Scientific report; 98 - 03

KNMI

Expermimenting with a similarity measure for atmospheric flows

R.A. Pasmanter and X.L. Wang

De Bilt, 1998

Scientific report = wetenschappelijk rapport; WR 98-03

De Bilt, 1998

PO Box 201
3730 AE De Bilt
Wilhelminalaan 10
De Bilt
The Netherlands
Telephone +31(0)30-220 69 11
Telefax +31(0)30-221 04 07

Authors: R.A. Pasmanter and X.-L. Wang

UDC: 551.511.33
551.511.6
551.55

ISSN: 0169-1651

ISBN: 90-369-2139-2

Experimenting with a similarity measure for atmospheric flows

Internal report
by R.A.Pasmanter and Xue-Li Wang

KNMI
January 1998

Abstract

The data related to four atmospheric pressure-levels during 14 winters are used in order to check a recently introduced measure of the "separation" between two atmospheric states. The fields of variables considered are: horizontal velocities, density and temperature. It is found that this new measure performs, in some respects, better than the geopotential-height or the kinetic-energy measures. In the jet regions, the new measure is almost identical to the kinetic-energy measure.

I. Introduction

In meteorology we are often confronted with the following question: given two atmospheric states how similar are they? how "close" are they to each other? A closely related question would be: given three atmospheric states which pair of states are the more similar to each other? Variants of these questions arise, e.g., when we want to compare model predictions and the actual weather, i.e., in model validation; when we want to calculate transient growth and optimal perturbations; when we try to determine the maximal density of results from Monte-Carlo simulations; in data assimilation, etc.
An atmospheric state or configuration at time t, can be defined by 1) the three-dimensional velocity field $\vec{v}(\vec{R},t)$, 2) the temperature field $T(\vec{R},t)$, 3) the density field $\rho(\vec{R},t)$ and 4) possibly other relevant variables like humidity which we will ignore for the moment; these fields are defined on all positions $\vec{R} \equiv (x,y,z)$. In order to compare different states, we have to determine the differences in density, temperature and velocities. If we had to deal with only one type of variable, e.g., if we could consider only the velocity and ignore the other variables, then it would be easy to determine the "distance" between different states. In order to circumvent the problem, very often the "distance" is computed in this way, e.g., by taking the difference only in geopotential height at each position, squaring and summing over all positions. Since all physical quantities can be derived from the basic fields $(\vec{v},T,\rho)$, it follows that an approach based on them[1] discriminates better than an approach based on only one quantity (like geopotential).
The definition of a "distance" between two states should be such that the results obtained from it are independent of the choice of physical variables used to describe the system; this self-evident prerequisite has often been ignored.

We have proposed a way of computing such a "distance"[2] directly from the dynamically relevant quantities $(\vec{v},T,\rho)$; its formula is given in the next section. We want to gain some experience in its computation and to evaluate the results one obtains from it; ideally, we would also like to find an example which clearly advocates the convenience of using it. In this report, we present the results of our first trials in this direction.

II. Definition of the "distance"

Suppose that at two different times t_1 and t_2, we are given two configurations which are characterized, respectively, by fields $\vec{v}_i(\vec{R}) \equiv \vec{v}(\vec{R},t_i)$, $T_i(\vec{R}) \equiv T(\vec{R},t_i)$ and by $\rho_i(\vec{R}) \equiv \rho(\vec{R},t_i)$ with $i = 1$ and 2. Define $\Delta\rho(\vec{R}) \equiv \rho_2 - \rho_1$, $\Delta T(\vec{R}) \equiv T_2 - T_1$, etc. Using some thermodynamic and statistical considerations, $(\Delta L)^2$, the "distance" squared between these two states can be approximated by

$$(\Delta L)^2 = \int d^3R\,\rho \left[\left(\frac{\Delta\rho}{\rho}\right)^2 + \frac{3}{2}\left(\frac{\Delta T}{T}\right)^2 + \frac{m}{kT}\Delta\vec{v}\cdot\Delta\vec{v} \right] \qquad (1)$$

[1] Or on any system of coordinates related one-to-one to the $(\vec{v},T,\rho)$-system.

[2] The details of the derivation can be found in the Los Alamos e-archives on chaotic dynamics (`http://xyz.lanl.gov`) nr. chao-dyn 9602016.

whenever $\Delta\rho \ll \rho$, $\Delta T \ll T$ and $m\Delta\vec{v}\cdot\Delta\vec{v} \ll kT$ (these inequalities are valid for the cases presented in this report). The quantities ρ, $\Delta\rho$, T, ΔT, $\vec{v}$ and $\Delta\vec{v}$ in (1) are all functions of the position $\vec{R}$. The constants in (1) are: Boltzmann's constant k (1.38066×10^{-23} Joule/Kelvin) and m, the mass of an 'air molecule' ($28.9644\times1.67265\times10^{-27}$ Kgram). Notice that $(\Delta L)^2$ is a dimensionless quantity; its value is independent of the spatial coordinate system **and** of the coordinates system used to describe the dynamical variables.

