Inferring instantaneous, multivariate and nonlinear sensitivities for the analysis of feedback processes in a dynamical system: Lorenz model case study

Filipe Aires
Department of Applied Physics, Columbia University, NASA Goddard Institute for Space

Studies, New-York

William B. Rossow
NASA Goddard Institute for Space Studies, New-York

Short title: FEEDBACK ANALYSIS

Abstract. A new approach is presented for the analysis of feedback processes in a nonlinear dynamical system by observing its variations. The new methodology consists of statistical estimates of the sensitivities between all pairs of variables in the system based on a neural network modeling of the dynamical system. The model can then be used to estimate the instantaneous, multivariate and nonlinear sensitivities, which are shown to be essential for the analysis of the feedbacks processes involved in the dynamical system. The method is described and tested on synthetic data from the low-order Lorenz circulation model where the correct sensitivities can be evaluated analytically.

1. Introduction

Feedback processes are present in dynamical systems that involve nonlinear relationships among many variables integrated over time. A formalism from electrical circuit theory [*Bode*, 1945] has been used to study feedback processes in climate models [e.g., *Hansen*, 1984; *Schlesinger*, 1985] using the sensitivities of the system (first derivatives of one variable by another). This approach characterizes the change of the equilibrium state of the system after the introduction of an external forcing.

Such an approach is valid in a theoretical model where the instantaneous sensitivities can be evaluated directly from the equations, however its application where the underlying equations are unknown (or only partially known) and all we have are observations of the system behavior over time is more questionable. This is the situation faced in the study of climate. There are various ways to estimate these sensitivities in experimental studies. With numerical models, one approach is to introduce a perturbation of one variable at a time and then evaluate the impact on other variables. The problems associated to this approach are numerous. First, since the initial perturbation is limited to only one variable, the inter-dependence (i.e., non-linearity) of the sensitivities is not taken into account. Even if many variables are perturbed, it is difficult to be sure that the estimate of the multi-variate sensitivities is complete. Second, in most cases the sensitivities are estimated by finite differences between (usually equilibrium) states of the system. These differences can be generated by geographical location differences or by time differences, the result is dependent on the strategy adopted [*Slingo et al.*, 2000]. As we will see, this simplistic approach can be highly misleading because it provides only the space and/or time averaged sensitivities that may not actually represent the system dynamics. Another problem with this approach is that the technique estimates sensitivities that are already "polluted" by the feedback processes in the numerical model. For example, if the effect of a feedback is to damp the impact of one sensitivity, the sensitivity measured will be an under-estimate (the same problem happens for an amplifying feedback process). To avoid under or over-estimation of the sensitivities, their estimation needs to be done at a time

3

scale sufficiently small to neglect the impact of the feedback processes. Moreover, such analyses can only be performed on models, not on observations of a real system, so the feedback estimates are only those what were introduced into the model, which does not provide model validation. Third, the sensitivities are often estimated by finite differences averaged over many geographical locations (e.g. global), which suppresses all the possible non-local processes. A similar effect is produced if the model outputs (or observations) are time averaged before analysis.

So in many studies, the hypotheses at the base of the feedback analysis are too crude: linear model, constant sensitivities, mutual independence of sensitivities, mono-variable conception of the forcing and response. The conclusions from these studies are questionable: a single feedback factor (i.e., a scalar) is supposed to explain the nonlinear multivariate processes integrated over time. If this number is positive, the process is said to have an amplifying effect on the initial perturbation. If the number is negative, the effect is a damping. We think that this over-simplification is misleading and that a time and space analysis of feedbacks may be required to understand such complex phenomena as climate.

Even if the sensitivities were correctly estimated, this classical approach only characterizes the change of the equilibrium state of the system after the introduction of an external forcing. The transient period, between the beginning of the forcing and the equilibrium is not described and the time needed to reach the new equilibrium remains undetermined. So the results of this kind of analysis are often insufficient even for comparison between numerical models and observations.

We first refine the terminology required to perform feedback analysis with an emphasis on the discrete formulation of dynamical systems, which is better adapted to prediction, to the description of the cause and effect relations underlying the feedback processes, and to the use of observations in the analysis.

Our approach uses a statistical modeling of the dynamical system to infer, from the observed behavior of the system, the sensitivities of the variables of the system. These sensitivities are the key concept for the feedback analysis. They give the inter-dependences

4

of one variable on another that cause the feedback loops in a dynamical system.

Because such empirical sensitivities provide the relationships among the variables they can be used more directly to understand the physical processes involved in the system dynamics (a model or observations) and to provide a more informative comparison of models to observation or one model to another. A stability analysis of the dynamical processes involved can also be performed using these sensitivities. For prediction purposes, the temporal propagation of errors onto the state of the system can also be analyzed [*Smith*, 1997].

As a test of the validity of this approach, we apply it to the output (observations) of the Lorenz low-order dynamical model where the equations are known and theoretical sensitivities can be calculated directly. We also evaluate classical feedback parameters and compare their usefulness as descriptors of the system dynamics to that of the instantaneous, multivariate and nonlinear sensitivities.

2. Feedbacks in a dynamical system

There are two general ways of formulating a dynamical system: the continuous and the discrete approaches. We prefer in this study the discrete formulation because it is simpler to describe the cause and effect relationships between variables. Furthermore, the discrete approach is more practical for prediction when no theoretical physical evolution model is available. We adopt the discrete formalism in the following, but will refer sometimes to the continuous case. The goal of this section is to show how time integration of dynamical relationships leads to feedback processes and to highlight the role of the sensitivities.

a. Dynamical systems

The object of this study is the analysis of a physical dynamical system by observing the time variations of the quantities defining the state of the system. A dynamical system is often described by a set of Ordinary Differential Equations (ODE) which come from the physics of the problem. For practical considerations or because these ODEs are not known,

5

the dynamical system is often discretized in the form:

$$X(t + 1) = \mathcal{A}(\boldsymbol{P}(t)) + \boldsymbol{\varepsilon}(t) \tag{1}$$

where $\boldsymbol{X}(t)$ is the p-dimensional vector of observable variables (defining the state of the system) at time t, $\boldsymbol{P}(t)$ is the d-dimensional vector of variables defining the system behavior (predictors) which can include $\boldsymbol{X}(t)$, $\boldsymbol{\varepsilon}(t)$ is noise (instrumental or model errors), and $\mathcal{A}$ is a mapping, possibly nonlinear (vectors and matrices are indicated in bold characters). This kind of model is often used in atmospheric and oceanic sciences to perform, for example, climatological predictions.

The determination of the good predictors $\boldsymbol{P}(t)$ is a crucial point for the quality of the model. This determination uses all the physical a priori knowledge about the model. If $\boldsymbol{P}(t) = \boldsymbol{X}(t)$ the system is said to be auto-regressive. Sometimes, the prediction of $\boldsymbol{X}(t+1)$ requires the knowledge of $\boldsymbol{X}(t), \boldsymbol{X}(t-1), \ldots, \boldsymbol{X}(t-q)$ because the system dynamics has an inertia that requires the knowledge of previous steps. Then, the system is said auto-regressive with memory q, denoted an AR(q) model. However, defining a new state variable $\boldsymbol{X}'(t) = (\boldsymbol{X}(t), \boldsymbol{X}(t-1), \ldots, \boldsymbol{X}(t-q))$, one can rewrite this AR(q) system as an AR(1) model with memory 1.

If the dynamical system (1) is linearized near $\boldsymbol{P}(t_0)$, we obtain:

$$\boldsymbol{X}(t_0 + \Delta t) - \boldsymbol{X}(t_0) = \boldsymbol{G}(\boldsymbol{P}(t_0)) \cdot \Delta \boldsymbol{P}(t_0) + \boldsymbol{\varepsilon}(t) \tag{2}$$

where

$$\boldsymbol{G}(\boldsymbol{P}(t_0)) = \frac{\partial \boldsymbol{X}(t_0 + \Delta t)}{\partial \boldsymbol{P}(t_0)} \tag{3}$$

is the Jacobian or sensitivity matrix of the mapping $\mathcal{A}$ at state $\boldsymbol{P}(t_0)$.

The uncertainty $\boldsymbol{\varepsilon}(t)$ is often neglected, so the discretized system of equation (2) is entirely defined by the sensitivities $\boldsymbol{G}(\boldsymbol{P}(t))$ of the dynamical system and by an initial state $\boldsymbol{P}(t_0)$.

b. General feedback analysis: sensitivities integration

For simplicity of notation, we suppose, as is true in most of the cases, that the system is auto-regressive, i.e. the responses are equal to the predictors, $\boldsymbol{X}(t) = \boldsymbol{P}(t)$

Even if the mapping $\mathcal{A}$ of equation (1) is linear, the global response of the dynamical system is not linear since the mapping is integrated in time. If in a system $\boldsymbol{X}(t+1) = \boldsymbol{A} \cdot \boldsymbol{X}(t)$ the matrix $\boldsymbol{A}$ is diagonal, the variables of the system are independent, so each variable X_i evolves independently as $X_i(t_0 + k\,\Delta t) = (A_{ii})^k \cdot X_i(t_0)$. The absolute value $|A_{ii}|$ is the damping (if $|A_{ii}| < 1$) or the amplifying (if $|A_{ii}| > 1$) coefficient of the variable X_i.

If matrix $\boldsymbol{A}$ is non-diagonal, i.e. some of the variables of the system are dependent on other variables, the system is more complex: an initial perturbation of one variable will be propagated into all the variables that are directly or indirectly dependent on this initially perturbated variable. After k time steps, the state of the system is given by: $\boldsymbol{X}(t_0 + k\Delta t) = \boldsymbol{A}^k \cdot \boldsymbol{X}(t_0)$. So, the responses, $\boldsymbol{X}(t)$, at any time, t, are still a linear combination of the predictors at time, t_0, but the impact of an initial perturbation is more complex than the previous case because feedback loops have mixed the initial perturbation into each linked variable.

If the mapping $\mathcal{A}$ of equation (1) is nonlinear, the Jacobians are dependent on the state $\boldsymbol{X}(t)$. So, even if we linearize the mapping $\mathcal{A}$ using its Jacobians, $\boldsymbol{G}(\boldsymbol{X}(t))$, after k time steps, the state of the system is given by: $\boldsymbol{X}(t_0 + k\Delta t) = \left[\prod_{l=1}^{k} \boldsymbol{G}(\boldsymbol{X}(t_0 + l\,\Delta t)) \right] \cdot \boldsymbol{X}(t_0)$ which is even more complex, where $\prod$ is the product symbol.

To define a feedback process in the discrete formulation of a dynamical system, we need at least two time steps to describe the feedback loops involved. If an initial perturbation $\Delta \boldsymbol{X}(t_0)$ is introduced into the system at time, t_0, the response of the system at time, $t_0 + \Delta t$, is approximated to first order by:

$$\Delta \boldsymbol{X}(t_0 + \Delta t) \simeq \boldsymbol{G}(t_0, t_0 + \Delta t) \cdot \Delta \boldsymbol{X}(t_0), \tag{4}$$

where $\boldsymbol{G}(t_0, t_0 + \Delta t)$, the gain of the system from t_0 to $t_0 + \Delta t$, is the Jacobian matrix

$G(X(t))$ of the mapping $\mathcal{A}$ between $[t_0, t_0 + \Delta t]$: matrix $G(t_0, t_0 + \Delta t)$ has elements $\frac{\partial \mathcal{A}_i(t_0 + \Delta t)}{\partial X_j(t_0)} = \frac{\partial X_i(t_0 + \Delta t)}{\partial X_j(t_0)}$ at coordinates (i, j). An initial perturbation $\Delta X_j(t_0)$ on variable X_j at time t_0 is then propagated at time $t_0 + \Delta t$ to each variable X_i that is linked to X_j via off-diagonal terms in $G(t_0, t_0 + \Delta t)$. But the resulting perturbations $\Delta X_i(t_0 + \Delta t)$ are the direct impact of the initial perturbation, so there is no feedback during $[t_0, t_0 + \Delta t]$.

