA}})_1 - (F_{\mathrm{E}} + F_{\mathrm{OH}}), \tag{5}$$

$$c_0 \frac{\mathrm{d}T_2}{\mathrm{d}t} = (F^{\downarrow}_{\mathrm{TOA}})_2 + (F_{\mathrm{E}} + F_{\mathrm{OH}}). \tag{6}$$

The TOA model is based on the explicit use of Equations (5) and (6), with the quantities on the right-hand side parametrized in terms of T_1 and T_2. It is implicitly assumed that $(F^{\downarrow}_{\mathrm{SFC}})_1$ and $(F^{\downarrow}_{\mathrm{SFC}})_2$ respond in such a way that Equations (3) and (4), and thus Equations (1) and (2), are always satisfied. The SFC model is based on the direct use of Equations (1) and (2), with the different quantities on the right-hand side now parametrized in terms of T_1 and T_2. In this case it is implicitly assumed that $(F^{\downarrow}_{\mathrm{TOA}})_1$, $(F^{\downarrow}_{\mathrm{TOA}})_2$ and F_{E} respond in such a way that Equations (3) and (4), and thus Equations (5) and (6), are always satisfied. Thus, both models provide energetically consistent representations of the idealized climate system: atmospheric energy balance is always observed, and there is no energetic inconsistency for either model at either the surface or the top of the atmosphere. The difference between the models consists in the choice of which quantities are considered

DOI: 10.1002/qj

as playing a determining role and which are considered as being a diagnostic consequence.

The equilibrium climates of the TOA and SFC models satisfy, respectively, Equations (5) and (6) and Equations (1) and (2), with $\frac{\mathrm{d}}{\mathrm{d}t} = 0$. For a description of the equilibrium climates based on observations, see (Bates, 1999). We are here concerned with small perturbations about the observed equilibrium climate with F_{OH} fixed at its observed equilibrium value. Observed equilibrium values will be denoted by an overbar and perturbations about these values by a prime.

2.1. Perturbation equations

2.1.1. The TOA model

In deriving the perturbation forms of Equations (5) and (6), we omit any variations in $F^{\downarrow}_{\mathrm{TOA}}$ due to short-wave radiation and consider only variations in this quantity due to changes in outgoing long-wave radiation (OLR). The OLR per unit area is parametrized as $A + BT$, T being the SST in degrees Celsius, as is customary in models of the Budyko–Sellers type (e.g. North *et al*, 1981). We choose different values of B in the tropical and extratropical zones, denoted by B_1 and B_2 respectively. Perturbations in F_{E} are parametrized as in (Bates, 1999), whereby

$$F'_{\mathrm{E}} = a_{\mathrm{E}}(a_{\mathrm{Z}1}T'_1 - a_{\mathrm{Z}2}T'_2). \tag{7}$$

Here, a_{E}, $a_{\mathrm{Z}1}$ and $a_{\mathrm{Z}2}$ are observationally-derived quantities found by relating the seasonal mean values of F_{E} to the difference between the tropical and extratropical 500 hPa heights, with $a_{\mathrm{Z}1}$ and $a_{\mathrm{Z}2}$ representing the sensitivities of these heights to SST changes.

The perturbation forms of Equations (5) and (6) can thus be written as

$$c_0\frac{\mathrm{d}T_1}{\mathrm{d}t} = -\hat{b}_1T'_1 - (\hat{d}_1T'_1 - \hat{d}_2T'_2) + (F'_{\mathrm{T}})_1, \tag{8}$$

$$c_0\frac{\mathrm{d}T_2}{\mathrm{d}t} = -\hat{b}_2T'_2 - (\hat{d}_1T'_1 - \hat{d}_2T'_2) + (F'_{\mathrm{T}})_2, \tag{9}$$

where $\hat{b}_i = \pi a^2 B_i$, $\hat{d}_i = a_{\mathrm{E}}a_{\mathrm{Z}i}$, and $(F'_{\mathrm{T}})_1$ and $(F'_{\mathrm{T}})_2$ represent the external forcings at the top of the atmosphere in the respective zones.

The coefficients $\hat{b}_1$ and $\hat{b}_2$ in the above equations, which are both positive, represent the local stabilizing effects of OLR on SST perturbations as seen from the top of the atmosphere, while the coefficients $\hat{d}_1$ and $\hat{d}_2$ represent the interzonal dynamical effects.

2.1.2. The SFC model

In deriving the perturbation forms of Equations (1) and (2), we omit any variations in $F^{\downarrow}_{\mathrm{SFC}}$ due to short-wave radiation or sensible heat flux, and consider only variations in this quantity due to net (upward minus downward) long-wave radiation (F_{I}) and latent heat flux (F_{L}). The perturbation in net long-wave radiation at the surface is parametrized as

$$F'_{\mathrm{I}} = \gamma_{\mathrm{I}}T' - F'_{\mathrm{G}}, \tag{10}$$

where the first term on the right-hand side includes variations in the upwelling Stefan–Boltzmann flux and the downwelling back radiation from the atmosphere, and the second term represents the external radiative forcing at the surface. The coefficient γ_{I} is negative (giving a locally destabilizing tendency), since downwelling long-wave radiation at the surface, with the water-vapour changes accompanying temperature changes taken into account, increases faster with temperature than the upwelling Stefan–Boltzmann flux: for a discussion of this point, see (Hartmann, 1994, chapter 9), (Bates, 1999) or (Lindberg, 2003).

The perturbation in latent heat flux, following the bulk aerodynamic formula, can be written as

$$F_{\mathrm{L}} = \overline{F}_{\mathrm{L}}\left(\frac{\Delta q'}{\Delta\overline{q}} + \frac{V'}{\overline{V}}\right). \tag{11}$$

Here, $\Delta q'/\Delta\overline{q}$ is a humidity factor (q being the specific humidity) and $V'/\overline{V}$ is a wind factor (V being the magnitude of the surface wind). The humidity factor is approximated, as in (Bates, 1999), by assuming the air–sea temperature difference and low-level relative humidity to be fixed; this leads to the Clausius–Clapeyron expression

$$\frac{\Delta q'}{\Delta\overline{q}} = \frac{L}{R_{\mathrm{v}}\overline{T}^2}T', \tag{12}$$

where L is the latent heat of vaporization of water and R_{v} is the gas constant for water vapour. The wind factor is approximated, as in (Bates, 1999), by assuming balance between angular momentum and torque and taking the ratio $u'/\overline{u}$ (u being the zonal wind component) as representative of $V'/\overline{V}$; thus, for small perturbations about equilibrium in either zone,

$$\frac{V'}{\overline{V}} = \frac{a_{\mathrm{M}}}{2\overline{F}_{\mathrm{M}}}(a_{\mathrm{Z}1}T'_1 - a_{\mathrm{Z}2}T'_2). \tag{13}$$

Here, a_{M} is an observationally-derived quantity found by relating the seasonal mean values of the eddy angular momentum transport at 30° (F_{M}) to the difference between the tropical and extratropical 500 hPa heights.

Using Equations (10), (11), (12) and (13), we see that the perturbation forms of Equations (1) and (2) become

$$c_0\frac{\mathrm{d}T_1}{\mathrm{d}t} = -b_1T'_1 - (d_{11}T'_1 - d_{12}T'_2) + (F'_{\mathrm{G}})_1, \tag{14}$$

$$c_0\frac{\mathrm{d}T_2}{\mathrm{d}t} = -b_2T'_2 - (d_{21}T'_1 - d_{22}T'_2) + (F'_{\mathrm{G}})_2, \tag{15}$$

where

$$b_i = \left(\overline{F}_{\mathrm{L}}\frac{L}{R_{\mathrm{v}}\overline{T}^2} + \gamma_{\mathrm{I}}\right)_i, \tag{16}$$

DOI: 10.1002/qj

and

$$d_{ij} = (\overline{F}_{\mathrm{L}})_i \frac{a_{\mathrm{M}} a_{\mathrm{Z}j}}{2\overline{F}_{\mathrm{M}}}. \tag{17}$$

As in the TOA case, the coefficients b_1 and b_2 in the above equations represent local stabilizing effects (it being assumed that the Clausius–Clapeyron evaporative term in Equation (16) predominates over the negative long-wave radiation term, so that b_1 and b_2 are both positive), while the coefficients d_{ij} represent the interzonal dynamical effects. In contrast to Equations (8) and (9), where the sign preceding the dynamical terms changes from one equation to the other, the sign remains the same in Equations (14) and (15). This is because in the TOA model an increase in the tropical–extratropical 500 hPa difference is explicitly reflected in an increase in poleward eddy heat transport (which cools the tropical zone and heats the extratropical zone), while in the SFC model it is explicitly reflected in an increase in poleward angular momentum transport (which cools the SST in both zones, thanks to an increase in surface wind speed in both).

