Local and hemispheric dynamics of the North Atlantic Oscillation, annular patterns and the zonal index

Geoffrey K. Vallis [*], Edwin P. Gerber

GFDL/Princeton University, AOS Program, Princeton, NJ 08544, United States

Available online 17 January 2008

Abstract

In this paper we discuss the atmospheric dynamics of the *North Atlantic Oscillation* (NAO), the zonal index, and annular patterns of variability (also known as annular modes). Our goal is to give a unified treatment of these related phenomena, to make explicit how they are connected and how they differ, and to illustrate their dynamics with the aid of an idealized primitive equation model. Our focus is on tropospheric dynamics.

We first show that the structure of the empirical orthogonal functions (EOFs) of the NAO and annular modes follows, at least in part, from the structure of the baroclinic zone. Given a single baroclinic zone, and concomitantly a single eddy-driven jet, the meridional structure of the EOFs follows from the nature of the jet variability, and if the jet variability is constrained to conserve zonal momentum then the observed structure of the EOF can be explained with a simple model. In the zonal direction, if the baroclinic zone is statistically uniform then so is the first EOF, even though there may be little correlation of any dynamical fields in that direction. If the baroclinic activity is zonally concentrated, then so is the first EOF. Thus, at the simplest order of description, the NAO is a consequence of the presence of an Atlantic storm track; the strong statement of this would be that the NAO *is* the variability of the Atlantic storm track. The positive phase of the NAO corresponds to eddy momentum fluxes (themselves a consequence of wave breaking) that push the eddy-driven jet polewards, separating it distinctly from the subtropical jet. The negative phase of the NAO is characterized by an equatorial shift and, sometimes, a weakening of the eddy fluxes and no separation between sub-tropical and eddy-driven jets. Variations in the zonal index (a measure of the zonally averaged zonal flow) also occur as a consequence of such activity, although the changes occurring are not necessarily synchronous at different longitudes, and the presence of annular modes (i.e., the associated patterns of variability) does not necessarily indicate zonally symmetric dynamics.

The NAO, is not, however, a consequence of purely local dynamics, for the storm tracks depend for their existence on patterns of topographic and thermal forcing of near hemispheric extent. The Atlantic storm track in particular is a consequence of the presence of the Rocky mountains, the temperature contrast between the cold continent and warm ocean, and the lingering presence of the Pacific storm track. The precise relationship between the NAO and the storm tracks remains to be determined, as do a number of aspects of storm track

* Corresponding author. Tel.: +1 609 258 6176; fax: +1 609 258 2850.
  E-mail address: gkv@princeton.edu (G.K. Vallis).

dynamics, including their precise relation to the stationary eddies and to the regions of largest baroclinicity. Similarly, the influences of the stratosphere and of sea-surface temperature anomalies, and the causes and predictability of the inter-annual variability of the NAO remain open problems.

> *It has been many times remarked, that the weather in Greenland is just the reverse to that in Europe*; *so that when the temperate climates are incommoded with a very hard winter, it is here uncommonly mild, and vice versa. David Cranzt, The History of Greenland*

1. Introduction

The *North Atlantic Oscillation* (NAO) is the name given to the apparent dipolar nature of the weather between Greenland and Europe; the *zonal index* is the name given to the changes in the zonally averaged zonal wind; the appellation *annular modes* is sometimes given to certain hemispheric patterns of variability. In this paper we discuss these various phenomena, with the goal of trying to understand their underlying dynamics and the degree to which they are related—indeed the degree to which they are all just aspects of the same phenomenon.

The first known written record of the North Atlantic Oscillation appears to come from David Cranzt's *The History of Greenland*, quoted above (Crantz, 1767). The text, first published in Dutch in 1765, came at a time when Europe's interest in Greenland was growing for the second time. The Norwegian Hans Egede launched a mission to the Greenland with the support of the Danish in 1721, in search of Viking colonists. He found none (the Viking colonies had been subsumed by native Inuit) but, and of more relevance to us, he and his son Paul kept detailed records of the weather and climate. Paul notes that "that in the well known cold winter [in Europe] between 1739 and 1740 it was so mild in Disco-creek [on the west coast of Greenland], that the wild geese fled from the temperate to this frigid zone to seek warmth in January." Situated a few degrees above the Arctic circle, the thought of any creature seeking warmth in Disco Bay in January is worthy of note. In England in that same year the Thames froze, one of only 20 such occurrences since Roman times. In North America, in that same year, the River Delaware was, also quite unusually, frozen until March 15 (Ashmead, 1884). Evidently, the NAO has been with us, and known about, for some time. [For more historical background, see van Loon and Rogers (1978), Stephenson et al. (2003) and Wanner et al. (2001).]

These days, the NAO is often defined as a measure of the difference in sea-surface pressure between a site in Iceland and another in the Azores (which is in the subtropical high pressure zone of the Atlantic), generally after filtering variations that occur on a timescale of less than 10 days or so. The NAO 'index' is conventionally constructed by subtracting the mean and normalizing the ensuing time series by its standard deviation or similar. Alternatively, and in effect largely equivalently, the NAO and the NAO index are respectively defined as the first empirical orthogonal function (EOF) of variability (often of surface pressure, and again often filtered to eliminate timescales of less than about 10 days) in the North Atlantic region, and the corresponding principal component (see Fig. 1). To the extent that an EOF pattern does capture the leading mode of variability, it is clear that the seasonal weather between Europe and Greenland will often indeed differ; furthermore, they differ *in concert*; when one is warm, the other is cool. An
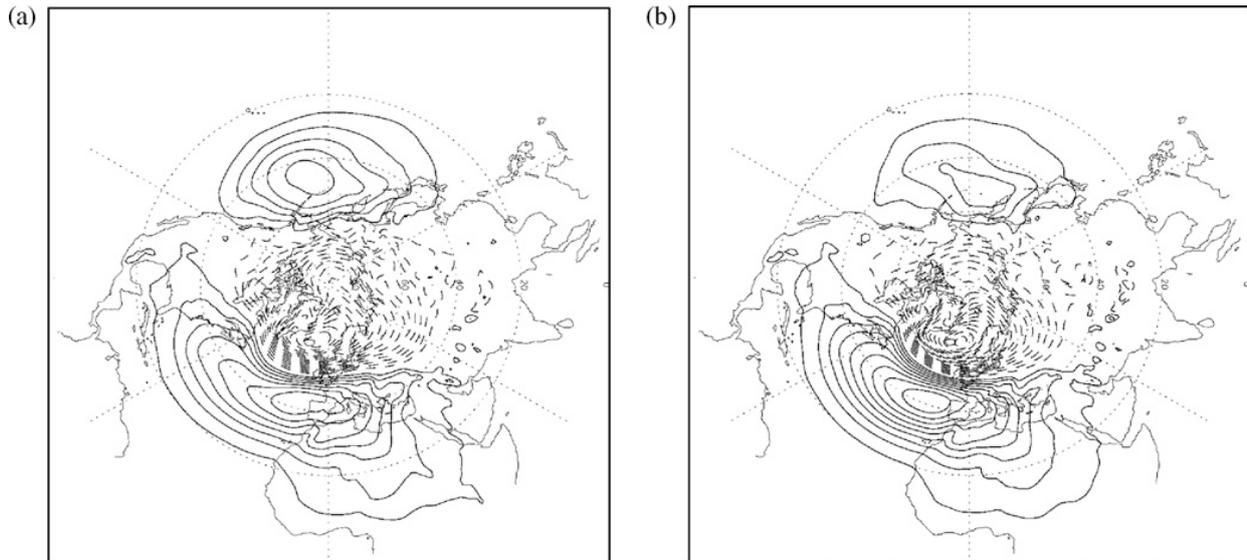
Fig. 1. Left: First EOF of winter (DJFM) sea level pressure (accounting for 25% of the variance). Right: first EOF of winter sea level pressure, based on variability in the Euro-Atlantic sector only. (Adapted from Ambaum et al., 2001.).