III. Simplifications

Most available data are presented on isobaric surfaces; this turns out to be convenient also for our calculations. This is because when passing from the space coordinates, $\vec{R}\equiv(x,y,z)$, to pressure coordinates (x,y,p) the following simplifications take place:
1) The volume element times the density, $d^3R\,\rho \equiv dx\,dy\,dz\,\rho$, appearing in the integral (1) becomes $g^{-1}dx\,dy\,dp$ (g being the gravity constant);
2) At fixed pressure, density and temperature are inversely proportional to each other so that the first term in the integral (1) can be approximated as

$$\left(\frac{\Delta\rho}{\rho}\right)^2 \cong \left(\frac{\Delta T}{T}\right)^2,$$

as long as $\Delta\rho \ll \rho$, $\Delta T \ll T$. Then the first two terms in the integral (1) can be combined into one, i.e.,

$$\left(\frac{\Delta\rho}{\rho}\right)^2 + \frac{3}{2}\left(\frac{\Delta T}{T}\right)^2 \cong \frac{5}{2}\left(\frac{\Delta T}{T}\right)^2.$$

We get then

$$(\Delta L)^2 = g^{-1}\int dxdydp\left[\frac{5}{2}\left(\frac{\Delta T}{T}\right)^2 + \frac{m}{kT}\Delta\vec{v}\cdot\Delta\vec{v}\right]. \qquad (2)$$

For the sake of simplicity, we shall call the first term "the temperature contribution".
Notice that now the differences do *not* correspond to same position (x,y,z) at different times but to the same (x,y,p) at different times, i.e., the height coordinate z will, in general, be different at t_1 than that at t_2. In eq. (1), instead of working with $\Delta\rho$ and ΔT we could work with Δp and ΔT, and since $\Delta\rho/\rho = (\Delta p/p) - (\Delta T/T)$, one obtains

$$(\Delta L)^2 = \int d^3R\,\rho\left[\left(\frac{\Delta p}{p}\right)^2 - 2\frac{\Delta p}{p}\frac{\Delta T}{T} + \frac{5}{2}\left(\frac{\Delta T}{T}\right)^2 + \frac{m}{kT}\Delta\vec{v}\cdot\Delta\vec{v}\right]. \qquad (3)$$

When working on surfaces of equal pressure, we are impossing $\Delta p = 0$. It follows then that eq. (2) underestimates the exact expression (3) whenever Δp and ΔT are, on the average, uncorrelated or anti-correlated or correlated but such that

$$0 < 2\left\langle\frac{\Delta p}{p}\frac{\Delta T}{T}\right\rangle < \left\langle\left(\frac{\Delta p}{p}\right)^2\right\rangle,$$

where the pointed brackets indicate an ensemble average. On the other hand, eq. (2) can overestimate (3) only if Δp and ΔT have, on the average, a sufficiently strong positive

correlation such that

$$2\left\langle \frac{\Delta p}{p}\frac{\Delta T}{T}\right\rangle > \left\langle \left(\frac{\Delta p}{p}\right)^2\right\rangle .$$

The first case seems more plausible, i.e., *eq. (2) will, on the average, underestimate the combined contribution of the first two terms in eq. (1).*
For later use, we introduce also the definition of a "distance" in terms of the geopotential height $Z(x,y,p,t)$:

$$(\Delta L)^2_H := g^{-1}\int dxdydp\,[\Delta Z]^2, \tag{4}$$

where $\Delta Z \equiv Z(x,y,p,t_2) - Z(x,y,p,t_1)$. The dimensions of this quantity are mass $\times$ length^2.

All the results presented in this report are computed on fixed pressure levels, i.e., the integral over p in the integrals (2) and (4) were *not* performed. Moreover, the constant of gravity g has been ignored and the results of integrating over the horizontal coordinates (x,y) have been divided by the total area of integration. Consequently, in the results presented below, $(\Delta L)^2$ is dimensionless while $(\Delta L)^2_H$ has dimensions length^2.

