Course 12.812, General Circulation of the Earth's Atmosphere
Prof. Peter Stone
Section 2: Analysis Techniques

Pressure Coordinates:

Most meteorological measurements are made in pressure coordinates, e.g., rawinsondes give T, $\vec{v}$ etc at a given pressure level, measured directly, and therefore it is simplest to carry out any analyses in pressure coordinates. This avoids any errors that would be introduced by trying to calculate the height above sea-level for a given pressure level at a given time, location, etc. The form of the equations is also greatly simplified by the transformation $(x, y, z) \rightarrow (x, y, p),$ provided we can assume hydrostatic equilibrium (H.E.) e.g., in x, y, z coordinates the continuity and zonal momentum equations are:

$$\boxed{\frac{\partial \rho}{\partial t} + \nabla \bullet \left(\rho \vec{v}\right) = 0}$$

and

$$\boxed{\frac{\partial p}{\partial z} = -\rho g}$$

$$\boxed{\rho \frac{du}{dt} = \rho\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z}\right) = \rho f v - \frac{\partial p}{\partial x} + \rho F_x}$$

(neglecting the coriolis term associated with w), etc.

Consider an element of mass δM

$= \rho \delta x \delta y \delta z$. If we assume hydrostatic equilibrium, then $\delta p = -\rho g \delta z$

$\therefore \delta M = \dfrac{-\delta x \delta y \delta p}{g}$ is conserved;

$$\therefore \frac{d}{dt}\left(\delta M\right) = 0 = \frac{d}{dt}\left(\delta x \delta y \delta p\right) = \delta y \delta p \frac{d}{dt}\left(\delta x\right) + \delta x \delta p \frac{d}{dt}\left(\delta y\right) + \delta x \delta y \frac{d}{dt}\left(\delta p\right) = 0;$$

$\dfrac{d}{dt}\left(\delta x\right) = \delta \dfrac{dx}{dt} = \delta u,$ etc.

let $\omega = \frac{dp}{dt}$; $\therefore$ in pressure coordinates

dividing by $\delta x \delta y \delta p$ we have

$$\frac{\delta u}{\delta x} + \frac{\delta v}{\delta y} + \frac{\delta \omega}{\delta p} = 0,$$

Or, taking the limit,

$$\boxed{\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial \omega}{\partial p} = 0}$$

To transform the momentum equation, we note that

$$p(x,z) = \text{constant} \Rightarrow$$
$$dp = \left(\frac{\partial p}{\partial x}\right)_z dx + \left(\frac{\partial p}{\partial z}\right)_x dz = 0$$

$$\therefore \left(\frac{\partial p}{\partial x}\right)_z = \frac{-\left(\partial z / \partial x\right)_p}{\left(\partial z / \partial p\right)_x} = -\left(\frac{\partial p}{\partial z}\right)_x \left(\frac{\partial z}{\partial x}\right)_p$$

$$= \rho g \left(\frac{\partial z}{\partial x}\right)_p, \text{ etc}$$

Substituting into the u equation, dividing by ρ, and noting that x, y, p are independent, we have

$$\frac{du}{dt} = \frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + \omega\frac{\partial u}{\partial p} = \frac{\partial u}{\partial t} + \nabla \cdot (u\vec{v})$$

$$= fv - g\frac{\partial Z}{\partial x} + F_x,$$ where now it is understood

that $\vec{v} = (u, v, \omega)$, $\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial p}\right)$, and Z = z = height of the pressure level.

Similarly for the v equation, and hydrostatic equilibrium becomes

$$\frac{\partial Z}{\partial p} = -\frac{1}{\rho g} = -\frac{RT}{pg}$$

Note that the triple product terms in the equations in rectangular coordinates (e.g. ρu^2) are replaced by quadratic terms (e.g. u^2) and $\therefore$ in our analyses we only have to worry about double correlations, not triple correlations.