NAO index based on seasonally averaged pressure fields (e.g., over December–January–February) may also be defined, and this may be particularly useful if inter-annual or decadal variability is to be studied, as in Hurrell (1995), but at the same time it cannot reflect the synoptic variability of the underlying pressure fields.

The zonal index has, superficially, little relation to the NAO. It is a measure of the strength of the mid-latitude westerly winds and is sometimes defined as the horizontal pressure difference in mid-troposphere between $35°$ and $55°$ latitude, or the corresponding wind. Such a definition slightly obscures the two main ways in which the mid-latitude westerlies can vary: the winds can vary in magnitude ('pulsing') and the jet can vary in position ('wobbling' or 'meandering'). Identification of these two effects go back to Rossby and Collaborators (1939) and Namias (1950), respectively, as discussed by Wallace (2000); see also James et al. (1994), Lee and Feldstein (1996), Robinson (1996) and Vallis et al. (2004). Of course, a defining a 'zonal' index also sidesteps the issue as to whether a zonally averaged measure of the zonal wind is a dynamically meaningful concept at all, given the dominance of synoptic weather systems in midlatitudes with a length-scale of at most a few thousand kilometers, much less than the hemispheric scale itself.

Finally, 'annular modes' are another measure of hemispheric-scale patterns of variability. Annular modes are usually defined as the first (spatially varying) empirical orthogonal function (EOF) of the surface pressure in a given hemisphere, and the annular mode index is the corresponding (temporally varying) principal component. In the Northern Hemisphere, both the EOF and its principal component (i.e., the annular mode and its index) correspond closely to the NAO; that is to say, the hemispheric modes of variability in the Northern Hemisphere seem to be dominated by variability in the Atlantic sector, consistent with the patterns of Fig. 1. This leads to the question as to whether the NAO is a local manifestation of a hemispheric mode or variability (for example, a wave-mean flow interaction involving the zonally symmetric flow), or whether the annular modes are just the hemispheric projection of a phenomenon that is really more local, occurring primarily in the Atlantic sector; this dichotomy was raised explicitly by Wallace (2000) and further discussed by Thompson et al. (2003) and Feldstein and Franzke (2006). In the Southern Hemisphere the annular mode (the first EOF) is more zonally symmetric, but the dynamical dichotomy still exists. Put simply, are the zonally symmetric patterns of variability dynamically meaningful?

Our goal in his paper is to provide a coherent and mostly dynamical discussion of these various phenomena, namely the North Atlantic Oscillation, annular modes and the zonal index. While recognizing that both the stratosphere and the ocean may play influential roles in the variability of these phenomena, the essential dynamics lies in the troposphere, and that is where our emphasis lies.

2. A simple model for the spatial patterns

It is generally a good idea to seek simple explanations where possible, and in this section we put forward a null hypothesis, albeit a dynamical null hypothesis, for the spatial structure of the EOFs and correlation patterns. The hypothesis is that this gross spatial structure is largely a consequence of the quasi-random variations in the zonal wind, constrained by the conservation of angular momentum and by the fact there is but one baroclinic zone. If observations do not reject the null hypothesis, then more sophisticated theories would need to improve on its predictions in order to have any real value. Before elaborating on this, let us first see what the observed patterns are.

2.1. The observed patterns

As noted in Section 1, one way of defining patterns in the atmosphere is by way of an EOF analysis, and Fig. 1 [adapted from Ambaum et al. (2001) where many other analysed fields are to be found] shows the first EOF of the surface pressure for the Northern Hemisphere, both when the pressure field for all the hemisphere is used and when the data are confined to the Atlantic sector. From this figure, and Figs. 2 and 3 showing eddy statistics, we may note:

1. The similarity of the hemispheric EOF pattern with the Atlantic pattern, and the general dominance of the Atlantic sector in the hemispheric field.
2. The dipolar nature of the EOF fields.
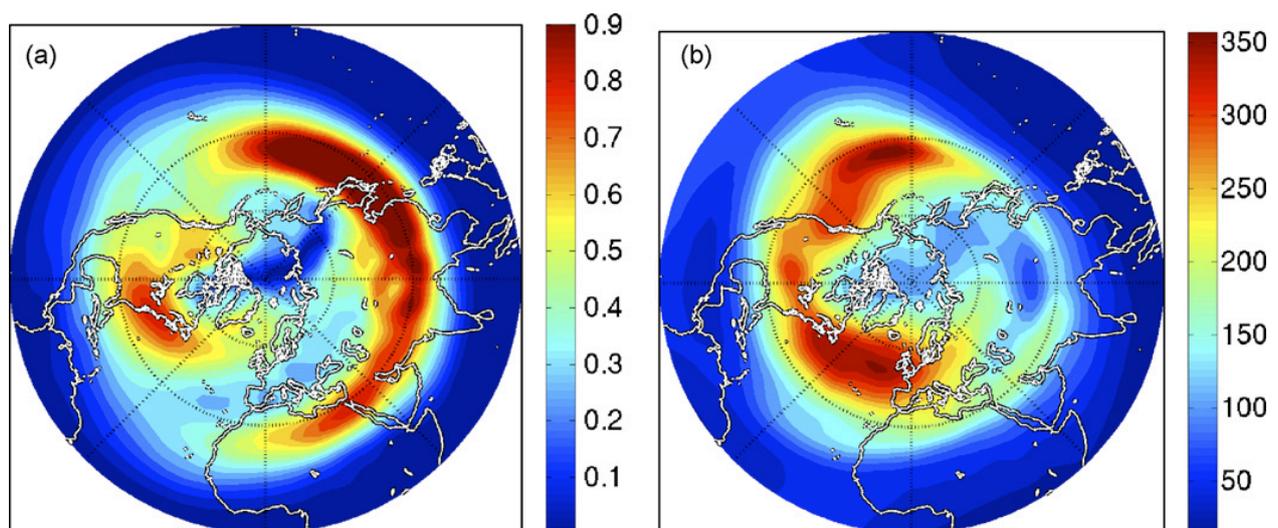
Fig. 2. Left: the Eady growth rate, σ_E, at 500 hPa in units of 1/days. Right: The average eddy kinetic energy at 250 hPa in units of (m/s)^2. Both are for the Northern Hemisphere winter (DJF), computed from the NCEP/NCAR re-analysis. The maxima in EKE are downstream of the maxima in growth rate, and the Pacific storm track does not fully decay before the beginning of the Atlantic storm track. The prime meridian (Greenwich) is at 6 O'clock.
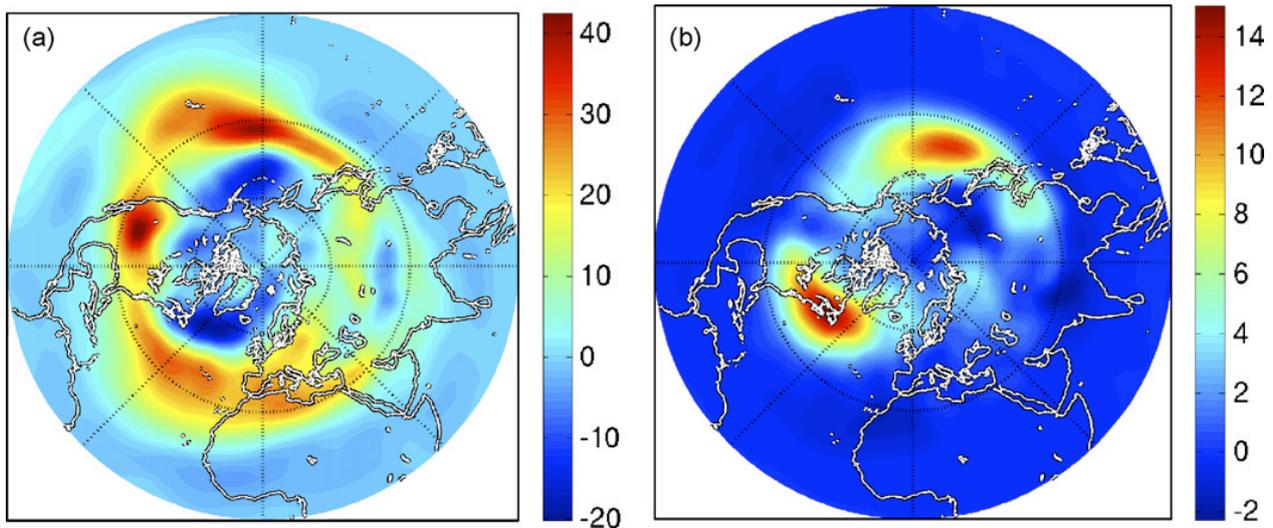
Fig. 3. Left: the eddy momentum fluxes at 250 hPa (m/s)^2. Right: the eddy heat fluxes at 500 hPa (mK/s), for the Northern Hemisphere winter (DJF). Both sets of data are band-pass filtered, allowing variability from 2 to 10 days, from the NCEP/NCAR re-analysis. Red values are large, blue values weak or negative.

