THE ROLE OF EDDY VARIABILITY IN THE EXTRATROPICAL RESPONSE TO SEA SURFACE TEMPERATURE ANOMALIES

by

Phillip George Kurimski

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

Geosciences

at

The University of Wisconsin-Milwaukee

August 1999

THE ROLE OF EDDY VARIABILITY IN THE EXTRATROPICAL RESPONSE TO SEA SURFACE TEMPERATURE ANOMALIES

by

Phillip George Kurimski

The University of Wisconsin-Milwaukee, 1999

Under the Supervision of Dr. Kyle Swanson

The problem explored in this thesis is the response of an idealized GCM to El Niño-like and how to causally link the model's response to its forcing. Once the forcing is determined, the model can then be corrected toward some designed goal, e.g. an improved simulation of El Niño. Since this approach implicitly includes the effects of transients, it provides a better understanding of how transients affect the overall climate response to SST anomalies and their importance in better overall climate simulations.

The first experiment places a Gaussian cooling anomaly in the central Pacific centered at 180^0W and the Equator. The anomaly itself extends from 140^0E to 140^0W and 20^0S to 20^0N. The GCM is integrated forward in time for 1000 days with this cooling anomaly to generate a climate anomaly. Then, the adjoint is integrated with this anomaly introduced at each time step and run backward in time to determine the sensitivity to that particular climate pattern. The sensitivity generated by the adjoint is centered roughly on the date line and has an overall character quite similar to the imposed Gaussian cooling anomaly.

In the second experiment, the GCM was integrated forward in time for 1000 days to determine the errors associated with the model as measured by departures from the DJF climate from the NCEP reanalysis. The model was then run backwards using the adjoint to determine the forcing that would cancel this climate anomaly. The model can then be adjusted to account for the error and run once again to examine if an improvement in the model climate is obtained. The thermal forcing increments associated with this iteration process seek to improve the climate by

adjusting the tropical/extratropical heating contrasts.

The results obtained appear to provide a useful way to correct climate models, and future research will focus on implementing this approach in a full physics GCM.

TABLE OF CONTENTS

LIST OF FIGURES

1. Introduction

During an El Niño event, the normally warm western Pacific turns cooler as the subtropical easterlies relax and the buildup of warm water ceases. The normally cool eastern Pacific warms up as the weaker easterlies reduce upwelling and the warmer water from the western Pacific migrates eastward in the form of a Kelvin wave pulse. Coincident with this change in Sea Surface Temperature (SST), the precipitation pattern also changes. Due to the cooler SST, the western Pacific does not experience as much convection as it normally does during a non-El Niño year. Conversely, the eastern Pacific has much more convection than normal due to the warmer SST.

These changes in the SST and accompanying changes in atmospheric heating associated with convection in the tropical Pacific Ocean also change weather patterns across the globe. Changes in the SST force what appear to be large-scale atmospheric Rossby waves, propagating from the subtropics into higher latitudes. The increased convection in the eastern Pacific causes a divergence pattern aloft that is not normally present. The decrease in convection in the western Pacific creates a surface high pressure in the central Pacific that is stronger than normal. Coincident with these changes in the tropical Pacific, the

Aleutian low deepens and high pressure over North America strengthens. This series of alternating high and low pressure systems has the structure of a wave train arcing over the Pacific and North America, and is called the Pacific-North American (PNA) teleconnection pattern.

The seasonal to interannual time scales associated with El Niño appear to be predictable (Trenberth et al. 1998); the natural question to ask is can we predict the accompanying climate changes in the extratropics? Global Circulation Models (GCMs) are generally used to model the response of the extratropical atmosphere to El Niño. However, GCMs each have their own bias and as such predict different responses to El Niño forcing. Gates et al. (1999) examined a number of these models within the context of the Atmospheric Model Intercomparison Project (AMIP) project and compared them to what actually occurs in the atmosphere to better understand model systematic errors. AMIP took monthly averaged sea-ice distributions and global SST from 1979-88 with a standardized carbon dioxide concentrations and solar constant values as boundary conditions for the climate model runs. Several generalizations are possible from this study. While GCMs model primary climate variables fairly well when given these reasonable boundary conditions, they do have some

common systematic errors that are relevant to understanding and predicting the El Niño response. In general, each model's output ended up looking more like the other model's output instead of what is actually observed. While GCMs handle the seasonal migration of precipitation fairly well, they do not handle the amount of precipitation that actually falls. GCMs are typically too dry in the large precipitation areas and too wet in the major desert regions for both DJF and JJA, with precipitation errors in the tropics on the order of 20 to 40% of the observed value. There is a dipole of error estimate in the Pacific Ocean with the model underestimating the amount of precipitation in the western Pacific by more than 20% and overestimating the amount of precipitation in the eastern Pacific from 200 to more than 300%. These findings are significant in that the errors in precipitation show up as errors in the divergence field in the upper atmosphere, which is the source for the extratropical climate anomalies.

The interannual variability in the Sea-Level Pressure (SLP) is modeled well in the tropics, with a fairly small standard deviation and the AMIP ensemble and NCEP-NCAR Reanalysis being relatively close together. However, the models do not do well in the middle latitudes. The AMIP ensemble and the NCEP-NCAR reanalysis are not close

together and the standard deviation is large. A model performance diagram shows that the models act more like each other then what is actually observed when modeling SLP anomalies in the mid-latitudes. If the middle latitude response to El Niño cannot be modeled correctly on a diagnostic basis, it is questionable whether they can accurately predict the response to an El Niño SST anomaly in forecast mode.

To understand the errors associated with the models in simulating the extratropical response to El Niño, it is necessary to first understand the dynamics associated with these responses. Warm SST anomalies in the Central Pacific associated with El Niño are found to enhance convection in the Pacific. Associated with the increased convection, a couplet of high pressure areas is found to the north and south of the anomaly in the models, with the anomaly usually centered near the equator (Sardesmukh and Hoskins 1988). Emanating away from the high pressure regions, a wave train of alternating geopotential lows and highs and stream function anomalies extends into the extra-tropics.

In modeling the atmosphere it has been found that a positive temperature anomaly in the tropics, coupled with a divergence field aloft will create a symmetric Rossby wave source (Sardeshmukh and Hoskins 1988). The origin of the

Rossby wave source is found in the divergent terms of the shallow water vorticity equation:

$$S = - v_\chi \bullet \nabla\zeta - \zeta D \qquad (1.1)$$

$$(A) \quad (B)$$

The Rossby wave source S is caused by the divergent wind v_χ blowing over a time mean vorticity gradient $\nabla\zeta$, along with the product of divergence D and absolute vorticity ζ. Even assuming a fixed divergent flow v_χ, the Rossby wave source can change with time if ζ changes (Sardeshmukh and Hoskins 1988). In general, the term A dominates the Rossby wave source, particularly away from the deep tropics.

The extratropical response is not directly caused by the upper level tropical divergence, which acting alone is relatively ineffective. According to equation 1.1 there can be an ample amount of divergence and divergent wind, but without vorticity or a vorticity gradient there will be no Rossby wave source. A divergent outflow coupled with subtropical convergence and the local Hadley circulation, which cause meridional overturning away from the Sea Surface Temperature (SST) anomaly, along with the presence of vorticity gradients are both essential in initiating Rossby wave forcing. For interaction with strong vorticity gradients, the divergence needs to interact with the mid-

latitude westerlies or within a westerly jet stream in the equatorial region. These westerly jet streams are regions of strong mean vorticity gradients. Once within these jet streams, the Rossby wave forcing can generate a disturbance that can propagate into higher latitudes (Trenberth et al., 1998).

The effect of such a Rossby wave source can be most easily seen by considering the barotropic vorticity equation using a Mercator projection of the sphere (e.g., Hoskins and Karoly 1981):

$$x = a\lambda, \tag{1.2}$$

$$y = a \ln[(1 + \sin\phi)/\cos\phi]. \tag{1.3}$$

Then

$$\frac{1}{a \cos \phi} \frac{\partial}{\partial y} = \frac{1}{\cos \phi} \frac{\partial}{\partial x},$$

$$\frac{1}{a} \frac{\partial}{\partial \phi} = \frac{1}{\cos \phi} \frac{\partial}{\partial y}, \tag{1.4}$$

$$\nabla^2 = \frac{1}{\cos^2 \phi}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right), \tag{1.5}$$

$$\cos\phi = \operatorname{sech} y/a, \tag{1.6}$$

$$\sin\phi = \tanh y/a. \tag{1.7}$$

The Mercator basic zonal velocity

$$\bar{u}_M = \bar{u} / \cos\phi \tag{1.8}$$

is proportional to the angular velocity. The equation for the horizontal stream function perturbation Ψ, on multiplying by $\cos^2\phi$, takes the form

$$\left(\frac{\partial}{\partial t}+\bar{u}_M\frac{\partial}{\partial x}\right)\left(\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}\right)+\beta_M\frac{\partial\psi}{\partial x}=0, \qquad (1.9)$$

where

$$\beta_M=\frac{2\Omega}{a}\cos^2\phi-\frac{d}{dy}(\cos^2\phi\bar{u}_M) \qquad (1.10)$$

is cos ϕ times the meridional gradient of the absolute vorticity on the sphere (Hoskins and Karoly 1981).

A barotropic model such as this seeks to mimic the contribution of the external mode in a continuously stratified model to the stationary wave field. To properly do this in this model, the wind blowing over topography must be larger than the actual surface wind, but smaller than the wind at the equivalent barotropic level (about 300 – 500mb). It should be noted that the group velocity of the stationary external mode in realistic vertical shear is found to be a great deal larger than that of the stationary Rossby wave in this barotropic model (Held et al. 1985).

