Climate, Control theory, Feedback. Does it make sense?

Richard Saumarez

Designing a machine using control and systems methods includes construction of a mathematical model of the device, analysis of the model and then building the device itself. With depressing regularity, the machine doesn't work properly. After eliminating simple mistakes, the mathematical model may be found to be wrong or the techniques to analyse it may be inappropriate. These errors are interesting because they may be quite subtle misapplications of theory and to address them, one has to go back to basics and re-examine things which one normally takes for granted. Using simple conceptual models and mathematics, one can analyse the nature of feedback as applied to the dynamics of climate. I will take the basic equation used by Spencer and Braswell [1] and Dessler [2] as a starting point and try to see whether it implies feedback and discuss what constraints exist on determining if a system has, or does not have, feedback.

A control system is one that preserves a controlled variable according to an input "demand" in the presence of perturbations and noise. This clearly distinguishes between the static gain of a system defined by feedback in a *steady state* and the time varying, or frequency dependent, gains in a *dynamic system.* However, control systems are designed to maintain an output signal in response to a demand, which probably doesn't exist in the atmosphere. Control theory nevertheless gives us mathematical methods to analyse systems with feedback.

The equation [1,2] relating radiative fluxes, ocean heat capacity *Cp*, feedback λ to temperature change, ΔT :

$$d\Delta T / dt = \frac{1}{Cp}\Sigma Fluxes(t) + \frac{\lambda}{Cp}\Delta T(t)\,,$$

can be written more generally as a first order differential equation:

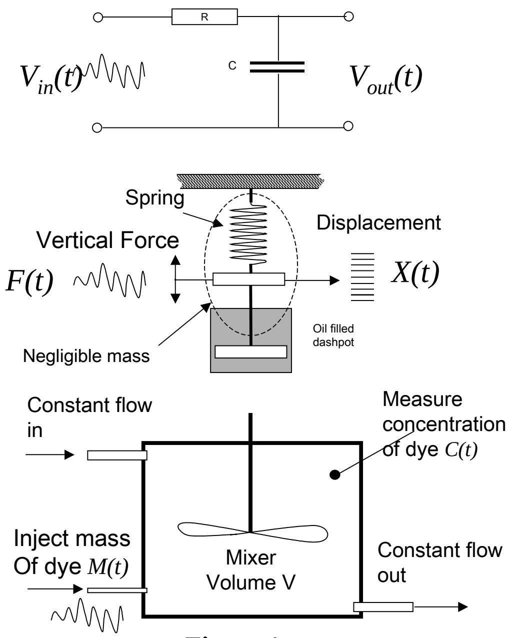

Figure 1

$$x' = kx + F(t)$$

What does this equation imply as a physical system? Could one actually make this system? All it means is that the system contains a storage element and a dissipative element [3].

Three examples of such a system, provided k is negative, a capacitor*/ resistor network, a spring*/dashpot[1] and a Volume*/flow mixer are shown in figure1.

One can draw three conclusions from this:

* Storage

1 A dashpot produces a force proportional to minus its velocity

1

> Once you have written a differential equation describing a system, you have made a *precise* statement about its physical behaviour.

Arguments made about systems with different physical structures but the same mathematical form use *analogous variables*. If one is going to use a generalised form of maths to analyse a system, one has to be extremely careful about ensuring that a variable that system really corresponds to that in another. For example, calculating V_{in} and V_{out} in the resistive/capacitor network, the basic equation should be written in current, not voltage.

2

> The system *does not* contain feedback.

The term k (or λ in [1]&[2]) is dissipation.

3

Finally, could *k* be positive? At first sight, this is unlikely as it might "blow up", but is it conceivable that other negative feedbacks could damp down this process or the system behaves in such a way that it might appear to be positive?

Any linear system[2] is defined by its output when the input is an impulse, $\delta(t)$, an abstraction of an infinitely short pulse, with an infinite amplitude and an area of 1. In this case, we have

$$x' = kx + \delta(t); \quad x(t) = e^{kt} \qquad (1)$$

Provided k<0, this negative exponential response shows that the system returns to zero, and it is stable in the sense that the integral of the modulus of signal is finite.

The relationship between the input, *x(t)*, and output, *y(t)*, signals are easily calculable from the system impulse response, which is usually described as *h(t)*.