3. Stability, sensitivity and feedbacks

The governing equations (Equations (8), (9), (14) and (15)) for our two-zone models can be written in the generic forms:

$$c_0 \frac{\mathrm{d}T_1}{\mathrm{d}t} = -\alpha_1 T_1' - \alpha_2 T_2' + F_1', \tag{18}$$

$$c_0 \frac{\mathrm{d}T_2}{\mathrm{d}t} = -\alpha_3 T_1' - \alpha_4 T_2' + F_2', \tag{19}$$

where for the TOA model

$$\left.\begin{array}{l} \alpha_1 = \hat{b}_1 + \hat{d}_1 \\ \alpha_2 = -\hat{d}_2 \\ \alpha_3 = -\hat{d}_1 \\ \alpha_4 = \hat{b}_2 + \hat{d}_2 \\ F_i' = (F_{\mathrm{T}}')_i \end{array}\right\} \tag{20}$$

and for the SFC model

$$\left.\begin{array}{l} \alpha_1 = b_1 + d_{11} \\ \alpha_2 = -d_{12} \\ \alpha_3 = d_{21} \\ \alpha_4 = b_2 - d_{22} \\ F_i' = (F_{\mathrm{G}}')_i \end{array}\right\}. \tag{21}$$

By elimination of one or other of the dependent variables, Equations (18) and (19) can be written as two second-order equations in the independent variables T_1' and T_2', as follows:

$$M \frac{\mathrm{d}^2 T_1'}{\mathrm{d}t^2} + \beta \frac{\mathrm{d}T_1'}{\mathrm{d}t} + k T_1' = r_1', \tag{22}$$

$$M \frac{\mathrm{d}^2 T_2'}{\mathrm{d}t^2} + \beta \frac{\mathrm{d}T_2'}{\mathrm{d}t} + k T_2' = r_2', \tag{23}$$

where

$$M = c_0^2, \tag{24}$$

$$\beta = c_0(\alpha_1 + \alpha_4), \tag{25}$$

$$k = \alpha_1 \alpha_4 - \alpha_2 \alpha_3, \tag{26}$$

$$r_1' = \left(c_0 \frac{\mathrm{d}}{\mathrm{d}t} + \alpha_4\right) F_1' - \alpha_2 F_2', \tag{27}$$

$$r_2' = \left(c_0 \frac{\mathrm{d}}{\mathrm{d}t} + \alpha_1\right) F_2' - \alpha_3 F_1'. \tag{28}$$

The quantities r_1' and r_2', which are functions of the forcings in both zones, represent the effective forcings on T_1' and T_2', respectively. Equations (22) and (23) are equivalent to the governing equation for a spring–mass–damper system in which M is the mass of the bob, β is the friction coefficient, k is the spring constant and r' is the external force on the bob (e.g. Dorf and Bishop, 2005, chapter 2).

We shall use our TOA and SFC models, and the zero-dimensional model to which they can be reduced, to study the various categories of feedback that were described in Section 1. These categories will be designated as follows:

- F1 – feedback according to the control-theory definition;
- F2 – feedback according to the electronics definition;
- F3 – stability-altering feedback as used in climate studies;
- F4 – sensitivity-altering feedback as used in climate studies.

3.1. A control-theory perspective on the two-zone models

At this point, we express our generic two-zone governing equations in the formalism of control theory, following Dorf and Bishop (2005). In control theory, the transfer function of a linear system is defined as the ratio of the Laplace transform of the output variable to the Laplace transform of the input variable, with all initial conditions assumed to be zero. A closed-loop feedback control system is conventionally represented as a block diagram as in Figure 1. Here, the signals are Laplace transforms of the system variables, these transforms being functions of the complex Laplace variable s. The input signal $R(s)$ is combined at the summing point (represented by the open circle in the figure) with the fed-back signal $B(s)$; in the case of a negative (positive) feedback, $B(s)$ is subtracted from (added to) $R(s)$, giving the actuating signal

$$A(s) = R(s) \mp B(s). \tag{29}$$

The output signal $Y(s)$ is related to the actuating signal by

$$Y(s) = G(s)A(s), \tag{30}$$

DOI: 10.1002/qj

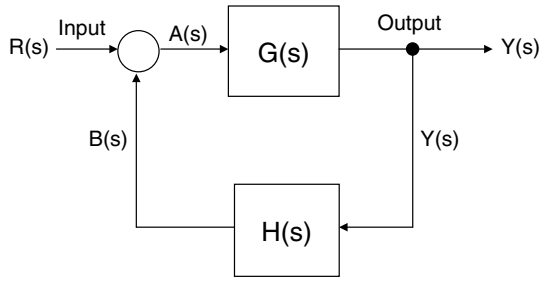

Figure 1. Block diagram for a single-loop feedback control system with forward gain $G(s)$ and feedback factor $H(s)$.

where $G(s)$ is the forward gain. The fed-back signal is the output signal multiplied by the feedback factor $H(s)$:

$$B(s) = H(s)Y(s). \tag{31}$$

Note that, while at the summing point the output is the difference or sum of the inputs, at the pick-off point (represented by the small full circle in Figure 1) the signal is transmitted undiminished in two directions. Using Equation (31) to eliminate $B(s)$ from Equation (29), and then using the result to eliminate $A(s)$ from Equation (30), we obtain:

$$Y(s) = G(s)\left(R(s) \mp H(s)Y(s)\right). \tag{32}$$

From this we obtain the standard expression for the closed-loop transfer function:

$$G_{\mathrm{F}}(s) \equiv \frac{Y(s)}{R(s)} = \frac{G(s)}{1 \pm G(s)H(s)}. \tag{33}$$

A positive sign in the denominator indicates that the sign of feedback F1 is negative, a negative sign that it is positive. Note that the sign of feedback F1 is independent of the form of the input signal and is not restricted in its connotation to a transient or steady-state response. In this respect, as we shall see, feedback F1 differs from feedbacks F2, F3 and F4.

To relate our model to the above control-theory formalism, we take the Laplace transforms of Equations (22) and (23). Assuming zero initial conditions (i.e. all perturbation quantities zero for $t < 0$), we then obtain

$$\left(Ms^2 + \beta s + k\right)\hat{T}_1(s) = \hat{r}_1(s), \tag{34}$$

$$\left(Ms^2 + \beta s + k\right)\hat{T}_2(s) = \hat{r}_2(s), \tag{35}$$

with

$$\hat{r}_1(s) = (c_0 s + \alpha_4)\hat{F}_1(s) - \alpha_2\hat{F}_2(s), \tag{36}$$

$$\hat{r}_2(s) = (c_0 s + \alpha_1)\hat{F}_2(s) - \alpha_3\hat{F}_1(s). \tag{37}$$

In the above, the Laplace transform $\mathcal{L}\{f(t)\} = \hat{f}(s)$ of a function of time $f(t)$ is defined by

$$\hat{f}(s) = \int_{0^-}^{\infty} \mathrm{e}^{-st} f(t)\,\mathrm{d}t, \tag{38}$$

and $\hat{T}_i(s)$, $\hat{r}_i(s)$ and $\hat{F}_i(s)$ represent the transforms of $T_i'(t)$, $r_i'(t)$ and $F_i'(t)$, respectively.

It is clear from Equations (34) and (35) that the variables T_1' and T_2' have the same transfer function $G_{\mathrm{F}}(s)$:

$$G_{\mathrm{F}}(s) = \frac{\hat{T}_1(s)}{\hat{r}_1(s)} = \frac{\hat{T}_2(s)}{\hat{r}_2(s)}, \tag{39}$$

where

$$G_{\mathrm{F}}(s) = \frac{1}{Ms^2 + \beta s + k}. \tag{40}$$

This transfer function can be written in the standard closed-loop form

$$G_{\mathrm{F}}(s) = \frac{G(s)}{1 + G(s)H(s)} \tag{41}$$

with six possible choices of $(G(s), H(s))$; for example, we can choose

$$G(s) = \frac{1}{Ms^2 + \beta s} \qquad H(s) = k. \tag{42}$$

Whichever one of the possible choices we take, the plus sign in the denominator in Equation (41) is always the relevant one as long as M, β and k are all positive. The parameter M is positive by definition, and we shall see below that β and k are positive if our system is stable. Assuming this to be the case, our generic two-zone system (Equations (18) and (19)) is dynamically equivalent to a negative-feedback control system: in other words, the sign of F1 for our system is negative. Note that Equations (39), (41) and (42), which were obtained by mathematical manipulation of the governing equations, are consistent with Equation (33), which was obtained by symbolic manipulations based on the block diagram of Figure 1.