The wave trains generated by a Rossby wave source in this barotropic model can be described by using ray tracing. An example is shown in Figure 1, taken from

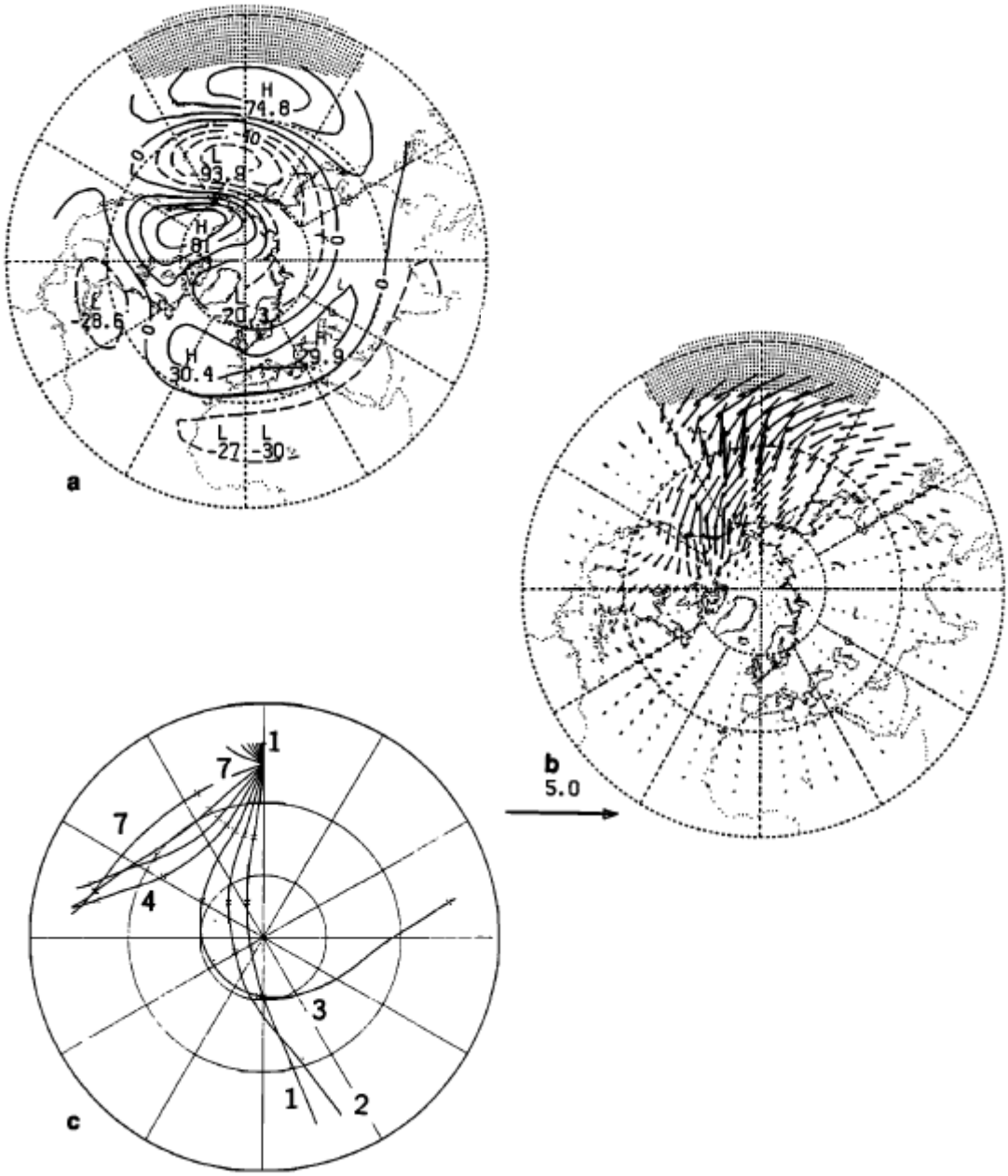

Figure 1. Linearized, steady state primitive equation model solution to thermal forcing in the tropics in northern hemisphere (NH) winter, showing (a) the height field anomalies at 200 hPa, contour interval 20m, (b) the associated horizontal flux of stationary wave activity, and (c) the barotropic Rossby wave rays for tropical forcing at 15^0N in a 300 hPa NH winter basic state; crosses indicate phases every 180^0 [from *Trenberth et al.*, 1998].

Trenberth et al. (1998). Ray tracing follows the wave packet of energy as it is forced out of the tropics and moves into the middle and higher latitudes. Figure 1a shows the alternating geopotential highs and lows emanating from a region of anomalous tropical heating in a linearized primitive equation model, denoted by the stippled area at 200 mbar. Figure 1b denotes the associated horizontal flux of stationary wave activity. Figure 1c shows the different rays, each corresponding to a different wave number K, which emanates out of the tropics. These rays reach their turning latitude, and then go back into the tropics, where (presumably) they are absorbed. The behavior of these rays can be explained by examining the dispersion relation for plane wave solutions exp*i(kx + ly - ωt)* of 1.9

$$\omega = \bar{u}_M k - \frac{\beta_M k}{k^2 + l^2}. \qquad (1.11)$$

Here $\bar{u}_M$ is the local zonal wind, k is the zonal wave number, β_M is the potential vorticity gradient, l is the meridional wave number, and the frequency, ω, is zero since we are considering stationary waves. The activity of almost-plane waves moves with the group velocity in the y direction

$$c_{gy} = \frac{\partial \omega}{\partial l} = \frac{2\beta_M kl}{(k^2 + l^2)^2} \qquad (1.12)$$

where k is constant. Therefore the variable that is changing the group velocity is l, so l must be isolated to determine the behavior of the waves. When l is isolated it yields the equation

$$l^2 = K_S^2 - k^2 \tag{1.13}$$

where the stationary wave number $K_S^2 = (\beta_M / \bar{u}_M)$. Rossby waves reach their turning latitude when $K_S = k$, as at that location their meridional group velocity vanishes (Hoskins and Karoly 1981).

As shown in Figure 1c, short waves (rays 4-7) reach their turning latitude very quickly and never make it into the extratropics. Since k is a fairly large for such waves, $K_S = k$ before the Rossby wave can make it through the jet stream to propagate into the higher latitudes. Long waves (rays 1 & 2) never reach their turning latitude and travel approximately along great circles. This occurs because k is so small that K_S never equals k. However, wavelengths that fall in between these two extremes (ray 3) correspond with the largest response in the extratropics due to El Niño forcing. These rays arc in an almost Great Circle path from the tropics and into the higher latitudes, spend a great deal of time in the vicinity of their turning latitude at about 60°N, and then return to the tropics. This occurs because $K_S \cong k$ in the middle of the jet stream,

allowing the Rossby wave to travel within the jet stream itself as if the jet were in wave guide. These wave trains seem to move most effectively in the middle latitudes along an equivalent barotropic level often found between 500 and 200 hPa (Hoskins and Karoly 1981).

The primary failing of this is that the response in GCMs does not translate with heating, as this barotropic model would suggest. Rather, response occurs in preferred locations, i.e., the PNA region. This is because there are resonant anomalies present in the zonally varying flow that need to be accounted for within the model. To this end, we next consider a barotropic model that contains a basic flow that varies with longitude as well as latitude.

Simmons et al. (1983) used such a barotropic model to determine if there are regions of the tropics that may excite a particularly large midlatitude response. Their results showed that Southeast Asia and the tropical northwest Pacific were two regions that excited a large response over the extratropical northeast Pacific. Interestingly, these two regions are areas of significant convective activity. Forcing of an opposite sign over the tropical central Pacific can create a weaker, though similar, response. Therefore, a PNA pattern response can be triggered by a tropical anomaly comprising an eastward

or westward shift of convective heating between the dateline regions and regions lying near 120^0E. This redistribution of convective heating is similar to the Southern Oscillation rainfall anomalies (Simmons et al. 1983). This significant result allows us to confidently focus our attention on tropical forcing in the Pacific.

The results of Simmons et al. (1983) show that zonal inhomogeneities in the time mean flow are important to the overall structure of the response. However, even with the zonal inhomogeneities this barotropic model still fails to correctly model the response to El Niño, primarily because transient eddies play a large role in that response. Held at al. (1989) found that transients are the dominant forcing mechanism in a linear diagnostic baroclinic model, particularly the anomalous upper tropospheric transient vorticity fluxes, and further that the direct response to anomalous diabatic heating was rather small in the extratropics. Transients respond in two primary ways to tropical heating: the movement of the Pacific storm track associated with the anomalous extratropical wave train, and changes in the penetration of Rossby waves into the tropics resulting from the modified tropical winds (Held et al. 1989). These results are illustrated in Figures 2-4 from Held et al. (1989).