Loosely, one can imagine a signal as composed of a set of impulses at Δt apart, weighted by the amplitude of the signal at each instant. Each has an impulse response and the output of the

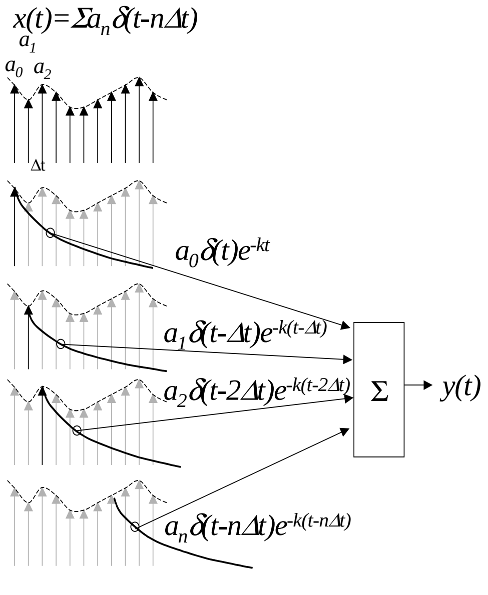

Figure 2

[2] A linear system is one described by linear differential equations. A more "signal processing" definition, that follows from this, is that it obeys proportionality – i.e.: the output is proportional to the input, it doesn't vary with time (stationarity) and it obeys superposition, .i.e.: if input A produces an output and B produces an output, the output of input A+B is the sum of individual outputs of A and B.

system is the sum of the impulse responses as shown in figure 2. In the limit that Δt tends to zero, the output becomes the convolution integral of the input and the system impulse response [3]:

$$y(t) = \int_{-\infty}^{+\infty} x(\tau)h(t-\tau)d\tau$$

The process of computing the output of linear system from its input is known as convolution. The process of working out the shape of the impulse response from the input and output is called deconvolution. We are attempting to define the structure of the system by observing radiation fluxes and temperature, so this is a problem of deconvolution. Alternatively, we can make a model of the process and use deconvolution/convolution to estimate the parameters of the model. This is more helpful because we can make a model that contains feedback and the use signal processing techniques to estimate the magnitude of the feedback.

Linear systems and linearisation.

Many real systems are non-linear and are difficult to analyse. One common strategy is to "linearise" the system, which is an assumption that for small changes, the system is approximately linear. This stems from Taylor's theorem:

$$f(x_0 + a) = f(x_0) + a\frac{df}{dx}_{|x=x_0} + \frac{1}{2!}a^2\frac{d^2 f}{dx^2}_{|x=x_0} + \frac{1}{3!}a^3\frac{d^3 f}{dx^3}_{|x=x_0} \,.........$$

showing that if we know *f(x)* and there is a **small** change in *x* from x_0 of *a*, the change in *f(x)* is approximately proportional to *a* because $a^3 <<< a^2 << a$. For, example, this allows treatment of the change in radiation according to the Stefan-Boltzmann equation as linear over small temperature excursions. Whether this is appropriate depends on one's choice of the range of *a* and that, for these excursions, the higher order terms really are negligible.

A very simple model of Cloud feedback

This assumption is used in the derivation of thermodynamic feedback in climate, and, using a linearised model, the feedback(s), *f*, can be written as a sum of functionals that relate flux at the top of the atmosphere F^{rad} to temperature T, by an intermediate internal variable *I*.[4]

$$f = \sum_j \frac{\partial F^{rad}}{\partial I_j} . \frac{dI_j}{dT}$$

Let us imagine a simple model of feedback due to clouds. Flux comes in, the temperature rise and so clouds form, which has an effect on temperature by changing flux, shown in figure 3.

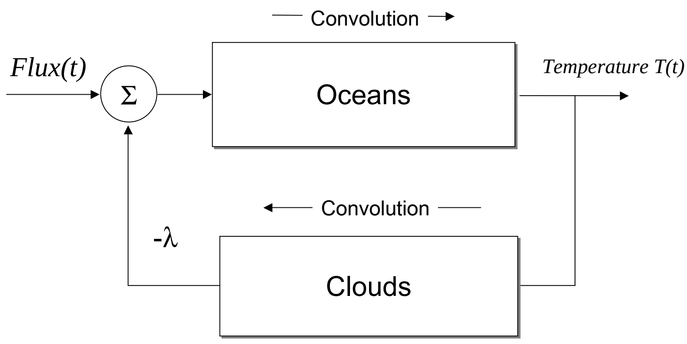

Figure 3

While this is definitely not a realistic model of climate, it can be used as an illustration of the process of how to construct a model from physical elements, which leads to conclusions about how one would try to confirm if it existed from observational data.