Of primary interest in control theory are the stability and time behaviour of a system. To examine these characteristics, an impulse-function input signal is normally used, and we follow this approach here by assuming a forcing of the form

$$F_i' = (\Delta F_i)\delta(t), \tag{43}$$

where $\delta(t)$ is the delta function (zero everywhere except at $t = 0$). Since $\mathcal{L}\{\delta(t)\} = 1$, we have

$$\hat{F}_i(s) = \Delta F_i, \tag{44}$$

and Equations (36) and (37) then give

$$\hat{r}_1(s) = (c_0 s + \alpha_4)\Delta F_1 - \alpha_2\Delta F_2, \tag{45}$$

$$\hat{r}_2(s) = (c_0 s + \alpha_1)\Delta F_2 - \alpha_3\Delta F_1. \tag{46}$$

DOI: 10.1002/qj

Hence, Equations (39) and (40) give

$$\hat{T}_1(s) = \frac{(c_0 s + \alpha_4)\Delta F_1 - \alpha_2 \Delta F_2}{Ms^2 + \beta s + k}, \tag{47}$$

$$\hat{T}_2(s) = \frac{(c_0 s + \alpha_1)\Delta F_2 - \alpha_3 \Delta F_1}{Ms^2 + \beta s + k}. \tag{48}$$

After some manipulation, the inverse transforms of Equations (47) and (48) give the solutions

$$T_1'(t) = \frac{1}{c_0^2(R_\mathrm{F} - R_\mathrm{S})}\left[\{(c_0 R_\mathrm{F} - \alpha_4)\,\Delta F_1 + \alpha_2 \Delta F_2\}\exp(-R_\mathrm{F} t) - \{(c_0 R_\mathrm{S} - \alpha_4)\,\Delta F_1 + \alpha_2 \Delta F_2\}\exp(-R_\mathrm{S} t)\right] \tag{49}$$

and

$$T_2'(t) = \frac{1}{c_0^2(R_\mathrm{F} - R_\mathrm{S})}\left[\{(c_0 R_\mathrm{F} - \alpha_1)\,\Delta F_2 + \alpha_3 \Delta F_1\}\exp(-R_\mathrm{F} t) - \{(c_0 R_\mathrm{S} - \alpha_1)\,\Delta F_2 + \alpha_3 \Delta F_1\}\exp(-R_\mathrm{S} t)\right], \tag{50}$$

where the characteristic rates of decay R_F and R_S are defined by

$$R_\mathrm{F} = \frac{1}{c_0}\frac{\alpha_1 + \alpha_4}{2}\left(1 + \sqrt{1 - x}\right), \tag{51}$$

$$R_\mathrm{S} = \frac{1}{c_0}\frac{\alpha_1 + \alpha_4}{2}\left(1 - \sqrt{1 - x}\right), \tag{52}$$

with

$$x = 4\frac{\alpha_1\alpha_4 - \alpha_2\alpha_3}{(\alpha_1 + \alpha_2)^2}. \tag{53}$$

These solutions satisfy the initial conditions $T_i' = \Delta F_i / c_0$ at $t = 0$, and it can easily be verified by substitution that they satisfy the original differential equations, Equations (18) and (19), for $t > 0$. (Note that R_F and R_S can also be viewed as the decay rates of the model's fast and slow normal modes: see (Bates, 1999)).

The conditions for asymptotic stability of the solutions given by Equations (49) and (50) are that the real parts of R_F and R_S be positive. These conditions are satisfied if

$$\alpha_1 + \alpha_4 > 0 \qquad \alpha_1\alpha_4 - \alpha_2\alpha_3 > 0. \tag{54}$$

The above conditions can also be expressed as

$$\beta > 0 \qquad k > 0; \tag{55}$$

in other words, the stability conditions are that the 'friction coefficient' and 'spring constant' both be positive. It is assumed everywhere in this paper that these conditions are satisfied (and when numerical values are assigned they are chosen so that this is the case).

Feedback F3 is defined with respect to the asymptotic response of a system to an impulse-function input such as Equation (43), being negative if a specified interaction gives increased asymptotic rates of decay of the impulsively-induced perturbation and positive if it gives decreased asymptotic rates of decay. Clearly, this is a different concept of feedback from F1, though it describes an aspect of the system's behaviour that is of interest in control theory. The sign of feedback F3 in our simple climate models will be considered later.

Another aspect of a system's performance of interest in control theory is its response to a step-function input. We study this aspect here by taking

$$F_i' = (\Delta F_i)1(t), \tag{56}$$

where $1(t)$ is the unit step function, which has the value zero for $t < 0$ and the value 1 for $t \geq 0$. The complete time-dependent response of our generic two-zone system to such an input can be obtained using the method of Laplace transforms, as for the impulse-function input. Here, however, we confine our attention to the steady-state response to the forcing given by Equation (56) after the transients have died out. The steady-state response can be simply obtained by setting all time derivatives to zero in Equations (22) and (23), whence we have

$$T_i'(\infty) = \frac{1}{k} r_i'(\infty) = G_\mathrm{F}(0) r_i'(\infty), \tag{57}$$

where $r_1'(\infty)$ and $r_2'(\infty)$ (the values at $t = \infty$) are given by Equations (27) and (28) with F_i' replaced by ΔF_i, and $G_\mathrm{F}(0)$ is given by Equation (40). Thus, the steady-state temperature increment in each zone in response to a step-function input is the steady-state value of the effective forcing in that zone divided by the 'spring constant', or multiplied by the steady-state transfer function.

Feedback F4 is defined in terms of the steady-state response of the climate system to a step-function forcing such as Equation (56), being negative if a specified interaction causes a decreased response and positive if it causes an increased response. Again, this is a different concept of feedback from F1, though it also describes an aspect of the system's behaviour that is of interest in control theory. The sign of feedback F4 in our simple climate models will be considered later.

To investigate the sign of feedback F2 in our generic two-zone system, we take a sinusoidal input

$$F_i' = (\Delta F_i)\mathrm{e}^{\mathrm{i}\omega t}, \tag{58}$$

where ω is the frequency. It is easily shown that the steady-state response of our system governed by Equations (22) and (23) to this input is an output signal of the same frequency satisfying

$$T_i' = G_\mathrm{F}(\mathrm{i}\omega) r_i', \tag{59}$$

where r_1' and r_2' are given by Equations (27) and (28), F_1' and F_2' by Equation (58), and $G_\mathrm{F}(\mathrm{i}\omega)$ by Equation

DOI: 10.1002/qj

(40). Choosing $G(s)$ and $H(s)$ as given by Equation (42), $G_F(i\omega)$ can be written:

$$G_F(i\omega) = \frac{G(i\omega)}{1 + G(i\omega)k}. \tag{60}$$

In the above, electronics theory is seen as a special case of control theory. However, the field of electronics has its own convention for defining the sign of feedback. According to the electronics definition (Smith, 1987, chapter 15), an amplifier provides a positive or negative feedback for a sinusoidal input signal of frequency ω according as the magnitude of the steady-state gain with feedback is greater or less than the magnitude of the steady-state gain without feedback. Applying this convention here, our system provides a positive or negative feedback F2 according as $G_F(i\omega)$ is greater or less in magnitude than $G(i\omega)$. For a system with $G(s)$ of the form given by Equation (42), it is easily seen from Equation (60) that

$$\frac{|G_F(i\omega)|}{|G(i\omega)|} = \left(\frac{M^2\omega^4 + \beta^2\omega^2}{M^2\left(\omega^2 - \omega_n^2\right)^2 + \beta^2\omega^2} \right)^{\frac{1}{2}}, \tag{61}$$

where $\omega_n = (k/M)^{1/2}$. (In terms of the spring–mass–damper analogy, ω_n is the natural frequency of oscillation of the system.) From Equation (61) it can be seen that the feedback F2 for our system is positive or negative according as ω is greater or less than $\omega_n/\sqrt{2}$. But we have seen that for our generic two-zone model the feedback F1 is negative for all input signals. Thus, we have an example of a system in which the feedbacks F1 and F2 can be of opposite signs.