Figure 2a shows the 300 hPa eddy geopotential height obtained from the GCM integration in Lau (1985). Figure 2b shows the total linear response of 300 hPa geopotential to the difference in heating between the two composites plus the anomalous transient eddy forcing in the momentum and temperature equations. Figure 2c is the linear response to

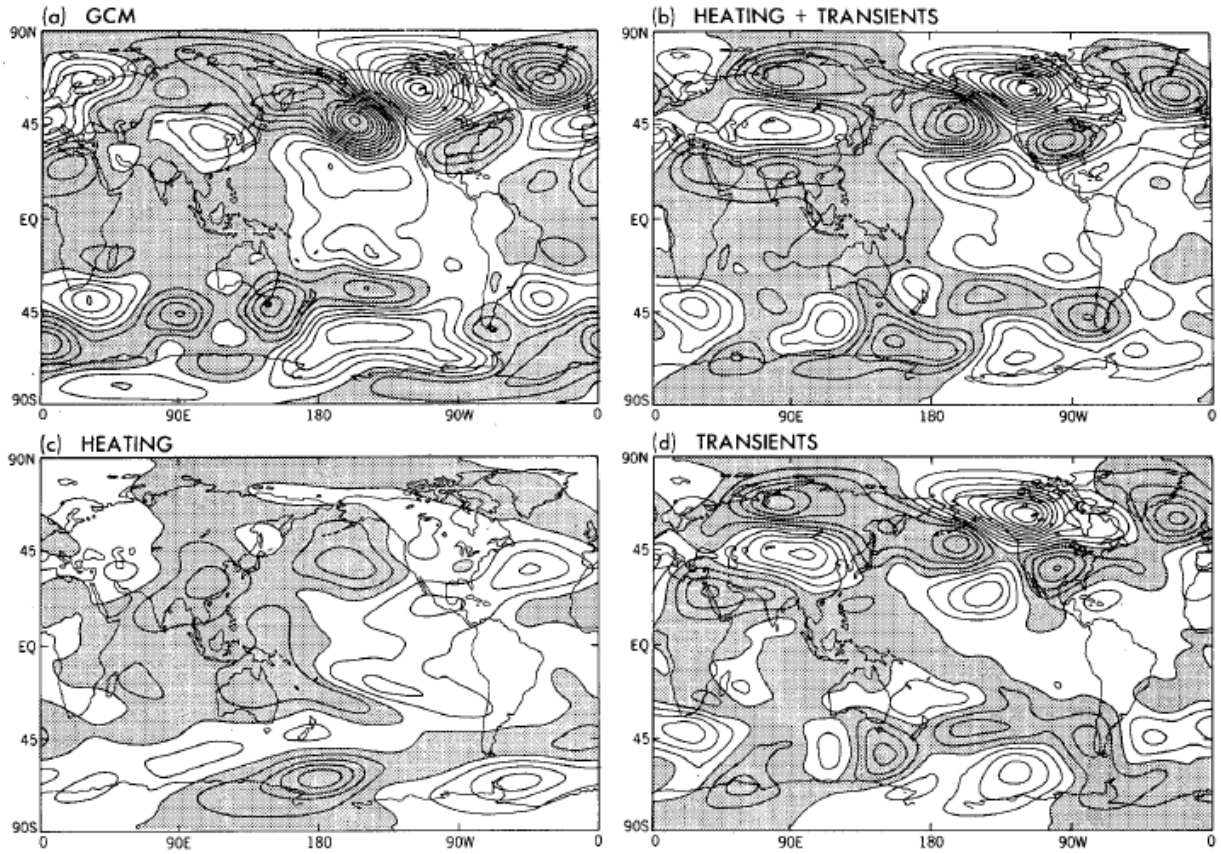

Figure 2. (a) The 300 hPa eddy geopotential height obtained from the GCM integration in Lau (1985) by compositing three El Niño winters and subtracting an analogous composite for three anti-El Niño winters; (b) the linear response of the 300 hPa geopotential to the anomalous diabatic heating and forcing by transient eddies; (c) the linear response to anomalous diabatic heating; (d) the linear response to the anomalous forcing by transients. The contour interval is 10 m. Negative values are shaded [from *Held et al.*, 1989].

anomalous diabatic heating and 2d is the linear response to the anomalous forcing by transients. The heating alone forces an extratropical wave train that is much smaller than that generated by the GCM while the response to the anomalous transients is clearly dominant (Held et al. 1989).

Figure 3 is similar to Figure 2 except the eddy zonal wind at 300 hPa is examined. This figure shows that

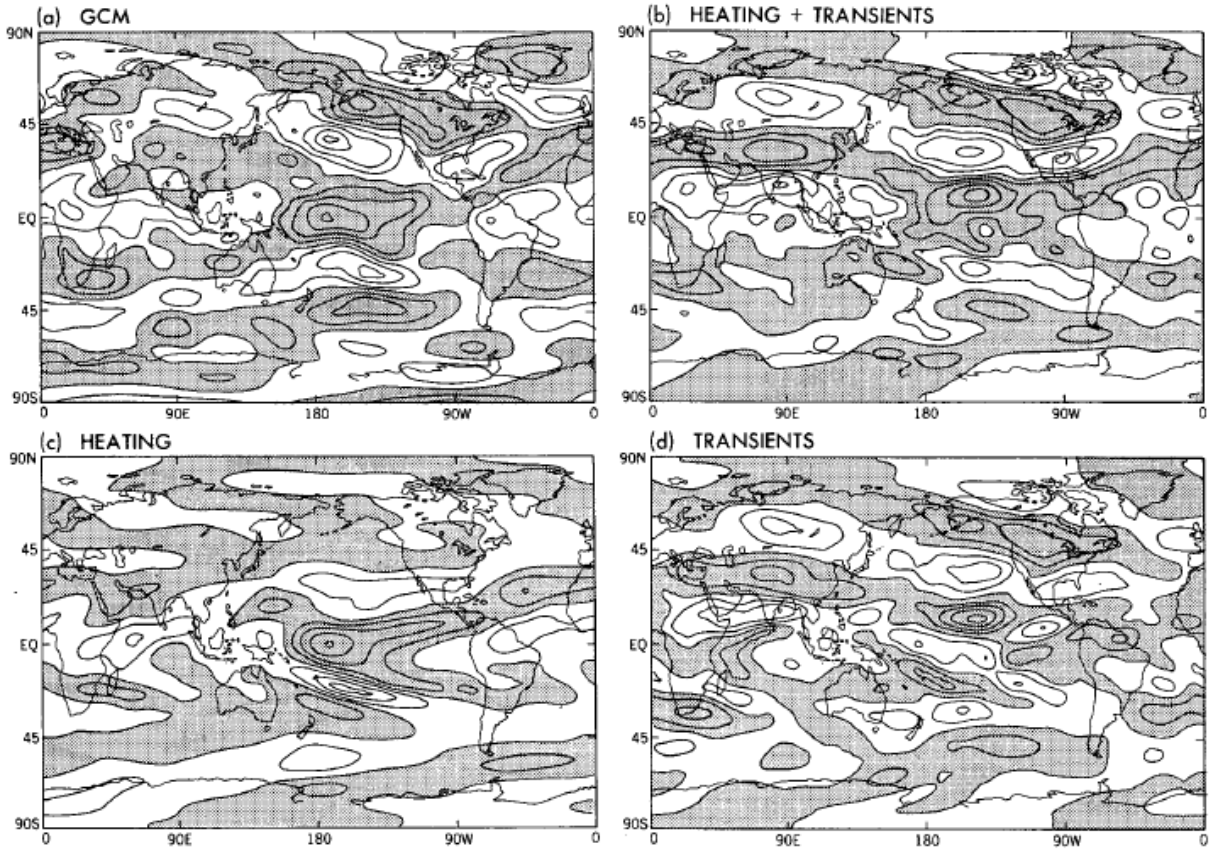

Figure 3. As in Figure 2, except for the eddy zonal wind at 300 hPa. The contour interval is 2 m s^{-1}. negative values are shaded [from *Held et al.*, 1989].

heating alone produces easterlies that do not penetrate far enough into the northern subtropics, along with a subtropical acceleration that is weak and displaced equatorward. The forcing by transients appears to be needed to extend the easterly anomaly further northwards as well as to generate the bulk of the acceleration of the midlatitude winds in the central Pacific (Held et al. 1989).

Figure 4 decomposes the transient response further to determine which part of the transient forcing predominates. Figure 4a shows the 300 hPa eddy geopotential produced by the anomalous transients in the vorticity equation. Figures 4b and 4c decompose this into the lower troposheric transients ($p < 600$ hPa) and upper tropospheric transients of which the upper tropospheric transients were found to provide the bulk of the response (Held et al. 1989).

The linear diagnosis of the GCMs response to El Niño clearly demonstrates that the model's extratropical wave train is not simply the direct response to anomalous tropical heating; the response to the anomalous transients is vitally important, particularly the response to anomalies in the upper tropospheric vorticity fluxes (Held et al. 1989).