The oceans are simply an integrator

$$T(t) = T_0 + \frac{1}{Cp}\int_0^t F^{Rad}(t)dt \quad (2)$$

As an internal variable, one could imagine cloud (C_l) formation being proportional to temperature so that $\frac{dI}{dT}$ can be represented as $\frac{dC_l}{dt} = k(T - T_0)$ and $\frac{\partial F^{rad}}{\partial C_l} = -\lambda$.

There are several important assumptions in these equations.

1) The first is that heat is perfectly distributed through a layer of ocean without delays, which probably isn't true.
2) The second assumption is that the feedback does not directly affect the output by "sucking" heat out of the ocean when clouds are formed. Technological systems are designed so that the feedback does not load the output, and in this case the latent heat of evaporation is assumed to be negligible compared to the ocean heat sink.

The feedback term in this model is convolution of the output, itself a convolution, and the impulse response of the feedback. Viewed in this way, analysis becomes increasingly clumsy.

There are more powerful (and computationally efficient) methods of solving these types of model than time domain convolution that are widely used in system and control theory. The idea is that a system and a signal can be *transformed* into another "domain", i.e.: expressed in terms of another variable, *s*. The consequence of this transformation, the *Laplace Transform,* is that operations in the time domain, which we perceive as involving a "calculus" operation can be treated as algebraic expressions in the *s* domain. Once an algebraic solution is obtained in the *s* domain, the time domain solution is calculated using the *inverse Laplace Transform*. This appears, at first sight, to be an incredible slight of hand, conceivably a "trick", but actually allows to one to see things about the system very intuitively.

$\mathcal{L}[f(t)]$ →

Time domain		S domain
$x(t)$	↔	$x(s)$
$\int x(t)dt$	↔	$\frac{1}{s}\, x(s)$
$\frac{dx(t)}{dt}$	↔	$sx(s)$
$\int x(\tau)h(t-\tau)d\tau$	↔	$x(s).h(s)$
$e^{-kt}.u(t)$	↔	$\frac{1}{k+s}$

← $\mathcal{L}^{-1}[f(s)]$

The transfer function, which is the Laplace transform of the impulse response, of a feedback system with *A(s)* in the forward path and *B(s)* in the feedback:

$$\frac{y(s)}{x(s)} = H(s) = \frac{A(s)}{1 + A(s)B(s)} \quad (3)$$

While this is a dynamic model, if *s=0*, it becomes the static gain equation or the "climate sensitivity".

The forward component is an integrator (equation 2) and is therefore $\frac{1}{Cp.s}$ and the feedback network is a linear first order system with a transform: $\frac{\lambda}{k+s}$. Hence the system response is:

$$H(s) = \frac{\frac{1}{Cp.s}}{1 + \frac{1}{Cp.s}\frac{\lambda}{k+s}} = \frac{1}{Cp.s + \frac{\lambda}{k+s}} = \frac{1}{S^2 + ks + \frac{\lambda}{Cp}} \quad (4)$$

This result is very significant because the response is a ***quadratic** equation in s* which when rewritten so that output is related to the input becomes:

$$s^2 y(s) + ksy(s) + \frac{\lambda}{Cp} y(s) = x(s)$$

Since integration in the Laplace transform is division by *s* and differentiation is multiplication by *s*, this means that this equation, in *s*, corresponds to a *second order* differential equation in time:

$$\frac{d^2y}{dt^2} + k\frac{dy}{dt} + \frac{\lambda}{Cp} y = x(t)$$

Therefore, the equation used to define "feedback" used by Spencer and Braswell and by Dessler, which is first order, *cannot* contain a term explicitly relating to feedback

Positive Feedback

Suppose there is an element that increases temperature exponentially with an impulse of flux, (i.e.: *k* in equation 1 is positive,) of the form T(t)=$be^{\alpha t}$ will this is lead to run away temperature? Strictly this is an unbounded element, but it could be created by positive feedback.

The Laplace transform is:

$A(s) = \frac{b}{s - \alpha}$. Note the transform of be^{-at} is $A(s) = \frac{b}{s + \alpha}$.

What matters is the sign of α, if it negative, the signal is unbounded, i.e.: has an infinite integral with respect to time and, if positive, it is bounded. If there is constant negative feedback, G, the transfer function of the system, *H(s)* is simply:

$$H(s) = \frac{A(s)}{1+GA(s)} = \frac{b}{s-\alpha+Gb}$$

Provided $G>\alpha/b$, the constant component of the denominator is positive and the system is stabilised, so converting its impulse response from a positive exponential to a negative exponential. Therefore in debate over whether clouds create negative or positive feedback, and, if positive is the system unstable, it is possible that they could do both with different time courses.