For future use, we note that in spherical coordinates:

$\phi =$ latitude,
$\lambda =$ longitude,
r= radius = a+z, z«a;

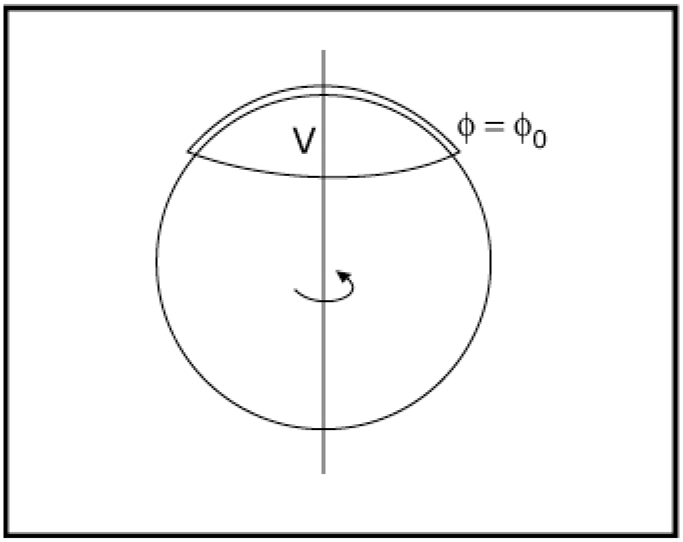
$u = a\cos\phi\,\dot{\lambda}$,

$v = a\dot{\phi}$,

$\omega = \dot{p}$, and continuity is

$$\frac{1}{a\cos\phi}\frac{\partial u}{\partial \lambda} + \frac{1}{a\cos\phi}\frac{\partial}{\partial \phi}\left(v\cos\phi\right) + \frac{\partial \omega}{\partial p} = 0$$

For another example of the convenience of p - coordinates, consider the mass of the atmosphere contained in a polar cap with volume V bounded by a wall at latitude ϕ_0.



The mass contained within the cap is $\int_V \rho dV$, and its time changes are

$$\frac{\partial}{\partial t}\int \rho dV = \int \frac{\partial \rho}{\partial t} dV = \int -\nabla \cdot \rho\vec{v} dV = -\int \rho\vec{v} \cdot \hat{n}\, dA$$ (by Gauss' theorem)

where now the integral is carried out over the latitude wall. The area integral of course just represents the flow of mass out of the polar cap. In the long term mean there is no net accumulation of mass (neglecting volcanic eruptions, etc.) Thus
P_0= surface pressure is independent of t and

$$\int \rho\vec{v} \cdot \hat{n}\, dA = 0 = \int_0^\infty dz \int_0^{2\pi} \overline{\rho v} a \cos\phi\, d\lambda = \frac{a \cos\phi}{g} \int_0^{P_0} dP \int_0^{2\pi} \bar{v}\, d\lambda = 0$$

where $(\bar{u}) = \frac{1}{\tau}\int_0^{\tau} (u)\, dt$

i.e. there is no net mass flux across a latitude circle. On a monthly time scale there can be small changes in the mass. These can be simply assessed from the surface pressure, P_0, since the

mass in a column of the atmosphere = $\int_0^\infty \rho dz \cong \frac{1}{g}\int_0^{P_0} dP = \frac{P_0}{g}$

Thanks to hydrostatic equilibrium seasonal (3 month mean) changes in P_0 are $\lesssim 20$ mb ~ 2% of total mass. Corresponding vertical mean velocities are $\lesssim 0.04$ m / s, not entirely negligible. However for stationary states, such as seasonal extremes and annual means, the mass flux is essentially zero. This result is built into the continuity equation in pressure coordinates. Take the time and zonal mean $[\bar{\ }]$ of the equation:

$$\therefore 0 + \frac{1}{a\cos\phi}\frac{\partial}{\partial\phi}\cos\phi\left[\overline{v}\right] + 0 - \cancelto{0}{\overline{\omega}_s} = 0;$$

$$\therefore \ \cos\phi\left[\overline{v}\right] = \text{constant} \to 0 \ \text{(boundary condition at pole)}$$