IV. Hydrostatic and geostrophic balances

We can express $(\Delta L)^2$ in terms of the gradients of the geopotential $\Phi(x,y,p,t) = gZ(x,y,p,t)$ as follows:
Assuming hydrostatic balance, the temperature T on the p-surface is related to the vertical gradient by $T = -R^{-1}p(\partial\Phi/\partial p)$, therefore

$$\frac{\Delta T}{T} = \frac{\Delta(\partial\Phi/\partial p)}{\partial\Phi/\partial p}.$$

At midlatitudes, assuming geostrophic balance, the (horizontal) velocity is related to the horizontal gradient of the geopotential by $\vec{v} = f^{-1}\vec{z}\wedge\vec{\nabla}\Phi$ where f is the Coriolis parameter and one has

$$\frac{m}{kT}\Delta\vec{v}\cdot\Delta\vec{v} = -\frac{1}{f^2 p}\frac{\partial p}{\partial\Phi}\left|\Delta\vec{\nabla}\Phi\right|^2.$$

Consequently, when these assumptions are valid, a very small value of (2) with the volume integration restricted to the extra-tropical regions means that the corresponding geopotentials have very similar values over the 3D, extra-tropical atmosphere. Similarly, a very small (4) restricted to the extra-tropical areas implies a very small value of (2) restricted to the same extra-tropical area. However, horizontal and vertical gradients of $\Phi(x,y,p,t)$ contribute with different weights to (2); moreover, in tropical regions and whenever deviations from geostrophy are appreciable, it is not possible to relate $(\Delta L)^2$ to the geopotential gradients.

V. Data description

We used ECMWF MARS initialized analysis data; more precisely, the temperature, zonal wind velocity, meridional wind velocity[3] and geopotential height fields on the 850, 500, 250,

[3]The vertical velocity gives a negligible contribution to the integral (1).

100 hPa isobars for 14 Northern winters (December through February) corresponding to the years 1982 through 1995. These datasets contain 4 time steps per day at 00 GMT, 06 GMT, 12 GMT and 18 GMT each.

VI. Test 1

One of the first things we wanted to check was whether the three terms in eq. (1) are, statistically speaking, of the same order of magnitude. We took the fields of the first 15 days in December 1995 at 00 GMT and computed, separately, the two terms in eq. (2) for fields ranging from 6hs up to 360hs after the chosen day for each of these 15 days and for each isobar; finally, we averaged the results over the 15 initial days.
The results are presented in Fig. 1a for the tropical region, in Figs. 1b and 1c for the extra-tropical regions. As expected, the average distance squared grows, approximately monotonically, with the time difference $(t_2 - t_1)$; the results are compatible with a possible saturation being reached around a time difference of approximately 15 to 20 days; however, a systematic, weak trend seems to be present, e.g., the temperature term on the 250 and 100 hPa levels in the extra-tropical regions, see Figs. 1b and 2c[4]. One observes a rapid raise during the first one or two days, later on a slower increase. All this agrees with the theoretical expectations: at very short times ("ballistic" regime), the growth is dominated by the deterministic part of the dynamics and $(\Delta L)^2 \propto (t_2 - t_1)^2$; at later times, the time-correlation ("memory") is lost and the motion acquires a random-walk (or diffusive) character so that $d(\Delta L)^2/dt \propto 2D$ where D is a diffusion-like constant; finally, $|\Delta L|$ reaches the average distance between two arbitrary points on the attractor and saturation sets in.
The "temperature contribution" has approximately the same value on all pressure surfaces. On the lowest and highest pressure surfaces, i.e., on the 850 hPa and 100 hPa pressure surfaces, the square of "temperature term" $\frac{5}{2}\left(\frac{\Delta T}{T}\right)^2$ is comparable to the kinetic energy term $\frac{m}{kT}\Delta\vec{v}\cdot\Delta\vec{v}$. On the other pressure surfaces, the kinetic energy term is appreciably larger than the square of the "temperature term", the largest difference occurs on the 250 hPa pressure surface where the kinetic-energy term is one order of magnitude larger than the temperature term. We interpreted this as a manifestation of large velocity fluctuations in the jets. As it can be seen in Figs. 1a-1c, both contributions to $(\Delta L)^2$ are a factor 1/8-th to 1/6-th smaller in the tropics than in the extratropics, the only exception being the 100 hPa pressure surface where they are approximately 1/2 (kinetic-energy term) to 1/4-th (temperature term) smaller than in the extra-tropics. All the differences between tropical and extra-tropical regions, as well as those between Northern and Southern extra-tropical regions, are compatible with associating stronger fluctuations, in particular velocity fluctuations, with the jet areas.