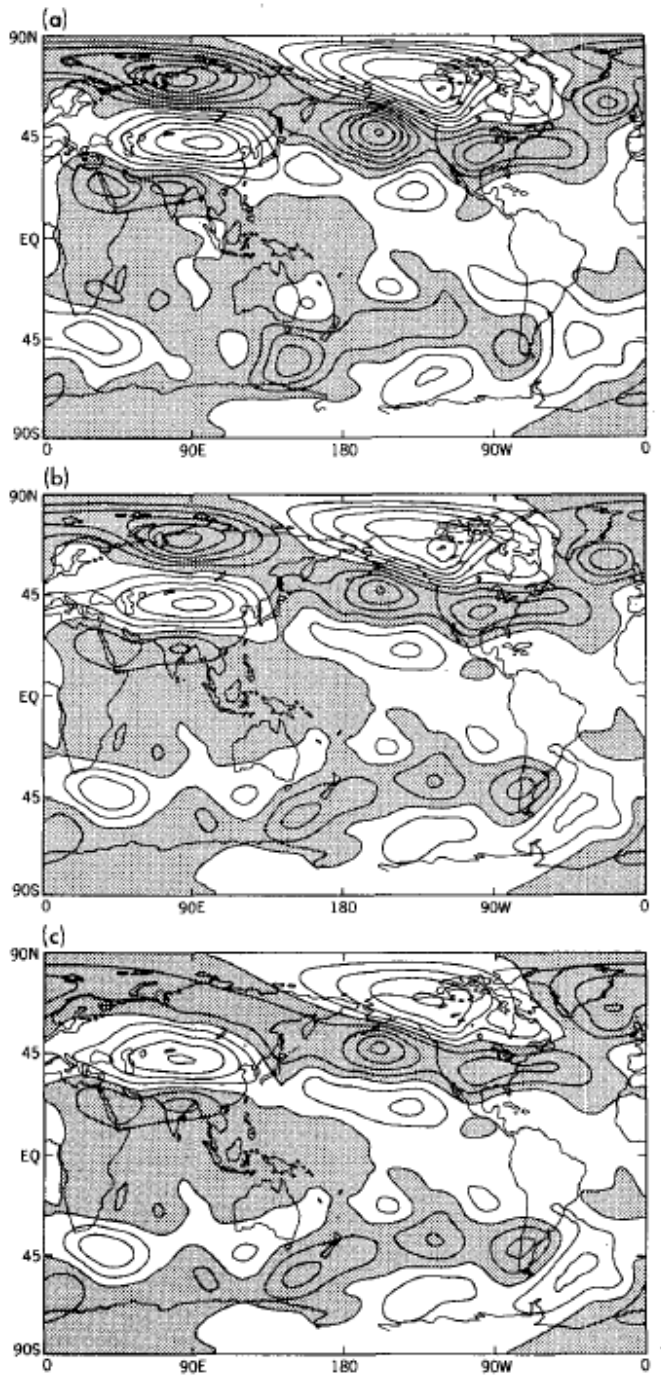

Figure 4. The linear response of 300 hPa geopotential height to (a) the anomalous vorticity tendency due to transients, (b) the anomalous vorticity tendency due to transients in the upper troposphere ($p < 600$ hPa), and (c) the anomalous vorticity tendency due to transients in the upper troposphere, with high latitude forcing ($> 70^0$N) excluded. Contour interval is 10 m [from *Held et al.*, 1989].

Other problems in correctly modeling the extratropical response to El Niño include the processes involved in transforming the thermally forced internal mode in the tropics into an external, equivalent barotropic mode in the extratropics, along with the mechanisms that produce extratropical convergence anomalies associated with the Rossby wave source. There also needs to be further research into nonlinearity along with the role of transients. In the atmosphere, as well as in models, climate anomalies are due in large part to motions on a time scale of a month or less and how those motions are influenced by errors in physical parameterization. The importance of transient eddies to the overall response suggests that if a systematic technique is to correctly attribute climate model errors to errors in physical parameterizations, leading to better El Niño forecasts, it must include the effects of such eddies. Climatological mean planetary waves can produce internal sources of energy that can compete in magnitude with the original perturbation. Planetary waves are also important in that they can produce changes in jet streams and storm tracks, which can alter and reinforce the initial response. All of these parameters can be modeled; however, the signal needs to be sorted out from the naturally occurring high level of

variability that creates noise in the models.

The problem we will explore is the response to El Niño-like forcing in an idealized GCM and how to causally link the model's response to its forcing. This will be done within the context of control theory. Once the forcing is determined, in principle the model can then be corrected toward some designed goal. Since this approach implicitly includes the effects of transients, it should give us a better understanding on how transients affect the overall climate response to SST anomalies and how we can account for them to yield better overall climate simulations.

2. Description of the Model

An idealized spectral GCM, based on atmospheric primitive equations and incorporating simple physical parameterizations, will be used to study the atmosphere's response to El Niño forcing. This model is a standard hydrostatic, σ-coordinate, semi-implicit, spectral transform model, in the vorticity-divergence form. The equations that will be used in the model are as follows:

$$\frac{\partial V}{\partial t} = \ldots - k_v(\sigma) v$$

$$\frac{\partial T}{\partial t} = \ldots - k_T(\phi, \sigma)[T - T_{eq}(\phi, p)]$$

$$T_{eq} = \max\left\{200K, \left[315K - (\Delta T)_y \sin^2\phi - (\Delta\theta)_Z \log\left(\frac{p}{p_0}\right)\cos^2\phi\right]\left(\frac{p}{p_0}\right)^{\kappa}\right\}$$

$$k_T = k_a + (k_s - k_a)\max\left(0, \frac{\sigma - \sigma_b}{1 - \sigma_b}\right)\cos^4\phi$$

$$k_v = k_f \max\left(0, \frac{\sigma - \sigma_b}{1 - \sigma_b}\right)$$

where

$\sigma_b = 0.7$ $\quad k_f = 1 \quad day^{-1}$

$k_a = 1/40 \quad day^{-1}$ $\quad k_s = 1/4\, day^{-1}$

$(\Delta T)_y = 60K$ $\quad (\Delta\theta)_Z = 10K$

$p_0 = 1000hPa$ $\quad \kappa = \frac{R}{c_p} = \frac{2}{7}$ $\quad c_p = 1004 Jkg^{-1}K^{-1}$

$\Omega = 7.292\times10^{-5} s^{-1}$ $\quad g = 9.8 ms^{-2}$ $\quad a_e = 6.371\times10^{6} m,$

is the usual force and energy balance terms in the momentum

and energy equations, respectively.

Newtonian relaxation of the temperature field to a zonally symmetric state and Rayleigh damping of low-level winds to represent boundary-layer friction is used for the forcing and dissipation. Forcing GCMs in this way is common, especially in two-layer models (Held and Suarez 1994).

To start the model, the following parameters must be specified: the gas constant (R), the specific heat of air at constant pressure (c_p), the acceleration of gravity (g), the radius of the sphere (a_e), the total mass of the atmosphere (p_0/g), and the forcing parameters. A simple linear damping of the velocities is the only specified dissipation. The strength of the damping k_v is a function of p/p_s, where p is the pressure and p_s is the pressure at the surface. In order for the boundary layer to follow the topography, σ is used instead of pressure. Layers near the surface ($\sigma > 0.7$) are the only place where the damping is nonzero (Held and Suarez 1994).

Temperatures are relaxed to create a radiative equilibrium T_{eq}, a function of latitude and pressure. This radiative equilibrium is given some static stability, which is large in the tropics and dissipates to zero at the poles. This tropical static stability is an artifact that

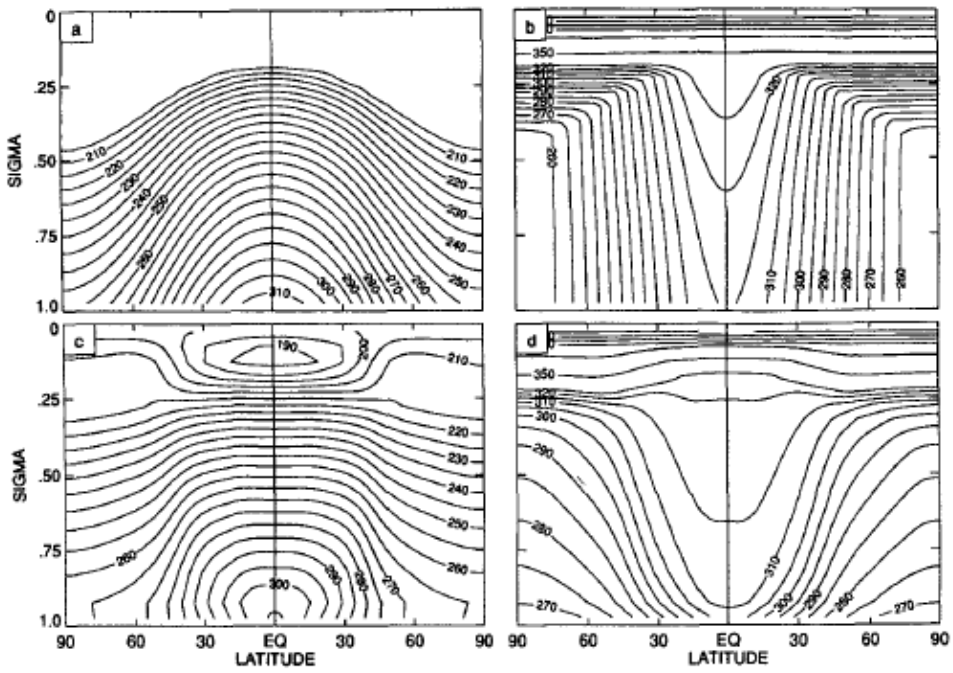

Figure 5. The upper panels contain the prescribed radiative equilibrium temperature (a) and potential temperature (b) distributions. The lower panels contain 1000-day averages of the zonal mean temperature (c) and potential temperature (d) distributions produced by the G72 gridpoint model [Held and Suarez 1994].

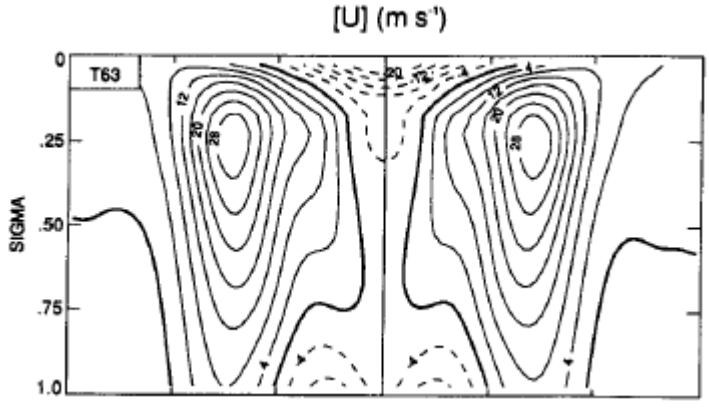

Figure 6. The 1000-day mean zonal-mean zonal wind produced by the T63 spectral model. Since the forcing is symmetric about the equator, differences between the hemispheres are indicative of sampling errors [Held and Suarez 1994].

helps minimize the occurrence of gravitational instability. The temperatures and the corresponding potential temperatures are shown in Figures 5a,b. The radiative relaxation, also a function of latitude and σ, creates an unrealistic thin cold layer near the surface, particularly in the tropics, when a long relaxation time is used everywhere. To reduce this effect, a shorter relaxation time is used in the tropics. Figures 5c,d show the effect that radiative relaxation have on the time-mean temperature and potential temperature distribution. The potential temperatures show that this static stability layer is supported by the dynamics in the extratropics well above its radiative equilibrium value (Held and Suarez 1994).