Statistics of the meteorological fields:

In this course we are not interested in weather, but rather in the "climate" of the atmosphere, i.e., in the average state over some time, i.e.

$$\overline{A} = \frac{1}{\tau}\int_0^{\tau} A\,dt\,,$$

where τ might be a season, a year, or some multi-year period. The fluctuating part of the field is defined as

$$A' = A - \overline{A}$$

and thus by definition, $\overline{A'} = 0$. This does not however mean that the fluctuating part of the field is irrelevant to climate. For example, if we are interested in kinetic energy per unit mass, $\frac{1}{2}\left(u^2 + v^2 + w^2\right)$, the time mean of one component would be

$$\frac{1}{2}\overline{u^2} = \frac{1}{2}\overline{\left(\overline{u} + u'\right)\left(\overline{u} + u'\right)} = \frac{1}{2}\left(\overline{u}^2 + \overline{u'^2}\right)$$

i.e., the fluctuating part has a time mean energy.

Climate generally varies more strongly with latitude than with longitude (as we will see) and therefore much attention has been devoted to the zonal mean climate and how it is maintained. In this course we will also be focusing on the zonal mean state. Thus we also define a zonal mean and its deviation,

$$\left[A\right] = \frac{1}{2\pi}\int_0^{2\pi} A\,d\lambda$$

$$A^* = A - \left[A\right], \qquad \left[A^*\right] = 0.$$

We also define covariances in time:

$\overline{AB} = \overline{A}\,\overline{B} + \overline{A'B'}$ and in space:

$$[AB] = [A][B] + [A^*B^*]$$

We can also decompose a field in both time and space (here longitudinally). The result depends on whether we average in time or zonally first. As an example, consider the covariance of v and T (which is a meridional transport of sensible heat, if multiplied by c_p.)

$$[vT] = [v][T] + [v^*T^*] = [v][T] + [v]'[T]' + [v^*T^*]$$

= transport by the steady mean meridional circulation
+ transport by the transient mean meridional circulation
+ transport by the spatial eddy circulation;

Alternatively, we have

$$[\overline{vT}] = [\bar{v}\,\bar{T}] + [\overline{v'T'}] = [\bar{v}][\bar{T}] + [\bar{v}^*\bar{T}^*] + [\overline{v'T'}]$$

Noting that

$$[v^*T^*] = [\bar{v}^*\bar{T}^*] + [v^{*\prime}T^{*\prime}]$$

and

$$[v'T'] = [v]'[T]' + [v'^*T'^*]$$

we see that the two results are identical.

Thus one needs to be careful in choosing the order of averaging; e.g., in the latter decomposition, the "transient eddy" term includes a contribution from fluctuations in the mean meridional circulation. "Eddies" can refer to time variations or zonal variations.

Also note that the division between the two eddy terms, $[\bar{v}^*\bar{T}^*]$ and $[\overline{v'T'}]$, depends on τ. If τ is very short so the motions are essentially steady over the interval τ, $[\overline{v'T'}] \to 0$. But as τ increases, if v* and T* are fluctuating, then some of the "stationary" eddy component will appear as transient eddies.

Let us consider a simple example, calculating the contributions to linear momentum transport.

$$\left[\overline{uv}\right] = \left[\bar{u}\right]\left[\bar{v}\right] + \left[\bar{u}^* \bar{v}^*\right] + \left[\overline{u'v'}\right]$$

Stations A and B are 180° longitude apart, and are at the same pressure level.