A similar study was performed using $(\Delta L)^2_H$, the geogeopotential-height definition of the "distance" as in equation (4); the results are shown in Fig. 2. For this computation we used ECMWF MARS initialized analysis of geopotential height on 850, 500, 250 hPa pressure levels for the same period, i.e., December 1995. In this case, tropical squared "distances" are, approximately, twenty times smaller than those in the extratropics; the 250 hPa pressure surface, which strongly overlaps with the jets, gives a squared "distance" six times larger than the 850hPa surface. The most striking differences observed between the geopotential-height

[4]This trend could be due to seasonal effects.

results and those obtained with (2) are: 1) when using the geopotential-height distance $(\Delta L)^2_H$, the lack of saturation even after 17days[5] is particularly evident on the 500 hPa and 250 hPa pressure surfaces in the extra-tropics, 2) in the tropical area, the tidal oscillation is more clearly visible with the geopotential-height definition and 3) the difference between tropical and extra-tropical regions is stronger with the geopotential-height distance $(\Delta L)^2_H$.

VII. Test 2

It was suggested to us that the "distance" $(\Delta L)^2$ could be used in order to search in a long series of atmospheric states for good analogues of a given state, the idea being that two such analogue states will, on the average, evolve rather similarly in time. Following this advice, we undertook the search of analogues, i.e., first we chose a particular state and next we found other states in the long time-series which minimize the $(\Delta L)^2$. For this purpose, we used 14 Northern winter (DJF) seasons, years 1982 through 1995, from the ECMWF MARS daily analysis data on the 850 hPa pressure surface, i.e., the pressure surface on which both contributions to $(\Delta L)^2$ in eq. (2) are comparable. This reduction in the size of the data set, i.e., from four times a day to daily and from four pressure surfaces to one, made possible the time-consuming search. The days chosen as reference days were the 1st of December of each year, alltogether we had 14 reference days; in order to eliminate any short-time correlations, the ten days immediately after each reference were not included in the search. The next Table presents the two best-analogue pairs, i.e., the states with the smallest $(\Delta L)^2$ with respect to the 14 reference days:

Reference day	Smallest $(\Delta L)^2$	Best analogue's date
1 Dec. 1992	1.2072E-03	22 dec. 1987
1 Dec. 1984	1.2543E-03	05 Dec. 1983

Table 1

The smallest $(\Delta L)^2$, 1.207×10^{-3}, was found between the reference day 1 Dec. 1992 and 22 Dec. 1987. The upper row in Fig. 3 shows the geopotential fields while the lower row shows the geopotential anomalies of the fields at 850 hPa corresponding to these two dates; the anomalies have been computed with respect to the 14-year Northern winter mean.
We proceeded then with a similar search, this time based on the geopotential-height $(\Delta L)^2_H$ which led to the following best-analogue pairs:

Reference day	Smallest $(\Delta L)^2_H$ in m^2	Best analogue
1 Dec. 1984	3282.94	23 Feb. 1984
1 Dec. 1983	3297.01	06 Jan. 1995

Table 2

This time, the best analogues were 1 Dec. 1984 and 23 Feb. 1984; their geopotentials and

[5]We extended the maximum time-lag to 20 days and still did not find signs of saturation on the 250 hPa level. We have some indications that this may be due to a sistematic seasonal trend.

anomalies can be seen in the upper and lower rows of Fig. 4, respectively.
In order to assess the quality of the analogues, we then computed the correlation coefficients between the anomalies for each pair of analogues. For the best analogue pair based on $(\Delta L)^2$, the correlation coefficients of the anomalies in temperature, meridional velocity, zonal velocity and geopotential turned out to be:

Variable	Correlation coefficient
T	0.29
U	0.30
V	0.33
Z	0.33

Table 3 : Anomaly correlation coefficients for 01/12/92 *and* 22/12/87.

For the best-analogue pair based on $(\Delta L)^2_H$, the anomaly correlations are listed in Table 4:

Variable	Correlation coefficient
T	0.18
U	0.11
V	0.18
Z	0.31

Table 4 : Anomaly correlation coefficients for 01/12/84 *and* 23/02/84.

For this pair, the correlation of the geopotential anomaly is 0.31, i.e., slightly smaller than it is for the best analogues found using $(\Delta L)^2$; notice that this is no contradiction. Some expectations are confirmed: the geopotential distance $(\Delta L)^2_H$ leads to smaller anomaly-correlations. On the other hand, in Table 6, one finds an example of two states with a small $(\Delta L)^2$ and (relatively) large correlation coefficients for temperature and wind anomalies but with a rather low correlation of geopotential height anomalies.
The correlation coefficients for the best analogues seem to be small. However, if we compare these values with the correlation coefficients between anomalies of an arbitrary day and a reference day, then we realize that they belong to the most extreme correlation values. The probability distributions of the correlation coefficients for temperature, zonal velocity, meridional velocity and geopotential height anomalies are shown in Fig. 5. We see that the total area under the tails corresponding to correlation coefficients larger, in absolute value, than 0.3 is extremely small; therefore, the analogues we have found are, in fact, very special. The time series is not sufficiently long in order to find better analogues with higher correlation coefficients.

