A formal introduction of Feedback gain in non-linear models

La secte ZOOM
Laboratoire de Météorologie Dynamique et d'autres
CNRS - Institut Pierre Simon Laplace - Paris - France

13 avril 2005

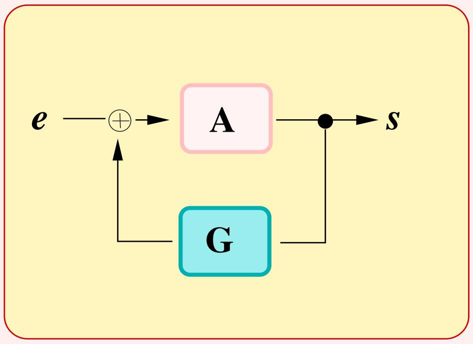
Corresponding author's address :
Alain Lahellec,
Laboratoire de Météorologie Dynamique - UMR 8539
Case courrier 99 Tour 25 - 4, place Jussieu
75252 Paris Cedex 05
France E-mail : alain@lmd.jussieu.fr fax : +33 01 44 27 62 72

Résumé

L'intelligibilité des systèmes numériques complexes requiert la construction d'outils conceptuels pour interpréter les phénomènes d'ensemble qu'ils manifestent. Le concept de feedback est aujourdhui largement utilise en climatologie sans qu'une définition rigoureuse en ai été donnée. Cette lettre a pour but d'introduire formellement ce concept en illustrant les possibilites d'application. Un exemple simple est donne sur le modele de Stommel.

1 The original concept of feedback gain

Hansen *et al.* (1987) quote Bode (1945) as originally introducing the Feedback (FB) concept in the fields of electric circuitry. It is a concept widely used also in control theory (e.g. Luenberger 79), as representing a cannonical element of fully controllable, or observable systems. Basically, it is a composition of two transfer functions in a loop as in Fig 1. The bloc diagram represents a single input – single output linear system. Given a function of time as entry e, the output signal s is defined by the FB transfer function given in Fig. 1. The diagram is classically viewed in a symbolic space, where p is the symbolic variable of a transformation : Laplace or Borel, Fourier, for continuous time systems, or z-transformation for discrete time systems. Hence, the FB gain is fundamentally a dynamic view of a particular system. The gain A is the transfer function when the FB loop is cut (G is set to 0). We define as s_{ins} as the open loop response, i.e. *insensitive* to feedback. When the loop is closed, the FB factor may increase the insensitive response (when AG is positive) or decrease the original gain (when the FB factor AG is negative). The FB sign is thus important when analysing the stability of a system. Using the



$$s(p) = H(p) = \frac{A(p)}{1 - A(p)G(p)} e(p)$$

H: effective gain
A: gain
G: feedback gain
AG: feedback factor

FIG. 1 – Bloc diagram of Feedback loop

relation between the output $s(p)$ and, the output when the system is*insensitive* to the feedback mecanism : $s(p)_{ins} = A(p)e(p)$, one can define the gain factor as

$$AG(p) = \frac{s(p) - s(p)_{ins}}{s(p)} \tag{1}$$

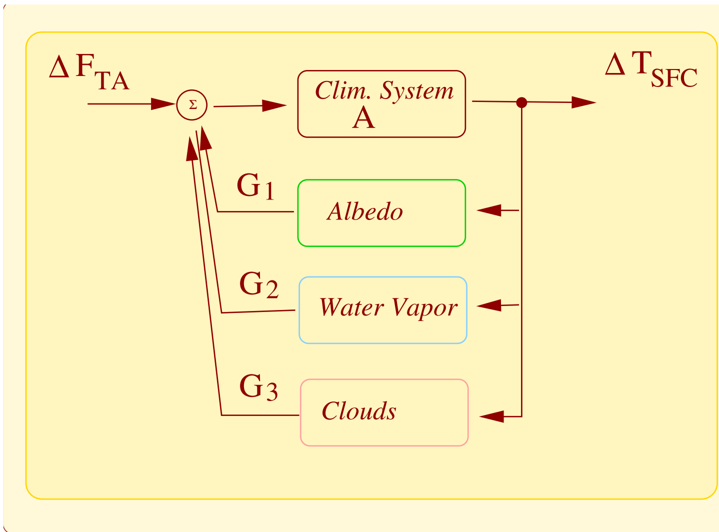
In control theory, G is sought as the FB solution of a control problem. When a system is fully controllable, the general placement theorem states that there is always a G function that can be determined such as the effective gain has a given characteristic polynomial. This means that any part of the output state space can be reached by a proper choice of G.