	Station A		Station B			
	u	v	u	v	[u]	[v]
Day 1	0	-2	6	6	3	2
Day 2	2	-4	4	8	3	2
Time mean	**1**	**-3**	**5**	**7**	**3**	**2**

By direct calculation:

$$\left[\overline{uv}\right] = \frac{0 - 8 + 36 + 32}{4} = 15$$

and

$$\left[\bar{u}\right]\left[\bar{v}\right] = 3 \times 2 = 6$$

Now calculate the eddy components:

	Station A		Station B	
	u*	v*	u*	v*
Day 1	-3	-4	3	4
Day 2	-1	-6	1	6
Time mean	**-2**	**-5**	**2**	**5**

$$\therefore \left[\bar{u}^* \bar{v}^*\right] = \frac{10 + 10}{2} = 10$$

$\therefore$ the transient eddy component should be -1:

	Station A		Station B	
	u’	v’	u’	v’
Day 1	-1	1	1	-1
Day 2	1	-1	-1	1
$\overline{u'v'}$	-1		-1	

$\therefore \left[\overline{u'v'}\right] = -1$

QED

Analysis techniques I: Rawinsonde-based

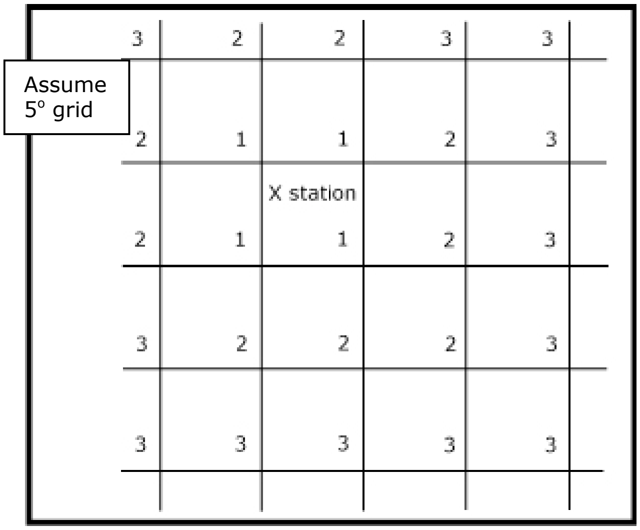
In order to calculate statistics like those illustrated above, the observations need to be analyzed, i.e., put on a grid suitable for easy mathematical manipulation. For many decades the primary source of upper air data was rawinsondes, and most of the early, classical analyses, e.g., those by Starr and his students, were based on these. In particular, most of the analyses presented in Peixoto and Oort use a particular technique to grid the rawinsonde network plus surface station data. Many of these analyses which we will look at have never been updated using the more modern data assimilation techniques which we will also describe later. Thus it is worth spending some time discussing the technique used by Peixoto and Oort and its limitations.

Most of these analyses are based on the data in the MIT general circulation (G.C.) library from the 10 year period from May 1963 through April 1973. Figure 5.3 in Peixoto and Oort (1992) shows the network of rawinsonde stations which were the main data source. (The distribution of surface stations is very similar). Ship data was also used, but this seems to have destorted the 1000mb analyses. We note that there are substantial areas where the data is sparse, particularly over the oceans and the Southern Hemisphere. In these regions the grided analyses will necessarily be dependent on the analysis technique. Also the measurements are not homogeneous - different instruments are used in different countries, and from time to time changes are made in the instruments and in how the data are recorded. Table 1 lists the changes that have affected rawinsonde data in the US during 1960-73. Note particularly that during the 10 year period of the Peixoto and Oort analyses there were changes that affected the humidity measurements. Also many stations did not report for all 10 years.

Table 1. Chronology of changes in U.S. radiosonde

Date	Change
1960	Introduced white-coated temperature elements.
1965	Carbon humidity element, began reporting low humidities.
1972	Redesigned humidity ducts to reduce solar effects.
1973	Stopped reporting relative humidity below 20%.