Trying to improve on the above results, we looked for hemispheric analogues, the idea being that one should have better chances of finding good analogues by restricting the analysis to a smaller area. The best hemispheric analogues found were:

	Northern Hemisphere	Southern Hemisphere
$(\Delta L)^2$	01/12/92 & 22/12/87	01/12/89 & 11/12/89
Geopotential $(\Delta L)^2_H$	01/12/89 & 25/01/92	01/12/82 & 02/01/90

Table 5 : Dates of best hemispheric analogues.

The corresponding anomaly correlation coefficients were found to be:

		North hemisphere	South hemisphere
	Z	0.15	0.37
$(\Delta L)^2$	T	0.40	0.46
	U	0.32	0.33
	V	0.36	0.42
	Z	0.70	0.29
Geopotential $(\Delta L)^2_H$	T	0.12	-0.21
	U	0.50	0.19
	V	0.36	0.02

Table 6 : Correlation coefficients for hemispheric results.

The correlation coefficients are higher than in the global search but not dramatically so. The NH best geopotential analogue achieves a geopotential-anomaly correlation of 0.70, this is the highest correlation coefficient we have observed. For this pair, also the zonal component U of the horizontal velocity, has a relatively high correlation coefficient while the temperature correlation coefficient is small. Similarly, it is worthwhile notiicng that 1) the best NH analogue pair has a rather low geopotential-anomaly correlation and 2) the best SH geopotential analogue has a *negative* correlation of temperature anomalies and also low velocity correlations. All this is in agreement with the analysis presented in Section IV.

VIII. Conclusions

a) The definition of a "distance" as given by eq. (1) was obtained by considerations about laminar flows. In spite of this, we have found that it gives very reasonable results also in the atmosphere. More specifically:

Even in the jet regions, the kinetic-energy term is, at most (remember that some density fluctuations have been neglected by taking isobaric coordinates), 10 times larger than the other contributions.

The magnitude of the wind and temperature terms are of the same order on the 850 hPa level over the whole globe. The level on which the wind contribution is more dominant is the 250 hPa surface.

b) Extra-tropical regions contribute, per unit area, approximately 5 times the contribution of tropical ones. If one uses the geopotential-height measure then this difference grows to become a factor of approximately 25, i.e., the geopotential measure gives an overwhelming weight to the extra-tropical regions.

c) With both methods, the best global analogues do not have high correlation coefficients, i.e., the time series is not long enough in order to find them. There was only one North-hemispheric best analogue pair with a geopotential-anomaly correlation of 0.70; this case illustrates the fact that a small geopotential $(\Delta L)^2_H$ does not necessarily imply a small ΔL^2.

IX. Points for further consideration or improvement

1) Subtraction of the seasonal trend.
2) Do the calculations not on surfaces of constant pressure but in the space coordinates (x, y, z) (instead of (x, y, p)).
3) The simplest way to modify the distance definition so that it "works" also in the jet areas would be to weigh the kinetic energy term not with kT but with the characteristic velocity variance (a season and location dependent quantity).
4) After completion of these calculations, the distance definition has been extended to include humidity. One should check how the results are affected by the inclusion of this variable, especially in the tropical regions