3. The general, albeit not perfect, coincidence of the EOF patterns with the main storm tracks of the Altantic and Pacific (compare Figs. 1a and 2b). The storm tracks in turn are not coincident with the regions of maximum linear growth rate of baroclinic instability, generally being somewhat downstream of them.

4. Within the storm track, the eddy fluxes of heat are separated from those of momentum, with the latter generally being downstream of the former, although again the relationship is not perfectly clean (Fig. 3). If the eddy fluxes are not time-filtered, the patterns change quantitatively but not qualitatively.

Similar EOF patterns, with somewhat different meridional phases, arise in the zonal wind. The southern hemisphere patterns are more zonally symmetric, with no pronounced center of action, but are also dipolar in the meridional direction. If we zonally average the fields before calculating the EOFs, the first EOF is dipolar, the second EOF is tripolar and so on,
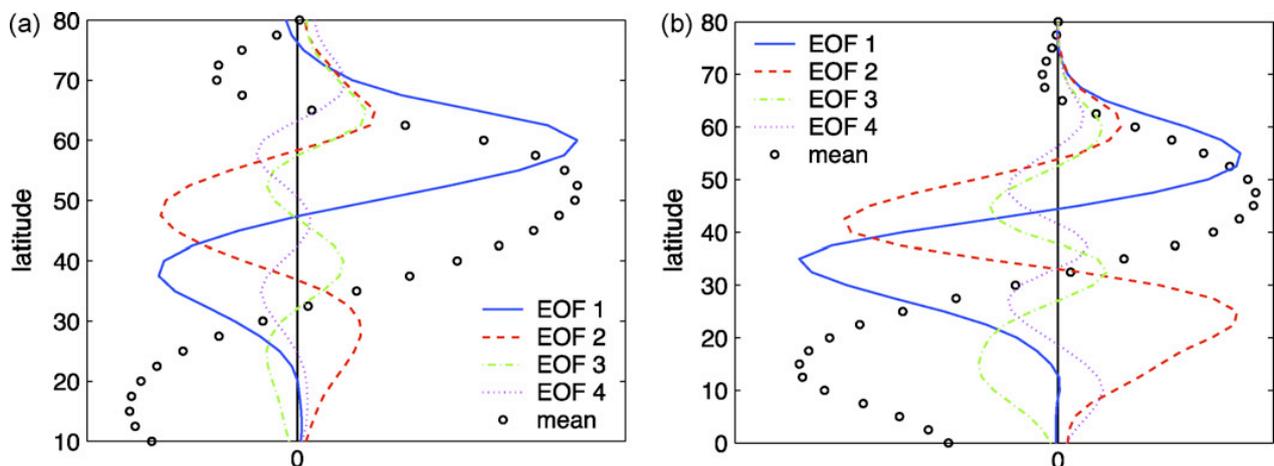

Fig. 4. The first four EOFs of the observed zonally averaged zonal winds (left), and of the axial angular momentum (right) for the Southern Hemisphere. The open circles, denoted 'mean', refer to the time averaged flow. The EOFs are normalized by the fraction of the variance for which they represent.

as illustrated in Fig. 4. Very similar patterns occur if we zonally average the two-dimensional EOFs.

Although the EOF patterns, especially in the Southern Hemisphere, are suggestive of hemisphere-wide influences, the observed correlation patterns (not shown) are not. These show little zonal correlation beyond about 90°, except on long timescales. In particular there is little correlation between the Atlantic sector and the Pacific sector except at very high latitudes. When higher (e.g., timescales less than a month) frequencies are filtered out, the correlation maps do show more significant correlations at larger scales.

2.2. Meridional structure and jet variability

The meridional structure of the EOFs has, in fact, a very simple origin, regardless of the dynamics that may be producing it. Consider first, following Wittman et al. (2005), a simple zonal jet, one for example, with a Gaussian structure as illustrated by the thick lines in the upper panels of Fig. 5. The jet itself is represented by

$$u(\phi, t) = U(t) \exp\left[-\left(\frac{\phi - \Phi(t)}{D(t)}\right)^2\right] \tag{2.1}$$

where the parameter Φ parameterizes the mean position of the jet, $U(t)$ its strength and $D(t)$ its width. Variations of latitudinal position, Φ, might be termed 'wobbling', variations of strength,