Provided the negative feedback is sufficiently large, it will constrain the positive feedback.

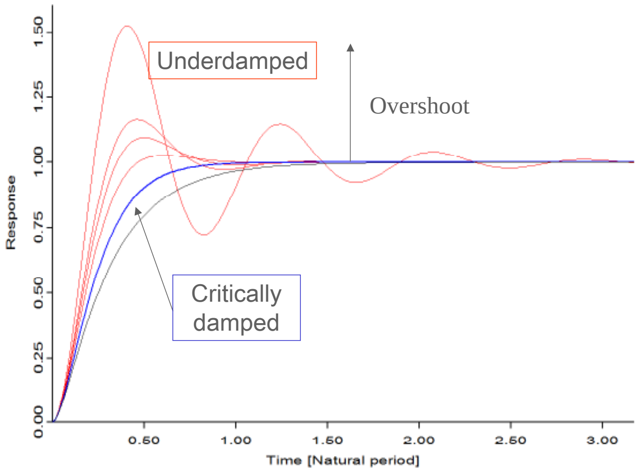

Figure 4

Equation 4 illustrates another feature of control systems that may be important in the “feedback” debate. The factor *k* in equation 4, determines the damping of the system. If it were zero, the system would oscillate. Were it small, the system in response to a step change in flux, would overshoot and perform an series of exponentially damped oscillations while settling to a new level (figure 4). As *k* becomes greater, the response of the system becomes non oscillatory and becomes two exponentials.

While major overshoot is unlikely, the possibility of a minor degree of overshoot should be borne in mind, since if incorrectly analysed, it could be interpreted as “positive feedback”.

What is the “delay” in the system?

One common misunderstanding is that a feedback system, such as that shown in figure 3, can be written with constants in the boxes. This is the solution of equation 4 when the frequency equals zero and is the static gain (or climate sensitivity). In reality, such a system cannot exist because all physical systems impose delay. What determines the dynamic behaviour of a feedback system is not so much the static gain but the frequency dependent delays imposed by the forwards and backwards components in the loop.

The Laplace transform is a generalised case of the Fourier Transform. Its variable, s, contains two variables, a damping coefficient, σ, and complex frequency $j\omega$. In general, the Fourier transform of the transfer function is calculable from the Laplace Transform[3]. Computationally, having derived equation 4, one can calculate an output from a known input and transfer function or, in principle, compute the transfer function from the input and output using a Discrete Fourier Transform (DFT)[3].

A first order system, with an exponential decay impulse response e^{-kt}, has a Fourier Transform of:

$$f(\omega) = \frac{1}{k + j\omega}$$

The response, $f(\omega)$, is a complex number and has both magnitude and **phase**. This means that if a cosine wave is passed through this system, it will be attenuated by magnitude ($G(\omega)$) of the frequency response and it will be shifted in phase $\phi(\omega)$ by the argument of the response:

$$G(\omega) = \frac{1}{\sqrt{k^2 + \omega^2}}, \quad \phi(\omega) = \tan^{-1}(\frac{-\omega}{k}).$$

A phase shifted cosine wave is simply a *delayed* cosine wave since:

$$\cos(\omega t + \phi) = \cos(\omega t - T), \text{ where } T = \phi/\omega.$$

This example shows that the delay of signal through a system is dependent on its frequency content isn't a delay at all but is a phase shift. Therefore what does one mean by a delay in the system and what, if anything is one measuring when one tries to measure it? Any system that stores energy will impose a non-linear frequency dependent phase shift rather that a pure delay, which has a linear change in frequency[4].

Therefore the attempt to define a system by the delay that it imposes on a signal is meaningless because delay will depend on the inputs.

The phase response of the system in equation 4 is plotted together with the phase responses of two different pure time delays (in green) in figure 5 and they are clearly completely different.

What is the relationship between the system response and lagged regression?