The above short introduction of FB gain explains why FB loop represents more than one among possible structures of bloc diagrams, and is more like a paradigm of the behaviour of systems, in spite of the strong limitation of linearity.

As for a definition of the g eff, one can remarq that g(p) is also the response to a step entry (citation Bode si on trouve ce f.. article, ou un autre).

2 How feedback gain is used in climatology

In climatology, the FB gain is a tentative approach to characterise the behaviour of a perturbed system. Peixoto (1982) for instance, analyses the effect of perturbing the extra-atmospheric visible radiation income on the terrestrial mean surface temperature (Fig 2). In such an analysis, the whole climate systems is analysed in terms of thought leading processes extracted from the global system. Considering the effect of the surface albedo for instance, one thinks of opening the corresponding loop. A decrease of F_{TA} leads to a decrease of T_{SFC}. Closing the loop by restoring the albedo effect, the decreasing temperature has the consequence of extending sea–ice areas on the long term, as well as continental ice–shells and also snow cover. These new white surfaces will reflect more of the visible incoming radiation (increase of the mean terrestrial albedo), and the insensitive decrease of temperature will be enhanced. One concludes that the albedo FB loop is associated with a positive feedback. The same is found for water-vapour and high–cloud cover feedbacks, while low clouds will have a negative feedback on the temperature.



$$\Delta T_{SFC} = \frac{A}{1 - \sum_i AG_i} \Delta F_{TA}$$

FIG. 2 – The climate system (Peixoto 1982)

Contrary to the original concept, these views of the climate system are purely conceptual, and does not correspond to well defined "components" of the climate. The one input one output diagram is obtained by considering space and time averages of the geophysical fields to begin with. Moreover, the climate system is by no means considered as a linear system. An implicit hypothesis underlying the FB view of the climate system is that when considering perturbations instead of state variables, the long–term behaviour of the system responds linearly to perturbation, provided its amplitude stays within some limits (no climate surprise). The main difference compared to the original FB concept is the loss of the dynamic aspect of the feedback function. Concerning the linear hypothesis, one more step is applied when featuring an explanation of the climate system. The leading processes are supposed well separated, i.e. non internally interacting, so that the overall effective transfer function at equilibrium is composed of a sum of individual FB factors (Fig. 2). This separation of different process is difficult to attain when using General Climate Models (GCM) to determine a set of FB factors (Schneider and Thompson, 1981). Coakley 77 and Wetherald and Manabe 88 give nevertheless a formal justification of feedback as follows. Suppose that the climate system perturbed by ΔF relaxes to a new equilibrium. Thus ΔR, the modified Output Longwave Radiation (OLR) of planet to space equals ΔF when the surface temperature change ΔT_s comes to the new equilibrium. One may write

$$\Delta F = \Delta R = \left(\frac{\partial R}{\partial T_s} + \sum_i \frac{\partial R}{\partial x_i} \frac{dx_i}{dT_s} \right) \Delta T_s$$

In this equation, the rhs first partial derivative is said to be the direct effect of the perturbation of surface temperature on the OLR. The added terms stand for indirect effects, or "going through x_i" feedback effects. The separation is implicitly adressed when writing total derivatives $\frac{dx_i}{dT_s}$. This implies in fact a limitation of the method : each feedback loop must be dependant on the surface temperature only. In other words, the ΔF perturbation must impact directly on T_s, which in turn is perturbing processes that will contribute to the new equilibrium. Again, this equation is purely conceptual and has no clear counterpoint in the models. It merely is an equivalent of Fig. 2 and a justification of the practical determination described now.