In the model, horizontal mixing is included on a very scale selective basis while vertical mixing, convective or diffusion adjustment, has been omitted altogether. The model can be integrated stably without enhanced vertical mixing to account for the gravitationally unstable regions that occur in the low latitudes. The zonal mean wind used in the model is illustrated in Figure 6.

The transform grid is chosen to ensure alias-free computation of quadratic products, in the usual way. The hydrostatic equation is integrated analytically assuming that temperature is constant within each layer, and the

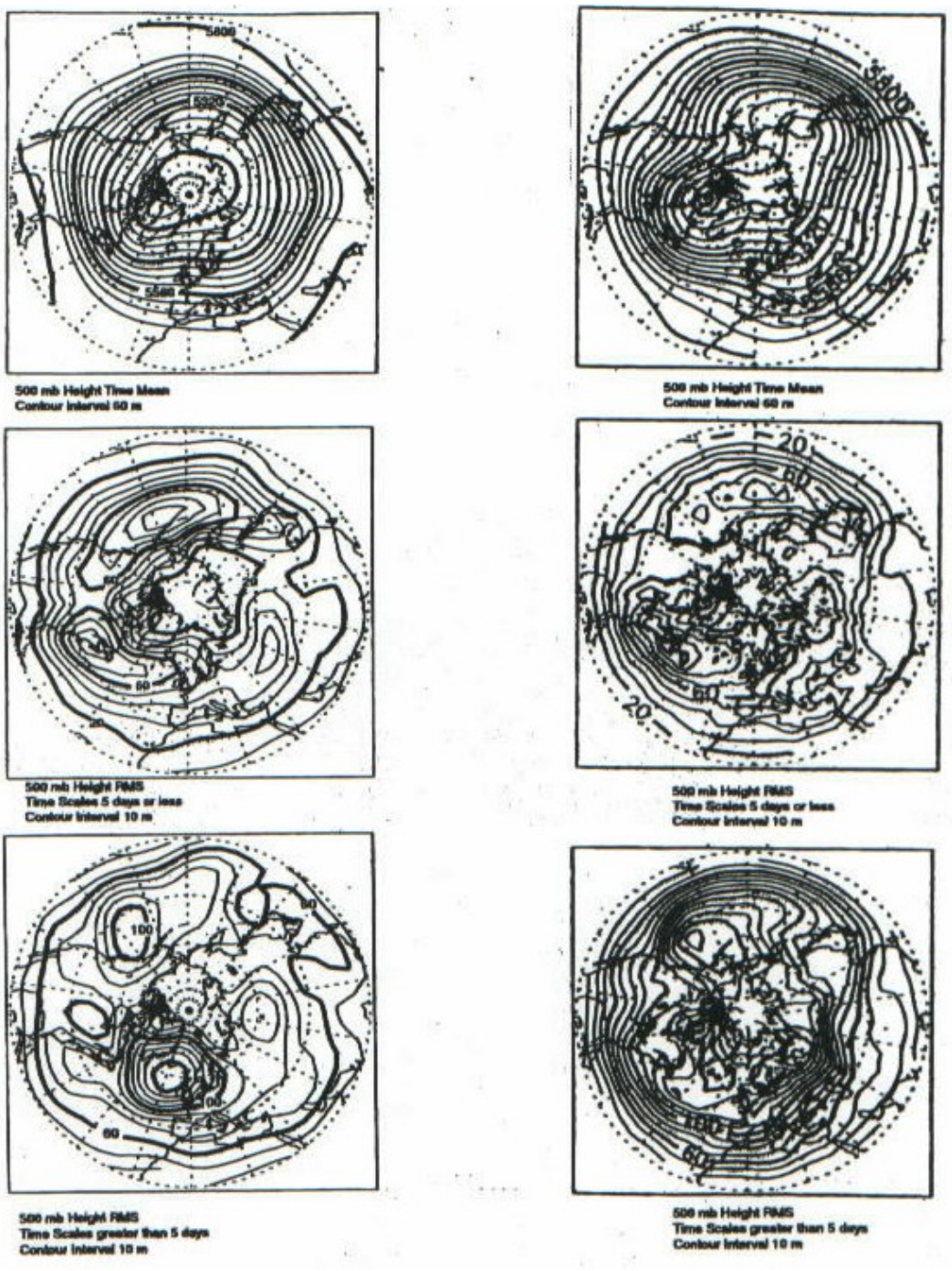

Figure 7. Model (left) and observed (right) 500 mb height time mean, high frequency variability RMS, and low frequency variability RMS for January. The modeled variability is calculated using the forcing resulting from 20 iterations of the forcing optimization technique, while the observed variability is calculated from the 1983-1993 ECMWF analyzed fields.

vertical differencing uses central differences. There are 5 vertical levels in the model, each equally spaced in sigma, with the top of the model at zero pressure. Time stepping is done using a leapfrog scheme, using a time filter to control the computational mode. The horizontal mixing of vorticity, divergence, and temperature takes the form of the Laplacian raised to the fourth power, with the strength set so that the e-folding time for the smallest wave in the system is always 0.1 days. The truncation is triangular.

The model climate used in the calculations is shown on the right hand side of Figure 7 while the observed climate is on the left hand side of Figure 7. The 500 mb height time mean in our model is too zonal and does not simulate the troughs and ridges that are present in the observed climate. The high frequency variability RMS shows that the model climate extends the storm tracks too far downstream does not contain enough low frequency variability RMS. Even with these deficiencies, this simulation is reasonable considering the simplicity of the model.

3. Description of Adjoint Methods

We are interested in the sensitivities of a large number of atmospheric output parameters, such as geopotential height at many different locations and times, with respect to one input field, namely heating. Practical implementation of the minimization of the scalar function will require repeated explicit determination of the gradient of the scalar function with respect to all the input parameters whose values are sought (Talagrand 1991).

Let us consider a finite algebraic process which, starting from some input vector u belonging to some open set $U \subset \mathbb{R}^n$ ($u = (u_I, i = 1,\ldots,n)$) produces an output vector v belonging to $\mathbb{R}^m$ ($v = (v_j, j = 1,\ldots,m)$). The process can be described by the equation

$$v = G(u) \tag{3.1}$$

where G is a differentiable function defined in U with values in $\mathbb{R}^m$. The perturbation δv on the output resulting from a perturbation δu on the input is given to first order by the corresponding *tangent linear equation*

$$\delta v = G' \delta u \tag{3.2}$$

where G′ is the local jacobian matrix of G. Let now $v \rightarrow J(v) \in \mathbb{R}$ be a differentiable scalar function defined on the image set G(U). The gradient of J with respect to u is

given by the chain rule

$$\frac{\partial J}{\partial u_i} = \sum_{j=1}^{m} \frac{\partial v_j}{\partial u_i} \frac{\partial J}{\partial v_j} \qquad i = 1,\ldots,n$$

or, in matrix notation

$$\nabla_u J = G'^* \nabla_v J \tag{3.3}$$

where G'* denotes the transpose of G' (Talagrand 1991).

The adjoint method simply consists in numerically computing $\nabla_u J$ through the formula (3.3). The main advantage of the adjoint method is that it is numerically very economical, at least in comparison with the other methods which could be considered for determining $\nabla_u J$. The numerical cost of one adjoint computation will generally be of the same order of magnitude as the cost of one direct computation (3.1). This is much less than the other methods (forward automatic differentiation or finite differences) which could be considered for numerically determining the gradient $\nabla_u J$. Two important remarks are in order.

1. If G is the composition of a number of more elementary processes,

   $$G = G_M \circ \ldots G_2 \circ G_1$$

   the transpose jacobian G'* will be the product of the elementary transposed jacobians, taken in reversed order

$$G'^* = G_1'^* \circ G_2'^* \ldots \circ G_M'^*$$

This suggests a systematic procedure for developing the adjoint of a given code: successively develop the adjoints of the various components of the basic code (these components can for instance be FORTRAN subroutines, each of which can itself be decomposed into more elementary components, down to the individual executable FORTRAN statements), and then connect the elementary adjoints in reversed order. In the adjoint code the input and output spaces of each component of the basic code will be reversed. These notions will be developed in detail later on (Talagrand 1991).

2. When the function J is nonlinear, the gradient $\nabla_u J$ will depend on the point u at which it is evaluated, and the adjoint computation (3.3) will require the knowledge of all quantities used in nonlinear operations in the direct computation of (3.1). It will therefore be necessary to store these results at the time of the direct computation. This is the price to be paid for the economy that the adjoint method allows in the amount of computation.

The adjoint method has just been described in the finite dimensional algebraic case, appropriate for computer codes. But the principle of the adjoint method is not

limited to algebraic cases. It will be useful for the sequel to describe more precisely what the adjoint method becomes when the basic computation (3.1) represents the integration of a dynamical system

$$\frac{dx}{dt} = F(x) \tag{3.4}$$

where, to fix ideas, we assume that x belongs to $\mathbb{R}^n$. The input u is now the initial condition $x(t_0)$ from which (3.4) is integrated, while the output v is the result $x(t_1)$ of the integration at a given time $t_1 > t_0$ (m = n in this particular case). For a given solution x(t) of (3.4), the analog of computation (3.2) is the integration, from $\delta u = \delta x(t_0)$, of the corresponding *tangent linear equation*

$$\frac{d\delta x}{dt} = F'(t)\delta x \tag{3.5}$$

where F'(t) denotes, for any t, the jacobian of the function F, taken at point x(t). As for the analog of the adjoint computation (3.3), it is the integration, from the "final" condition $\delta' x(t_1) = \nabla_v J$, of the adjoint equation

$$\frac{d\delta' x}{dt} = -F'^*(t)\delta' x \tag{3.6}$$

(Talagrand 1991).