At time, $t = t_0 + 2\Delta t$, the impact on the system is given to first order by:

$$\Delta X(t_0 + 2\Delta t)$$

$$\simeq G(t_0 + \Delta t, t_0 + 2\Delta t) \cdot \Delta X(t_0 + \Delta t) \tag{5}$$

$$\simeq G(t_0 + \Delta t, t_0 + 2\Delta t) \cdot G(t_0, t_0 + \Delta t) \cdot \Delta X(t_0) \tag{6}$$

$$\simeq G(t_0, t_0 + 2\Delta t) \cdot \Delta X(t_0). \tag{7}$$

The previouly propagated perturbations, $\Delta X_i(t_0 + \Delta t)$, resulting from $\Delta X_j(t_0)$ can then perturb $\Delta X_j(t_0 + 2\Delta t)$, completing a feedback loop. The initial perturbation $\Delta X_j(t_0)$ can be amplified or damped into $\Delta X_j(t_0 + 2\Delta t)$. We see in this simple example that feedback processes result from the time integration of the variable dependencies of the system. The term $G(t_0, t_0 + 2\Delta t)$, representing the evolution of the system in two time steps, includes these feedback loops.

c. Forcing / Response

We introduce in this section the concept of external forcing to formalize the perturbations of the variables of the system we have discussed in the previous example. It is important to note that the feedback processes are present in a dynamical system, even when no external forcing is applied (forcing and feedback are often confused).

An external forcing is a perturbation of the internal variables of the system (i.e. variables that define the state of the system). The external forcing has an impact on the internal variables, but the reverse is not true: the forcing is independent of the internal variables. There are many ways an external forcing could operate. The simplest model is the introduction of an unique perturbation at time t_0: $E(t) = E_0 \delta(t_0 - t)$, a time-localized

8

volcanic eruption for example. In this case, the impulsive initial perturbation will be propagated in time through the internal variables, following the inter-dependencies of the variables. This is the example discussed in the previous section.

The external forcing can also begin at time t_0 and remain constant in time: $E(t) = E_0, t \in [t_0, t_0 + \Delta t, \ldots]$. In this case, the relations (5)/(7) become more complex:

$$\Delta X(t_0 + 2\Delta t)$$

$$\simeq E(t_0 + 2\Delta t) + \frac{\partial X(t_0 + 2\,\Delta t)}{\partial X(t_0 + \Delta t)} \cdot E(t_0 + \Delta t) + \frac{\partial X(t_0 + 2\,\Delta t)}{\partial X(t_0 + \Delta t)} \frac{\partial X(t_0 + \Delta t)}{\partial X(t_0)} \cdot E(t_0) \quad (8)$$

$$\simeq E_0 + G(t_0 + \Delta t, t_0 + 2\,\Delta t) \cdot E_0 + G(t_0, t_0 + 2\,\Delta t) \cdot E_0 \quad (9)$$

If the gains of the system are constant, equal to a constant G (i.e. a linear dynamical system), then at time $t + k\,\Delta t$:

$$\Delta X(t_0 + k\,\Delta t) = (I + G + G^2 + \ldots + G^k) \cdot E_0 \quad (10)$$

$$= \frac{I - G^{k+1}}{I - G} \cdot E_0 \quad (11)$$

where I is the identity matrix. If the eigenvalues of matrix G have an absolute value lower than 1 (otherwise the system is unstable), the effect of the external forcing is stabilized, the dynamical system eventually reaches in time a stabilized state:

$$\Delta X(t_0 + k\,\Delta t) \simeq \frac{I}{I - G} \cdot E_0, \text{for} k \to +\infty \quad (12)$$

$$\simeq E_0 + \frac{G}{I - G} \cdot E_0. \quad (13)$$

For example, a mono-variable system with $G = 1/2$, $E_0 = 1$ and $X(t_0) = 0$ stabilizes at $\lim_{k \to +\infty} X(t_0 + k\,\Delta t) = 2$, the forcing has changed the equilibrium state of the system. Figure 1 shows the values of the stabilized solutions of this simple system for different values of G. If $-1 < G < 0$ then the system stabilizes, but oscillates. If G is positive and close to 0, the system stabilizes near E_0. The closer the gain of the system G to 1^- (i.e., lower but close to 1), the higher is the value at which it stabilizes, $\lim_{k \to +\infty} X(t_0 + k\Delta t)$. If the absolute value of G is bigger than 1, the system is unstable.

d. One particular case: the classical analysis of a parallel feedback configuration

The previous examples are very general since each variable of the system can be dependent on other variables. But in some cases, knowledge of cause and effect relationships provides a priori information about the ordering and the structure of dependencies. It is then possible, and recommended, to use this information. This kind of a priori information is used, for example, in the Cause and Effect Analysis technique [*Andronova and Schlesinger, 1991; Andronova and Schlesinger, 1992*].

In the classical feedback analysis [*Hansen, 1984; Schlesinger, 1985*], it is supposed that the external forcings E_{X_e} of the system act on only one variable X_e of the system, that all the other internal variables $\{X_i = X_i(X_d)\}$ are all dependent on one particular internal variable X_i (i.e., the diagnosed variable), that the impact of the external forcings is observed on this particular internal variable X_d, and that the feedbacks act in parallel (Figure 2). These assumptions are very strong cause and effect constraints: the diagnosed variable X_d is supposed to be more important than the others internal variables $\{X_i\}$, by hypothesis the $\{X_i\}$ are dependent only on X_d, and they are not directly dependent on the external forcing. In this case, the feedback processes are assumed to act in parallel, i.e., they do not interact.

Since the external forcing E_{X_e} of the system acts on only one variable, X_e, of the system, the general multi-dimensional expression in equation (8) becomes a scalar relation:

$$\Delta X_e(t_0 + 2\Delta t) \simeq E_{X_e}(t_0 + 2\Delta t) + \sum_i \frac{\partial X_e(t_0 + 2\ \Delta t)}{\partial X_i(t_0 + \Delta t)} \cdot \Delta X_i(t_0 + \Delta t)$$

$$+ \sum_i \sum_j \frac{\partial X_e(t_0 + 2\ \Delta t)}{\partial X_i(t_0 + \Delta t)} \frac{\partial X_i(t_0 + \Delta t)}{\partial X_j(t_0)} \cdot \Delta X_j(t_0) \tag{14}$$

We measure the effect of the constant external forcing, E_{X_e}, on X_d, the diagnosed variable (Figure 2). We then analyze the system $\Delta X_e \longrightarrow \Delta X_d$. So the perturbations $\Delta X_i(t_0 + \Delta t)$ and $\Delta X_j(t_0)$ are considered only for the diagnosed variable X_d. In other words, the impact of the perturbations of other variables than X_d are not taken into account in this classical

10

analysis (see Figure 2). Thus, the expression (14) becomes:

$$\Delta X_e(t_0 + 2\Delta t) \simeq E_{X_e}(t_0 + 2\Delta t) + \frac{\partial X_e(t_0 + 2\Delta t)}{\partial X_d(t_0 + \Delta t)} \cdot \Delta X_d(t_0 + \Delta t)$$

$$+ \sum_i \frac{\partial X_e(t_0 + 2\Delta t)}{\partial X_i(t_0 + \Delta t)} \frac{\partial X_i(t_0 + \Delta t)}{\partial X_d(t_0)} \cdot \Delta X_d(t_0) \tag{15}$$

Due to the hierarchical dependencies adopted, $X_e(t_0) \to X_d(t_0 + \Delta t) \to X_i(t_0 + 2\Delta t)$, see Figure 2, some of the partial derivatives in (14) are zero:

$$\frac{\partial X_e(t_0 + 2\Delta t)}{\partial X_d(t_0 + \Delta t)} = \frac{\partial X_e(t_0 + 2\Delta t)}{\partial X_e(t_0 + \Delta t)} = 0, \tag{16}$$

and the expression (15) simplifies to:

$$\Delta X_e(t_0 + 2\Delta t) \simeq \underbrace{E_{X_e}(t_0 + 2\Delta t)}_{\text{external forcing}} + \underbrace{\sum_{i \neq d, i \neq e} \frac{\partial X_e(t_0 + 2\Delta t)}{\partial X_i(t_0 + \Delta t)} \frac{\partial X_i(t_0 + \Delta t)}{\partial X_d(t_0)} \Delta X_d(t_0)}_{\text{feedback terms}} \tag{17}$$

$$?? \simeq E_{X_e}(t_0 + 2\Delta t) + \sum_{i \neq d, i \neq e} H_i(t_0, t_0 + 2\Delta t)\Delta X_d(t_0) \tag{18}$$

where the terms $H_i(t_0, t_0 + 2\Delta t)$ are the products of first derivatives describing the cause and effect relations. Expression (??) can be multiplied by the gain $G(t_0 + 2\Delta t, t_0 + 3\Delta t)$ of the system $\Delta X_e(t_0 + 2\Delta t) \longrightarrow \Delta X_d(t_0 + 3\Delta t)$:

$$\Delta X_d(t_0 + 3\Delta t) \simeq G(t_0 + 2\Delta t, t_0 + 3\Delta t) \cdot \Delta X_e(t_0 + 2\Delta t)$$

$$\simeq G(t_0 + 2\Delta t, t_0 + 3\Delta t) \cdot E_{X_e}(t_0 + 2\Delta t) + G(t_0 + 2\Delta t, t_0 + 3\Delta t) \cdot \sum_{i \neq d, i \neq e} H_i(t_0, t_0 + 2\Delta t)\Delta X_d(t_0) \tag{19}$$

If the system is in equilibrium, or if Δt, the time discretization, is sufficiently small, $\Delta X_d(t_0 + 3\Delta t) \simeq \Delta X_d(t_0)$, thus:

$$\left(1 + G(t_0 + 2\Delta t, t_0 + 2\Delta t) \cdot \sum_{i \neq d, i \neq e} H_i(t_0, t_0 + 3\Delta t)\right) \Delta X_d(t_0 + 3\Delta t) \simeq$$

$$G(t_0 + 2\Delta t, t_0 + 3\Delta t) \cdot E_{X_e}(t_0 + 2\Delta t) \tag{20}$$

$$\Rightarrow \Delta X_d(t_0 + 3\Delta t) \simeq \frac{G(t_0 + 2\Delta t, t_0 + 3\Delta t)}{1 - G(t_0 + 2\Delta t, t_0 + 3\Delta t)\sum_{i \neq d, i \neq e} H_i(t_0, t_0 + 3\Delta t)} E_{X_e}(t_0 + 2\Delta t) \tag{21}$$

$$\simeq \frac{G(t_0 + 2\Delta t, t_0 + 3\Delta t)}{1 - \sum_{i \neq d, i \neq e} f_i(t_0, t_0 + 3\Delta t)} E_{X_e}(t_0 + 2\Delta t) \tag{22}$$

11

where the terms $f_i(t_0, t_0 + 3\Delta t) = G(t_0 + 2\Delta t, t_0 + 3\Delta t) \cdot H_i(t_0, t_0 + 2\Delta t)$ are called the feedback factors. The gain with feedbacks is then defined by:

$$G_f(t_0 + 2\Delta t, t_0 + 3\Delta t) = \frac{G(t_0 + 2\Delta t, t_0 + 3\Delta t)}{1 - \sum_{i \neq d, i \neq e} f_i(t_o, t_0 + 3\Delta t)}. \tag{23}$$

The feedback f_i factors are dependent on **both** the variable X_e perturbed by the external forcing and the diagnosed variable, X_d, chosen in the beginning of the analysis. These feedback factors are time-dependent, but this expression is traditionally [*Peixoto and Oort,* 1991; *Curry and Webster,* 1998] given without time reference. This means that it is supposed that the system is in equilibrium or that the quantities are examined locally in time.