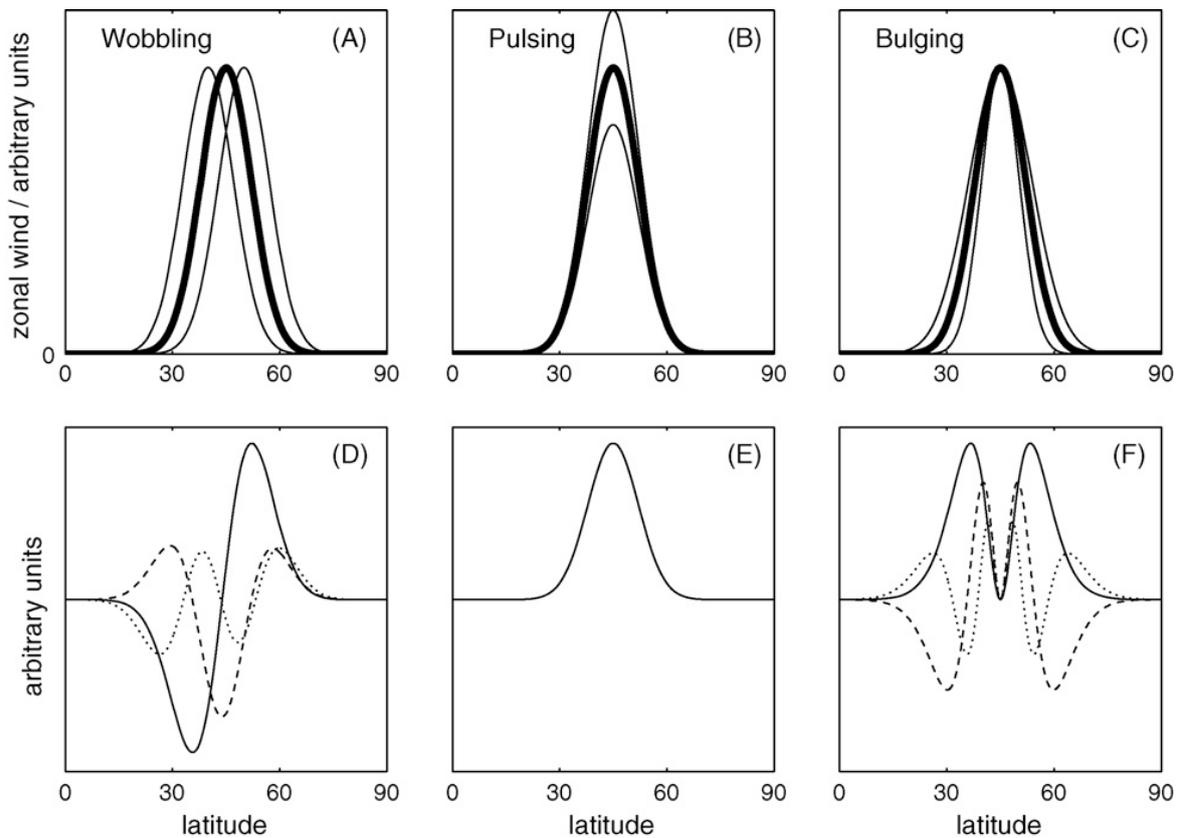Fig. 5. (A)–(C) Schematic illustration of three types of jet variability, resulting from the variation of a single parameter in (2.1) according to (2.2): wobbling refers to changes in $\Phi(t)$, pulsing to changes in $U(t)$ and bulging to changes in $D(t)$. (D)–(F) Corresponding EOFs. The first EOF is the solid line, the second is dashed, and the third is dotted (pulsing produces only one EOF). (From Wittman et al., 2005.)

U, may be termed 'pulsing', and variations of thickness, D, 'bulging'. Now allow the parameters to vary stochastically; thus, U evolves according to

$$dU = -\alpha(U - U_0)dt + \sigma\, dW, \tag{2.2}$$

and similarly for φ and D. The parameter α represents the strength of the relaxation of the jet back to its mean position, U_0, and W is a Wiener process (essentially random forcing, and dW/dt is white noise) with a strength parameterized by σ. The variations in U, φ and D are thus temporally uncorrelated in this model, but of course there is no reason to think that this is necessarily so in the real atmosphere.

The stochastic variability of each of the variables U, φ and D then produces qualitatively distinct eigenmodes, as shown in Fig. 5. In particular, variations in the position of the jet tend to produce a dipolar first EOF (Fig. 5D), and variations in the jet strength produce an EOF with a similar structure to the jet itself (Fig. 5E). Allowing all three parameters to vary with reasonably realistic amplitudes (taken from observations) leads to a dipolar first EOF and a monopolar second EOF (not shown). The dipolar EOF is much as is observed, but the second EOF is, in reality, tripolar, not monopolar. The difference is that the pulsing model above does not conserve momentum: had the pulsing in the jet core been accompanied by a corresponding variation on the jet flanks that ensured conservation of zonal momentum, the corresponding EOF would indeed have been tripolar. Nevertheless, the observed EOFs of the zonal wind correspond to straightforward types of variability of the wind as, perhaps, might be expected.

2.3. A random walk model

We may abstract the above description still further, and in a way that also allows us to consider the fundamental dynamics that gives rise to the various forms of jet variability. Specifically, let us consider the variability of the zonally averaged jet to be mimicked by a random walk, or more generally a Brownian motion, that obeys certain constraints. Consider a zonal flow, $u(y, t)$ in a channel, $0 < y < L$, and suppose that the zonal vanishes at the channel walls. Let us further suppose that the zonal flow can vary however it wants to, provided that it conserves its total zonal momentum (or the axial angular momentum on a sphere). Thus, at any time the zonal flow satisfies:

$$u(0, t) = 0 \tag{2.3a}$$

$$u(L, t) = 0 \tag{2.3b}$$

$$\int_0^L u(y, t)dy = C \tag{2.3c}$$

where C is a constant. Without real loss of generality we may consider $C = 0$, and then $u(y,t)$ represents deviations from its mean value, and we may nondimensionalize y so that $L = 1$.

We model this variability with a random walk between $y = 0$ and $y = 1$, in which we enforce the corresponding constraints

$$M(0, \omega_i) = 0 \tag{2.4a}$$
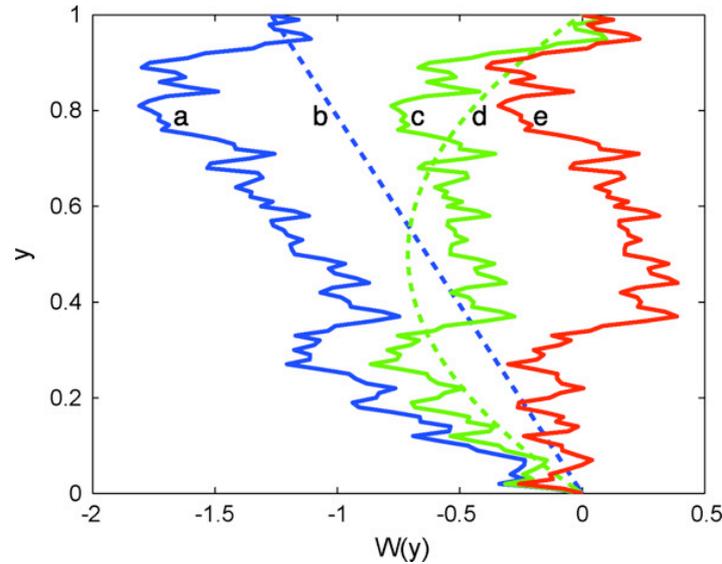
$$M(1, \omega_i) = 0 \tag{2.4b}$$

Fig. 6. Construction of a random walk model of zonal flow variability. The first realization is the random walk 'a'. This is detrended by subtracting line 'b' from it, to give 'c', a 'Brownian bridge'. Curve 'd' is the expected path of all Brownian bridges with the same mean as 'c', and if we subtract this from 'd' we obtain line 'e'. The collection of all random walks like 'e' represents variability of the zonal flow, subject to the constraints (2.3) or (2.4).

$$\int M(y, \omega_i)\mathrm{d}y = 0. \tag{2.4c}$$

where ω_i marks the realization of the process. For simplicity, consider a discrete random walk, with N steps to cross the channel. Referring to path 'a' in Fig. 6, the walker begins at $y = 0$ with a location $M(0, \omega_i) = 0$, where the location $M(y, \omega_i)$ corresponds to the value of $u(y,t)$ in the zonal flow. The walker moves one step in the y-direction, at the same time moving one lateral step either to the right or the left—with the lateral steps corresponding to changes in u. Eventually, the walker reaches the other side of the channel, but in general he reaches the other side some distance to the right or left of his original location; that is, the walker does not satisfy the constraint that $u(y = 1, t) = 0$. In the limiting case in which the walk becomes continuous, we may ensure that the walker does satisfy that constraint by detrending the walk, that is by subtracting the linear trend from startpoint to endpoint. This gives us path 'c', which is known as a 'Brownian bridge', and denoted $B(y,\omega_i)$. This path still does not satisfy constraint (2.4c), but we make it do so by additionally subtracting the expected path taken by all Brownian bridges that have the same integral at $B(y,\omega_i)$. The resulting random walk, path 'e', is known as a *mean zero Brownian bridge*. We may construct a large ensemble of such walks, and analyse their statistical properties. (In the discrete case, we may equivalently construct a large number of random walks like 'a', and then select only that subset that satisfies the mean-zero Brownian bridge property.) Each walk then corresponds to an independent realization of the zonal flow, and the variability of the walks represents the variability of the jet subject to the constraints (2.3). That is, the random walks mimic random variations of the zonal flow, subject only to the aforesaid constraints.