Our model results show that transferring the gain formula of electronics (which refers to the steady-state response of a system to a sinusoidal input) to the climatic context (where it is meant to apply to the steady-state response to a step-function input) is not as straightforward as is often assumed. When $\omega \neq 0$, we see from Equation (60) that $G_F(i\omega)$ can be written in the closed-loop form $G/(1 - f)$ (in the climate literature, the loop gain GH is usually replaced by the quantity f, which is called the feedback); however, $G_F(i\omega)$ is then complex and not equal to the model's steady-state gain for a step-function input, $G_F(0)$. In the limit as $\omega \to 0$, $G_F(i\omega)$ becomes real and equal to the steady-state gain for a step-function input, but then it does not have the form $G/(1 - f)$. It is possible, of course, to define a G and an f independently of any governing equations in such a way that $G_F(0)$ *can* be written as $G/(1 - f)$, but any direct relationship with control theory is then lost and calling G a forward gain and f a feedback becomes purely figurative.

3.2. The zero-dimensional model

If the dynamical interaction terms are omitted and the local stabilizing coefficients are regarded as having the same (positive) value in both zones, our generic two-zone system of Equations (18) and (19) reduces, with an assumption of interhemispheric symmetry, to a simple equation for the global average SST perturbation, of the form

$$c_0 \frac{dT'}{dt} = -bT' + F'. \tag{62}$$

A model represented by Equation (62) is usually referred to as a zero-dimensional model. The TOA version of this equation is of the form used in conceptual discussions of climate feedbacks in the IPCC context (IPCC, 2001, chapter 9). The considerations of this subsection refer equally to the TOA and SFC versions. We here examine in the context of this model the various concepts of feedback defined earlier.

We begin, once more, by adopting a control-theory perspective. Taking the Laplace transform of Equation (62), and assuming zero initial conditions, we obtain

$$c_0 s\hat{T}(s) = -b\hat{T}(s) + \hat{F}(s), \tag{63}$$

where $\hat{T}(s)$ and $\hat{F}(s)$ represent the transforms of $T'(t)$ and $F'(t)$ respectively. Thus our transfer function $G_F(s) \equiv \hat{T}(s)/\hat{F}(s)$ is given by

$$G_F(s) = \frac{1}{c_0 s + b}. \tag{64}$$

We see that $G_F(s)$ can be written in the closed-loop form of Equation (41) with two possible choices of $(G(s), H(s))$; for example, we can choose

$$G(s) = \frac{1}{c_0 s} \qquad H(s) = b. \tag{65}$$

Whichever choice we make, the plus sign in the denominator of Equation (41) is the relevant one as long as c_0 and b are positive. The parameter c_0 is positive by definition, and b is positive if the system represented by Equation (62) is stable (see below). Assuming this to be the case, the zero-dimensional model is dynamically equivalent to a negative-feedback control system; i.e. the sign of F1 for the model is negative.

If we take an impulse-function input $F' = (\Delta F)\delta(t)$, we have $\hat{F}(s) = \Delta F$ and Equation (64) gives

$$\hat{T}(s) = \frac{\Delta F}{c_0} \frac{1}{s + R}, \tag{66}$$

where $R = b/c_0$. The inverse transform gives the solution

$$T'(t) = \frac{\Delta F}{c_0} \exp(-Rt), \tag{67}$$

showing that the system is stable if $b > 0$.

If we take a step-function input $F' = (\Delta F)1(t)$, the steady-state solution can be seen directly from Equation (62) to be

$$\Delta T = \frac{\Delta F}{b}, \tag{68}$$

DOI: 10.1002/qj

where we have set $T'(\infty) = \Delta T$. Clearly, any interaction between physical processes that *increases* b will increase the asymptotic rate of decay of a perturbation caused by an impulsive forcing and decrease the magnitude of the steady-state perturbation caused by a step-function forcing; in other words, the feedbacks F3 and F4 resulting from the interaction are both negative. Similarly, in the case of an interaction that *decreases* b, the feedbacks F3 and F4 are both positive. Thus, for the zero-dimensional model, the common assumption regarding the equivalence in sign of feedbacks F3 and F4 is valid. It is clear, however, that for this system the sign of these feedbacks can be opposite to that of feedback F1: feedbacks F3 and F4 can be of either sign, depending on the change in b, but feedback F1 remains negative for any change in b as long as b remains positive.

From Equation (67) it is clear that in the zero-dimensional model the initial and asymptotic tendencies $\mathrm{d}T'/\mathrm{d}t$ of an impulsively-forced perturbation are of the same sign. (For a perturbation that is impulsively forced at $t = 0$, 'initial values' and 'initial tendencies' will mean the values and tendencies at $t = 0^+$.) Furthermore, any changes in the initial and asymptotic tendencies caused by a change in b are of the same sign: in other words, for this model, the sign of feedback F3 (which is defined in terms of changes in asymptotic tendencies) can be correctly derived from changes in initial tendencies. We shall later present examples of cases where these features do not carry over to a two-zone model.

To investigate the sign of feedback F2 in the zero-dimensional model, we take a sinusoidal input $F' = (\Delta F)\exp(\mathrm{i}\omega t)$. The steady-state response to this input signal can be seen from Equation (62) to be

$$T' = G_{\mathrm{F}}(\mathrm{i}\omega)F', \tag{69}$$

where $G_{\mathrm{F}}(\mathrm{i}\omega)$ is given by Equation (64). Using Equation (65) (or the other possible choice of $(G(s), H(s))$), it is easily seen that $|G_{\mathrm{F}}(\mathrm{i}\omega)| < |G(\mathrm{i}\omega)|$ for all ω when $b > 0$; thus in the zero-dimensional model the sign of feedback F2 is negative, coinciding with the sign of F1. For this model we therefore have F1 and F2 of the same sign *and* F3 and F4 of the same sign, but the signs of these two pairs of feedbacks do not necessarily coincide. Thus, it is not possible in the zero-dimensional model to equate precisely the stability-altering and sensitivity-altering concepts of feedback with the control-theory and electronics concepts.

In what follows, we show that when one moves from the zero-dimensional to the two-zone models, one introduces further complexity in the relationships between the signs of our four selected feedbacks.

3.3. Stability-altering and sensitivity-altering feedbacks

3.3.1. The TOA model

The sign of feedback F3 in the two-zone models is determined by examining how the dynamical interaction terms influence the magnitudes of the real parts of the characteristic decay rates R_{F} and R_{S} defined by Equations (51) and (52). The sign of feedback F4 is determined by examining how the dynamical interaction terms influence the magnitude of the steady-state hemispheric mean SST perturbation caused by a step-function forcing. Here and in the remainder of this section, we shall take the step-function forcing as being that due to a doubling of the atmospheric CO_2 content. For greater physical transparency, we expand Equation (57) into the more explicit forms

$$\Delta T_1 = \frac{\alpha_4 \Delta F_1 - \alpha_2 \Delta F_2}{\alpha_1\alpha_4 - \alpha_2\alpha_3}, \tag{70}$$

$$\Delta T_2 = \frac{\alpha_1 \Delta F_2 - \alpha_3 \Delta F_1}{\alpha_1\alpha_4 - \alpha_2\alpha_3}, \tag{71}$$

where we have set $T_i'(\infty) = \Delta T_i$.

For the TOA model, we choose $(B_1, B_2) = (1.8, 1.6)\ \mathrm{Wm^{-2}K^{-1}}$. (These are close to the value 1.7 $\mathrm{Wm^{-2}\ K^{-1}}$ calculated by Nakamura *et al.* (1994) using a 1D radiative–convective model, and used in (Bates, 1999).) We choose $(a_{Z1}, a_{Z2}) = (28, 16)\ \mathrm{mK^{-1}}$. (These are the observational values used in (Bates, 2004).) The dynamical coefficient a_{E}, which determines the rate of poleward transport of moist static energy by baroclinic eddies between the zones, is allowed to vary between 0 and 0.006 $\mathrm{PW\,m^{-1}}$ ($1\,\mathrm{PW} = 10^{15}\,\mathrm{W}$), the upper limit corresponding to the observational value used in (Bates, 1999). Our coefficients on the right-hand side of Equations (8) and (9) thus have the values $(\hat{b}_1, \hat{b}_2) = (0.229, 0.204)\ \mathrm{PW\,K^{-1}}$, and with the value of a_{E} set at its upper limit, $(\hat{d}_1, \hat{d}_2) = (0.168, 0.096)\ \mathrm{PW\,K^{-1}}$.