How are FB factors determined from GCMs ? The common practice uses thee trajectories : an open loop perturbed trajectory is run with a modified version of a GCM, in which some process is frozen : for instance, impeding a change in sea-ice surfaces. One obtains in the long term, after eventual chaotic transients generated by the perturbation have died, the insensitive trajectory of Fig. 3. This determines ΔT_{ins} as the difference between the insensitive trajectory and the reference one. The third trajectory is determined with the full degree of freedom of the GCM restored (closed–loop). The feedback gain factor is determined as

$$AG(t) = \frac{\Delta T(t) - \Delta T(t)_{ins}}{\Delta T(t)} \tag{2}$$

in analogy to equation (1), with the essential difference that here, real trajectories are considered. The linearity is checked by considering the ratio of the two ΔT differences variation in time : if it is slowly varying, one considers that the long term climate model responds linearly to the perturbation and the FB factor is valuable. This practice is applied in numerous tentatives

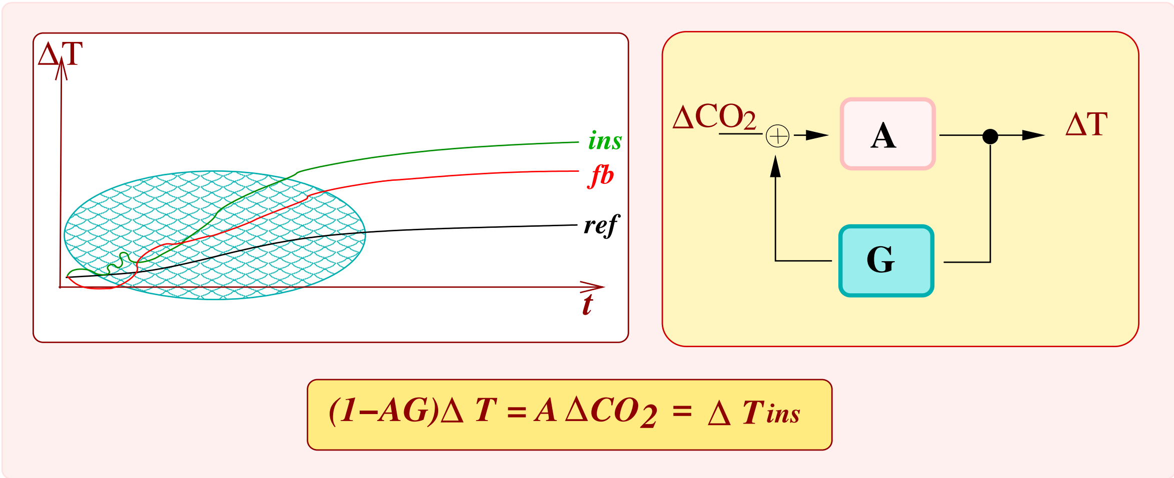


FIG. 3 – The three trajectories FB determination : change in terrestrial surface temperature resulting from a perturbation of the atmospheric CO_2 content

to sketch the climate as an intelligible system (see Carney 77, Hansen et al. 84, Schlesinger et al. 87, Ramanathan 89, Cess 89, Lindzen 93, Dufresne et al. 2003 ...). It is however a conceptual view projected on models, which means that the models themselves are inadequate to provide the values. Moreover, the difficulty of properly separating different processes, and freezing some of them without violating physical laws render questionable such pratice, as claimed by Rossow et al. (2002). These authors concluded that the FB gain concept should be abandoned as lacking of rigorousness.

In summary, it appears that the climate researchers are desperately looking for some tools that can restore some intelligibility on the general behaviour of such complex numerical models as the climate system. The Feedback loop is considered as a paradigm of such a conceptual tool, but remains out of reach of scientific rigorousness. It is thus a challenge to build such indispensable tools while keeping a scientific objectivity. We show that it is possible in the next sections.

```
-> conclusions : 1/ pb de structure
                 2/ linearisation
                 3/ aspect dynamiques important
                    (convolution -> diagnostic de rechauffement cliomatique)
                 3/ definition generale de gain effectif
                    g_eff = sensibilite a un step
```

3 Feedback and structure of the model

Si on veut pouvoir étudier les inter-influences entre deux systèmes, il est nécessaire de définir ces deux systèmes comme, la fois, séparés, et couples. Considérons un modèle couple ocean-atmosphere ; on a deux systèmes arbitrairement séparés (l'océan d'une part, l'atmosphere de l'autre). Chacun d'eux est bien défini des qu'on lui adjoint des conditions limites, et le modèle mathématique associe doit alors conduire a un problème bien pose.