The principle of the adjoint method can be further extended to infinite dimensional cases. In such examples,

the analog of equation (3.1) will often be the integration of a set of partial differential equations, u denoting initial or boundary conditions whose specification uniquely defines the solution of the equations, and v denoting that solution. The adjoint computation (3.3) will then represent the integration of a set of linear partial differential equations, whose output along the initial/lateral boundaries will define how one must act on these boundaries in order to make the function J(u) vary in a prescribed way. In that context, the adjoint equations can be used, either as a theoretical tool for obtaining the mathematical solution of a given control problem, or a numerical tool for obtaining the (discretized) numerical solution of a control problem. Variational assimilation of observations is in fact a discretized version of an infinite dimensional control problem with respect to the initial conditions of an atmospheric model (Talagrand 1991).

We will use the adjoint method to determine heating anomalies which, when superimposed on given meteorological fields, will project onto patterns of interest, e.g. PNA and model climate errors. This question is of the utmost importance for what is now a major problem in climate modeling, namely how do we estimate the error associated

with the climate models in order to improve upon them. The evolution of the flow being governed by a differential equation of the form (3.4), the time evolution of a perturbation δh_0 superimposed on a solution x(t) of (3.4) at time t_0 will be described, to first order, by the corresponding tangent linear equation (3.5).

4. Coding the Adjoint

To code the adjoints the components of the adjoint code must be in one-to-one correspondence with the components of the direct code: to each subroutine of the adjoint code, performing a well defined task, will correspond a subroutine of the adjoint code, performing the adjoint task (with inversion of the inputs and outputs). Similarly, to each FORTRAN statement in the direct code will correspond one (or several) statement in the adjoint code.

These principles can be illustrated on the FORTRAN statement

$$a = b*c \tag{4.1}$$

to be considered here as the analog of equation (3.1). The input to this statement is made up of b and c, while its output is made up, not only of a, but also of b and c, which, in a FORTRAN code, will of course still be available, after statement (4.1) has been executed, for possible future use. The corresponding "tangent linear" statement, analogous to (3.2), reads

$$\delta a = \delta b*c + b*\delta c \tag{4.2}$$

which, for given b and c (*i.e.* for a given basic computation), defines a linear operator, with input

$(\delta b,\delta c)^T$, and output $(\delta a,\delta b,\delta c)^T$. The corresponding matrix is the 3x2 matrix

$$\begin{matrix} c & b \\ 1 & 0 \\ 0 & 1 \end{matrix}$$

The corresponding adjoint computation, which leads from a 3-vector $(\delta' a,\delta' b,\delta' c)^T$ to a 2-vector $(\delta'' b,\delta'' c)^T$, therefore reads

$$\begin{aligned} \delta'' b &= \delta' + c * \delta' a \\ \delta'' c &= \delta' c + b * \delta' a \end{aligned} \qquad (4.3)$$

We recall the meaning of the different variables: $\delta' a,\delta' b$ and $\delta' c$ are the partial derivatives of some output function of the basic code with respect to the variables contained in the respective addresses a, b and c after statement (4.1) has been executed, while $\delta'' b$ and $\delta'' c$ are the partial derivatives of the same output function with respect to the variables contained in addresses b and c before statement (4.1) has been executed (Talagrand 1991).

Equations (4.2) and (4.3), and the meaning associated to $\delta' b$, $\delta' c$, etc, suggest to use the same FORTRAN name for $\delta' b$ and $\delta'' b$, and for $\delta' c$ and $\delta'' c$. Indeed, experience shows that it is more convenient to

use the same FORTRAN name for a linear code, and for the corresponding adjoint variable in the adjoint code. As for the "nonlinear" quantities which have been saved form the basic computation, they will be named according to a simple transparent rule, for instance be adding a prefix b in front of their basic name. With these conventions, eqs (4.2) and (4.3) respectively become

$$a = b*bc + bb*c$$

and

$$b = b + bc*a$$
$$c = c + bb*a \qquad (4.4)$$

The form of statements (4.4) clearly shows that all adjoint variables must be set equal to 0 in the adjoint code before they are used for the first time. Similarly, statement (4.1) defines in the basic code a new content for the address a. In order to avoid any possible interference with a possible previous use of address a in the basic code, it is necessary, in the adjoint code, to set the content of address a to zero after statements (4.4) have been executed. The adjoint statements corresponding to statement (4.1) therefore finally read

$$b = b + bc * a$$
$$c = c + bb * a$$
$$a = 0$$

The general rules which have just been demonstrated can be automatically implemented on a FORTRAN code. It is seen that the tangent linear code is a necessary intermediary for developing the adjoint code. The tangent linear code is also extremely useful for checking the adjoint code. The only basic choice to make, when developing either the tangent linear or the adjoint code, is whether the quantities appearing in nonlinear computations in the basic code will have to be recomputed in the course of the tangent linear or adjoint computation (Talagrand 1991).

There is a very simple method to test whether or not the adjoint works and if you chose the right matrix. In the tangent linear model $\frac{M(x+\delta x)-M(x)}{M\delta x} \to 1$ in order to insure that the correct matrix M was chosen. In order to determine if the adjoint worked correctly the condition $<M\delta x, \delta x>=<\delta x, M^{*}\delta x>$ should be met where M^{*} is the transpose of M. This condition was placed at the beginning of the program with M coming from the subroutine tanint and M^{*} coming from adjint and then multiplied by the corresponding x and y variables. For

the program to continue the condition has to be met within double precision accuracy.

5. Experiments

Our project has two objectives. The first of which is to show we can reproduce forced anomalies through adjoints. Specifically, given an El Niño anomaly we want to find the forcing responsible for creating this anomaly. The second objective is to determine the error associated with the model and reduce it systematically. Before either of these can be dealt with the tangent linear growth needs to be understood along with the model error itself.

To understand how to account for model error we will take a simple matrix propagator corresponding to the system $\delta x_{t+1} = M \delta x_t$ where

$$M = \begin{matrix} 3 & 1/3 & 1/3 \\ 0 & 1 & 1/3 \\ 0 & 0 & 1/2 \end{matrix}$$

The way this matrix is set up the middle and lower term act are neutral or decaying while the top term is the growth term. In this matrix the middle and lower term do not feel any effects from the upper term. The adjoint to this simple system simply involves the transpose of the matrix,

$$M^T = \begin{matrix} 3 & 0 & 0 \\ 1/3 & 1 & 0 \\ 1/3 & 1/3 & 1/2 \end{matrix},$$

with the corresponding sensitivity evolution equation $\delta x_t^* = M^T \delta x_{t+1}^*$. Multiplying M^T and M by themselves 10 times yields:

$$M^{10} = \begin{matrix} 59049 & 9841.\bar{3} & 9185.29 \\ 0 & 1 & .666 \\ 0 & 0 & 9.77*10^{-4} \end{matrix} \quad (M^T)^{10} = \begin{matrix} 59049 & 0 & 0 \\ 9841.3\bar{3} & 1 & 0 \\ 9185.29 & .666 & 9.77*10^{-4} \end{matrix};$$

in the $(M^T)^{10}$ matrix, the growth of the middle and lower terms fshows that this operator "feels" the growth of the upper term, i.e. errors in these decaying terms can lead to errors in the growing term in the future for this simple matrix operator. However, in the M^{10} matrix the middle and lower terms are unaffected by the growth of the upper term. In order to find the leading vector multiply both results by the unit vector A = [1 0 0]

$$M^{10} * A = \begin{matrix} 59049 \\ 0 \\ 0 \end{matrix} \qquad (M^T)^{10} * A = \begin{matrix} 59049 \\ 9841.3\bar{3} \\ 9185.29 \end{matrix}$$

Growth in sensitivity occurs in all degrees of freedom; in contrast with the tangent linear M^{10} matrix, where only the unstable direction amplifies. In general time dependence of M means non-trivial growth of $M^T \delta x$ will

occur in an arbitrary basis. This growth of the sensitivity must be handled for application of inverse techniques over climatic time scales.

Generalizing the above, we define a non-linear model M that operates on the true state vector x_n. Then, the actual system evolution at time t_n is

$$x_{n+1} = M(x_n) + e_n^q, \tag{5.1}$$

where e_n^q is the model error at time t_n. Introducing the tangent linear approximation yields

$$M(x_n + \delta x_n) \approx M(x_n)\delta x_n,$$

where $M(x_n) = \partial M(x)/\partial x\big|_{x=x_n}$ is the tangent linear model of M evaluated at x_n and δx_n is the variation of x_n (Ménard and Daley 1996).

To handle the growth of the sensitivity within this tangent linear system, we begin with an adjoint equation of the form

$$\begin{aligned} &\lambda_n = M^T(x_n^*)\lambda_{n+1} - R_n^{-1}[z_n - x_n^*], \\ &1 \le n \le N-1, \end{aligned} \tag{5.2}$$

subject to the adjoint final condition,

$$\lambda_N = -R_n^{-1}[z_n - x_n^*], \tag{5.3}$$

where Q_n^{-1} is the model uncertainty, R_n^{-1} is the observed error covariance, and z_n is the observations.

These equations involve x_n^* which has not been determined. Variation of 5.1 on e_n^q assuming Q_n^{-1} is non-singular yields the evolution equation,

$$x_{n+1}^* = \mathscr{M}(x_n^*) - Q_n \lambda_{n+1}, \quad 0 \le n \le N-1, \tag{5.4}$$

subject to the *initial condition*,

$$x_0^* = x_0^a - P_0^a M^T(x_0^*)\lambda_1, \tag{5.5}$$

where x_0^a is the analysis produced by observations for $t \le t_0$ (Ménard and Daley 1996). The system (5.2-5.5) forms the *discrete-time Euler-Lagrange* equations of the variational problem. This system of equations is coupled and cannot be solved with a simple direct integration.