The asymptotic stability properties of the TOA model for these parameter values are illustrated in Figure 2, where the characteristic decay rates R_{F} and R_{S} given by Equations (51), (52) and (20) are plotted as functions of a_{E}. It can be seen that R_{F} and R_{S} (both real here) increase monotonically with a_{E} throughout its range. Since the interzonal dynamical transport coefficients $\hat{d}_1$ and $\hat{d}_2$ are proportional to a_{E}, we see that the interaction between the dynamical transport terms and the local stabilizing terms increases the asymptotic stability of the system relative to the zero-interaction case; i.e. the sign of feedback F3 provided by this interaction is negative.

To illustrate the sensitivity properties of the TOA model, we need to assign values to the prescribed steady forcing at the top of the atmosphere in zones 1 and 2 corresponding to a CO_2 doubling. Values for the zonally-averaged forcing as a function of latitude from the NCAR Atmospheric General Circulation Model for such a doubling have been presented by Harvey (2000) (see his figure 7.3). Approximating from his graph of the forcing at the tropopause, we take $(\Delta F_1, \Delta F_2)_{\mathrm{T}} = \pi a^2[(4, 3)\ \mathrm{Wm^{-2}}]$. The sensitivity properties of the TOA model with this forcing, and the remaining parameters as before, are illustrated in Figure 3, where the equilibrium temperature increments ΔT_1 and ΔT_2 given by Equations (70) and (71) and their mean ΔT_{m} are plotted as functions

DOI: 10.1002/qj

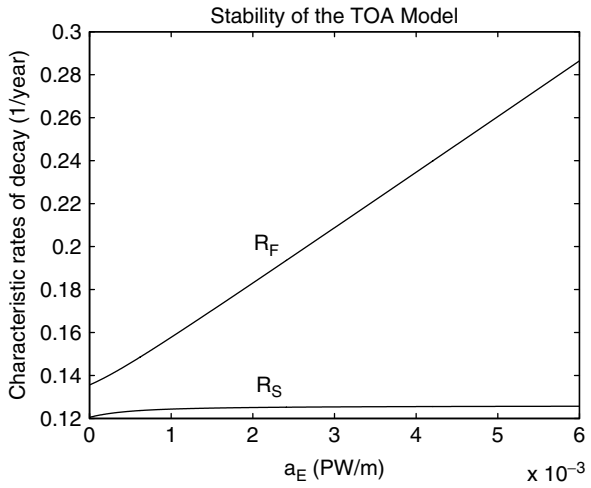

Figure 2. Characteristic rates of decay of impulsively-forced perturbations in the TOA model as functions of the dynamical interaction coefficient a_E, which varies between 0 and 0.006 PW m^{-1}. The remaining parameters have the values $(B_1, B_2) = (1.8, 1.6)\,\mathrm{Wm^{-2}K^{-1}}$, $(a_{Z1}, a_{Z2}) = (28, 16)\,\mathrm{mK^{-1}}$, and $D = 100\,\mathrm{m}$.

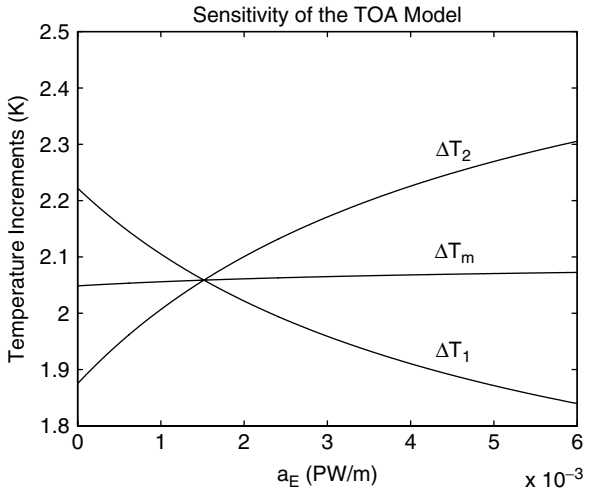

Figure 3. Equilibrium temperature increments ΔT_1 and ΔT_2 for a CO_2 doubling in the TOA model, and their mean ΔT_m, as functions of the dynamical interaction coefficient a_E, which varies between 0 and 0.006 PW m^{-1}. The top-of-the-atmosphere radiative forcings are taken as $(\Delta F_1, \Delta F_2)_T = \pi a^2[(4, 3)\,\mathrm{Wm^{-2}}]$, and the remaining parameters are as in Figure 2.

of a_E. It can be seen that ΔT_1 decreases as a_E increases over its allowed range, while ΔT_2 increases. The mean temperature increment ΔT_m shows a slight monotonic increase over the range. Thus, the interaction between the dynamical transport terms and the local stabilizing terms increases the sensitivity of the system as measured by ΔT_m relative to the zero-interaction case for all values of a_E considered; i.e. the sign of feedback F4 provided by the interaction is positive.

3.3.2. *The SFC model*

For the SFC model, we choose $(b_1, b_2) = (0.17, 0.19)$ PW K^{-1}; these are to be compared with the observational estimates $(b_1, b_2) = (0.56, 0.15)\,\mathrm{PW\,K^{-1}}$ used in (Bates, 2004). The dynamical coefficient a_M, which determines the rate of poleward transport of angular momentum by baroclinic eddies between the zones, is allowed to vary between 0 and 0.13 Hadley m^{-1} (1 Hadley $= 10^{18}\,\mathrm{kg\,m^2\,s^{-2}}$), the upper limit being the observed value found by Alexeev and Bates (2000). We choose $((\overline{F}_L)_1, (\overline{F}_L)_2) = (13, 11)$ PW, $\overline{F}_M = 30$ Hadley, and $(a_{Z1}, a_{Z2}) = (28, 24)\,\mathrm{mK^{-1}}$. Thus, the d_{ij} vary between 0 and the values $(d_{11}, d_{12}, d_{21}, d_{22}) = (0.789, 0.676, 0.667, 0.572)\,\mathrm{PW\,K^{-1}}$. These are to be compared with the observational estimates $(d_{11}, d_{12}, d_{21}, d_{22}) = (0.98, 0.56, 0.30, 0.17)\,\mathrm{PW\,K^{-1}}$ used in (Bates, 2004).

The asymptotic stability properties of the SFC model for these parameter values are illustrated in Figure 4, where the characteristic decay rates R_F and R_S given by Equations (51), (52) and (21) are plotted as functions of a_M. It can be seen that, apart from a small region to the left of the figure, the real parts of R_F and R_S (both now complex) increase monotonically with a_M and exceed the values at $a_M = 0$. Since the interzonal dynamical transport coefficients d_{ij} are proportional to a_M, we see that for almost the whole range of variation of these coefficients the interaction between the dynamical transport terms and the local stabilizing terms increases the asymptotic stability of the system; i.e for most of the range the sign of feedback F3 provided by the interaction is negative.

To illustrate the sensitivity properties of the SFC model, we need to assign values of the prescribed steady forcing at the surface in zones 1 and 2 corresponding to a CO_2 doubling. These are again taken from (Harvey, 2000). Approximating from Harvey's graph of the surface forcing, we take $(\Delta F_1, \Delta F_2)_G = \pi a^2[(1, 2)\,\mathrm{Wm^{-2}}]$. The sensitivity properties of the SFC model with this forcing, and the remaining parameters as above, are illustrated in Figure 5, where the equilibrium temperature increments

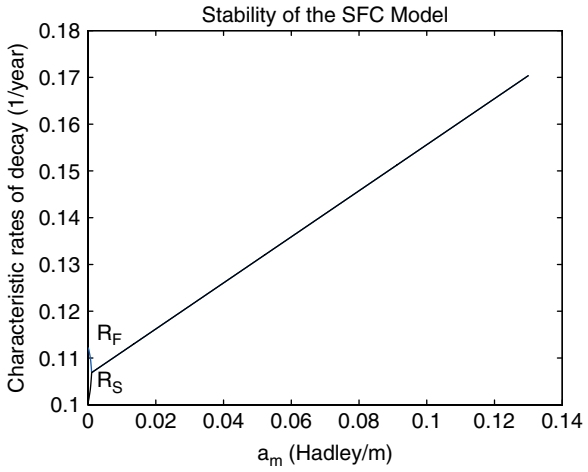

Figure 4. Characteristic rates of decay of impulsively-forced perturbations in the SFC model as functions of the dynamical interaction coefficient a_M, which varies between 0 and 0.13 Hadley m^{-1}. The remaining parameters have the values $(b_1, b_2) = (0.17, 0.19)\,\mathrm{PW\,K^{-1}}$, $(a_{Z1}, a_{Z2}) = (28, 24)\,\mathrm{mK^{-1}}$, $((\overline{F}_L)_1, (\overline{F}_L)_2) = (13, 11)$ PW, $\overline{F}_M = 30$ Hadley, and $D = 100\,\mathrm{m}$. This figure is available in colour online at www.interscience.wiley.com/qj