Given the above random walk, it is possible to calculate the EOFs and correlation or covariance function of the walks, and these are given in Figs. 7 and 8. (For more details, see Gerber and Vallis, 2005; these authors also explored the consequences of conserving mass instead of momentum.) The EOFs should be compared to those calculated from the observed zonally averaged zonal winds given in Fig. 4. The structure of the EOFs of the mean-zero random walk is very similar to that of the observations: the first EOF is (in both cases) a dipole, the second EOF a tripole and
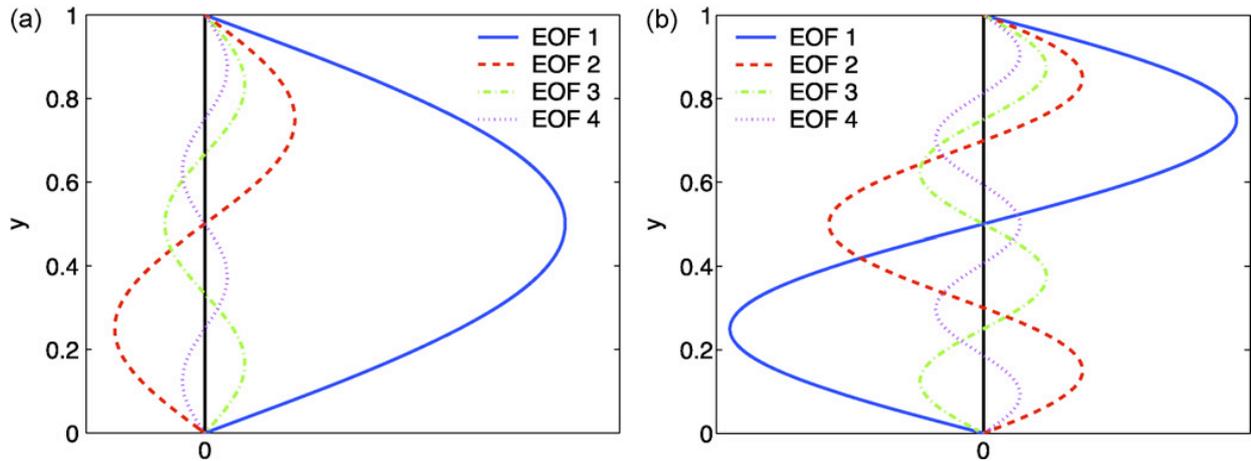
Fig. 7. First four EOFs of Brownian bridge (a) and of mean-zero Brownian bridge (b). The right panel should be compared with Fig. 4.

so on. The constraint that the walk have zero mean – reflecting the conservation of momentum – is crucial, for as can be seen by comparing the two panels in Fig. 7, in the absence of the constraint the first EOF is a monopole. With the constraint, the dipolar pattern is the gravest mode of variability allowed, and appears as the first EOF. Furthermore, the 'variance accounted for' in the model resembles the observations: in the northern hemisphere winter, the variance accounted for by the first three EOFs is 32, 22, and 12%; in the southern hemisphere 37, 21 and 10%; in the constrained random walk model it is 38, 19 and 10%. The close correspondence must be regarded as somewhat coincidental, because in a randomly forced model of the barotropic vorticity equation on the sphere – a system to which we might also expect the random walk model to – apply the corresponding variance of the first three EOFs was found to be 51, 27, and 11%. The spatial structure of the EOFs in the batotropic model was, however, quite similar to both the random walk model and the observations.

The covariance structure of the random walk (Fig. 8b) is also qualitatively similar to the meridional covariance structure of observations and numerical experiments. The amplitude at any

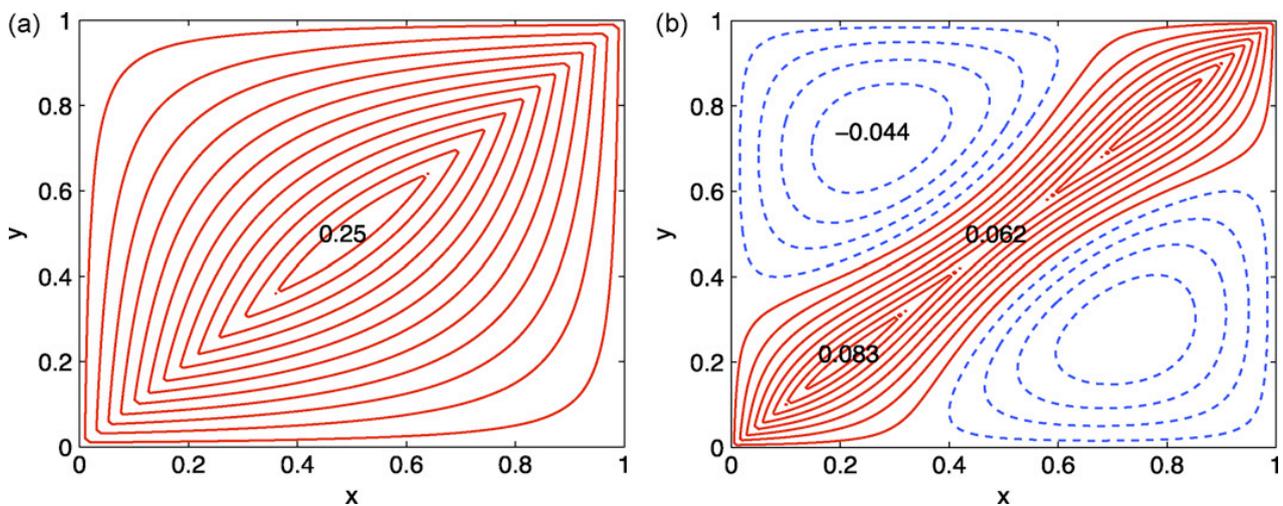Fig. 8. Covariance function of Brownian bridge (a) and of mean-zero Brownian bridge (b). A given point (x, y) gives the covariance between the two points. The auto-covariance at each point is obtained by traversing the diagonal line from $(0,0)$ to $(1,1)$. The covariance between a given latitude and all other latitudes is given by taking a horizontal or vertical slice at the chosen latitude.

given low latitude is negatively correlated with that at high latitude, simply because on average a positive anomaly at a low latitude must be accompanied by a negative anomaly at high latitudes in order to satisfy the mean-zero constraint. On the other hand, if the random walk does not have mean zero, a positive anomaly at some latitude is generally accompanied by a positive anomaly at all latitudes, which is not in accord with observations.

We may conclude, therefore that producing a dipolar EOF is rather a weak hurdle to surmount for any theory of zonal wind variability. It corresponds to wobbling of the zonal jet, but almost any mode of variability of a jet that preserves momentum – that is, that preserves the meridional integral of the zonal velocity, $\int u \, dy$, or its spherical coordinate analog – will produce such an EOF. Similarly, a simple covariance structure is also a consequence of constrained variability.