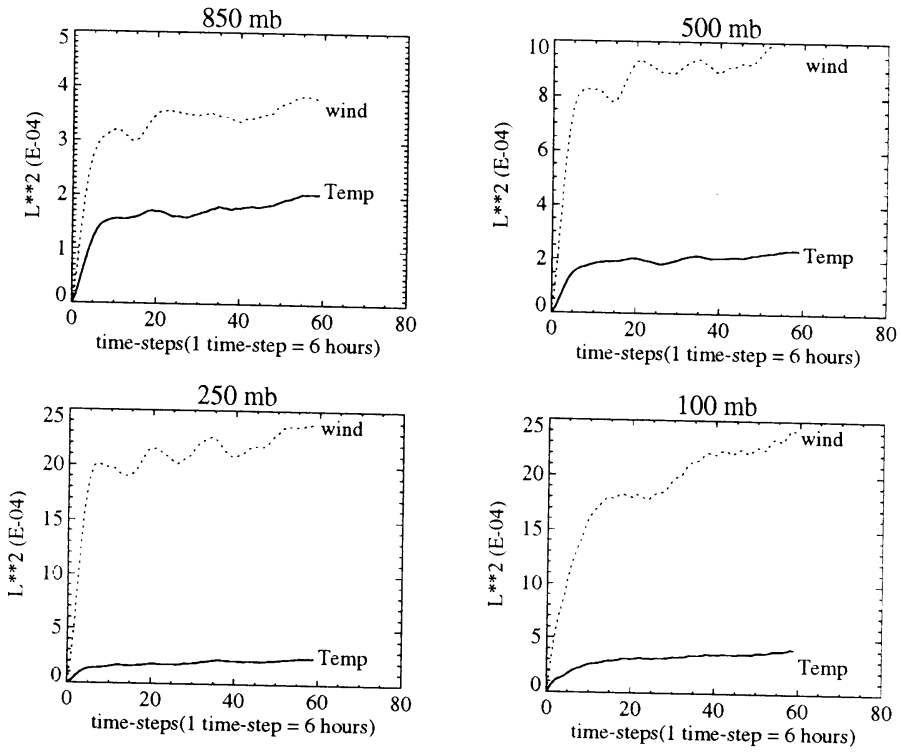

Figure 1a: Wind and temperature term in the tropical area, on the 850, 500, 250, 100 hPa.

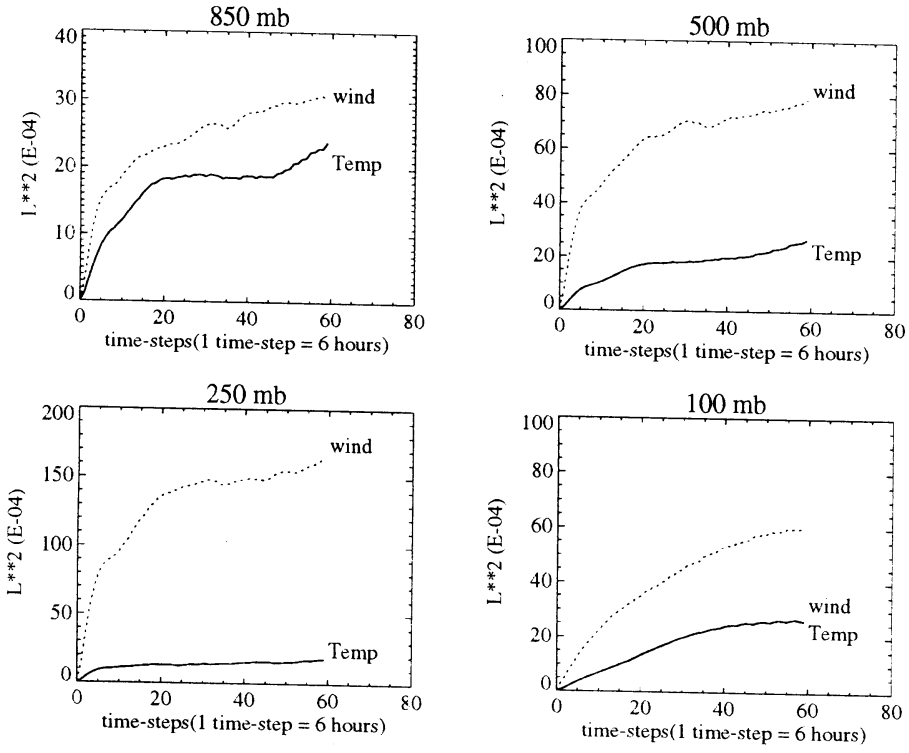

Figure 1b: As in Figure.1a, but for Northern latitudes.