DOI: 10.1002/qj

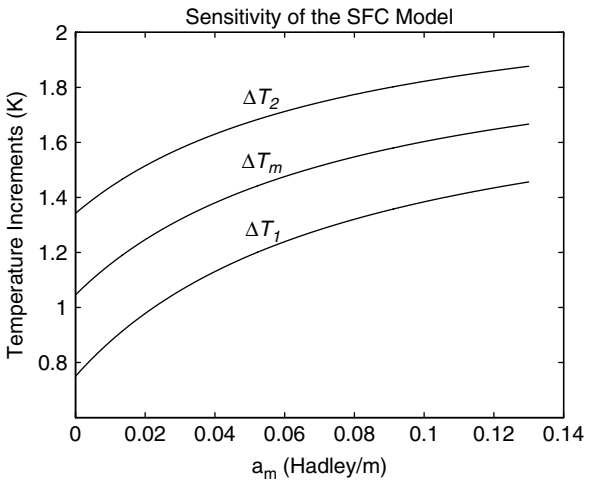

Figure 5. Equilibrium temperature increments ΔT_1 and ΔT_2 for a CO_2 doubling in the SFC model, and their mean ΔT_m, as functions of the dynamical interaction coefficient a_M, which varies between 0 and 0.13 Hadley m^{-1}. The surface radiative forcings are taken as $(\Delta F_1, \Delta F_2)_G = \pi a^2[(1, 2)\,\mathrm{Wm}^{-2}]$, and the remaining parameters are as in Figure 4.

ΔT_1 and ΔT_2 given by Equations (70) and (71) and their mean ΔT_m are plotted as functions of a_M. It can be seen that ΔT_1, ΔT_2 and ΔT_m all show a monotonic and significant increase as a_M increases over its allowed range. Thus, for the full range, the sign of feedback F4 provided by the dynamical interaction is positive.

Thus the same conclusion emerges from both the TOA and the SFC models: for certain regions of parameter space, the interaction between the interzonal dynamical transport and the local stabilizing processes provides feedbacks F3 and F4 that are of opposite signs. The result is more clear-cut in the case of the SFC model, where, for most of the parameter range considered, the interaction more clearly stabilizes the model and more clearly increases its sensitivity relative to the zero-interaction case.

An aspect of the sensitivity of the SFC model that has not been discussed here is the effect on ΔT_m of different distributions of the external forcing with latitude. This question has been examined in some detail in (Bates, 2004); it is shown there that, for a given hemispheric average forcing, the value of ΔT_m is quite sensitive to how the forcing is distributed between the tropical and extratropical zones. Such a result has also been found in the GCM experiments of Alexeev *et al.* (2005).

3.4. Initial versus asymptotic tendencies in the two-zone models

Here we examine the relationship between initial and asymptotic tendencies of impulsively-forced perturbations in the TOA and SFC models, as well as considering whether initial tendencies provide reliable indications of the sign of feedback F3. An impulsive forcing of the form of Equation (43) is assumed. As noted earlier, this gives the initial perturbation $T_i' = \Delta F_i/c_0$ and the solution given by Equations (49) and (50) for $t > 0$.

We first consider the TOA model, with the parameters chosen as in Section 3.3.1 above. The model is then asymptotically stable and the initial perturbation decays asymptotically to zero. However, it is possible under these circumstances to choose the form of the initial perturbation so that the hemispheric mean SST perturbation T_m' is initially growing though asymptotically decaying. To see this, we take the sum of the governing equations (Equations (8) and (9)) for the TOA model and divide by two; thus, for $t > 0$,

$$c_0 \frac{dT_m'}{dt} = -\frac{1}{2}(\hat{b}_1 T_1' + \hat{b}_2 T_2'). \tag{72}$$

Clearly, $T_m' > 0$ and $dT_m'/dt > 0$ initially for any initial conditions that satisfy $T_1' < 0$, $T_2' > 0$ and

$$\frac{\hat{b}_2}{\hat{b}_1} T_2' < \left|T_1'\right| < T_2'.$$

Since our chosen parameters are such that $\hat{b}_2 < \hat{b}_1$, there is a range of initial conditions that satisfy the above requirements. It is to be noted, however, that the individual perturbations T_1' and T_2' both show initial decay when the above conditions are satisfied. (This is easily seen from Equations (8) and (9) using only the conditions $T_1' < 0$ and $T_2' > 0$.) Thus, for the TOA model, it is possible to have initial growth and asymptotic decay for T_m', though not simultaneously for T_1' and T_2'. An example illustrating the above considerations is shown in Figure 6. It is also possible, with the same parameters, to have initial growth and asymptotic decay for one, but not both, of the individual perturbations T_1' and T_2', but then it is not possible to have it for T_m'. Clearly, in the TOA model there is no simple relationship between initial and asymptotic tendencies such as exists for the zero-dimensional model.

We next consider the SFC model, with the parameters chosen as in Section 3.3.2 above. With these values, the SFC model is asymptotically stable. Here it is possible to choose initial conditions for which all three of T_1', T_2' and T_m' show initial growth and asymptotic decay. To see this, consider the case where initially both $T_1' > 0$ and $T_2' > 0$. From Equations (14) and (15) we then see that, provided $d_{22} > b_2$ (which is satisfied with our chosen parameters for the upper part of the range of a_M), we have both $dT_1'/dt > 0$ and $dT_2'/dt > 0$ initially, provided that

$$\frac{T_1'}{T_2'} < \frac{d_{12}}{b_1 + d_{11}}, \tag{73}$$

$$\frac{T_1'}{T_2'} < \frac{d_{22} - b_2}{d_{21}}. \tag{74}$$

It is easy to choose initial conditions such that Equations (73) and (74) are both satisfied, giving the required result regarding the initial growth of both T_1' and T_2'. But when both of these quantities are positive and have initial growth, their mean T_m' is also positive

DOI: 10.1002/qj

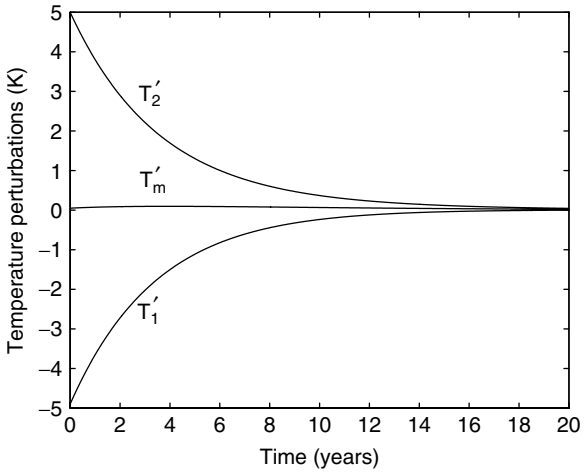

Figure 6. An example of an impulsively-forced perturbation in the TOA model showing initial growth and asymptotic decay of T'_m. The corresponding evolution of T'_1 and T'_2 is also shown. The initial conditions are $(T'_1(0), T'_2(0)) = (-4.9, 5.0)$ K. The parameters are as in Figure 2, with $a_E = 0.006\,\mathrm{PW\,m^{-1}}$.