2.4. A two-dimensional random walk model

Let us now extend the random walk model to two-dimensions – zonal and meridional – by allowing a sequence of random walks to occur around a line of latitude, so that each walk occurs at a different longitude. Each random walk then represents the variability of the zonal wind at that longitude. For simplicity let us consider the discrete problem with, say, N walks around a latitude line, each walk denoted m_j where j marks the location in longitude, and runs from 1 to N. Now, the variability of the real zonal wind is correlated in longitude, and we may mimic that by specifying a correlation between neighboring random walks. For specificity, let us build a correlation structure in which each walk is correlated with its neighbors to either side, but uncorrelated with all the other walks. To do this, we suppose that the walks m_j are initially uncorrelated with each other. Then we can build in a correlation structure by defining a new set of random walks, n_j say, such that

$$n_j = \sqrt{1/2}(m_j + m_{j+1}) \qquad (2.5)$$

with a minor modification at the end points to make it cyclic. The walks n_j then have the same variance as the initial set m_j, and each walk has correlation of 0.5 with its immediate neighbors but no correlation with any other walk. We may then construct the EOFs of this set of walks, and the first EOF is illustrated in the left panel of Fig. 9. It is *zonally symmetric*. By construction, however, there is no zonal correlation of hemispheric extent, indeed no correlation between non-adjacent random walks. We may therefore conclude that the presence of a zonally symmetric EOF *does not imply the presence of zonally symmetric dynamics*. But nor, of course, does the presence of such EOFs militate against the presence of zonally symmetric dynamics.

We can extend this random walk model to produce a simple stochastic model of the North Atlantic Oscillation (NAO). The NAO is characterized by a longitudinally localized storm track, over the Atlantic, and we can model this by supposing that the random walks of (2.5) have increased variance over some region. For specificity, let us suppose that the fourth and fifth random walks have increased variance, and that instead of (2.5) they are calculated by

$$n_4 = \sqrt{1/2}m_3 + \sqrt{a - 1/2}m_4, \qquad n_5 = \sqrt{a - 1/2}m_4 + \sqrt{1/2}m_5 \qquad (2.6)$$

where any value of a greater than 1 will act to increase the variance in the regions corresponding to walks 4 and 5. The first EOF calculated with $a = 2$ is illustrated in Fig. 9 (right panel), and it shows a localized dipole. The observed structure of the first EOF in the Northern Hemisphere also shows a localized dipole.
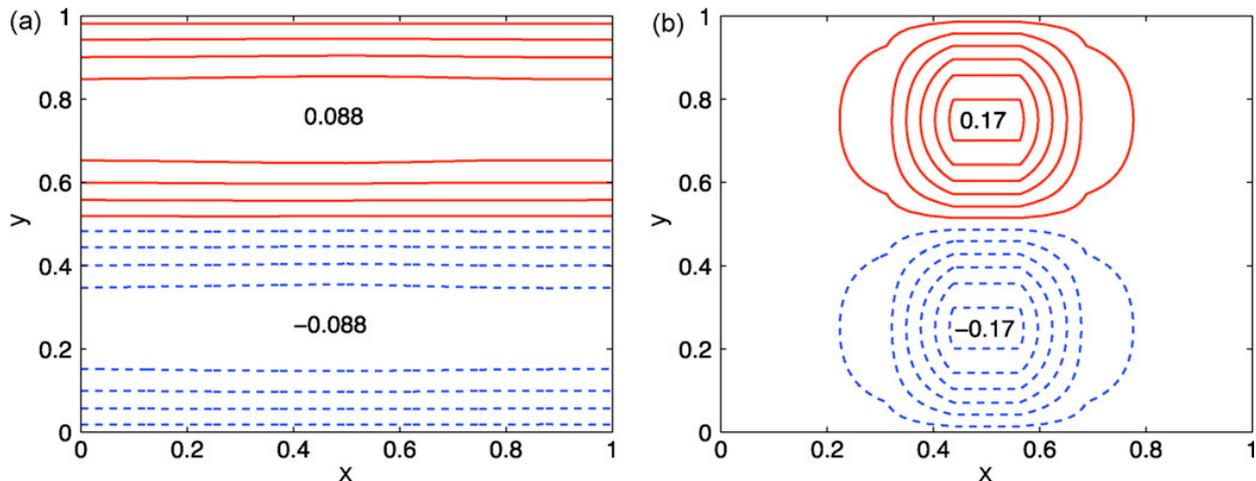
Fig. 9. Left: the first EOF of the random walk model with a zonal correlation constructed using (2.5) (and so the zonal correlation of the dynamics producing the EOF extends no more than 0.125 x units in either direction from a given point). With infinite sampling, the EOF is perfectly uniform in x and sinusoidal in y. Right: the first EOF from a simulation with enhanced variability in one region, specifically as described by (2.6) with $a = 2$.

2.5. Conclusions from the random walk models

We may draw the following conclusions from our random walk model:

(i) Variability in the meridional direction that conserves the meridional integral of the variable (corresponding, say, to a conservation of zonal momentum, with $\int \bar{u}\,dy = $ constant) will produced a first EOF with a dipole structure, the second with a tripole, etc. (as in Fig. 7).

(ii) A collection of random walks around a latitude circle with nearest neighbor correlations (either with or without long range zonal correlations) will produce a zonally symmetric EOF.

(iii) If the random walk has an increased variance over some longitudinal region, then the first EOF becomes correspondingly zonally localized.

Item (i) is relevant to the common occurrence of dipolar EOFs in both Northern and Southern Hemispheres, and suggests that these EOFs correspond to momentum-conserving variations in the zonal flow—perhaps more evocatively, to wobbles in the zonal flow. Item (ii) tells us that zonally symmetric EOFs (such as the first EOF of pressure or zonal flow) such as approximately occur in the Southern Hemisphere do not necessarily correspond to zonally symmetric dynamics, a conclusion that was also reached by Cash et al. (2002) and Vallis et al. (2004) using dynamical numerical models. And item (iii) tells us that if the zonal flow has a higher variance in some particular band of longitudes, such as might occur in a storm track, then the EOF will no longer be a zonally uniform dipole, but will be localised in regions where the variance is highest. This, crudely, corresponds to the EOF structure of North Atlantic Oscillation.

The random walk model does not, of course, attempt to model the real dynamics of zonal flow and its variability in the atmosphere. There is, for example, no precise analogy in the random walk of the distinctive meridional patterns of eddy momentum fluxes seen in the atmosphere, with zonal momentum converging in the vicinity of the jet. However, this is not to say that the variability of random walk is random at each point, and there are two features of the walk that are essential and lead to correlation structures similar to those of the real atmosphere. The first is that

the random walk is a *bridge*, starting and ending at the same level. This ensures that the maximum variance of the walk, corresponding to the jet variability, can be expected to occur approaching the middle of the channel, and that the EOFs are zero at either end of the channel. Second, the variations are constrained to have mean zero. It is this that leads a correlation structure shown in Fig. 8, and to the first EOF having a dipolar structure, as in Fig. 7. However, the random walk model tells us little about the *dynamics* of the zonal index or the NAO (and nothing about their temporal structure) so let us turn our attention to that.

3. Dynamics of the zonal index

There are two main mechanisms that give rise to zonal flow in the atmosphere. The first is simply that, by thermal wind balance, a meridional temperature gradient between the equator and pole is accompanied by a shear in the zonal wind. Friction and topographic drag will tend to limit the value of the wind at the surface, and the consequence is eastward flow aloft. The meridional temperature gradient tends to be largest in the subtropics, at the edge of the Hadley Cell, and this is the location of the subtropical jet. The jet is *baroclinic*, because it is associated with a temperature gradient and a shear.

In midlatitudes an eastward jet is produced for another reason, because of a momentum convergence in the region of baroclinic instability [see e.g., Vallis (2006) for more background]. A key aspect of the mechanism is that in barotropic Rossby waves the direction of the momentum transport is opposite to that of the group velocity. Thus, if Rossby waves are excited in mid-latitudes and propagate meridionally away from the source region, then eastward zonal momentum will converge on the region of excitation, generating an eddy-driven, eastward zonal flow. Westward momentum is deposited in the region of Rossby wave breaking, and overall the zonal momentum is preserved. The mechanism depends, of course, on the interaction of Rossby waves; as described above it is weakly nonlinear, 'weakly' because the nature of the momentum transport depends on the dispersion relation for Rossby waves, and if turbulence is so dominant that the dispersion relation ceases to have real meaning then the momentum flux due to Rossby waves will be negligible compared to that of the turbulent flow.