After some quality control on the observations, the observations are converted to a grid in the following way (described in Rosen et al. (1979)). First an initial guess for the field is picked, generally a pre-existing climatology, often just a zonal mean climatology. Then



the grid points are divided into 3 classes. The first class points are the points adjacent to an observing station. The initial guess values of the four surrounding points are interpolated to the station location (weighted inversely by distance) and compared to the observed value. The difference is then added to the initial guess at all class 1 points. If there is more than one nearby station, then differences are calculated for all stations, and the average difference is added to the class 1 points. 2nd class points are those that do not have a nearby station but are within a specified distance of a class 1 point, generally about 5°. The same correction is applied to these points as to the nearest class 1 points. Class 3 points are farther away from any stations. The values at these points are chosen to satisfy a Poisson equation, e.g. for temperature, T(x, y),

$$\nabla^2 T(x, y) = F(x, y),$$

written in finite difference form. The forcing function is chosen to be the Laplacian of the initial guess field, $F(x, y) = \nabla^2 T_0(x, y)$, and the boundary conditions are the values of T at the 2nd class points. In effect this means that in data sparse regions the curvature of the initial guess fields is preserved. Since the latter are climatologies, this generally means that eddies are likely to be missed out in data sparse regions. As a final step, the analyzed fields are smoothed to eliminate noise.

Note that all the analyses based on the MIT G.C. library include two non-standard levels, 950 and 900mb as well as 1000, 850, 700, 500, 400, 300, 200, 100, 50.

Also note that in a stationary state mass is conserved and thus the vertically integrated mass flux must equal zero:

$$\int_0^{P}\left(\frac{\partial \bar{u}}{\partial x}+\frac{\partial \bar{v}}{\partial y}\right)dp = -\int_0^{P}\frac{\partial \bar{\omega}}{\partial p}\,dp = \bar{\omega}\left(p_s\right) = \overline{\frac{\partial P_s}{\partial t}} = 0.$$

This would be the case for the annual mean or seasonal extremes. However, because of observational error when v is calculated from observations this is not generally so. Thus an additional correction has to be made to guarantee mass conservation.

In the case of analyses carried out by Starr and his students in the 1960's and 1970's, they assumed vertically integrated divergences exceeding $6\times10^{-5}\mathrm{s}^{-1}$ were unphysical, i.e., at grid points where

$$\frac{1}{p_s}\left|\int_0^{p}\left(\frac{\partial\bar{u}}{\partial x}+\frac{\partial\bar{v}}{\partial y}\right)dp\right| > 6\times10^{-5}\mathrm{s}^{-1}$$

$\bar{u}$ and $\bar{v}$ (the monthly means) were adjusted so that the integral was zero. At other points where the divergences were smaller the values were accepted, and thus mass was still not completely conserved, e.g., see Table A3 in Oort and Peixoto (1983) with the results for $\left[\bar{v}\right]$ calculated this way. Note that the results are particularly bad at mid and high latitudes where the vertical mean of $\left[\bar{v}\right]$ is the same magnitude as $\left[\bar{v}\right]$ at individual levels. However at low latitudes, the imbalances are not so bad. Because of this problem Oort and Peixoto developed another balancing technique based on angular momentum conservation. This was used for most of the later analyses using the MIT general circulation library, particularly in Oort & Peixoto (1983) and most of Peixoto & Oort (1992). In particular, because of the large errors shown in Table A3 in Oort & Peixoto (1983) and Figure 6 in Oort (1978), they adopted the following method. For levels outside the boundary layer, (p < 875 mb), they retained the analysis described above for $[\bar{v}]$ for $20\mathrm{S}\le\phi\le20\mathrm{N}$, but for higher latitudes ($|\phi|\geq10^{o}$) an "indirect" method was used, with $[\bar{v}]$ being calculated from angular momentum balance. (The equation will be derived in the next lecture.) Between 10^{o} and 20^{o} a weighted average of the two methods was used.