Another way to find this expression, is to formulate this system as a "mono-variable" forced dynamical system $\Delta X_d(t_0) \rightarrow \Delta X_d(t_0 + \Delta t)$. The total gain of this system is defined as GH which represents the feedback loops plus the non-feedback gain, where:

- $G = \frac{\partial X_d(t_0 + \Delta t)}{\partial X_e(t_0)}$ is the gain without feedbacks of the system $\Delta X_e \rightarrow \Delta X_d$

- and $H = \sum_{i \neq d, i \neq e} \frac{\partial X_e(t_0 + 2\Delta t)}{\partial X_i(t_0 + \Delta t)} \frac{\partial X_i(t_0 + \Delta t)}{\partial X_d(t_0)}$, represents the feedbacks.

The forcing of the variable X_d is given by GE_{X_e}. In the limit of decreasing time steps, we could use the expression (12), to obtain:

$$\Delta X_d = \frac{G}{1 - GH} E_{X_e} \tag{24}$$

This expression converges to the continuous case as $\Delta t \rightarrow 0$. In the original field where this formalism was developed, i.e., the analysis of electrical circuits [*Bode,* 1945], the relation (15) is instantaneous since the electricity propagates continuously. In this continuous case, the time reference in equation (15) can be suppressed. The same remark holds if the system is in equilibrium, i.e., if the previous perturbations are constant in time. Thus, this analysis has to be done locally in time or at equilibrium.

The gain of the system, G_f, is very sensitive to the estimation of the factors, f_i. Furthermore, it is very important to estimate all these factors simultaneously since the

effect of one particular feedback is sensitive to the presence or or the absence of other feedbacks.

Figure 3 shows the gain of the system, G_f, as a function of the unique feedback factor, f, supposing that the gain of the system without feedback is $G = 0.5$. If $f < 0$, the gain with feedback is damped, $0 < G_f < G$. If $f = 0$, the gain with feedback is unchanged, $G_f = G$. If $0 < f < 1$, the gain of the system with feedback is increased, $G_f > G$, and $\lim_{f \to 1} = +\infty$ (the system becomes unstable). If $f > 1$, G_f is negative, so the system oscillates, and it is unstable if $G_f < -1$. We see in this figure how the effect of a feedback factor on the system can be sensitive and highly nonlinear. So the significance of a feedback factor is strongly dependent on the availability of the feedback factors of all variables: an isolated feedback factor can't characterize the behavior of the whole system.

e. One classical example

The following example as been intensively used in the literature. We suppose that the global mean net radiation flux (solar minus terrestrial) at the top of atmosphere (TOA) is in equilibrium ($\Delta F_{TOA} = 0$). The question is: if a forcing is introduced into the system, how will the system react ? The global mean surface temperature Ts is often taken as diagnosed variable since a lot of other internal variables of the system are dependent on this variable. Then, we can analyze the feedback process loops acting on Ts using the above formalism if all feedbacks are assumed to act in parallel.

A forcing $E_{X_{ext}}$ is introduced into an external variable, X_{ext} (i.e. the solar insolation, volcanic eruptions, etc). We analyze the system:

$$F_{TOA}(t + \Delta t) = F(X_{ext}(t), X_i(t), Ts(t)). \tag{25}$$

The terms X_i are the internal variables of the system (i.e. that depend on the surface temperature, $X_i = X_i(Ts)$) like the albedo, the water vapor, the lapse rate, the clouds, etc.

We suppose here that it is possible to express the external forcings, $E_{X_{ext}}$, in terms of perturbations of the net radiation flux, E_{TOA}. The forcing introduces perturbations into the variables of the system; the link between these perturbations can be expressed by, see

13

equation (17):

$$\Delta F_{TOA}(t_0 + 2\Delta t) = \underbrace{E_{TOA}(t_0 + 2\Delta t)}_{\text{external forcing}} + \underbrace{\sum_i \frac{\partial F_{TOA}(t_0 + 2\Delta t)}{\partial X_i(t_0 + \Delta t)} \frac{\partial X_i(t_0 + \Delta t)}{\partial Ts(t_0)}}_{\text{feedback loops}} \Delta Ts(t_0) \quad (26)$$

If the equilibrium state is reached, or if the sensitivities are instantaneous, the reference to time can be suppressed:

$$\Delta F_{TOA} = E_{TOA} + (\sum_i H_i)\Delta Ts \quad (27)$$

where the terms H_i are the products of first derivatives describing the cause and effect relations in equation (26). By multiplying this expression by the gain of the system without feedbacks, $G = \frac{\partial Ts}{\partial F_{TOA}}$, we finally obtain the following familiar expression:

$$\Delta Ts = \frac{G}{1 - G\sum_i H_i} E_{TOA} = \frac{G}{1 - \sum_i f_i} E_{TOA} \quad (28)$$

f. The classical analysis in a series feedbacks configuration

It is supposed again that the external forcings, E_{X_e}, of the system acts on only one variable, X_e, of the system. There are two diagnosed variables: X_{d1} and X_{d2}, X_{d2} being dependent on X_{d1}. Some of the internal variables $\{X_{i1} = X_{i1}(X_{d1})\}$ are dependent on X_{d1}, and some others $\{X_{i2} = X_{i2}(X_{d2})\}$ are dependent on X_{d2}. The impact of the external forcing is observed on diagnosed variable X_{d2} (Figure 4). This internal structure describes a dynamical system $X_e \rightarrow X_{d1} \rightarrow X_{d2}$, with feedbacks in series. In this case, the gain of the subsystems $X_e \rightarrow X_{d1}$ and $X_{d1} \rightarrow X_{d2}$ would be computed as in section 2.d. Then, the global gain of the system would be $G_f = G_{f_2} \cdot G_{f_1}$.

g. Comments on classical feedback analysis

We have seen in the two previous subsections that where particular cause and effect relations in the system are known, the time reference is required in the discrete case, but can be suppressed in two situations:

- In an **equilibrium** state: the perturbations are stabilized $\frac{\partial X}{\partial t} = 0$ (not to be confused with zero forcing), so they are the same at each time step. The feedback analysis is

14

then only a characterization of the equilibrium state. There is no estimation of the time required to reach the equilibrium and we can't predict transient behavior of the system. Furthermore, we don't know a priori the sensitivities in the equilibrium state, so we are required to assume (without evidence) that the sensitivities are constant and that we have a good estimate of them.

- When the sensitivities are **instantaneous**: the relations between the perturbations of each variable of the system are then valid without a time reference. But in this case, instantaneous estimates of the sensitivities are required and the feedbacks factors have to be computed at each time. To our knowledge, this approach has not yet been investigated since no technique was available to estimate these instantaneous, multivariate and nonlinear sensitivities.

The classical approach to feedback analysis from the electrical circuits theory [*Bode,* 1945] was first used on a theoretical energy balance model of the climate where instantaneous sensitivities are available. Even if the estimation of sensitivities was crude, the applicability of the technique was justified when the cause and effect relationships were supposed to be known. In more recent studies, and particularly with the analysis of observations, this approach to the estimation of sensitivities is highly questionable. In particular, the use of this characterization of the equilibrium state to predict the system response to an external forcing is difficult since the sensitivities used to produce the equilibrium state are unknown.

Some of the limitations of actual studies are:

- **Model used:** The hierarchical model of cause and effect relations, described by greatly simplified relations between sensitivities, is usually much too simple. For example, the fact that the forcing/gain/response system has to be mono-variable is a very strong simplification: such assumptions result in the suppression/neglect of some perturbations and some first derivatives in the system.

- **Estimation of sensitivities:** The sensitivities are often estimated by finite difference between two (usually equilibrium) states of the system. First, this approach measures only the coincidence of the changes in two quantities, but it does not mean that there is a cause

15

and effect relationship between these variables. The relationships might also be indirect (via the ordering of the dependencies). Second, this approach measures the changes in two quantities and the sensitivity is then estimated assuming that the other variables do not interact. This is a strong limitation since there are a lot of cross-linkages in the variables of the climate system. Third, the finite difference for the estimation of the sensitivities can be highly misleading if the sensitivities of the system are not constant in time.

- **Forcing process:** The forcing model is often not expressed: localized in time, constant, growing in time, cyclic, etc ? The ways the external forcing evolves in time are also important for the study of the transient response.

- **Better description:** Previous approaches to feedback analysis are often only a characterization of the equilibrium state of the system after the introduction of an external forcing. The transient period between the beginning of the forcing and the equilibrium state is not described, the time to reach the equilibrium is not estimated. This is a real drawback for the understanding of these phenomena. Furthermore, the gain of the system with feedback factors is highly dependent on the precision of the sensitivity estimates.

In conclusion, the classical feedback analysis is limited by some very strong assumptions like linearity (i.e., sensitivities constant in time), equilibrium, mono-variable cause and effect relationships, etc, and so does not seem at all appropriate for application to the climate system. Moreover, the resulting expressions for the feedback factors are products of the instantaneous sensitivities, so it would seem more straightforward to evaluate these sensitivities instead. To avoid the classical limitations, the analysis needs to employ a general feedback formulation to evaluate the nonlinear, multivariate and instantaneous sensitivities in both numerical models and observations is important.

3. A nonlinear regression scheme for estimation of sensitivities

To estimate the sensitivities of the dynamical model in equation (1), we use a multivariate nonlinear regression fit to the statistics produced by observing the behavior of the system over a time period long enough to provide a good sample of the different states of

the system. For this purpose, we introduce a neural network technique because of its ability to process large dimension data (which will be helpful for further experiments on numerical models) and its capacity to integrate a priori information about the problem [*Aires, 1999*]. Any other multivariate nonlinear regression technique, such as spline interpolation or ARMAX models, etc, could be used instead of the neural network technique.

a. The neural network model

The Multi-Layer Perceptron (MLP) network is a mapping model composed of parallel processors called "neurons". These processors are organized in distinct layers: the first layer (number 0) represents the input $P = (p_i \; ; \quad 0 \leq i \leq m_0)$ with m_0 the number of neurons in layer 0. The last layer (number L) represents the output mapping $X = (x_k \; ; \quad 0 \leq k \leq m_L)$. The intermediate layers ($0 < m < L$) are called the "hidden layers". These layers are connected via neuronal links (Figure 5): two neurons, i and j, between two consecutive layers have synaptic connections associated with a synaptic weight w_{ij}.

Each neuron, j, executes two simple operations: first, it makes a weighted sum of its inputs from the previous layer, z_i; this signal is called the activity of the neuron:

$$a_j = \sum_{i \in \text{Inputs}(j)} w_{ij} \cdot z_i. \tag{29}$$

Then, it transfers this signal to its output through a so-called "transfer function", often a sigmoidal function such as $\sigma(a) = tanh(a)$. The output z_j of neuron j in the hidden layer is then given by: $z_j = \sigma \left(\sum_{l \in \text{Inputs}(j)} w_{lj} \, z_l \right)$. Generally, for regression problems, the neurons in the output (last) layer have no transfer function. For example, in a one hidden layer MLP (Figure 5), the kth output, x_k, of the network is defined as:

$$x_k(y) = \sum_{j \in S_1} w_{jk} \, \sigma(a_j) = \sum_{j \in S_1} w_{jk} \, \sigma \left(\sum_{i \in S_0} w_{ij} \, p_i \right) \tag{30}$$

where σ is the sigmoidal function, a_j is the activity of neuron j and S_i is the ith layer of the network (with $i = 0$ for the input layer). We have deliberately omitted the usual bias term in this formula for clarity, but include it in the actual network.