The generation of Rossby waves occurs in baroclinic instability – in storm tracks – and these are naturally variable. Thus, the eddy-induced zonal wind will naturally vary, and analyses of both observations and GCM experiments indicate this is a major cause of the variability of the zonal index, with both high frequency transients and stationary waves having a role (Limpasuvan and Hartmann, 2000; Watterson, 2002). However, if the generation and subsequent breaking of Rossby waves occurs on the timescale of baroclinic systems – the so-called synoptic timescale, i.e., about a week – then one might expect the zonal index to vary on that same timescale. In fact, in both hemispheres the zonal index varies on a somewhat longer timescale (Fig. 10) and there are at least two potential reasons that might cause this:

 (i) Some kind of interaction or feedback between the zonal index and the momentum transport in Rossby waves.
(ii) A reddening of the zonal index by other factors, such as the slow variability of sea-surface temperatures, or the stratosphere.

Although external factors may certainly play a role in inter-seasonal and inter-annual variability, mechanisms internal to the troposphere are almost certainly important for item (i), and this is our focus.
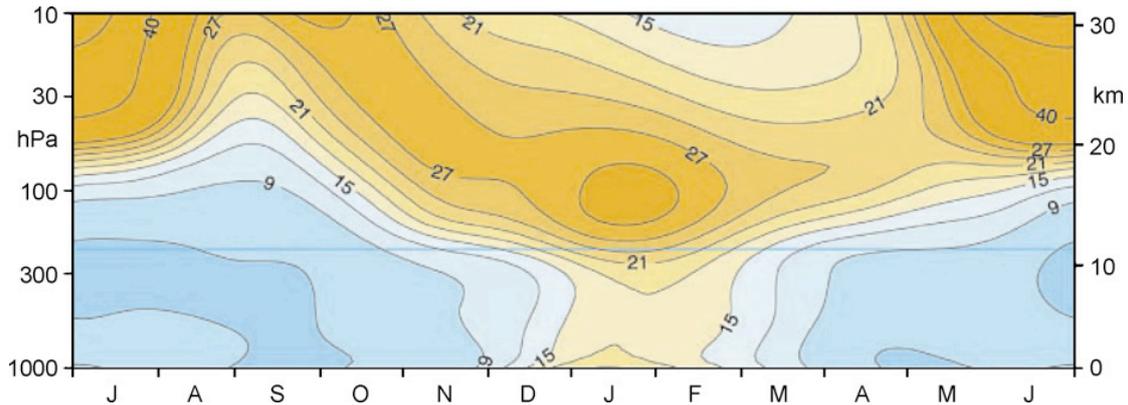
Fig. 10. Timescale of the Northern Annular mode as a function of season and height, in days. The NAM is the first EOF of the geopotential at each isobaric level, and the timescale is the e-folding time of the auto-correlation. (Adapted from Baldwin et al., 2003.).

3.1. Temporal structure of the zonal index

A useful framework to describe wave–mean flow interactions is that of the transformed Eulerian mean (TEM; Andrews and McIntyre, 1976). In the quasi-geostrophic limit, the zonal-mean zonal momentum equation may be written

$$\frac{\partial \bar{u}}{\partial t} - f_0 \bar{v}^* = \nabla \cdot \mathscr{F} + D \tag{3.1}$$

where D represents dissipative terms (e.g., friction in the Ekman layer), $\bar{v}^*$ is the meridional residual velocity and $\mathscr{F}$ is the Eliassen–Palm flux, given by

$$\mathscr{F} = \overline{v'q'} = \overline{v'\zeta'} + \frac{\partial}{\partial z}\left(\frac{\overline{v'b'}}{N^2}\right). \tag{3.2}$$

where b is the buoyancy. To see a typical Eliassen–Palm (EP) flux look ahead to Fig. 19a [this is from a GCM; observed fluxes are qualitatively similar—see, for example Edmon et al. (1980)]. The time-averaged EP flux converges in the upper troposphere somewhat equatorwards of its generation region; this convergence reduces the overall shear, and reflects the effects of pole-wards transport of heat in reducing the meridional temperature gradient. However, the upper level divergence (at about $50°$) also accelerates the flow eastwards, producing an eddy-driven jet and a region of surface westerlies somewhat polewards of the main subtropic jet. This is the region of momentum convergence due to Rossby waves mentioned above. Now, because the EP flux is a product of chaotic baroclinic eddies, it fluctuates in time and space, and this leads to fluctuations in the zonal flow, $\bar{u}$, as well as in the overturning circulation (e.g., Feldstein and Lee, 1998).

Consistent with observations, numerical experiments with the dry dynamical core of a primitive equation model of the atmosphere indicate that the zonal index fluctuates on somewhat *longer* timescales than typical baroclinic eddy turnover ('synoptic') timescales (Gerber and Vallis, 2007). These authors found that neither the model's thermal damping timescale (a relaxation back to a radiative equilibrium timescale) nor the momentum damping timescale (a Rayleigh drag near the surface) straightforwardly determined the timescale of the zonal index. That is to say, over a wide parameter range the timescale of variability varied largely independently of the thermal damping timescale, the momentum damping timescale and the eddy turnover time. Furthermore, in their model, there is no properly resolved stratosphere, and there is no equivalent of persistent

sea-surface temperature anomalies. These results suggest that some kind of internal dynamical timescale is operating, and we now describe one potential mechanism; other mechanisms have also been proposed, and the issue is not settled.

3.1.1. Stickiness of the zonal flow

The fluctuations in the zonal index and, relatedly, the structure of the first EOF of the zonally averaged zonal wind, are largely barotropic, or at least equivalent barotropic – meaning they have little change in phase with height; they do not, for example, tilt westward with height. This suggests that the perturbations that affect it are essentially barotropic. Let us consider how these perturbations might grow and decay and affect the jet itself.