The expression [5,6]:

$$R_{xy}(t) = \int_{-\infty}^{+\infty} x(\tau)y(t+\tau)d\tau \quad (5)$$

---

[3] The Laplace and Fourier Transforms are analytical concepts. The integration limits are between +- infinity in time and they are continuous in frequency. You cannot compute the Fourier Transform of a real signal! We observe signals over a limited time frame and the signals are usually sampled. In the computational modelling of transforms, one is not actually using true transforms but series representations of the transforms, which are repetitive in the time domain. The results are not continuous in frequency. This is often a source of misunderstanding and quite spectacular errors. Not all functions that have Laplace transforms have FTs. In our case, an integrator has an impulse of *u(t)*, which is not Fourier Transformable

[4] The Fourier transform of a signal f(t), that is delayed by T is simply

$$F(\omega) = \int f(t-T)e^{-j\omega t}dt = \int f(s)e^{-j\omega(s+T)}ds = e^{-j\omega T}F[f(t)]$$

i.e.: a phase shift proportional to frequency.

is very similar to the convolution integral. Writing this in a discrete form, assuming that the expected values of *x(t)* and *y(t)* are zero and normalising for the power in the signals, we get:

$$R'_{xy}(n) = \frac{\sum_K x(k).y(k+n)}{\sum_K x(k)^2 . \sum_K y(k)^2}$$

This is clearly a correlation coefficient, *r*, between a signal *x* and a signal *y* that has been shifted by n. Finding the value of n at which the cross-correlation is a maximum identifies the "delay" in the system for *one particular set of signals* only. How is this related to the system response *h(t)*? It is easy to show, by taking a Fourier Transform of equation 5, that:

$$R_{xy}(\omega)=Y(\omega)X^*(\omega)$$

Where $X^*(\omega)$ is the complex conjugate of $X(\omega)$. Since $Y(\omega)=H(\omega)X(\omega)$:

$$Y(\omega)X^*(\omega)=H(\omega)X(\omega)X^*(\omega) \text{ and so: } R_{xy}(\omega)=H(\omega)R_{xx}(\omega)$$

Therefore the cross-correlation between the two signals is the convolution between the system transfer function and the auto correlation of the input and the mathematical techniques used to extract system parameters from the signal should reflect this.

The way that lagged regression is used by both SB2011 and D2011 does not appear to reflect this relationship, since it does not represent time domain deconvolution and so system parameters are not easily recoverable. The technique seems to originate from static climate sensitivity calculations (Forster and Gregory [7]), but its use to calculate climate dynamics is therefore questionable. The model in equation 4 can be used to extract a "time delay" and a "gain" by regression using a band pass filtered random input sequence, and this immediately shows that the estimates vary with input bandwidth. In other words, the measured "parameters" of the system are not parameters.

How to measure the system Parameters

One has to be very clear in what one is trying to achieve with an analysis of highly variable data and should start with a formal statistical hypothesis.

Hypothesis: The means of the distributions of the system feedback parameters are zero and the system cannot be distinguished from a system without feedback.

Having raised this hypothesis, one can explore the difficulties in testing it. A very common problem is that one can have an elegant mathematical model of a process but the data in too restricted to distinguish one's model from a simpler one, in this case a model with feedback and the null hypothesis, one without feedback. This problem crops up in virtually every branch of science and generally results in an ill conditioned set of equations used to calculate the parameters of the model.

Another major problem is that if the model has an incorrectly modelled component, the real response of this component can "contaminate" the representation of another part of the system. In the model presented here, the ocean is represented as an integrator, simply because that is how SB2011 and D2011, D2010 choose to represent it. I practice, I bet it isn't and it really should be represented as a multi-compartment system with fluxes and diffusion between different compartments. However, using real data to perform a deconvolution, the real behaviour of the oceans will turn up spuriously in the "feedback", because it can't behave in a realistic way, with multiple time delays, in the model. This has been a common problem in systems identification in biology/medicine and would require careful analysis.

In climate because there is only one record of temperature, fluxes etc. and so one is dealing with essentially one observation, unlike classical systems analysis where one can "interrogate" a system using multiple input sequences and obtain a robust estimate of the system parameters. This may lead to formidable statistical problems in testing the hypothesis.

Looking at a Bode plot (figure 5) of the model in equation 4, clearly the best frequency range to resolve the feedback parameters is a region of 0.3-3 times the natural frequency of the system ω_0. In practice, this depends on the length of the record, which determines the fundamental frequency of the DFT of the record. In the time domain, this means that if the impulse response of the system is much longer than the record, one can't characterise it. A further problem is that the Flux signals are band limited and may not fully cover the range of the impulse response.