In order to decouple the Euler-Lagrange equations the sweep method has to be implemented which takes the form

$$x_n^* = x_n^a - P_n^a M^T(x_n^a)\lambda_{n+1}, \tag{5.6}$$

Applying the sweep method adds a damping term to the adjoint equation which, when not in place, causes the sensitivity to grow exponentially with time as noted above. The sweep method yields a modified adjoint equation of the form

$$\lambda_n = [I - R_n^{-1} P_n^a] M^T(x_n^a)\lambda_{n+1} - R_n^{-1}[z_n - x_n^a], \qquad (5.7)$$

subject to the final condition,

$$\lambda_N = -R_N^{-1}[z_N - x_n^a],$$

Unlike equations (5.1-5.2), the modified adjoint equations depend on x_n^a and P_n^a (Ménard and Daley 1996). The modified adjoint equations are useful in that the standard adjoint equations (5.2-5.3) cannot be used over very long time scales even when the model is perfect. Unfortunately, the analysis error covariance matrix P_n^a is expensive to calculate. Therefore, we approximate the matrix $I - R_n^{-1} P_n^a = \gamma I$ by a simple decay time scale, where $\gamma < 1$. This calculation is essentially free and recognizes that the effect of model error will be to limit the extent to which information propagates in the system.

Our first experiment using this modified adjoint system places a Gaussian cooling anomaly in the central Pacific centered at 180^0W and the Equator. The anomaly itself extends from 140^0E to 140^0W and 20^0S to 20^0N. We then run our GCM forward in time for 1000 days with this cooling anomaly to generate a climate anomaly. Then, the adjoint is integrated with this anomaly introduced

at each time step and run backward in time to determine the sensitivity to that particular climate pattern. In principle, this sensitivity should resemble the imposed Gaussian cooling anomaly.

The response to this cooling is an alternating wave train of 500 hPa geopotential highs and lows emanating from Southeast Asia, arcing into the north Pacific, then arcing back over the west coast of North America into

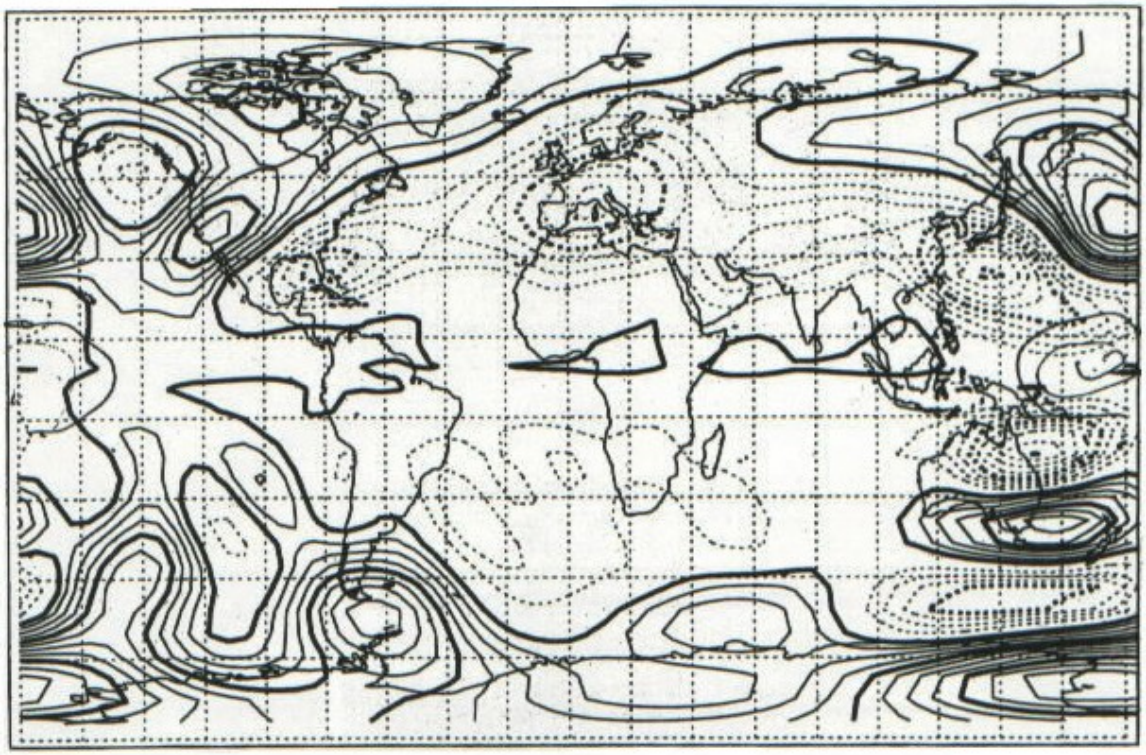

Figure 8. 500mb height response to a La Niña forcing with a maximum cooling of 4^{0}K/day and contour interval of 10m.

California, as shown in Figure 8. Although the wave train itself is not surprising, the fact that it does not emanate from the cooling maximum is surprising, and is indicative of the complexity of this system.

The adjoint model is run backwards in time to generate the sensitivity by summing contributions from the adjoint thermal sensitivity. The sensitivity generated by the adjoint is not markedly different from the imposed forcing. It is centered roughly on the date line, and although it has more structure in meridional direction, with maxima at 10^0N and 10^0S, still has an overall character quite similar to the imposed Gaussian cooling anomaly. In addition, there are heating anomalies to the north and south of the cooling anomaly in the extratropics, as shown in Figure 9. While this obviously differs from the imposed anomaly, it is consistent with an enhanced Hadley cell in this region.

Given the success of this initial experiment at reproducing a forcing anomaly for a given climate anomaly, a second experiment was attempted. In this experiment, the model was integrated forward in time for 1000 days to determine the errors associated with the model as measured by departures from the DJF climate

from the NCEP reanalysis. The model was then run backwards using the adjoint technique described above to determine the forcing that would cancel this climate anomaly. When this forcing is determined, the model can be adjusted to account for the error and run once again to examine if an improvement in the model climate is obtained. Figure 10 shows that as the iterations increase, the empirical climate model error decreases and the model behaves more like the actual observed climate. Not surprisingly, the thermal forcing increments associated with this iteration process seek

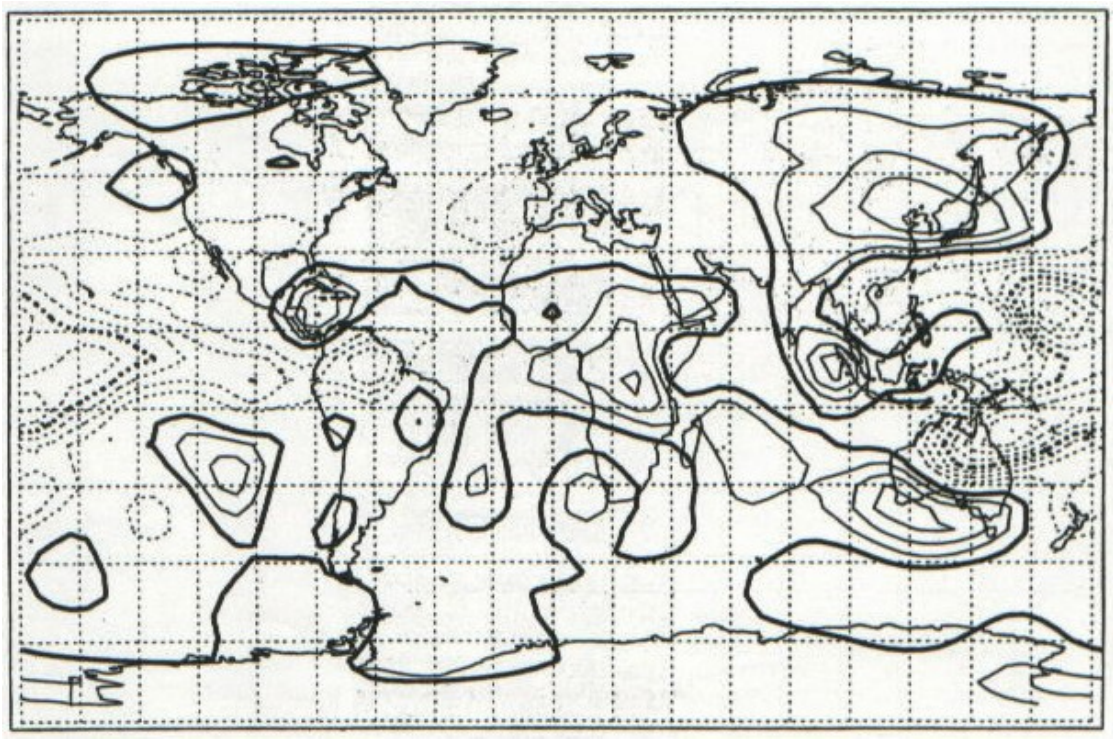

Figure 9. 500mb cooling from adjoint due to La Niña forcing. Contour is arbitrary.

to improve the climate by adjusting the tropical/extratropical heating contrasts, and as such, adjust the positions of the midlatitude jet streams. Figure 11 shows such an increment, showing the model heating contrasts over the subtropics in the vicinity of the storm tracks off the east coasts of Asia and North America. These contrasts act to tighten the jet, lending to a more realistic flow and correcting the too zonal tendency of the base climate shown in Figure 7.