In the boundary layer (p > 875 mb) they assumed $[\bar{v}]$ to be independent of height and calculated, the constant value so that mass was conserved, i.e.,

$$\int_0^{p_s}[\bar{v}]\,dp=\int_0^{875}[\bar{v}]_{anal}dp+\int_{875}^{p_s}v_0dp=0\Rightarrow v_0$$

<u>**Errors due to sparseness**</u>

One way of estimating the errors due to sparseness was carried out by Oort (1978). He took simulations with the GFDL AGCM, (11 σ levels $\left(\sigma=\dfrac{p-p_s}{p_T-p_s}\right)$, $2\frac{1}{2}^{\circ}$ resolution, two year simulation, with the first year discarded) for January and July, and used the output to test the adequacy of the Rawinsonde network. Note that errors in the model simulation are not an issue here. There are quantitative errors in the model simulation, compared to observation, but qualitatively the simulations are reasonable, and thus the model output is taken as "truth". The gridded model outputs were interpolated to the location of the rawinsonde stations, and then these interpolated values were used to produce a new gridded analysis using the technique described above. Then various statistics were calculated from the original GCM data (which is known exactly) and from the hypothetical rawinsonde analysis and compared. Results are shown in Figures 5, 6 and 12 in Oort (1978).

Figure 5 in Oort (1978) shows the results for $\left[\bar{u}\right]$. Note the jet stream. Here the analysis is reasonably good -- typical errors are 0 (10%). Figure 6 in Oort (1978) shows the result for the stream function for the meridional overturning circulation, ψ, conventionally defined as follows:

The continuity equation in spherical coordinates is:

a= earth's radius

$$\frac{1}{a\cos\phi}\frac{\partial u}{\partial\lambda}+\frac{1}{a\cos\phi}\frac{\partial\left(v\cos\phi\right)}{\partial\phi}+\frac{\partial\omega}{\partial p}=0\,;$$

Average zonally: $\frac{1}{a\cos\phi}\frac{\partial\left(\left[v\right]\cos\phi\right)}{\partial\phi}+\frac{\partial\left[\omega\right]}{\partial p}=0;$

Thus [v] and [ω] can be described by a stream function. Conventionally this is taken to be the mass stream function, with ψ defined to be zero at the top of the atmosphere:

$$\psi=\int_{z}^{\infty}dz\int_{0}^{2\pi}\left(\rho v\right)a\cos\phi\,d\lambda$$

$$=\frac{1}{g}\int_{0}^{p}dp\int_{0}^{2\pi}a\cos\phi\,v d\lambda=\frac{2\pi}{g}\int_{0}^{p}dp\,a\cos\phi\left[v\right]$$

$$\therefore\frac{\partial\psi}{\partial p}=\frac{2\pi a\cos\phi}{g}\left[v\right]$$

$$\therefore\frac{\partial\left[\omega\right]}{\partial p}=-\frac{1}{a\cos\phi}\frac{g}{2\pi a}\frac{\partial^2\psi}{\partial p\,\partial\phi}$$

Using the B.C. [ω] = 0 at p = 0, we have

$$\therefore\left[v\right]=\frac{g}{2\pi a\cos\phi}\frac{\partial\psi}{\partial p};\quad\left[\omega\right]=-\frac{g}{2\pi a^2\cos\phi}\frac{\partial\psi}{\partial\phi}$$

ψ calculated from the rawinsonde analysis is shown in Figure 6 in Oort (1978). Note however that Oort in this paper did not use the normal sign convention for ψ. We see that for this field the analysis in the Southern Hemisphere is completely unreliable and even in the Northern Hemisphere errors about 25% occur. This is basically, because $v \ll u$, so the relative errors are larger. Note that ω is too small to observe, and in practice ψ is always calculated from v.

Figure 12 in Oort (1978) shows the results for the different components of the transport of linear zonal momentum, integrated vertically. Again we see that the Southern Hemisphere results are unreliable. The Northern Hemisphere is better, but still the stationary eddy component seems unreliable.