On prend le cas d'un couplage simplifie où l'océan est spécifié par un flux provenant de l'atmosphere (φ), et celle-ci par la SST (θ). (ref : on pourra citer les papiers de XXX sur la coherence math du couplage). On examine d'abord le cas particulier du couplage unicolonne. beginverbatim Formulation par le TEF, avec explicitation d'un gain en fonction des 4 matrices.

Ensuite, montrer qu'on deux gains equivalents scalaires et deux "matrices de gains" // Fig 4

Dire qu'on peut definir l'aspect stabilisant ou non du couplage par ¡Delta2 Phi¿/¡DeltaPhi¿ =¿ determinant de [G]. (où Delta2 Phi = Delta Phi - Delta - Phi ins) Hors, ce determinant est caracteristique de la boucle. En deduire la definition du gain de la boucle et l'interet de le determiner sous forme scalaire.

on place la remarque sur la possibilite d'aller plus loin en analysant les structures singulieres de G et donc des champs associes a des gains. endverbatim Le système forme une boucle de rétroaction parce que, symboliquement :

$$\frac{d\eta}{d\zeta} = \frac{\partial\eta}{\partial\theta}\frac{\partial\theta}{\partial\zeta}; \qquad \frac{\partial\eta}{\partial\varphi} = 0 \tag{3}$$

et que

$$\frac{d\zeta}{d\eta} = \frac{\partial\zeta}{\partial\varphi}\frac{\partial\varphi}{\partial\eta}; \qquad \frac{\partial\zeta}{\partial\theta} = 0 \tag{4}$$

Autrement dit, la structure du système partage les *sensibilités* des sous-systèmes selon les variables d'interface. suite a refaire, mais je laisse parce qu'ilre ste ce que l'on dit a propos du

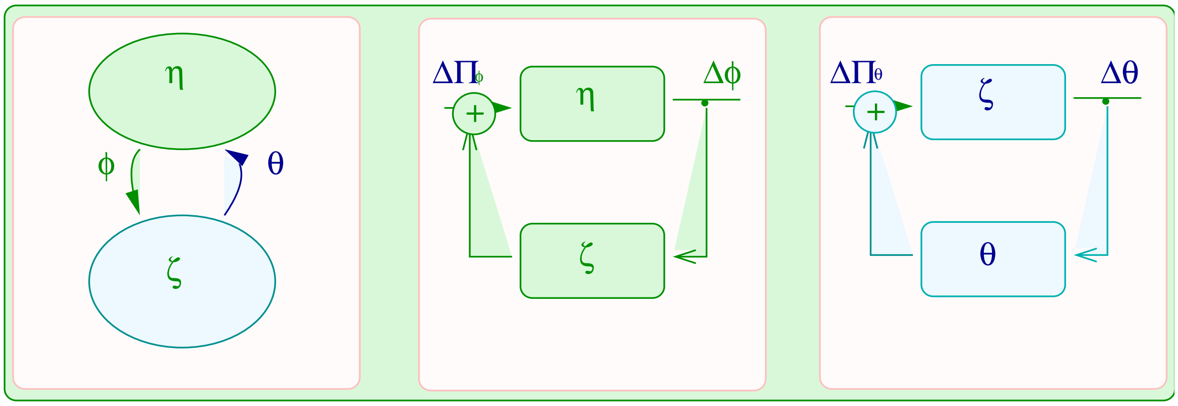


Fig. 4 – A Feedback loop between atmosphere and ocean and two feedback gains

dt !

Ces contraintes peuvent être prises dans leur sens le plus général pour définir la structure de boucle de rétroaction. On voit cependant que l'utilisation de symboles de dérivation qui expriment les interdépendances et leur média (variables d'interface) nécessitent une explicitation : quel est le rôle du temps, que représentent les variables etc. Ces points seront formellement élucidés dans la section 4. Néanmoins, la boucle de rétroaction de la figure 4 définit deux gains de rétroaction, suivant que l'on s'intéresse à l'effet de l'océan sur l'atmosphere ou réciproquement. Dans le premier cas par exemple, on considère une perturbation (un forçage) appliquée à la SST ($\Delta\Pi_\theta$). Cette perturbation modifie l'état de l'atmosphère par changement des conditions limites du système d'équations associées : c'est l'effet direct. De plus, les flux φ subiront les conséquences de ce changement, perturbant l'océan, de telle manière que la SST en sera affectée. Autrement dit la perturbation initiale $\Delta\Pi_\theta$ de θ sera in fine modifiée par la réponse complète de la boucle de rétroaction ($\Delta\theta$). Pour rester dans le cadre de la définition de gain comme fonction scalaire, on suppose que dans l'exemple du couplage océan-atmosphère, on ne considère qu'une colonne verticale du modèle maillé, pour laquelle θ et φ sont deux scalaires. On peut alors caractériser un gain effectif de rétroaction entre ces deux variables, et écrire symboliquement :