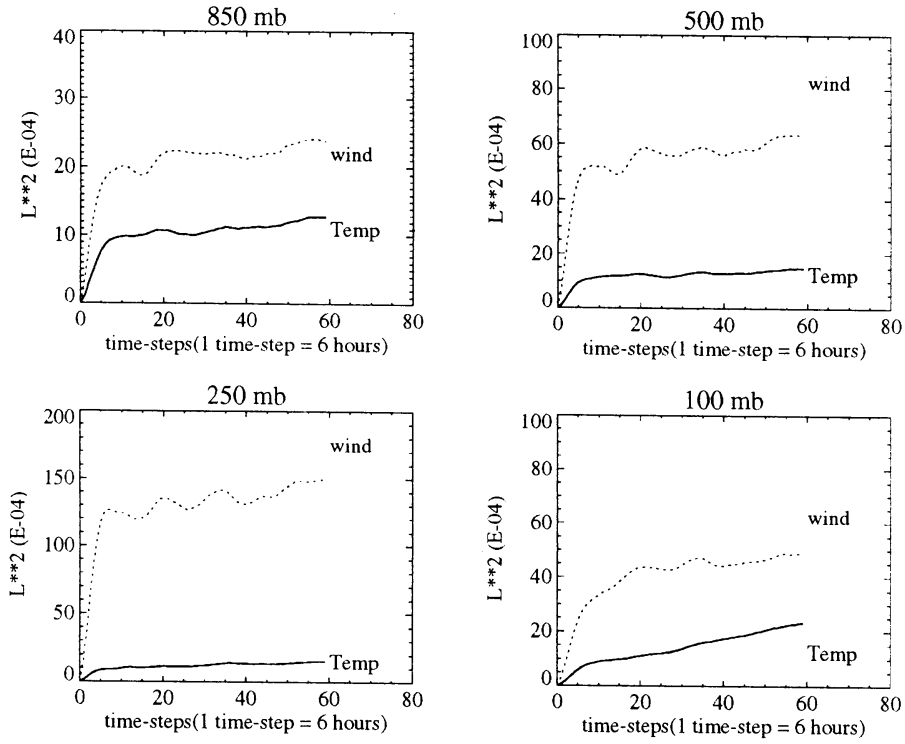

Figure 1c: As in Figure.1b, but for Southern latitudes.

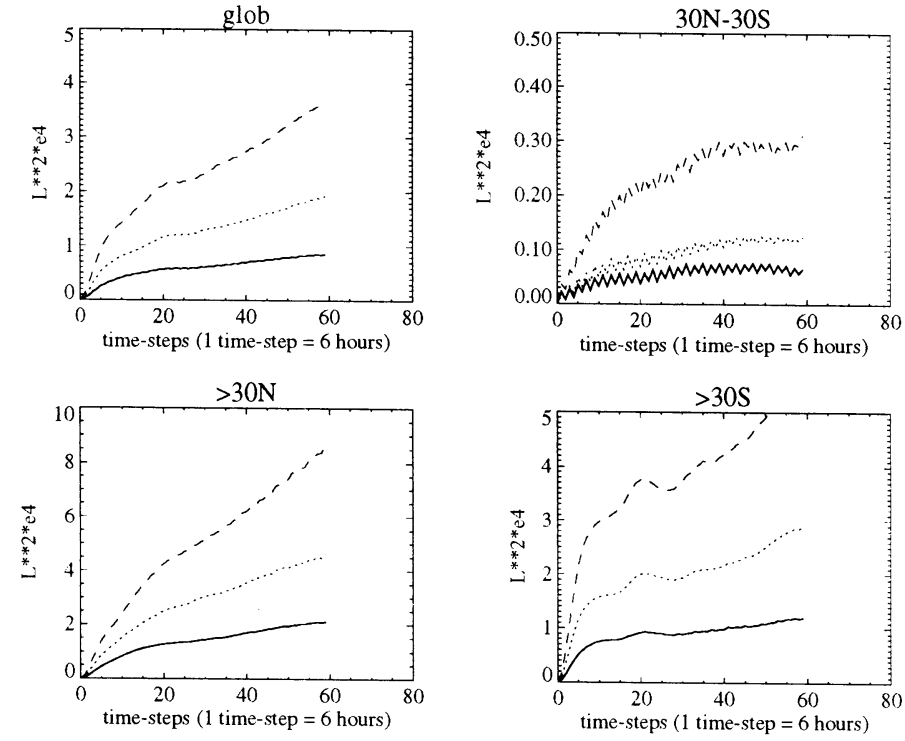

Figure 2: L_H^2 in 10^4 m^2, global, tropical, Southern and Northern in clockwise order. The solid line is for 850 hPa. The dotted line is for 500 hPa. The dashed line is for 250 hPa.

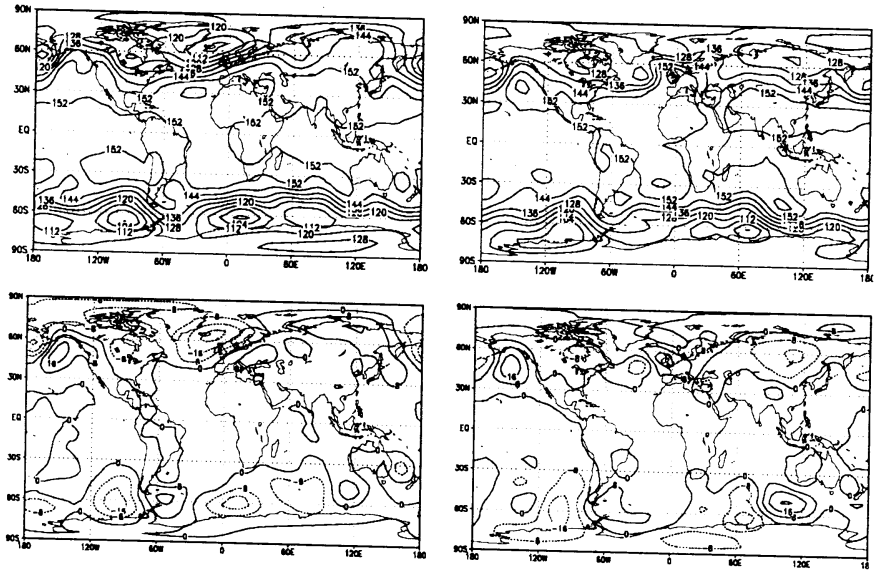

Figure 3: Fields of the best-analogues pair 1 Dec. 1992 - 22 Dec. 1987. Upper row: geopotential heights; lower row: geopotential height anomalies.