The key to our analysis is that any continuous function can be represented by a one-hidden layer MLP with this kind of sigmoid function [*Hornik et al.*, 1989; *Cybenko*, 1989]. Hence the process of training the MLP to fit the observed multi-variate, nonlinear relationship statistics is equivalent to deriving a multi-variate, nonlinear function that behaves in a similar fashion as the dynamic system in question. The key advantage of the neural network approach over some other methods is that the Jacobians (i.e., sensitivities) can be evaluated directly from the MLP (see next section).

b. The learning algorithm

Given a neural architecture (number of layers, neurons and connections, type of transfer functions), all the information of the network is contained in the set of synaptic weights w_{ij}. The learning algorithm is an optimization technique that estimates the network parameters $W = \{w_{ij}\}$ by minimizing a loss function, $C(W)$, needed to fit the desired function defined by observations as closely as possible. The criterion usually used to adjust W is the mean square error in network outputs:

$$C(W) = \frac{1}{2} \sum_{k=1}^{m_L} \int \int (x_k(\boldsymbol{P}; W) - t_k)^2 H(t_k/\boldsymbol{P}) H(\boldsymbol{P}) dt_k d\boldsymbol{P} \tag{31}$$

with t_k the kth desired output component, x_k the kth neural output component, $H(t_k/\boldsymbol{P})$ the probability function of output t_k given the input $\boldsymbol{P}$, and $H(\boldsymbol{P})$ the probability density function of input data, $\boldsymbol{P}$. If specific a priori information about the probability distribution functions is available, other quality criteria than least-squares could be used. For example, criteria involving higher-order statistics have been defined [*Aires et al.*, 2000]. Practically, $C(W)$ is approximated by the classical least square criterion:

$$\overline{C}(W) = \frac{1}{2E} \sum_{e=1}^{E} (x_k(\boldsymbol{P}; W) - t_k)^2 \tag{32}$$

The Error Back-Propagation algorithm [*Rumelhart et al.*, 1986] is used to minimize $\overline{C}(W)$. It is a stochastic steepest descent (i.e. Newtonian minimization procedure) very well adapted to the MLP hierarchical architecture because the computational cost is linearly related to the number of parameters.

18

c. The neural Jacobians

The important feature of neural network is that the adjoint model of the neuronal model is directly available [*Aires et al., 1999*]. The computation of this adjoint model (or neural Jacobians) is accurate and very fast. Since the neural network is nonlinear, these Jacobians are dependent on the situation x. For example, the neural Jacobians in the previous example of equation (30) (a MLP network with one hidden layer) are:

$$\frac{\partial x_k}{\partial p_i} = \sum_{j \in S_1} w_{jk} \frac{\partial \sigma}{\partial a} \left(\sum_{i \in S_0} w_{ij} p_i \right) w_{ij}. \tag{33}$$

where $\frac{\partial \sigma}{\partial a}$ is the derivative of the transfer function σ. For a more complex MLP network with many hidden layers, there still exists a back-propagation algorithm for efficiently computing these neural Jacobians.

The neural Jacobians concept is a very powerful tool because it allows for the direct statistical evaluation of the multivariate and nonlinear sensitivities of the dynamical system under study.

d. Regularization

If a priori information about the dynamical model under study is available, it is possible and recommended to use this knowledge in the neural network analysis. This a priori information could be introduced in the three distinct components of a neural network:

- **Dataset**: The quality and the representativeness of the dataset used for the training of the neural network is directly responsible for the quality and the generality of the nonlinear regression obtained. We will comment further on this during the construction of the dataset for our application.

- **Architecture**: A lot of information could be used to define the neural network architecture: ordering of variables, neighborhood system between variables, dependencies structure, etc. A particularly promising development would be to use the ordering of dependencies discussed in Section 2d to define the neuronal links of the neural model.

- **Training**: If the sensitivity between an input and an output variable is already known or if it is known that this sensitivity is constant, it is possible to specify this a priori information as a penalty term added to the quality criterion (32) used to train the neural network. Other kinds of solution constraints can also be used: shape of the solution distribution, noise in the measurements, particular dependencies between variables. Such an approach has been used, for example, in [*Aires et al.*, 1999] in the atmospheric radiative transfer field.

4. Analysis of the discrete Lorenz model

To test the definitions and the technique previously presented, we apply it to a simple nonlinear, multivariate, chaotic, non-stationary and forced dynamical model for which the sensitivities are known analytically. We choose here a discrete form of the low-order Lorenz model [*Lorenz*, 1984]. This model is very general since it is not a mono-variable structure, as described in Sections 2d and 2f, and it exhibits very complex behavior. Nonetheless, we can define the time relationships directly from the equations of the model to test our ability to infer these relationships from the observed behavior (model output).

We have discretized the Lorenz continuous model for two reasons: first, the discrete formulation makes it easier to describe the cause and effect relations of the feedback processes. Second, we know exactly, in this case, the analytical sensitivities of the system, which allows for a better quantitative evaluation of our analysis technique.

a. Continuous Lorenz model

The low-order model used in this study was developed by Lorenz [*Lorenz*, 1984; *Lorenz*, 1990] to analyze the chaos and stability assumptions about the atmospheric circulation. This simple model is able to represent the Hadley circulation and is used to determine the stability or the instability of this circulation (stationary or migratory disturbance). This

model is defined by three Ordinary Differential Equations (ODE):

$$\begin{cases} \frac{dX(t)}{dt} & = -Y^2(t) - Z^2(t) - a\,X(t) + a\,F_1 \\ \frac{dY(t)}{dt} & = X(t)\,Y(t) - b\,X(t)\,Z(t) - Y(t) + F_2 \\ \frac{dZ(t)}{dt} & = b\,X(t)\,Y(t) + X(t)\,Z(t) - Z(t) \end{cases} \qquad (34)$$

where:

- t is the time (in units of about 1 day),

- X is the intensity of the symmetric globe-encircling westerly wind current and also the poleward temperature gradient (assumed to be in permanent equilibrium with it),

- Y is the cosine phase of a chain of superposed large-scale eddies, which transport heat poleward at a rate proportional to the square of their amplitudes,

- Z is the sine phase of a chain of superposed large-scale eddies, which transport heat poleward at a rate proportional to the square of their amplitudes,

- F_1 is a zonally symmetric thermal forcing on X,

- F_2 is a zonally asymmetric thermal forcing on Y.

The two forcings F_1 and F_2 are the values to which X and Y would be driven if the westerly current and the eddies were not coupled.

The discretization of these ODEs is a very delicate process, but the Runge-Kutta fourth-order technique can be used for that purpose. Figure 6 shows the integration of 34 from $t_0 = 0$ to $T = t_0 + N\Delta t$ using: $a = 0.25$, $b = 4$, $F_1 = 8$, $F_2 = 1$ and $\Delta t = 0.08$. The initial state of the system at time $t = 0$ is taken as: $X(0) = 1.312465072$, $Y(0) = 1.486416698$ and $Z(0) = 0.3487878144$. Lorenz has shown that this system with these parameter values has a chaotic behavior.

c. Discretization of the dynamical system

We are not interested in a perfect simulation of the Lorenz model; rather, we are interested in a representation of this system in a form like

$$\begin{pmatrix} X(t+1) \\ Y(t+1) \\ Z(t+1) \end{pmatrix} = \mathcal{A} \begin{pmatrix} X(t) \\ Y(t) \\ Z(t) \end{pmatrix},$$ (35)

as a test of our analysis technique.

By discretizing (34), we obtain:

$$\begin{cases} X(t+1) &= \Delta t \left[-Y(t)^2 - Z(t)^2 + a\, F_1 \right] &+ (1 - a\, \Delta t)\, X(t) \\ Y(t+1) &= \Delta t \left[-b\, X(t)\, Z(t) + F_2 \right] &+ (1 - \Delta t + \Delta t\, X(t))\, Y(t) \\ Z(t+1) &= \Delta t\, b\, X(t)\, Y(t) &+ (1 + \Delta t\, X(t) - \Delta t)\, Z(t) \end{cases}$$ (36)

where Δt is the discrete time step.

The size of the time step Δt needs to be sufficiently small so that the linearization of the system during a time interval is accurate in order that means that the hypothesis that the Jacobians of the system are constant during the time interval is true.

The time discretization is also directly related to the regularity of the Jacobians of the system: high complexity requires small time steps to ensure a good description of the evolution of the Jacobians. We take $\Delta t = 0.08$.

The differences between the fourth-order Runge-Kutta integration of (34) and the discrete Lorenz model 36 are shown in the Figure 7. We see that these differences are small at the beginning of the period, but that the amplification of these little difference becomes important over time since the system is chaotic. The behavior of the continuous and the discrete systems seems to be the same, in particular the same amplitudes for the minima and maxima are observed. All that is required to test our analysis is that the discrete model exhibit complex, nonlinear behavior: we take the discrete model to be the truth and test whether we can infer the correct the relationships with our neural network technique.

22

d. Sensitivities of the dynamical system

The Jacobian matrix of the discrete system is:

$$
G \begin{pmatrix} X(t+1) \\ Y(t+1) \\ Z(t+1) \end{pmatrix} = \begin{pmatrix} \frac{\partial X(t+1)}{\partial X(t)} & \frac{\partial X(t+1)}{\partial Y(t)} & \frac{\partial X(t+1)}{\partial Z(t)} \\ \frac{\partial Y(t+1)}{\partial X(t)} & \frac{\partial Y(t+1)}{\partial Y(t)} & \frac{\partial Y(t+1)}{\partial Z(t)} \\ \frac{\partial Z(t+1)}{\partial X(t)} & \frac{\partial Z(t+1)}{\partial Y(t)} & \frac{\partial Z(t+1)}{\partial Z(t)} \end{pmatrix} \tag{37}
$$

$$
= \begin{pmatrix} 1 - a\,\Delta t & -2\,\Delta t\,Y(t) & -2\,\Delta t\,Z(t) \\ -\Delta t\,b\,Z(t) + \Delta t\,Y(t) & 1 - \Delta t + \Delta t\,X(t) & -b\,\Delta t\,X(t) \\ \Delta t\,b\,Y(t) + \Delta t\,Z(t) & \Delta t\,b\,X(t) & 1 + \Delta t\,X(t) - \Delta t \end{pmatrix} \tag{38}
$$

These Jacobians are dependent on the state of the system, so they are also dependent on time and the hypothesis of constant Jacobians, as in classical feedback analysis, can not be used to understand this system.

e. Theoretical feedback analysis

The two external forcing, aF_1 on X, and F_2 on Y, are continuous and constant in (34). In the discrete formalization, this is simulated by a constant impulsive forcing:

$$
\begin{cases} \Delta X(t_0 + k\,\Delta t) = \Delta t\,a\,F_1 \\ \Delta Y(t_0 + k\,\Delta t) = \Delta t\,F_2 \end{cases} \quad \text{for } k = 0, \ldots, N \tag{39}
$$

If the beginning state of the simulation is chosen as:

$$
\begin{cases} X(t_0) = 0 \\ Y(t_0) = 0 \\ Z(t_0) = 0 \end{cases} \tag{40}
$$

then, the state of the system at the next time step is given by:

$$
\begin{cases} X(t_0 + \Delta t) = \Delta t\,a\,F_1 \\ Y(t_0 + \Delta t) = \Delta t\,F_2 \\ Z(t_0 + \Delta t) = 0 \end{cases} \tag{41}
$$

23

and, for the next time step:

$$
\begin{cases}
X(t_0 + 2\,\Delta t) & = 2\,\Delta t\; a\; F_1 - \Delta t^3\; F_2{}^2 - \Delta t^2\; a^2\; F_1 \\
Y(t_0 + 2\,\Delta t) & = 2\,\Delta t\; F_2 - \Delta t^2\; F_2 + \Delta t^3\; a\; F_1\; F_2 \\
Z(t_0 + 2\,\Delta t) & = \Delta t^3\; a\; b\; F_1\; F_2
\end{cases}
\tag{42}
$$

and so on. We analyze the impacts of the external forcings, $a\,F_1$ and F_2, in the diagnosed variable, chosen here to be X.