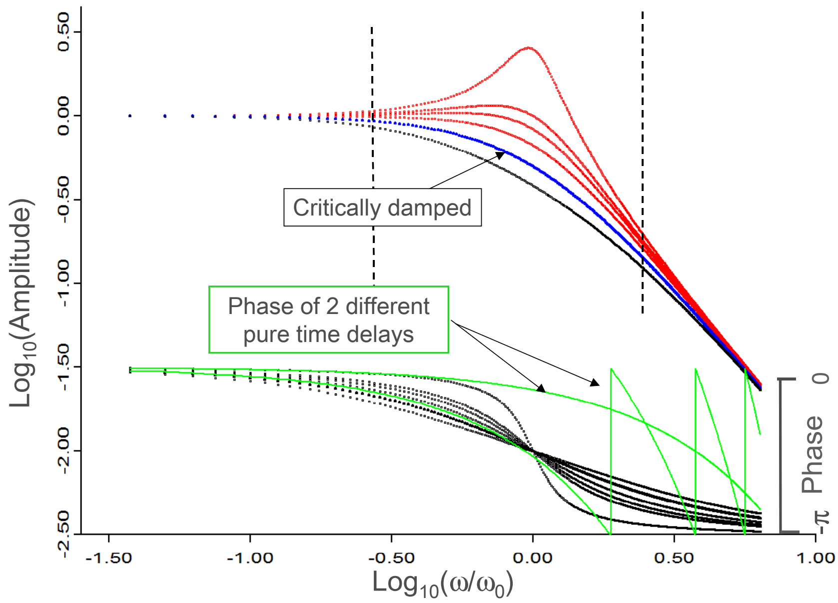

Figure 5

Alternatively, if the impulse response is relative short, say a month, one will not be able to detect it with a record in monthly samples because it will be aliased[5]. For some reason, filtering a signal with a rectangular impulse response and then

[5] Aliasing means that the sampling frequency is too low to capture the full frequency content of a signal. In this case, the signal is irretrievably corrupted and cannot be analysed. This is an error that has bitten many uncritical researchers in the hindquarters and given a sampled signal, the first question should be "is it aliased?".

decimating the signal is commonly used in climate “to reduce noise”. This is highly undesirable because it may force aliasing on previously correctly sampled signal.

If we are lucky and the impulse response is relatively short compared with the length of the record, the record can be segmented into lengths that contain the impulse response and the parameters estimated from each record, allowing a better statistical estimate of the system response to be obtained.

An alternative approach, which I would use, is a parameter fitting technique. Although there are many schemes, the simplest is to use the model to create an output as function of the parameters and determine the integral squared error between the predicted and actual outputs. Minimisation of this integral with respect to the parameters of two competing models may determine whether the “best fit” model is one with or without feedback and allow the hypothesis to be tested.

The thesis I have advanced is that if one is to model a process, great care should be taken to ensure that the model actual conforms to the physical hypothesis, which should be stated explicitly and the assumptions underlying the model should be scrutinised. The data analysis methods should reflect the mathematics of the model and lead directly to estimation of its parameters. The use of multiple stages of analysis can become blind alleys and so one should ask what each processing step means in physical terms. I am slightly sceptical that a systems approach of using deconvolution to imply feedback would give an unambiguous answer, although it may be possible given some more accurate models of ocean temperature dynamics and cloud formation, coupled with a careful error analysis. I am, however, in no doubt that analysis of flux and temperature data, even at the level presented here, is a formidable problem and requires a substantial body of highly critical thought.
.

References

1. Spencer RW and Braswell WD. “On the Misdiagnosis of Climate Feedbacks from Variations in the Earth’s Radiant Energy Balance. Remote Sensing 2011. Internet preprint.
2. Dessler AE. “Clouds Variation and the Earth’s Energy Budget”. GRL 2011. Internet preprint.
3. Oppenheim AV, Willsky AS, Young IT. “Signals and Systems” Prentice/Hall International 1983. Chapters 2,3, pp39-120
4. Curry JA and Webster PA. “Thermodynamics of Atmospheres and Oceans”, Academic Press 1999. Chapter 13,pp 352-358.
5. Bendat JS and Piersol AG. “Random Data. Analysis and measurement procedures”. 2nd Edition, John Wiley & Sons 1986. Chapter 6, pp164-187
6. Papoulis A. “Probability, Random Variables and Stochastic Processes” McGraw-Hill, KogaKusha 1965. Chapter 10, pp 336-351.
7. Forster, PMDEF and Gregory JM. The Climate Sensitivity and Its Components Diagnosed from Earth Radiation Budget Data. J Climate 2006; 19: 39-52.