Analysis Techniques II: Data Assimilation:

In this technique in effect an atmospheric general circulation model (AGCM) is used to fill in the data-sparse regions. Again an initial guess is needed, but in this case the initial guess is a forecast (usually a 6-hour forecast) from an earlier state. The forecast is corrected using all the observations available from the intervening period. At a given gridpoint, a number of observations, typically 5 to 10, that are best for that gridpoint, in terms of time and location, are used to correct the prediction at that gridpoint. Generally an optimum interpolation scheme is used. If A is any scalar, then the forecast value, A_p, is corrected to find the new adjusted analysis value, A_a, by using the difference between the selected observations, $A_{o,i}$, $i = 1 \ldots N$, and the predicted observations with appropriate weights, w_i:

$$\frac{A_a - A_p}{E_p} = \sum_{n=1}^{N} w_i \frac{A_{a,i} - A_{p,i}}{E_{p,i}} \ .$$

The E_p's are the root mean square errors of the predictions, estimated from the errors in the previous analysis. The weights are chosen so as to minimize the root mean square difference over the whole grid between the new analysis and the observations. Normalizing by the estimated errors gives the strongest weight to the observations where the predictions are most accurate, and therefore the error most significant. In this analysis, in data sparse regions the model's own forecast is kept. In addition to rawinsonde data, ship, plane, and satellite data may be assimilated. A balance constraint (geostrophic or otherwise) may be enforced in the analyzed corrections. In practice the ECMWF analyses are pretty well balanced, but the NCEP analyses are not. A 6-hour assimilation cycle is generally used. The resolution is higher than in the earlier analyses, and continually increasing. See Bengtsson et al. (1982) and Rabier et al (2000a, b, c).

Recently 4D data assimilation techniques have been adopted. In these the difference between the observation and the model forecast are minimized over a time interval (generally 6 hours) rather than at a given time.

Due to the huge computational requirements for doing this, the minimization, i.e., the calculation of the weights, is generally done with a lower resolution model with simplified physics, rather than the high resolution, best physics model used for the forecasts.

Importance of balancing

Consider the following example based on data from Oort & Peixoto (1983) for the annual mean at 45N.

Let $\frac{1}{p_s}\int_0^{p_s}[\overline{v}]\,dp = v_0$ and assume the data is not balanced, i.e., $v_0 \neq 0$.

Consider $\frac{1}{p_s}\int_0^{p_s}[\overline{v}][\overline{T}]dp = H =$ mean circulation transport .

The contribution to H by v_0 is just $\frac{v_0}{p_s}\int_0^{p_s}[\overline{T}]\,dp = v_0 T_0$ where $T_0 =$ mean temperature. In this example without balancing $v_0 = -0.251$ m/s and $T_0 = 248.9$ K, and thus the contribution to H by v_0 is –62.5 (m/s)K. The total H calculated without balancing is –58.6 (m/s)K. Thus the transport is totally dominated by the contribution from the unbalanced wind. One simple way to correct for v_0 is to simply subtract v_0T_0 from the calculated H. This would give H = 3.9 (m/s)K. This is equivalent to subtracting v_0 from $[\overline{v}]$ at each level, a method that is sometimes used (Rosen et al, 1985). When $[\overline{v}]$ is calculated using Oort & Peixoto's "indirect" [v] the result is H = 3.4, a difference of about 15%. The difference is an indication of the error remaining in H after [v] has been balanced. Note however that the vertical distributions of $[\overline{v}]$ calculated by the different methods are quite different as shown in the table. Note that the total annual mean eddy transport at 45N in Peixoto and Oort's same analysis is 9.4 (m/s)K.