The perturbation at time $t_0 + \Delta t$, $\Delta_X(t_0 + \Delta t) = \Delta t\; a\; F_1$, is straightforward. At time $t_0 + 2\,\Delta t$, without feedbacks, the forcing would simply be added:

$$
\Delta X(t_0 + 2\,\Delta t) = 2\,\Delta t\; a\; F_1
\tag{43}
$$

With feedbacks, the true perturbation is given by:

$$
\Delta X(t_0 + 2\,\Delta t) = E_X(t_0 + 2\,\Delta t) + \frac{\partial X(t_0 + 2\,\Delta t)}{\partial X(t_0 + \Delta t)} E_X(t_0 + \Delta t)
$$

$$
+ \frac{\partial X(t_0 + 2\,\Delta t)}{\partial Y(t_0 + \Delta t)} E_Y(t_0 + \Delta t)
$$

$$
= 2\Delta t a F_1 - \Delta t^3 F_2{}^2 - \Delta t^2 a^2 F_1
\tag{44}
$$

Comparing expression (43) and expression (44), we note the presence of two correction factors giving the contribution of the feedback processes: so far there are feedbacks caused only by the integration of the variables over time. This expression for the perturbation is in agreement with first equation in (42).

For the description of the indirect feedbacks, three time steps are required. At time $t_0 + 3\,\Delta t$, the integration of the external forcings is even more complex:

$$
\Delta X(t_0 + 3\,\Delta t) = \underbrace{E_X(t_0 + 3\,\Delta t)}_{\text{external forcing}} + \underbrace{\frac{\partial X(t_0 + 3\,\Delta t)}{\partial X(t_0 + 2\,\Delta t)} E_X(t_0 + 2\,\Delta t) + \frac{\partial X(t_0 + 3\,\Delta t)}{\partial Y(t_0 + 2\,\Delta t)} E_Y(t_0 + 2\,\Delta t)}_{\text{direct feedbacks}}
$$

$$
+ \underbrace{\left[\frac{\partial X(t_0 + 3\,\Delta t)}{\partial X(t_0 + 2\,\Delta t)} \frac{\partial X(t_0 + 2\,\Delta t)}{\partial X(t_0 + \Delta t)} + \frac{\partial X(t_0 + 3\,\Delta t)}{\partial Y(t_0 + 2\,\Delta t)} \frac{\partial Y(t_0 + 2\,\Delta t)}{\partial X(t_0 + \Delta t)} + \frac{\partial X(t_0 + 3\,\Delta t)}{\partial Z(t_0 + 2\,\Delta t)} \frac{\partial Z(t_0 + 2\,\Delta t)}{\partial X(t_0 + \Delta t)} \right] E_X(t_0 + \Delta t)}_{\text{indirect feedbacks}}
$$

$$
+ \underbrace{\left[\frac{\partial X(t_0 + 3\,\Delta t)}{\partial X(t_0 + 2\,\Delta t)} \frac{\partial X(t_0 + 2\,\Delta t)}{\partial Y(t_0 + \Delta t)} + \frac{\partial X(t_0 + 3\,\Delta t)}{\partial Y(t_0 + 2\,\Delta t)} \frac{\partial Y(t_0 + 2\,\Delta t)}{\partial Y(t_0 + \Delta t)} + \frac{\partial X(t_0 + 3\,\Delta t)}{\partial Z(t_0 + 2\,\Delta t)} \frac{\partial Z(t_0 + 2\,\Delta t)}{\partial Y(t_0 + \Delta t)} \right] E_Y(t_0 + \Delta t)}_{\text{indirect feedbacks}}
\tag{45}
$$

24

We note in this expression some terms that do not appear in the classical analysis formalism. For example, the direct feedbacks terms (due to time integration of the variables) are suppressed in the classical analysis. Furthermore, we see that in this expression both forcings (on variable X and on variable Y) are taken into account, which is not possible in the classical approache.

Integrating the system for one more time step would be highly complex, this is the reason why prediction of this kind of dynamical system is a difficult problem. To perform prediction, the model needs to represent the sensitivities with a high degree of precision. Otherwise, an error at one time step is rapidly amplified in the next time steps.

The classical formalism for the feedback analysis is not well adapted to the analysis of the Lorenz system since here is no preferred variable on which the other two variables of the model depends. So we see in this simple example how limited the assumptions used in the classical feedback analysis formalism are and how such an analysis could be misleading. Again, it is clear that evaluation of the sensitivities is more straightforward than evaluation of feedback factors, which are products of sensitivities. However, for illustrative purpose we will also use the classical formalism to calculate feedback factors because they are more familiar. If we choose the variable Y as the variable affected by the external forcing and X as the diagnosed variable, the gain of the system $E_y \rightarrow \Delta X$ is given by, see equation (24):

$$\Delta X = \frac{H}{1 - GH} E_Y = G_f \cdot E_Y \tag{46}$$

$$= \frac{G}{1 - \sum_i f_i^{XY}} E_Y \tag{47}$$

where:

- $G = \frac{\partial X}{\partial Y}$ is the gain without feedbacks of the system $E_Y \rightarrow \Delta X$,

- and $H = \frac{\partial Y}{\partial X} + \sum_i \frac{\partial Y}{\partial X_i} \frac{\partial X_i}{\partial X}$.

The three feedbacks factors for this mono-variable system are defined as:

$$f_X^{YX} = \frac{\partial X}{\partial Y} \frac{\partial Y}{\partial X} \tag{48}$$

25

$$f_Y^{YX} = \frac{\partial X}{\partial Y}\frac{\partial Y}{\partial Y}\frac{\partial Y}{\partial X} \tag{49}$$

$$f_Z^{YX} = \frac{\partial X}{\partial Y}\frac{\partial Y}{\partial Z}\frac{\partial Z}{\partial X} \tag{50}$$

Note that the sensitivities used in this relation still are dependent on time and have to be estimated precisely, so that the feedback factors are also time dependent. As we will show, this fundamental property of complex, nonlinear dynamical systems reduces the value of the classical (linear) feedback analysis for understanding the system behavior. Note again that the above quantities are not the true feedback factors for the Lorenz model since they are defined using invalid assumptions.

5. Experimental results

a. Construction of the dataset

The quality of the dataset used to evaluate the sensitivities is a crucial point. For example, using data from a system in equilibrium or from a system during a transient change will not give the same results in the analysis. Ideally, a good dataset would be one including all ranges of variability for all combinations of the variables of the system. The more situations that are included in the dataset, the larger will be the range of validity of the sensitivity estimates. This situation parallels that in climate analysis where the range of validity is limited by the range of climate states actually observed.

The discrete dynamical version of the Lorenz model stabilizes more rapidly onto its attractor than the continuous version. So to create a dataset closer to the behavior of the continuous system, we choose 200 noisy states of the continuous system as initial states for 200 trajectories of 1000 times steps of the discrete system in equation (36). The final dataset is then composed of $N = 200,000$ couples $\{(I^k, O^k)\ ;\ k = 1, \ldots, N\}$, where $I^k = (X(t_0 + k\ \Delta t), Y(t_0 + k\ \Delta t), Z(t_0 + k\ \Delta t))$ is an $N \times 3$ matrix of the inputs of the system and $O^k = (X(t_0 + (k+1)\ \Delta t), Y(t_0 + (k+1)\ \Delta t), Z(t_0 + (k+1)\ \Delta t))$ is an $N \times 3$ matrix of the outputs. Each couple is linked by: $O^k = \mathcal{A}(I^k)$.

The parameters for the Lorenz model are the same as previously: $a = 0.25$, $b = 4$,

$F_1 = 8$, $F_2 = 1$ and $\Delta t = 0.08$, but we have introduced a Gaussian noise $\mathcal{N}(0, 0.001)$ at each time step and in each variable. Figure 8 shows the resulting noisy trajectories included in the dataset.

b. Linear and nonlinear regressions

If a priori information is available about what are the good predictors, the dynamical system can be described as a linear model. In the Lorenz case, the good predictors, $P(t)$, of the general model (1) can be determined from the theoretical model (36):

$$P(t) = (X(t), Y(t), Z(t), Y^2(t), Z^2(t), X(t)Y(t), X(t)Z(t), F_1, F_2). \tag{51}$$

In this configuration, the dynamical system of equation (36) becomes:

$$\begin{pmatrix} X(t+1) \\ Y(t+1) \\ Z(t+1) \end{pmatrix} = A \cdot (X(t), Y(t), Z(t), Y^2(t), Z^2(t), X(t)Y(t), X(t)Z(t), F_1, F_2) \tag{52}$$

where the constant matrix A is given by:

$$A = \begin{pmatrix} 1 - a\,\Delta t & 0 & 0 & -\Delta t & -\Delta t & 0 & 0 & a\,\Delta t & 0 \\ 0 & 1\Delta t & 0 & 0 & 0 & -\Delta t & -b\,\Delta t & 0 & \Delta t \\ 0 & 0 & 1 - \Delta t & 0 & 0 & b\,\Delta t & \Delta t & 0 & 0 \end{pmatrix} \tag{53}$$

A linear regression in this case would give a good estimate of the matrix A. This is a very general idea: all nonlinear dynamical systems could be simplified, and even linearized, if all of the good predictors are known.

In practise, this a priori information is not available, so choosing the good variables to predict system behavior is a key issue that has no general answer. Usually, then, the predictors are chosen as the state variables; model (1) becomes:

$$\begin{pmatrix} X(t+1) \\ Y(t+1) \\ Z(t+1) \end{pmatrix} = \mathcal{A} \begin{pmatrix} X(t) \\ Y(t) \\ Z(t) \end{pmatrix} \tag{54}$$

27

Now, a linear regression analysis approximates the nonlinear function $\mathcal{A}$ by a linear model: $\mathcal{A}$ is replaced in (54) by a 3×3 matrix $\boldsymbol{A}$. This matrix is estimated by minimizing the least squares criterion and is given by:

$$A = \frac{\mathcal{I}^t \cdot \mathcal{I}}{\mathcal{I}^t \cdot \mathcal{O}} \tag{55}$$

The use of this linear regression is already an improvement compared to classical approaches because it allows the simultaneous estimation of multivariate sensitivities.

For a nonlinear regression, we use an MLP network with one hidden layer. The architecture has three units in the input layer coding $\boldsymbol{I} = (X(t), Y(t), Z(t))$, 30 units in the hidden layer (this number was chosen by trial in the training phase) and three units in the output layer coding the prediction, $\boldsymbol{O} = (X(t+1), Y(t+1), Z(t+1))$.

For the training of the neural network (i.e. estimation of the parameters for the nonlinear regression), we have used 150,000 points randomly chosen from the data set previously constructed and for the test data (i.e. to measure the ability of the model to generalize to unknown data) we have taken the remaining 50,000 points.

In Figure 9, the theoretical (points) function $\mathcal{A}$ and its two estimates (by linear and neural network regressions) are illustrated. For display purpose, each plot represents one of the variables at time $t+1$, as a function of a variable at time t, supposing that the two other variables are equal to their mean values. It is clear that the neural network regression is very precise (differences with the theoretical function are undetectable) and useful for the nonlinear behavior ($X(t+1)$ as a function of $Y(t)$, for example), where the linear regression is very poor. This figure shows how important the nonlinear aspect is: the multivariate approach of the linear regression is not sufficient. These conclusions are confirmed in Figure 10 where the RMS error for the estimation of the functions is given. Here the errors of the linear regression are nearly as large as the variability of the quantities.

A dilemma that we will be faced with in applying this technique to a real case, numerical model or observations of the climate, is that we do not know the true answer as we do here for the Lorentz model. Hence, we must develop practical ways to assess the fidelity of the analysis results. One possibility is to conduct "prediction" experiments where

we pick many specific and different episodes in the observed record (preferably time periods not included in the original analysis), initialize the neural network at the beginning state, and calculate forward for a short time interval. The goal of such experiments is diagnostic, to test quantitatively whether the derived sensitivities used in the neural network can reproduce the observed system dynamics in cases not included in the original analysis. It is not our goal to propose that such a neural network be used for climate forecasts (i.e, be used as a statistical forecast model) in place of a physical model of the climate (see [*Yuval,* 1999] for a previous study on this subject). Rather, we are interested in whether the derived sensitivities can be used to understand the physical processes; at least the sensitivities of a model can be compared with observations. We have tested this idea by making prediction runs with our neural network representation of the Lorentz model: the calculation proceeds by calculating the state of the system at time step, $t + 1$, from the state and sensitivities of the system at time, t; the sensitivities are then calculated at time, $t + 1$, and used in the next cycle. Figure 11 shows the evolution of the rms error of the predictions based on the linear and our nonlinear statistical models against the actual model started at the same state (each time step = 0.08 units, about 2 hr in the scaling of the equations). As expected, the nonlinear regression by the neural network does much better than the linear regression, but the fact that the Lorentz system is chaotic (with the particular parameter values used) results in a relatively rapid increase of prediction error, even with an accurate approximation of the system dynamics. Figure 12 illustrates the time records from the prediction model and the actual model.

c. Analysis of sensitivities

We illustrate the retrieval of the variable sensitivities in the form of histograms of their distribution of values as a functions of $X(t)$. Similar figures (not shown) are obtained as function of $Y(t)$ or $Z(t)$. The standard-deviation of the theoretical sensitivities of the system are shown in Figure 13, indicating that the all sensitivities of the system, except for $\frac{\partial X(t+1)}{\partial X(t)}$, are not constant.