$$\Delta\varphi = \frac{\frac{\partial\varphi}{\partial\eta}\frac{\partial\eta}{\partial\theta}}{1 - \frac{\partial\varphi}{\partial\eta}\frac{d\eta}{d\zeta}\frac{\partial\zeta}{\partial\varphi}}\Delta\Pi_\theta \tag{5}$$

The converse feedback effective gain would appear as ... **si on a deja dit que meme determinant, on conclue que c'est le meme - ou dans l'ordre inverse, a voir.**

$$\Delta\theta = \frac{\frac{\partial\theta}{\partial\zeta}\frac{\partial\zeta}{\partial\varphi}}{1 - \frac{\partial\theta}{\partial\zeta}\frac{d\zeta}{d\eta}\frac{\partial\eta}{\partial\theta}}\Delta\Pi_\varphi \tag{6}$$

and would give the response of the SST to a perturbation externally applied to the fluxes entering the ocean ($\Delta\Pi_\varphi$).

Donc, en gros, on est debarrases du pb de structure et on passe au reste.

Mais il faut bien aborder ici l'aspect dynamique, et prendre l'exemple de l'equilibre avec Borel avant de sortir ?

4 A formal introduction of feedback gain in numerical models

On a definit le gain effectif (1/1-g ou a/1-ag) par une analyse de sensibilite associee a une structure de boucle. Hors, Cacuci etc, que l'on rappelle.

After twenty years of nuclear–reactor modelling at Oakridge, Cacuci provided what he called a general sensitivity theory (*in J. of Math. Physics*, Dec. 1981). We briefly recall this formal theory before calling for the FB concept.

The general sensitivity theory of Cacuci

Consider the following system of equations where small letters symbols are vectors :

je verrai mieux s'en tenir au meme modele Oc-Atm et simplifeir les notations au max.

$$\left\{ \begin{array}{lcl} \mathcal{N}(u(x),\alpha(x),x) & = & \mathcal{Q}(\alpha(x),x)\,;\; x\in\Omega \\ |\mathcal{A}(u,\alpha)-\mathcal{B}(\alpha)\,|_{\partial\Omega} = & & 0 \end{array} \right. \tag{7}$$

$\mathcal{N}$ stands for a non-linear operator operating on state vectors u (say in a square integrable Hilbert space). x are space-time coordinates, $\mathcal{Q}$ represents inhomogeneous sources, and the second equation is a general form for initial conditions and boundary limit conditions. α are model parameters.

The general sensitivity theory considers a functional response $\mathcal{R}(u,\alpha)$ whose variation – e.g. to a change in parameters α – is analysed. Cacuci based his definition of sensitivity on the Gâteaux-differential (G-differential) defined as

$$\mathcal{VR}(e) = \lim_{\epsilon\to 0}\frac{\mathcal{R}(e+\epsilon\Delta e)-\mathcal{R}(e)}{\epsilon}$$

for a given reference e, that may be defined as the reference trajectory. ϵ is a scalar and Δe is any perturbation vector of e (i.e. corresponding to a perturbed trajectory). A variation of $\mathcal{R}$ is related to the G-differential by :

$$\mathcal{R}(e+\Delta e)-\mathcal{R}(e) = \mathcal{VR}(e;\Delta e)+\Upsilon(\Delta e)$$

where

$$\lim_{\epsilon\to 0}\frac{\Upsilon(\epsilon\Delta e)}{\epsilon} = 0.$$

The use of the G-differential has the definite advantage that it does not require continuity in the vector variable e. The result of this operation on a system neither needs to be linear in Δe. For clarity, we will nevertheless consider only systems that are linear in Δe, which requires necessaires et suffisantes conditions (see cacuci) :