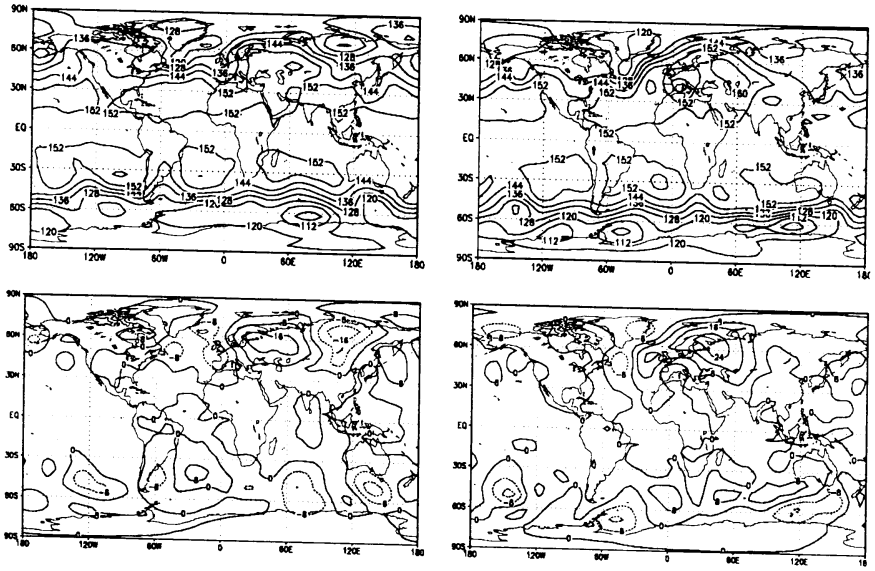

Figure 4: As in Figure.3, for the geopotential best analogue pair, 1 Dec. 1984 - 23 Feb. 1984

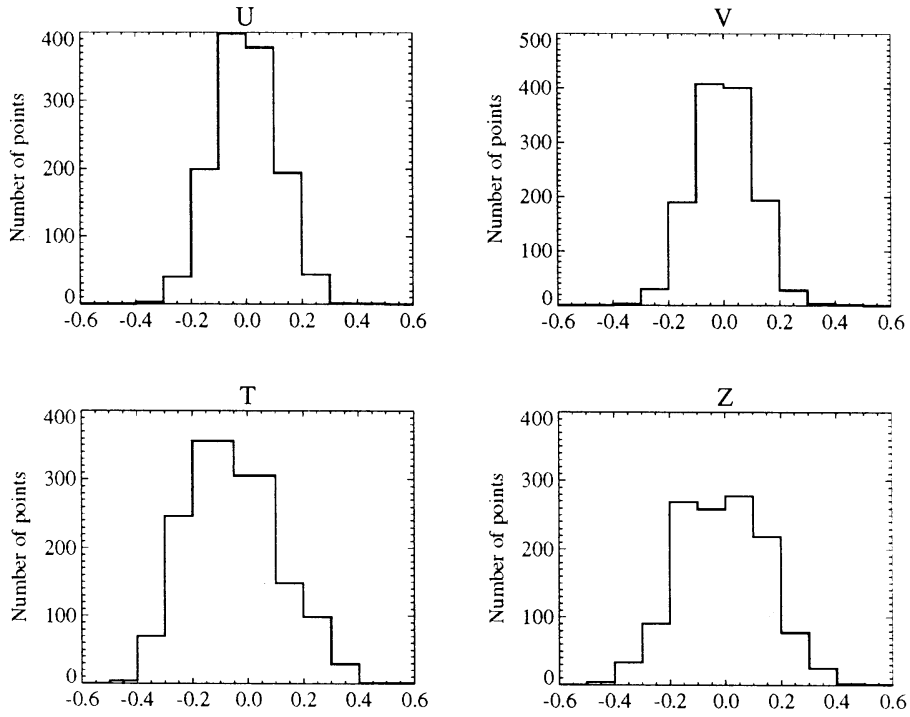

Figure 5: Histograms of the correlation coefficients of U, V, T, Z. The refference day is 01/12/92.