The classical approach for the estimation of sensitivities takes the finite difference in two variables between two (usually equilibrium) states of the system or two extreme events. For example, for the estimation of $\frac{\Delta X}{\Delta Y}$, two sets of extreme events of the variable Y could be selected in the observations and the averages of the state differences $< \Delta X >$ and $< \Delta Y >$ estimated. Then, the following approximation would be used:

$$\frac{\partial Y}{\partial X}(t) \simeq \frac{< \Delta X >}{< \Delta Y >}. \tag{56}$$

We see how this approach can go wrong because it is so dependent on the selection of data: at best, it gives a crude estimate of the mean sensitivity for the selected dataset of extremes. The results of this approach for the Lorenz model would be very poor.

The particular sensitivity $\frac{\partial X(t+1)}{\partial X(t)}$ is the only one that is constant, i.e. does not depend on the state of the system: in equation (38), $\frac{\partial X(t+1)}{\partial X(t)} = 1 - a \, \Delta t$ (the values in Figure 13 are not perfectly equal to zero due to numerical imprecision). The linear regression, for this particular sensitivity, is then a good estimation technique. So the results are good in this particular case, but for the eight other sensitivities, the results of the linear regression are insufficient. In a real world case, we would not know which results are correct, if any.

The neural network-based estimates of the sensitivities (Figure 14) are a considerable improvement in comparison to the linear regression-based ones (except for the constant sensitivity $\frac{\partial X(t+1)}{\partial X(t)}$ at extreme values of $X(t)$, but the differences are still negligible). Note that the magnitudes of the sensitivities are very different, yet our technique seems to be able to retrieve these different orders of sensitivity in the system. Furthermore, these results are good if we compare the rms errors with the natural variability described in Figure 13. These results are summarized in Table 1. Since linear regression-based sensitivities are constant by assumption, the rms errors of this representation are essential equal to the standard deviations of the sensitivities. The improvement of the neural network-based sensitivities is considerable with respect to the linear regression: standard-deviations errors is always (except for the constant sensitivity) smaller than the natural standard-deviation of the theoretical sensitivities by one and somtimes two orders of magnitude. Given the large range of sensitivity magnitude, it is notable that the RMS errors of the neural network are

30

uniformly distributed over the nine sensitivities, even if the variability of the sensitivities is quite different. Table 1 summarizes the improvement gained by use of the neural network Jacobians to estimate the instantaneous, multivariate and nonlinear sensitivities of the discrete Lorenz dynamical system.

Figure 16 shows an example of the evolution in time of the theoretical and neural network estimates of sensitivities. The differences between the theoretical sensitivities and the neural network-based estimates are undetectable in this figure. This figure also highlights the more complex role of the feedbacks processes: when the state of the system reaches some extreme value, the sensitivities change, even in their sign, taking to the system back towards a middle range of values and finally to stabilize the system on its attractor.

For example, using the theoretical sensitivities in equation (38), we can analyze the relation between the variables X and Y. If Y is large and positive, then the sensitivity $\frac{\partial X(t+1)}{\partial Y(t)} = -2\,\Delta t\,Y(t)$ becomes large and negative. So, if Y continues to increase, the variable X will decrease even more rapidly. But the auto-sensitivity $\frac{\partial Y(t+1)}{\partial Y(t)}$ (the most important sensitivity for the variable Y) is equal to $1 - \Delta t + \Delta t\,X(t)$, which will be lower than 1 (damping effect) when X is lower than 1. One consequence of this behavior is that particular sensitivities, even when they are small on average, can still have a strong impact on the behavior of the system. A linear regression analysis assuming that the sensitivities are constant in time, may provide some estimate of mean sensitivities from a dataset. For example, the sensitivity $\frac{\partial X(t+1)}{\partial Z(t)}$ is, on average, nearly zero. A linear analysis, in this case, might suggest neglecting this relationship in understanding the system. Figure 15 shows how wrong this approximation would be: this figure represents the discrete Lorenz model defined in equation (36) with and without this particular sensitivity. The two trajectories have a quite distinct behavior: the simulation without the sensitivity oscillates more strongly and with a different time scale. The behavior of the complete system is produced by oscillations of the particular sensitivity, depending on the state of the system, between a positive and a negative value, theby stabilizing the system dynamics.

These results are special features of the general tendency of the sensitivities to exhibit

similar shapes in their time records (Figure 16), which means that they are closely linked with each other. This type of nonlinear behavior prevents a linear, even multi-variate, regression analysis from extracting even approximate information about the system dynamics. Comprehension of the system seems to require a more accurate representation of the time evolution of the multi-variate sensitivities.

d. Feedback analysis

We have seen that the classical approach for the feedback analysis, which makes strong (and incorrect) hypotheses about the dynamical system, is not well adapted to the Lorenz model.

However, the feedback factors can still be computed for the theoretical function, the linear regression model and the neural network model according to equations (48)/(50). We suppose here that these expressions are applicable to show that these feedback factors evolve in time (Figure 17), in violation of one of the assumptions used to describe the expressions. The feedback factors (48)/(50) are not simple and do not improve our understanding of the system since their physical interpretation is confused since these feedback factors are products of the sensitivities. The sensitivities, themselves, seem to be the more fundamental quantities. Furthermore, as we showed in Section 2, without all the assumptions at the base of this formalism (linearity, constant sensitivities, hierarchical cause and effect relationships, constant forcing, equilibrium state, etc), the whole formulation in terms of feedback factors falls apart.

6. Concluding remarks

What we have learned with this study of the Lorenz model is that the feedback processes are dependent on some important particular properties of the dynamical system under study. First, the feedback processes appear in a dynamical system when multivariate sensitivities are integrated over time. Second, if the system is nonlinear (i.e. the dynamical operator in equation (1) is nonlinear), the sensitivities are not constant with time, which

32

means that the feedback processes evolve in time. Third, each feedback has a strong impact on the character and behavior of the dynamical system, even those that may have a small time-averaged magnitude can have a stabilizing effect that changes drastically the characteristics of the system. Without such feedbacks the dynamical system would have a tendency to destabilize when an external forcing is introduced. The feedback processes have a stabilization effect, so the system does not diverge too much from its initial equilibrium. But this new equilibrium state could be different, with for example a higher frequency of extreme events. This is a theory that has been discussed recently by [*Palmer*, 1999]. Such an effect might explain the increase of the frequency of ENSO events with an increase of CO_2.

We have shown that the classical technique to analyze climatological feedback processes, from the electrical circuit theory, is, by hypothesis, very limited in its validity when applied to highly nonlinear multi-variate systems. Its applicability to the climate problem is even more questionable. Furthermore, the results of this kind of classical analysis are no more than a "schematic" measure of feedback processes at equilibrium of the system, which may be very misleading.

In comparison, the multivariate, instantaneous and nonlinear sensitivity concept, is more generally applicable without these constraints, and appears to be a good way of understanding the behavior of a system with coupled feedback processes. This general technique allows the quantification of these processes both spatially and temporally. This dynamical information seems to be more useful than classical feedback factor (only one number per variable).

Our technique for statistically inferring the complex network of sensitivities is particularly efficient and its generality and simplicity allow for the use of important a priori information in real-world studies. The dataset used in our analysis technique needs to satisfy some statistical requirements. First, the space and time sampling needs to be adequate to the description of the space and time variability of the sensitivities that originate the process feedbacks, so that the assumption that the sensitivities are constant over one

time step is an account approximation. Using too coarse time sampling is equivalent to using time-averaged data, which mixes many physical processes and ruins the sensitivity estimates. Using space-averaged data is also dangerous; for example, a mean sensitivity equal to zero could be generated by two opposite regimes with non-zero sensitivity. In other words, even if we are studying the longer-term behavior of the system, we must resolve the dynamics appropriately or the nonlinear integration will be incorrect. A study on the space/time variability of the sensitivities is then a prerequisite for the definition of a dataset sampling for feedback analysis. Second, the dataset has to have a good space and time coverage in order to represent as many climatological situations as possible. In other words, the dataset should contain all possible combinations of the state variables. The more situations in the dataset, the better will be the "laws" inferred by the analysis. These two points are a major argument to use actual, very large, long-term, datasets instead of generating new ones, limited in time. Moreover, these comments mean that the dynamics of the system cannot be correctly deduced from datasets where individual quantities have been separatelly averaged on space and time.

Our technique has the advantage of being applicable to numerical model data as well as observations, which means that the important work of inter-comparison of models and of validation of models could be carried out with a meaningful measure: the sensitivities of the variables of the system. This diagnostic measure is particularly interesting because it concerns very intuitive and physical quantities. Comparisons of the sensitivity relationships could also be made with field experiment data to understand how physical processes product these sensitivities. Thus, our analysis approach provides a framework for a whole new attack on these problems.

The statistical model estimating the sensitivities can also be used to study the response of the system new equilibrium state, including the time to reach equilibrium after a small perturbation. This simplified model could also be used to analyze the propagation of uncertainties when predictions are performed. In other words, the neural network statistical model provides a better approximation of "small perturbation" behavior than attempts

to linearize the system by dropping relationships. The next step of these ideas is to use this new technique for more complicated climate systems involving real observations or numerical model outputs.

References

Aires, F., Schmitt, M., Scott, N.A., and Chédin, A., 1999: The Weight Smoothing regularization for Jacobian stabilization. *IEEE Transactions on Neural Networks*, **10**, 6, 1502–1510.

Aires, F., Problème inverses et réseaux de neurones: application à l'interféromètre haute résolution IASI et à l'analyse de séries temporelles, PhD thesis, 220 pp., Université Paris IX - Dauphine, March 1999.

Aires, F., Chédin, A., and Nadal, J.-P., 2000: Independent component analysis of multivariate time series. Application to the tropical SST variability. *Journal of Geophysical Research*, in press.

Andronova N.G. and Schlesinger, M.E., 1991: The application of Cause-and-Effect Analysis to Mathematical Models of Geophysical Phenomena - 1. Formulation and Sensitivity Analysis. *Journal of Geophysical Research*, **96**, D2, 941–946.

Andronova N.G. and Schlesinger, M.E., 1992. The application of Cause-and-Effect Analysis to Mathematical Models of Geophysical Phenomena - 2. Stability Analysis. *Journal of Geophysical Research*, **97**, D5, 5911–5919.

Bode, H. W., *Network Analysis and Feedback Amplifier Design,* 551 pp., Van Nostrand, New York, 1945.

Curry, J.A., and P.J. Webster, Thermodynamics of Atmospheres and Oceans, 467 pp, Academic Press, London, 1998.

Cybenco, G., 1989: Approximation by Superpositions of a Sigmoidal Function, *Math. Control Signals Systems*, **2**, 303–314.

Hansen, J., Lacis, A., Rind, D., Russel, G., Stone, P., Fung, I., Ruedy, R. and Lerner, J., 1984: Climate Sensitivity: Analysis for Feedback Mechanisms, *Climate Processes and Climate Sensitivity*, (Maurice Ewing Series, No 5), American Geophysical Union, Washington D.C., 120–163.