- a weak Lipschitz condition, that insure numerical stability on a perturbed trajectory
- and the linearizability condition

$$\mathcal{R}(e+\epsilon\Delta e_1+\epsilon\Delta e_2)-\mathcal{R}(e+\epsilon\Delta e_1)-\mathcal{R}(e+\epsilon\Delta e_2)+\mathcal{R}(e) = o(\epsilon) \tag{8}$$

in that case the G-variation is called a Gâteaux derivative (G-derivative). Applying the D-derivative to $\mathcal{R}(u,\alpha)$ yields

$$\mathcal{VR} = \mathcal{R}'_1\Delta u + \mathcal{R}'_2\Delta\alpha \tag{9}$$

Two terms appear in the above equation, the second rhs term giving the direct contribution of the perturbed parameters, while the first is depending on the modified trajectory. As the sensitivity of $\mathcal{R}$ is looked for as a (sensitivity) gradient to perturbation , one looks for methods to eliminate Δu, that is the trajectory dependence of $\mathcal{VR}$. This requires to forward integrate the system along the reference trajectory. One might apply the G-differential to (7) to get the Linear Tangent System (TLS) related to the reference trajectory

$$\left\{ \begin{array}{lcl} \mathcal{N}(u(x),\alpha(x),x) & = & \mathcal{Q}(\alpha(x),x) \\ |\mathcal{A}(u,\alpha)-\mathcal{B}(\alpha)\,|_{\partial\Omega} = & & 0 \end{array} \right. \quad \xrightarrow{TLS} \quad \left\{ \begin{array}{lcl} \mathcal{N}'_1\Delta u+\mathcal{N}'_2\Delta\alpha & = & \mathcal{VQ}(\alpha;\Delta\alpha) \\ |\mathcal{VA}(u,\alpha)-\mathcal{VB}(\alpha)\,|_{\partial\Omega} = & & 0 \end{array} \right. \tag{10}$$

A step by step forward integration of the system is used to replace the $\Delta u(t)$ vector in (9) by algebraic manipulations. This will provide the variation of $\mathcal{R}$ in a form explicitly dependent on the perturbed variable. Alternatively one can use the Adjoint System of the TLS (see Cacuci 82,83,84), and backward integrate. The adjoint variables are then used to eliminate the trajectory dependance on $\mathcal{VR}$. We will apply the direct method to reach the definition of feedback function.

A formal definition of feedback

On se contenterait alors de montrer comment on utilise l'adjoint pour que la reponse ne soit plus associee qu'aux perturbations.

Et alors, le gain eff = terme $< dw -_d j >$ dans la reponse ($R = \int_0^T < d.u > r(t)dt$).

retour sur Oc-Atm $< d| = 1/n - mailles - interface < 1,1,1,...1|$ donne le gain de la moyenne des u (temperature ou flux).

On a donc etendu la notion auxx systemes non-L, mais alors le gain eff dyn est lie a un instant diagnoctic (t=c) : g(tau,c) avec t=c-tau.

5 Application to the Stommel model of thermohaline circulation

on n'a plus qu'a donner un exemple avec des commentaires rassurants sur l'analyse du gain en faisant varier c.

6 Conclusive discussion

```
a/   definition formelle de gain dynamique
    -> necessite de calculer les Jacobiennes du systeme  ou de definir le TLS le long...
       ce qui permet le contexte lineaire
    -> obtenu par Borel dans le cas lineaire
    -> extension du sens des tempos caracteristiques obtenus par Borel
       au cas non-L // Cacuci en adjoint
    -> application pour systemes non-L avec instant fixé de diagnostic
       par sensibilite d'une fonctionnelle de reponse.

b/   il est possible d'utiliser le concep de gain dynamique de retro
pour les systemes complexes a condition d'utiliser l'adjoint.

c/   Conclusions generale sur les methodes pragamatiques de la litterature.
et sur la possibilite d'appplication numerique par Gateaux fini sans adjoint.
donc sur l'importance de fondements rigoureux pour prolonger et verifier
les conditions d'applications pratiques des gains.
```

Enfin prolongements prevus pour exploiter l'analyse de structures multi-dim associees a des gains // EOF.