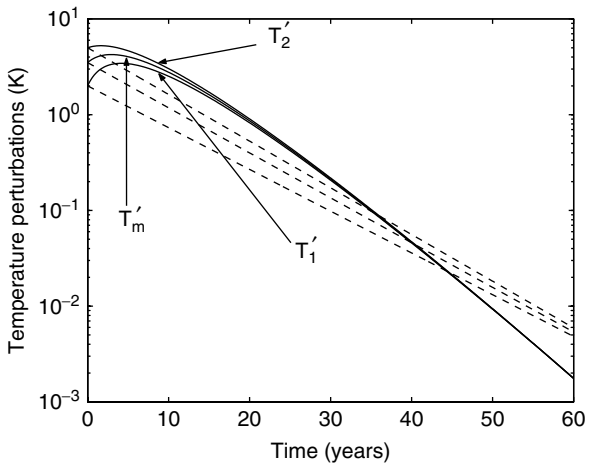

Figure 7. An example of an impulsively-forced perturbation in the SFC model. The initial conditions are $(T'_1(0), T'_2(0)) = (2.0, 5.0)$ K. The solid curves show the initial growth and asymptotic decay of T'_1, T'_2 and T'_m when the dynamical interaction is present with full strength ($a_M = 0.13\,\mathrm{Hadley\,m^{-1}}$); the dashed curves show the corresponding evolution of the perturbation when the dynamical interaction is absent ($a_M = 0$). The remaining parameters are as in Figure 4.

and has initial growth. Meanwhile, for the given initial conditions, asymptotic stability guarantees decay of all three quantities as $t \to \infty$. An example illustrating these considerations is shown by the solid curves in Figure 7, where a_M has been set to the maximum value in its assigned range ($a_M = 0.13\,\mathrm{Hadley\,m^{-1}}$) and initial conditions satisfying Equations (73) and (74) have been chosen. The initial growth and asymptotic decay of all three of T'_1, T'_2 and T'_m is clearly seen.

The solid curves in Figure 7 represent the case where the dynamical interaction terms are present with full strength, while the dashed curves show the corresponding results when the dynamical interaction terms are set to zero ($a_M = 0$). The latter curves represent the zero-interaction case, relative to which the effect of the dynamical interaction on the perturbation is measured. Comparing the solid with the dashed curves, we see that the asymptotic effect of the dynamical interaction is opposite to its initial effect. Asymptotically, the interaction is stability-enhancing, increasing the rate of decay of the perturbation and thus providing a negative feedback F3 (in accord with the results of Section 3.3.2 above). Judged in terms of initial tendencies, the interaction is stability-diminishing, giving growth where decay otherwise occurs. We therefore have an example showing that estimating the sign of feedback F3 on the basis of initial (or instantaneous) tendencies can give a false result. It is not possible to find such a clear illustration of this using the TOA model. (There, as can be seen from Equation (72), the dynamical interaction terms have no influence on the initial tendency of T'_m.)

3.5. Summary of results

We summarize in Table I the results of our investigation of the signs of our four selected forms of feedback in the simple climate models. Only situations where the models are stable have been considered. In the case of the zero-dimensional model, our results are quite general, applying throughout the full stable region of the model's parameter space. The same is true for the two-zone models as far as feedbacks F1 and F2 are concerned, but in discussing feedbacks F3 and F4 only a particular region of parameter space has been considered in each of the TOA and SFC models.

It can be seen that from the point of view of feedback the zero-dimensional model is relatively simple in that the signs of feedbacks F1 and F2 coincide and the signs of feedbacks F3 and F4 coincide. However, the signs of these two pairs of feedbacks do not necessarily coincide with each other. Thus, even in the case of the simplest model, one cannot draw any direct parallel between feedback as used in control theory or electronics and the two prototype concepts of feedback used in climate studies.

In the two-zone models, there is much greater complexity in the signs of the feedbacks. Here, differences in sign can arise both within and between the two pairs. In the case of the pair F3 and F4, our model parameters have been chosen specifically to demonstrate that, for both the TOA and SFC models, these two forms of feedback can be of opposite sign in the same region of parameter space.

The zero-dimensional model is also relatively simple in that the initial and asymptotic tendencies of an impulsively-forced perturbation are always of the same sign, and the sign of feedback F3 can always be derived from changes in initial tendencies. With a two-zone model, on the other hand, a perturbation can grow initially while decaying asymptotically and changes in initial tendencies can give an incorrect indication of the sign of feedback F3.

DOI: 10.1002/qj

Table I. Signs of the feedbacks in the simple models.

Category of feedback	Zero-dimensional model	Two-zone models (TOA and SFC)
F1 (control theory definition)	Negative	Negative
F2 (electronics definition)	Negative	Either sign[b]
F3 (stability-altering definition)	Either sign[a]	Negative for the chosen parameters[c]
F4 (sensitivity-altering definition)	Same as for F3	Positive for the same chosen parameters[c]

[a] Depends on whether an interaction between physical processes increases or decreases the parameter b relative to some specified zero-interaction case.
[b] Depends on the frequency of the input signal.
[c] The feedback considered here is that provided by the interaction between the interzone dynamical transport terms and the local stabilizing terms.

4. Discussion and conclusions

A conceptual study of climate feedbacks has been carried out using two simple two-zone climate models, called the TOA and SFC models, that incorporate dynamical interaction between the Tropics and the Extratropics through baroclinic eddy transports. The models are linear and admit of analytical solution. A zero-dimensional model, to which both of the two-zone models reduce under simplifying assumptions, is used as a reference for comparing the feedback properties of the two-zone models.

Our study is carried out against the background of control theory, which provides a general framework for the concept of feedback. In an automatic control system, a measure of the output is used as a feedback signal to control the system. Negative feedback is said to occur when this signal is fed back so that it subtracts from the input signal, positive feedback when it adds to the input signal. Mathematically, the operation of the system is modelled by a closed-loop transfer function that explicitly shows the sign of the feedback and that allows the output to be calculated for arbitrary forms of input. In control theory, the sign of the feedback is already defined before the system's response to any particular form of input is considered.

Historically, 'feedback' as a technical term originated in the field of electronics. Feedback in electronics again involves an actual signal being fed back to the input; but here a specific form of input is assumed, namely, a sinusoidal signal. The sign of the feedback is defined in terms of the system's steady-state response to such a signal. Negative feedback is said to occur when the magnitude of the system's gain with feedback is less than the magnitude of its gain without feedback, and vice versa for positive feedback. We have shown here, using our simple climate models, that a system whose feedback is negative in the control-theory sense (as judged by the form of its transfer function) can have a feedback that is positive in the electronics sense (depending on the frequency of the input signal).

In climate studies, feedback is frequently assumed to have a well-defined meaning, based on one or other of the above paradigms. Close examination, however, reveals different usages that can be in mutual conflict and that may not be in close accord with the assumed paradigms. Two prototype usages, stability-altering feedback and sensitivity-altering feedback, have been isolated for study. The following definitions of these, suggested by actual and widespread usage in the climate literature, are proposed:

- If an interaction between physical, chemical or biological processes alters the asymptotic stability of the global climate system or its components, the interaction is said to provide a *stability-altering* feedback. A negative feedback increases the stability of the system relative to the zero-interaction case, causing or contributing to the asymptotic decay of an initial impulsively-forced perturbation; and vice versa for a positive feedback.
- If an interaction between physical, chemical or biological processes alters the equilibrium deviation of the globally-averaged surface temperature from its initial equilibrium value under the influence of a step-function forcing, the interaction is said to provide a *sensitivity-altering* feedback. A negative feedback decreases the equilibrium deviation relative to the zero-interaction case; a positive feedback increases it.

The stability-altering and sensitivity-altering feedbacks are defined with respect to specific forms of forcing – an impulse function and a step function, respectively – and their signs are defined in terms of the system's response to the forcing in question. Thus, neither of them corresponds to the concept of feedback used in control theory, though each of them describes a specific aspect of a system's behaviour that is of interest in control theory. Neither do they correspond in any literal sense to the concept of feedback as used in electronics, though the amplification formula of electronics, $G_F = G/(1 - GH)$, for the special case where the loop gain $GH \equiv f$ is real,

DOI: 10.1002/qj

is often used figuratively in discussing sensitivity-altering feedback. The figurative transfer of an amplification formula from another field into the climate area must not be seen as implying that some general physical principle is being invoked.

The signs of the four feedbacks described above have been examined in the context of the two-zone models and the zero-dimensional model, on the basis of the actual governing equations rather than any figurative considerations. From the point of view of climate feedback, the situation in the zero-dimensional model is relatively simple. Here, the stability-altering and sensitivity-altering feedbacks always have the same sign. Furthermore, the initial and asymptotic tendencies of the model's dependent variable in response to an impulsive forcing are always of the same sign, and any changes in the initial tendencies due to a feedback interaction always give a correct indication of the sign of a stability-altering feedback. However, even for this simple model the signs of the prototype climate feedbacks can be opposite to those given by the control-theory and electronics definitions.