Hansen J. et al., 1997: Forcing and chaos in interannual to decadal climate change. *Journal of Geophysical Research*, **102**, 25679–25720.

Hornik, K., Stinchcombe, M. and White, H., 1989: Multilayer feedforward networks are universal approximators, *Neural Networks*, **2**, 359–366.

Lorenz, E.N., 1984: Irregularity: a fundamental property of the atmosphere. *Tellus*, **36A**, 98–110.

Lorenz, E.N., 1990: Can chaos and intransitivity lead to interannual variability ? *Tellus*, **42A**, 378–389.

Palmer, T.N., 1999: A Nonlinear Dynamical Perspective on Climate Prediction, *J. of Climate*, **12**, 575–591.

Peixoto, J., and A.H. Oort, 1991: Physics of Climate, New York: American Institue of Physics.

Rossow, W.B. and Schiffer R.A., 1999: Advances in understanding clouds from ISCCP, Bulletin of the American Meteorological Society, **80**, 11, 2261–2287.

Rumelhart, D., Hinton, G. and Williams, R., 1986: Learning representations by back-propagating error, volume I, chapter 8, 318–362.

Schlesinger, M.E., 1985: Analysis of results from Energy Balance and Radiative-Convective models, *in projection the climatic effects of increasing carbon dioxide*, december 1985, report DOE/ER-0237, United States Department of Energy, pp. 280–318.

Slingo, A., Pamment, J.A., Allan, R.P., and Wilson, P.S., 2000: Water vapour feedbacks in the ECMWF re-analyses and Hadley Centre climate model. *Journal of Climate*, submitted.

Smith, 1997: Uncertainty dynamics and predictability in chaotic systems. *Q. J. R. Meteorol. Soc*, **123**, 1–34.

Yuval, 1999: Neural network training for prediction of climatological time series; regularized by minimization of the Generalized Cross Validation Function. *Submitted to Monthly Weather Review.*

F. Aires, Department of Applied Physics, Columbia University, NASA Goddard Institute for Space Studies, 2880 Broadway, New-York, NY 10025, USA. (e-mail: filipe@sharki.giss.nasa.gov)

W.B. Rossow, NASA Goddard Institute for Space Studies, 2880 Broadway, New-York, NY 10025, USA. (e-mail: wrossow@giss.nasa.gov)
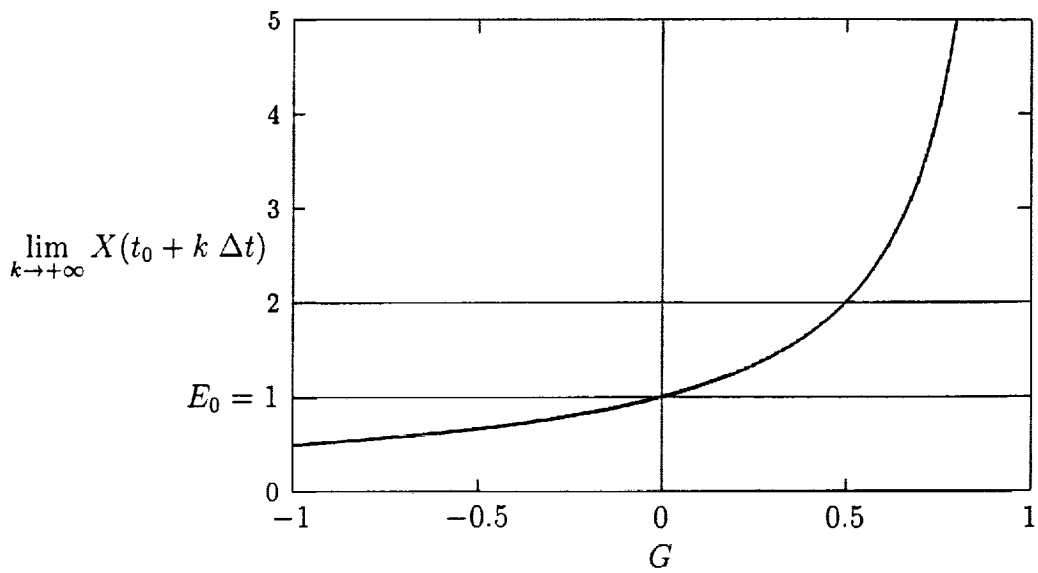
Figure 1. The stabilized values $\lim_{k \to +\infty} X(t_0 + k\,\Delta t)$ of a mono-variable linear system for different values G of the gain of the system and with external forcing $E_0 = 1$
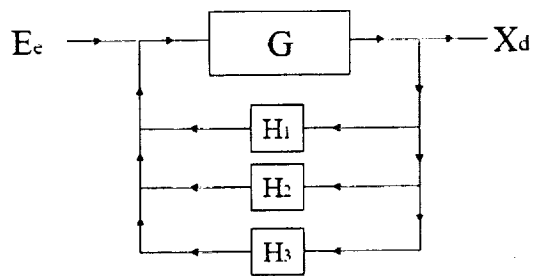
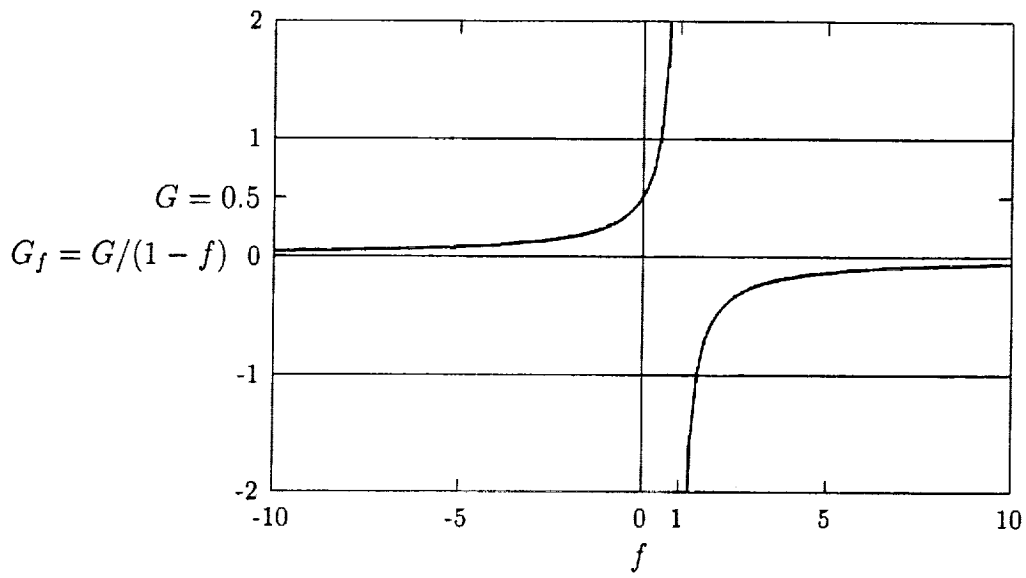**Figure 2.** Feedback loops system in parallel

Figure 3. Analysis of the gain G_f of the system as a function of the unique feedback factor f, with $G = 0.5$
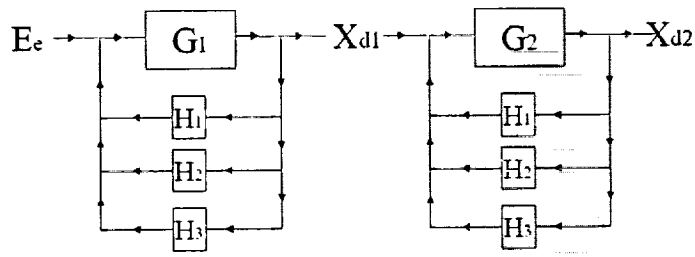
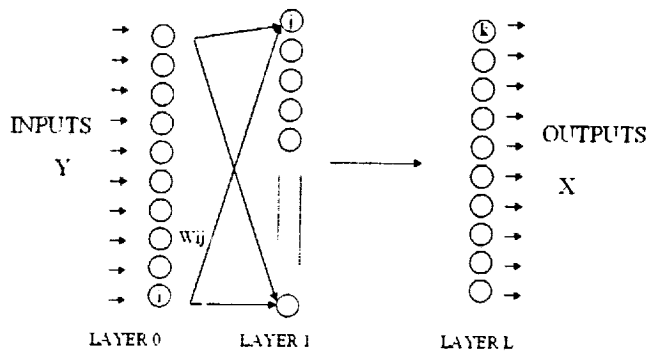**Figure 4.** Feedback loops system in series

Figure 5. Architecture of a MLP neural network with L layers, with inputs P and outputs X
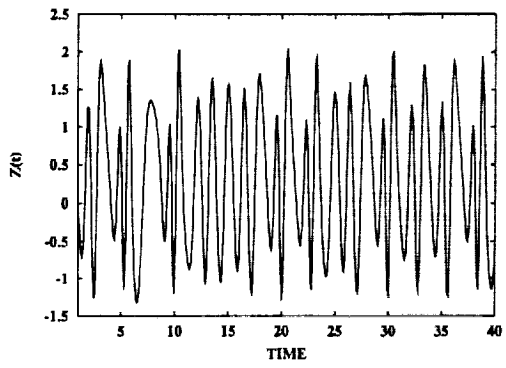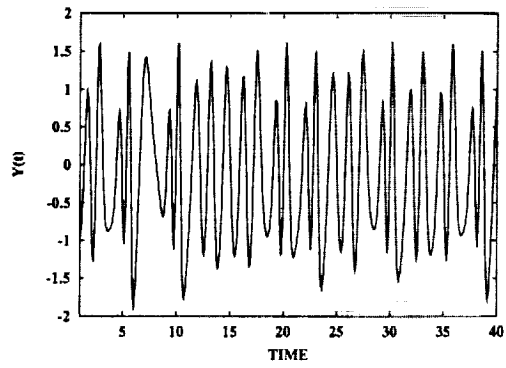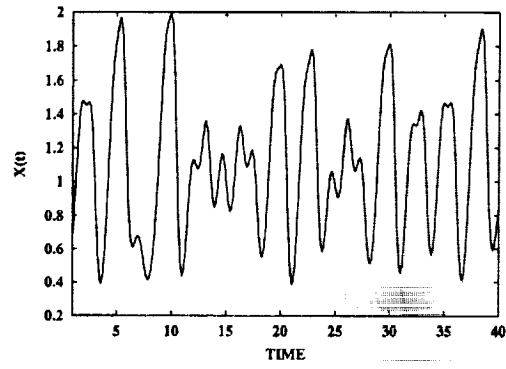
Figure 6. Lorenz model, with parameters $a = 0.25$, $b = 4$, $F_1 = 8$, $F_2 = 1$ and $\Delta t = 0.08$, simulated by fourth-order Runge-Kutta
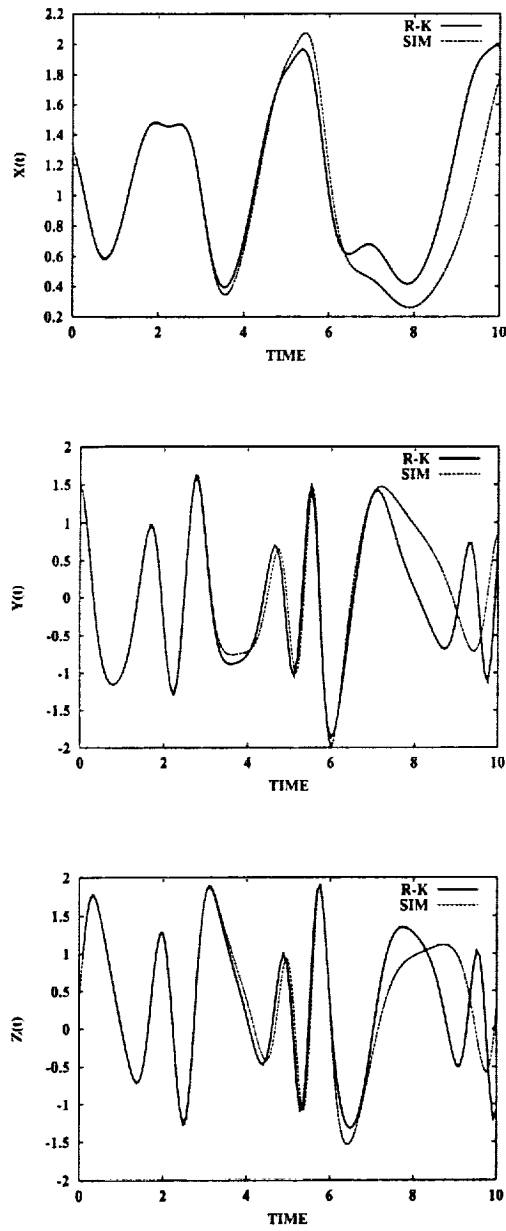
Figure 7. Fourth order Runge-Kutta (continuous lines) and discretized Lorenz model (dashed lines) with parameters $a = 0.25$, $b = 4$, $F_1 = 8$, $F_2 = 1$ and $\Delta t = 0.08$