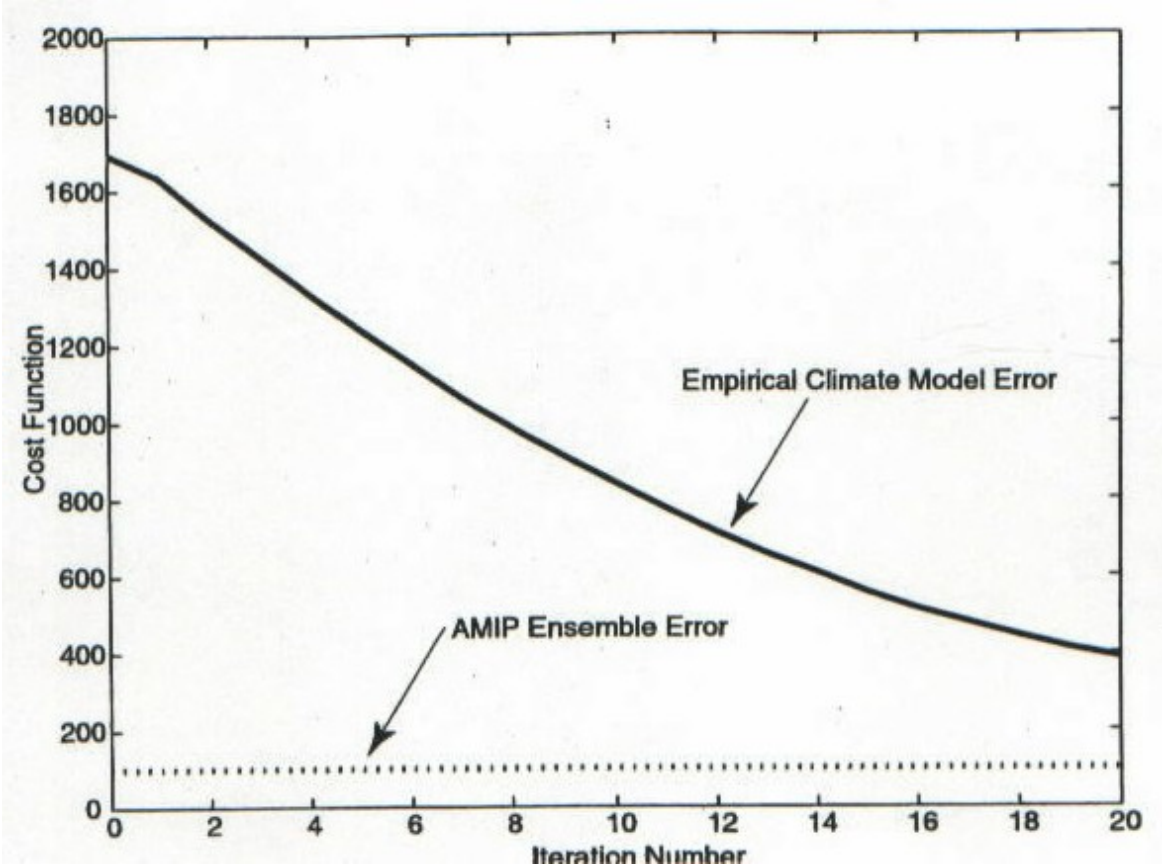

Figure 10. Reduction in the cost function measuring the difference the model's climate and the observed January climate. Also shown for comparison is the approximate error for the AMIP ensemble based upon Gates et al. (1999).

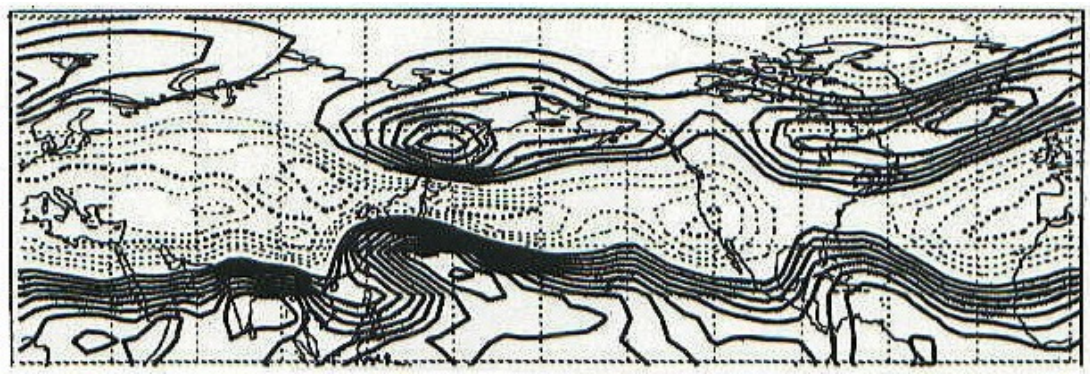

Figure 11. Thermal residual forcing increment on the σ = 0.3 surface. Contour interval is arbitrary, with negative contours dashed.

6. Conclusions

The goal of our research was to create a climate model that could improve its ability to model the atmosphere by self-adjustment of a thermal forcing anomaly. After determining that the central Pacific was the area which created the largest response in our climate model, a cooling similar to La Niña was used to test our ideas. We began the research with the simplest climatic model to determine its weaknesses and allowed the model to self-adjust to correct for its own deficiencies. In principle, this process could be continued until model skill on the order of a GCM was obtained.

One benefit of this approach is that the role of transients, so important in modeling the atmosphere, can be included in this model self-correction implicitly. This is advantageous for a number of reasons, primarily because one does not need a "theory" of transients to correct model errors, as would be the case for a linear or nonlinear stationary wave model (Held et al. 1989).

It is interesting to ponder the question of whether the results we obtained would be useful in correcting current climate models. Specifically, would our results

improve the current climate models to the point where they could forecast the atmosphere's response to El Niño more accurately? Since our models response greatly improved with the introduction of self-correcting forcing, we believe that the current climate models would also benefit from this type of adjustment.

With more time available, a Green's function evaluation on our model would offer us more evidence on how well our model handles the atmosphere's response to a SST anomaly. Further, it would have been interesting to study the model forcing that generate the North Atlantic Oscillation (NAO) pattern, and understand how that pattern affects global weather.

REFERENCES

Branstator, G. 1995: Organization of storm track anomalies by recurring low frequency circulation anomalies. *J. Atmos. Sci.*, **52**, 207-226.

Geisler, J. E., M. L. Blackmon, G. T. Bates, and S Munoz 1985: Sensitivity of January climate response to the magnitude and position of equatorial Pacific sea surface temperature anomalies. *J. Atmos. Sci.*, **42**, 1037-1049.

Giering, R., and T. Kaminski 1996: *Recipes for adjoint code construction*. Report No. 212, Max-Planck-Institut für Meteorologie.

Held, I. M., S. W. Lyons, and S. Nigam 1989: Transients and extra-tropical response to El Niño. *J. Atmos. Sci.*, **46**, 163-174.

Held, I. M., R. T. Pierrehumbert, S. T. Garner, and K. L. Swanson 1995: Surface quasi-geostrophic dynamics. *J. Fluid Mech.*, **282**, 1-20.

Held, I.M., and M. J. Suarez 1994: A proposal for the intercomparison of the dynamical cores of atmospheric general circulation models. *Bull. Amer. Meteor. Soc.*, **75**, 1825-1830.

Holton, J. R., 1992: *An Introduction to Dynamic Meteorology*, Academic Press, 511 pp.

Hoskins, B. J., and D. J. Karoly, 1981: The steady linear response of a spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci.*, **38**, 1179-1196.

Karoly, D. J., 1983: Rossby wave propagation in a barotropic atmosphere. *Dyn. Atmos. Oceans*, **7**, 111-125.

Lau, N.-C., 1988: Variability of the observed midlatitude storm tracks in relation to low-frequency changes in the cirulation pattern. *J. Atmos. Sci.*, **45**, 2718-2743.

Ménard, R., and R. Daley, 1996: The application of Kalman smoother theory to the estimation of 4DVAR error statistics. *Tellus*, **48A**, 221-237.

Philander, S. G. H., 1990: *El Niño, La Niña, and the Southern Oscillation*, Academic Press, 293 pp.

Pitcher, E. J., R. C. Malone, V. Ramanathan, M. L. Blackmon, K. Puri, and W. Bourke, 1983: January and July simulations with a Spectral General Circulation Model. *J. Atmos. Sci.*, **45**, 580-604.

Rassmusson, E. M., and T. H. Carpenter, 1982: Variations in tropical sea surface temperature and surface wind fields associated with the Southern Oscillation/El Niño. *Mon. Wea. Rev.*, **110**, 354-384.

Sardeshmukh, P. D., and B. J. Hoskins, 1988: The generation of global rotational flow by steady idealized tropical divergence, *J. Atmos. Sci.*, **45**, 1228-1251.

Simmons, A. J., J. M. Wallace, and G. Branstator, 1983: Barotropic wave propagation and instability, and atmospheric teleconnection patterns. *J. Atmos. Sci.*, **40**, 1363-1392.

Shukla, J., and J. M. Wallace, 1983: Numeric simulation of the atmospheric response to equatorial Pacific sea surface temperature anomalies. *J. Atmos. Sci.*, **40**, 1613-1629.

Talagrand, O., 1991: *The use of Adjoint Equations in Numeric Modelling of the Atmospheric Circulation*. Laboratoire CNRS 168, Université de Nice, France.

Trenberth, K. E., G. W. Branstator, D. J. Karoly, A. Kumar, N.-C. Lau, and C. Ropelewski, 1998: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures, *J. Geophys. Res.*, **103**, 14291-14324.

Webster, P. J., 1981: Mechanisms determining the atmospheric response to sea surface temperature anomalies. *J. Atmos. Sci.*, **38**, 554-571.

Webster, P. J., 1982: Seasonality in the local and remote atmospheric response to sea surface temperature anomalies, *J. Atmos. Sci.*, **39**, 41-52.