In the case of the two-zone models, the above simple relationship between the signs of the two prototype forms of climate feedback no longer has any general validity. It has been shown that in certain regions of parameter space the stability-altering and sensitivity-altering feedbacks can be of opposite signs. In the same region of parameter space, the response to an impulse function input (or imposed initial conditions) can exhibit initial growth while exhibiting asymptotic decay, and changes in initial tendencies can give an incorrect indication of the sign of a stability-altering feedback. Furthermore, for these models there is no simple relationship between the signs of the prototype climate feedbacks and the signs of the feedbacks as defined in control theory or electronics.

The stability-altering and sensitivity-altering feedbacks studied in our two-zone models are both dynamical, the interaction providing the feedback being that between the interzonal atmospheric eddy transport processes and the local stabilizing processes. It seems likely that similar results could be found using any form of dynamical interaction between zones – for example, east–west interactions involving the Walker circulation in the Tropics, or north–south interactions involving the thermohaline circulation in a combined atmosphere–ocean model. The present results give no direct indication of whether similar conflicts could occur between the sign of stability-altering and sensitivity-altering feedbacks in the context of the processes directly governing the radiative energy exchange with space. However, it is clear that the two categories of feedback are conceptually distinct, and that information about the sign of one cannot be assumed to apply to the other, except in the context of a zero-dimensional model.

In this paper we have not considered feedbacks in a transient forced climate-change situation (such as one with a time-dependent CO_2 forcing, or with a step-function CO_2 forcing where the system has not reached its new equilibrium); the only form of transience considered is that due to an initial impulsive forcing. The complexities that can arise in considering feedback in a transient forced situation have been examined by Hallegatte *et al*. (2006).

Many usages of the term 'feedback' occur in the climate literature that do not fall within either of the definitions given here. An example is that of Aires and Rossow (2003). They define feedbacks in terms of the interactions that occur between the state variables of a dynamical system when the dynamics are resolved in time. Feedbacks so defined are present and active even when the system is in equilibrium and no external forcing is applied.

In conclusion, the present paper emphasizes that the term 'climate feedback' has no universal and obvious meaning. It also stresses that assumptions about feedback concepts that are valid for a zero-dimensional model may not carry over to more complex models. Some existing glossary definitions of feedback fail to describe the dominant usages in the literature, and none appear to recognize that there are common usages that can be in conflict. It is hoped that the definitions suggested here, and the surrounding discussion, will help to clarify some conceptual issues of climate feedback and contribute usefully to the ongoing debate in this area.

Acknowledgements

The author is most grateful to Annraoi de Paor for extensive discussions of the concept of feedback as used in control theory. Thanks also to William Ingram and two anonymous reviewers for constructive criticisms of a first version of the manuscript, and to Peter Lynch, Paul Curran and Nigel Sealey for useful comments.

References

Aires F, Rossow W. 2003. Inferring instantaneous, multivariate and nonlinear sensitivities for the analysis of feedback processes in a dynamical system: Lorenz model case-study. *Q. J. R. Meteorol. Soc.* **129**: 239–275.

Alexeev VA. 2003. Sensitivity to CO_2 doubling of an atmospheric GCM coupled to an oceanic mixed layer: a linear analysis. *Clim. Dyn.* **20**: 775–787.

Alexeev VA, Bates JR. 2000. *A Dynamical Stabilizer in the Climate System: An Observational Study of the Underlying Parameterizations*. DCESS Report No. 1. Department of Geophysics, University of Copenhagen, Denmark.

Alexeev VA, Langen PL, Bates JR. 2005. Polar amplification of surface warming on an aquaplanet in 'ghost forcing' experiments without sea ice feedbacks. *Clim. Dyn.* **24**: 655–666.

AMS. 2000. *Glossary of Meteorology* (2nd edition). American Meteorological Society.

Bates JR. 1999. A dynamical stabilizer in the climate system: a mechanism suggested by a simple model. *Tellus* **51A**: 349–372.

Bates JR. 2004. On climate stability, climate sensitivity and the dynamics of the enhanced greenhouse effect. In *Paleoclimate and the Earth Climate System*, Proceedings of the Milutin Milankovitch Anniversary Symposium, Belgrade, September 2004, 27–46. Serbian Academy of Sciences. Also available at *www.dclimate.gfy.ku.dk*.

Black HS. 1977. Inventing the negative feedback amplifier. *IEEE Spectrum* December 1977: 55–60.

Bony S, Colman R, Kattsov VM, Allan RP, Bretherton CS, Dufresne J-L, Hall A, Hallegatte S, Holland MM, Ingram W, Randall DA, Soden BJ, Tselioudis G, Webb MJ. 2006. How well do we understand and evaluate climate change feedback processes? *J. Climate* **19**: 3445–3482.

DOI: 10.1002/qj

Caballero R. 2001. Surface wind, subcloud humidity and the stability of the tropical climate. *Tellus* **53A**: 513–525.
Charney J, Quirk WJ, Chow H-S, Kornfield J. 1977. A comparative study of the effects of albedo change on drought in semi-arid regions. *J. Atmos. Sci.* **34**: 1366–1385.
Dorf RC, Bishop RH. 2005. *Modern Control Systems* (10th edition). Pearson Educational International.
Hallegatte S, Lahellec A, Grandpeix J-Y. 2006. An elicitation of the dynamic nature of water vapor feedback in climate change using a 1D model. *J. Atmos. Sci.* **63**: 1878–1894.
Hartmann DL. 1994. *Global Physical Climatology*. Academic Press.
Harvey LD. 2000. *Global Warming: The Hard Science*. Prentice Hall.
IPCC. 2001. *Climate Change 2001: The Scientific Basis*. Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.
Jayne SR, Marotzke J. 1999. A destabilizing thermohaline circulation–atmosphere–sea ice feedback. *J. Climate* **12**: 642–651.
Lindberg K. 2002. A study of climate stability and sensitivity using a simple atmosphere–ocean model. PhD thesis, University of Copenhagen, Denmark.
Lindberg K. 2003. Supporting evidence for a positive water vapor/infrared radiative feedback on large scale SST perturbations from a recent parameterization of surface longwave irradiance. *Meteorol. Atmos. Phys.* **84**: 285–292.
Lindzen RS, Chou M-D, Hou A. 2001. Does the Earth have an adaptive infrared iris? *Bull. Am. Meteorol. Soc.* **82**: 417–432.
Manabe S, Wetherald RT. 1967. Thermal equilibrium of the atmosphere with a given distribution of relative humidity. *J. Atmos. Sci.* **24**: 241–259.
Marotzke J. 1996. Analysis of thermohaline feedbacks. In *Decadal Climate Variability: Dynamics and Predictability*, Anderson DLT, Willebrand J (eds). *NATO ASI Series*, 144, 333–375. Springer.
Maxwell JC. 1868. On governors. *Proc. R. Soc. London* **16**: 270–283; Reprinted in *Selected Papers on Mathematical Trends in Control Theory*, Bellman R, Kalaba R (eds). Dover: New York, 1964, 3–17.
Nakamura M, Stone PH, Marotzke J. 1994. Destabilization of the thermohaline circulation by atmospheric eddy transports. *J. Climate* **7**: 1870–1882.
North GR, Cahalan RF, Coakley JA. 1981. Energy balance climate models. *Rev. Geophys. Space Phys.* **19**: 91–121.
NRC. 2003. *Understanding Climate Change Feedbacks*. National Research Council: Washington, DC.
Peixoto JP, Oort AH. 1992. *Physics of Climate*. American Institute of Physics.
Rahmstorf S, Willebrand J. 1995. The role of temperature feedback in stabilizing the thermohaline circulation. *J. Phys. Oceanogr.* **25**: 787–805.
Ramanathan V, Collins W. 1991. Thermodynamic regulation of ocean warming by cirrus clouds deduced from observations of the 1987 El Niño. *Nature* **351**: 27–32.
Smith RJ. 1987. *Electronics: Circuits and Devices* (3rd edition). John Wiley & Sons.
Stephens GL. 2005. Cloud feedbacks in the climate system: a critical review. *J. Climate* **18**: 237–273.
Stommel H. 1961. Thermohaline convection with two stable regimes of flow. *Tellus* **8**: 224–230.
Wetherald RT, Manabe S. 1988. Cloud feedback processes in a general circulation model. *J. Atmos. Sci.* **45**: 1397–1415.

DOI: 10.1002/qj