Figure 8. Noisy trajectories of the dataset from the discrete Lorenz dynamical system

Figure 9. Representation of the theoretical Lorenz Dynamical operator (continuous line), its neural network estimate (dotted lines), and its linear regression estimate (dashed lines)

Figure 10. RMS error for the estimation of the dynamical Lorenz operator: neural network regression (continuous lines), and linear regression (dashed lines)

Figure 11. Prediction RMS error for the Neural Network regression (continuous line) and the Linear regression (dashed line) as a function of the forecast range (in time steps $\Delta t = 0.08$)

Figure 12. An example of prediction with a forecast range of 24 time steps

Figure 13. Standard-Deviation of the sensitivities of the discrete Lorenz dynamical system

Figure 14. Root Mean Square error for the sensitivities estimates: neural network-based estimates (continuous lines) and linear regression-based estimates (dashed lines)

Figure 15. Discrete Lorenz model (continuous line) and discrete Lorenz model minus the sensitivity $\frac{\partial X(t+1)}{\partial Z(t)}$ (dashed line)

Figure 16. Jacobians evolution through time: theoretical Jacobians (continuous line), linear regression based estimates (dotted lines), and neural network based estimates (dashed lines)

Figure 17. Feedback factors evolution through time

Table 1. Statistics on true and retrieved sensitivities

Sensitivity	Statistics	Theoretical	Linear	Neural Network
$\frac{\partial X(t+1)}{\partial X(t)}$	Mean	0.980	0.973	0.981
	Std-Dev	0.000	0.000	0.003
	RMS Error		0.007	0.003
$\frac{\partial X(t+1)}{\partial Y(t)}$	Mean	-0.077	-0.025	-0.076
	Std-Dev	0.133	0.000	0.132
	RMS Error		0.144	0.004
$\frac{\partial X(t+1)}{\partial Z(t)}$	Mean	-0.057	-0.064	-0.057
	Std-Dev	0.146	0.000	0.145
	RMS Error		0.147	0.004
$\frac{\partial Y(t+1)}{\partial X(t)}$	Mean	-0.077	0.014	-0.077
	Std-Dev	0.297	0.000	0.297
	RMS Error		0.310	0.003
$\frac{\partial Y(t+1)}{\partial Y(t)}$	Mean	0.955	0.979	0.956
	Std-Dev	0.048	0.000	0.043
	RMS Error		0.054	0.003
$\frac{\partial Y(t+1)}{\partial Z(t)}$	Mean	-0.141	-0.133	-0.141
	Std-Dev	0.192	0.000	0.193
	RMS Error		0.193	0.003
$\frac{\partial Z(t+1)}{\partial X(t)}$	Mean	0.184	0.259	0.184
	Std-Dev	0.281	0.000	0.281
	RMS Error		0.291	0.004
$\frac{\partial Z(t+1)}{\partial Y(t)}$	Mean	0.141	0.226	0.142
	Std-Dev	0.192	0.000	0.192
	RMS Error		0.210	0.003
$\frac{\partial Z(t+1)}{\partial Z(t)}$	Mean	0.955	0.962	0.955
	Std-Dev	0.048	0.000	0.048
	RMS Error		0.049	0.003

Figure 1. The stabilized values $\lim_{k \to +\infty} X(t_0 + k\ \Delta t)$ of a mono-variable linear system for different values G of the gain of the system and with external forcing $E_0 = 1$

Figure 2. Feedback loops system in parallel

Figure 3. Analysis of the gain G_f of the system as a function of the unique feedback factor f, with $G = 0.5$

Figure 4. Feedback loops system in series

Figure 5. Architecture of a MLP neural network with L layers, with inputs P and outputs X

Figure 6. Lorenz model, with parameters $a = 0.25$, $b = 4$, $F_1 = 8$, $F_2 = 1$ and $\Delta t = 0.08$, simulated by fourth-order Runge-Kutta

Figure 7. Fourth order Runge-Kutta (continuous lines) and discretized Lorenz model (dashed lines) with parameters $a = 0.25$, $b = 4$, $F_1 = 8$, $F_2 = 1$ and $\Delta t = 0.08$

Figure 8. Noisy trajectories of the dataset from the discrete Lorenz dynamical system

Figure 9. Representation of the theoretical Lorenz Dynamical operator (continuous line), its neural network estimate (dotted lines), and its linear regression estimate (dashed lines)

Figure 10. RMS error for the estimation of the dynamical Lorenz operator: neural network regression (continuous lines), and linear regression (dashed lines)

Figure 11. Prediction RMS error for the Neural Network regression (continuous line) and the Linear regression (dashed line) as a function of the forecast range (in time steps $\Delta t = 0.08$)

Figure 12. An example of prediction with a forecast range of 24 time steps

Figure 13. Standard-Deviation of the sensitivities of the discrete Lorenz dynamical system

Figure 14. Root Mean Square error for the sensitivities estimates: neural network-based estimates (continuous lines) and linear regression-based estimates (dashed lines)

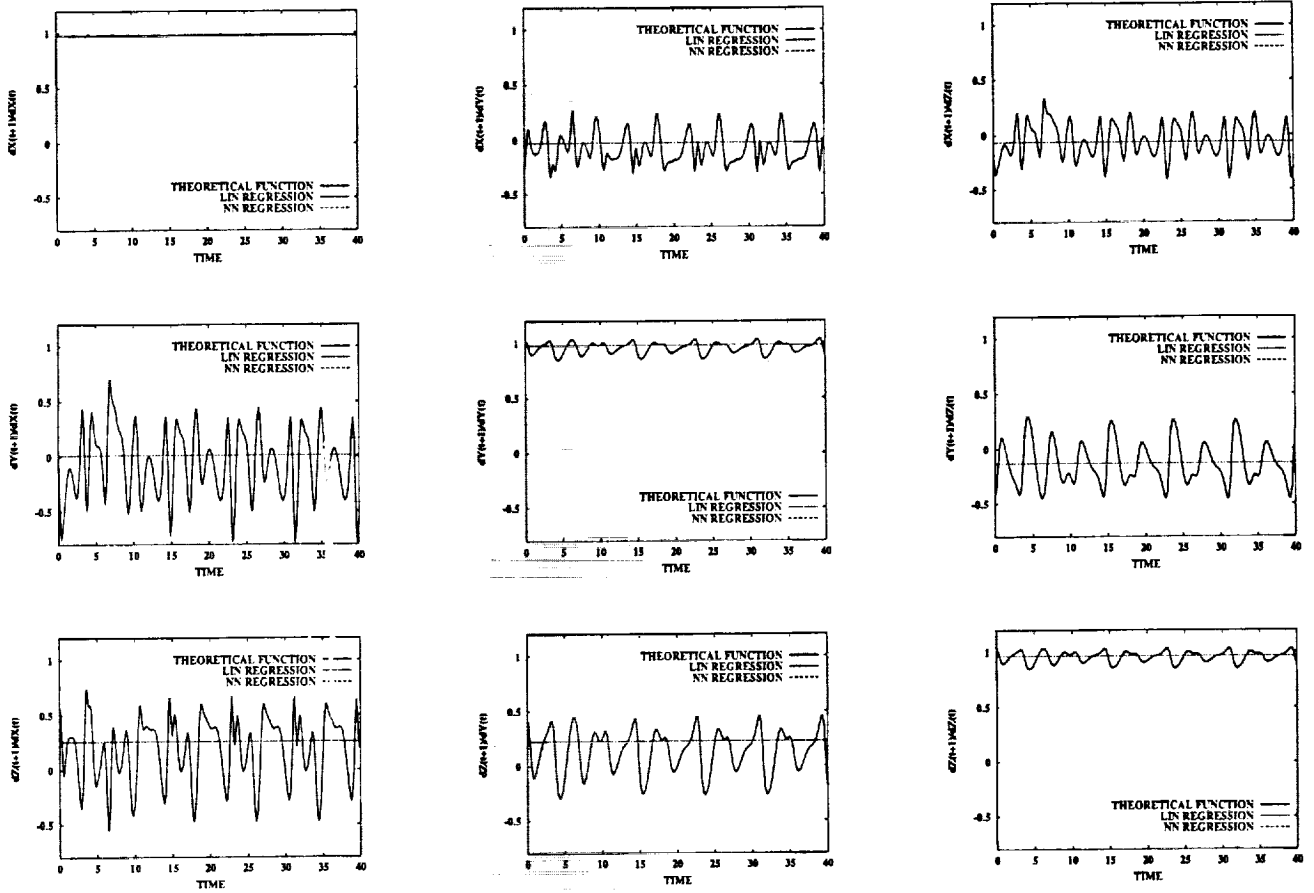**Figure 15.** Discrete Lorenz model (continuous line) and discrete Lorenz model minus the sensitivity $\frac{\partial X(t+1)}{\partial Z(t)}$ (dashed line)

Figure 16. Jacobians evolution through time: theoretical Jacobians (continuous line), linear regression based estimates (dotted lines), and neural network based estimates (dashed lines)
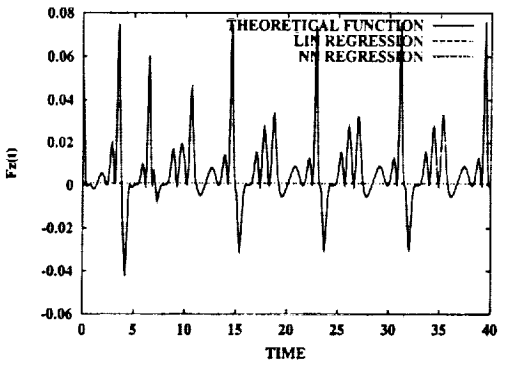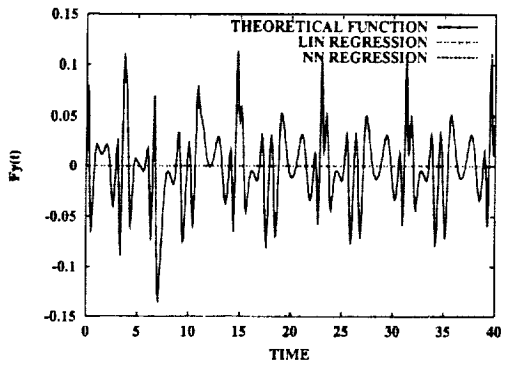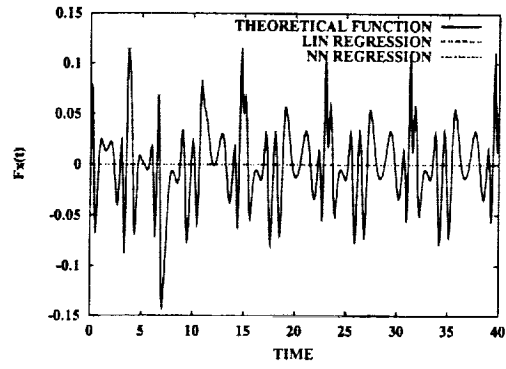
Figure 17. Feedback factors evolution through time