Annual Mean at 45N

level	$[\overline{v}]_{dir}$	$[\overline{v}]_{ind}$	$[\overline{v}]_{dir} - v_0$
50	0.1	0	0.35
100	-0.1	0	0.15
200	-0.7	-0.2	-0.45
300	-0.6	-0.2	-0.35
500	-0.4	-0.1	-0.15
700	-0.2	0	0.05
850	0.2	0	0.45
1000	0	0.7	0.25
v_0 (m/s):	-0.25	0	0

H {(m/s)K}: -58.6 3.4 3.9

Comparisons of Analyses:

Rosen et al (1985) compared results using the station based analyses with data assimilation analyses. The results are shown in Figures 2, 6, 8, and 11 in Rosen et al (1985). Note the analyses are based on data for 1979 from the First GARP Global Experiment, when conventional data was supplemented (by special efforts) from other sources (ships, aircrafts, satellites).

Mean zonal wind, $[\overline{u}]$, for January and June (Figure 2 in Rosen et al (1985)). The station analysis was carried out for two different initial conditions, a climatology and a zonal mean. The latter shows stronger [u], but the comparison with the data assimilation result doesn't favor either I.C. However, the differences are not large, ~ 3 m/s, 0 (10%).

Mean meridional overturning circulation, $[\overline{\psi}]$, (Figures 6a and 6b in Rosen et al (1985)). Note that all the analyses had to be balanced. They simply subtracted the mean from each level. In the station analyses the different initial guess introduces a somewhat larger difference than in [u], ~ 20% in January. Here Rosen et al compared with two different data assimilation schemes, GFDL and ECMWF. Those two schemes gave very different results, although they bracket the station based analyses. The differences in July (Figure 6b in Rosen et al (1985)) are considerably larger in %, because the circulations are much weaker. Note that ECMWF omitted the 950/900 mb levels in their analysis. GFDL did not. Thus the GFDL analysis should be superior. GFDL had 12 levels, ECMWF had 9 levels.

Eddy statistics (Figures 8 and 11 in Rosen et al (1985)). Here the initial condition was zero. Thus the Poisson equation reduced to a Laplace equation, $\nabla^2 A = 0$, so the station data are just linearly interpolated in data sparse areas. Figure 8 in Rosen et al (1985) shows the results for $[\overline{u'v'}]$. We note that the station analysis gives larger values. Recall that sparseness in the Northern Hemisphere did not seem to have much effect on $\int [\overline{u'v'}]dp$ in Oort (1978). Also note that GCM expts indicate that the resolution of the GFDL model $(1\frac{7}{8}^{\circ})$ should be adequate to resolve $[\overline{u'v'}]$. Thus the reason for the difference between the $[\overline{u'v'}]$ analyses is obscure. Figure 11 in Rosen et al (1985) shows the comparison for $[\overline{v'T'}]$. The agreement here seems much better, although there is some discrepancy in the upper troposphere where the GFDL analysis gives larger $[\overline{v'T'}]$. We shall revisit this problem later in the term when we discuss the global heat balance.

More recent data assimilation schemes are presumably better than the ones used in the Rosen et al (1985) analysis, e.g., they now typically have ~30 levels in the vertical and ~1° horizontal resolution. However, they are still subject to model error, and we shall also revisit this later. Many people in the field assume that the modern data assimilation techniques are superior, but it is not obvious that this is so, and there is no obvious way to test it. They can be tested in numerical forecast models, and forecasts are improving. However this does not assure the credibility of the results in data sparse areas where forecast verification is largely absent. In any case, in general the analyses that have been based on modern data assimilation methods do not have the interesting detail that the earlier station-based analyses have, as represented by Peixoto and Oort's book. Where possible, we will compare.

Re-analyses: Until recently, attempts to use data from data assimilation techniques were plagued by the fact that NWP centers were continually upgrading their models (faster computers $\rightarrow$ higher resolution, etc.), and therefore the data from year to year were not calculated in a consistent way. In recent years NCEP and ECMWF froze their analysis techniques, and then went back and re-assimilated and re-analyzed all the data so as to produce homogeneous data sets. One can now access the NCEP/NCAR re-analyses and ECMWF re-analyses on the NCAR and ECMWF websites.