Let us consider a fairly broad baroclinic jet – wider than the first deformation radius, say – so that an eddy might equally well grow on the flanks of the barotropic jet as in the core. Let us examine what happens in these two cases. Suppose first that a new eddy grows in the jet core, as illustrated in panels (a) through (c) of Fig. 11. Panel (a) shows the EP flux in the initial, baroclinically growing, part of the eddy lifecycle (see Simmons and Hoskins (1978) for more discussion of life cycles themselves). The EP flux is composed primarily of a heat flux, which corresponds to the vertical component of the EP flux (dashed lines). These reduce the shear by
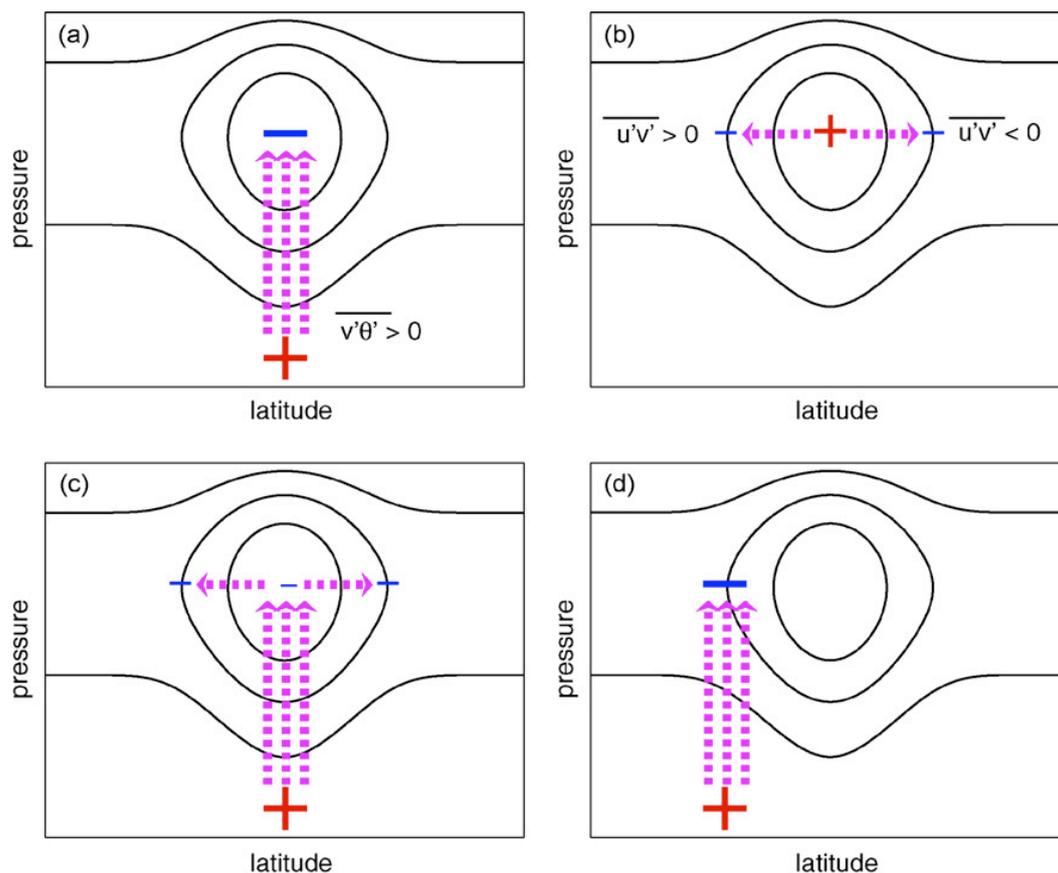

Fig. 11. A schema of the eddy mean flow feedback. Solid contours mark the position of the zonal mean winds. Dashed arrows denote the propagation of wave activity (EP fluxes). Plus and minus signs indicate divergence and convergence of the E–P fluxes, and so westerly or easterly torque on the mean flow. Panels (a) and (b) break down the impact of eddy heat and momentum fluxes of an eddy formed in the jet core, leading to the net forcing of the flow pictured in (c). Momentum fluxes in upper levels spread the easterly torque over the jet, maintaining the shear in the core. For an eddy generated on the flank of the jet, shown in panel (d), weak upper level winds limit meridional propagation, and the shear tends to be reduced only locally. (Adapted from Gerber and Vallis, 2007.)

drawing momentum from the upper to lower levels, and thus act to barotropize the flow. This transfer is indicated by the minus and plus signs in panel (a), which indicate the sign of the zonal flow acceleration.

In the next stage of the baroclinic life cycle, and before the eddy decays, meridional propagation occurs, as shown in panel (b). Linear theory suggest that particle displacements, η, scale as $\psi/(\bar{u} - c)$, where ψ is the streamfunction perturbation. In critical layers (where $\bar{u} = c$), particle displacements become unbounded even for infinitesimal perturbations, leading to nonlinearity and/or damping. (In practice, nonlinear effects become relevant before waves reach their critical latitude, and typically most of the dissipation may occur 10–20° away from the critical layer (Randel and Held, 1991).) In the example illustrated, the wave activity propagates away from the latitude at which the eddy was generated before being damped on the flanks of the jet, where u is weaker and $\bar{u} - c$ is closer to zero. The resulting divergence of wave activity in the jet core and convergence on the shoulders leads to an upgradient transfer of momentum at upper levels, into the jet core region. This is denoted by the plus (+) sign in panel (b).

The net effect on the eddy life cycle is indicated in panel (c). There is an overall weakening of the upper flow in the both the jet core and on its flanks. At lower levels there is a divergence of EP flux and hence a flow acceleration, but this acceleration is limited by the surface friction, and the overall baroclinicity at the latitude of the jet core is more-or-less maintained. Thus, in sum, the tendency of the eddy heat transport to reduce the baroclinic circulation is spread over the entire jet by the lateral EP flux in upper levels, so maintaining baroclinicity in the jet core. Thus, the structure of the jet is maintained.

Now consider an eddy that grows on the shoulders of the jet, as illustrated in panel (d). In this case heat fluxes associated with eddy growth also lead to vertical propagation of EP fluxes, and a deceleration of the jet aloft, and a transfer of this momentum to the surface. However, the meridional propagation of the EP flux aloft is limited, because the eddy is closer to the critical latitude. Only Rossby waves with phase speed less than $\bar{u}$ can exist; the waves, being nearer their critical latitude, and damped more quickly than eddies in regions of stronger zonal flow. Thus there is a tendency of the heat fluxes to reduce the shear *locally*, that is at the latitude where the eddy initially forms, rather than over the entire baroclinic zone. The reduction of baroclinicity makes it less likely for a second eddy to form at this latitude. Thus, the net effect of all this is that in a broad baroclinic zone (i.e., one wider than the scale on which an eddy initially develops) there is a tendency for eddies to continue to form at the jet core, so reinforcing the current position of the jet. Rather than being a positive feedback between the jets and the eddies, we might say that the mechanism discourages the movement of the jet position.

Some numerical experiments illustrating this mechanism are given by Gerber and Vallis (2007), but nevertheless we must regard it as rather speculative, vis-à-vis its applicability to the real atmosphere as the mechanism is rather delicate, and is sensitive, for example, to the location of the critical lines. This may be important because in the real atmosphere the sub-tropical jet and the eddy-driven jet are not distinct, with the result that the upper level westerly jet is some 10° equatorward of the region of maximum surface westerlies. The mechanism is similar to that of Robinson (1996, 2000), although it differs in detail; in his mechanism, friction plays an important role in reducing the near surface flow and developing a baroclinic flow in the region of momentum flux convergence, thus maintaining the jet.

In summary, variations of the zonally averaged zonal flow take place on timescales somewhat longer than those associated with baroclinic weather systems. Plausible mechanisms, involving solely tropospheric dynamics, have been proposed that might explain these timescales (e.g., Robinson, 1996; Lorenz and Hartmann, 2001; Gerber and Vallis, 2007). This is not to say that

in reality other mechanisms involving the stratosphere or surface forcing are not as or more important.

4. The spatial structure of the NAO

We now turn our attention to the spatial structure of the North Atlantic Oscillation, its relation to the storm tracks and to the more zonally symmetric annular patterns of variability. The underlying questions are: (i) What is the relationship of the NAO to the North Atlantic storm track? (ii) To what extent are the two phenomena regional, as opposed to hemispheric? Let us first consider the question of what produces a storm track, from a rather idealized point of view [see Chang et al. (2002) for a comprehensive review].

4.1. Storm tracks

A storm track is, loosely, a connected region in which the storminess is greater than the average. The 'storminess' may be defined as, for example, the eddy kinetic energy, possibly after undergoing some form of filtering in time (e.g., a high-pass filter allowing timescales of less than 10 days). Fig. 2 shows one measure of the northern hemisphere storm tracks, and the Atlantic and Pacific storm tracks stand out, although the Pacific storm track does not have a well-defined end to it and the storm track picks up again in the Atlantic without ever having really died. In the Southern Hemisphere the storm track (not shown) is more zonally symmetric, although it is not completely zonally uniform.

A storm track ultimately owes its existence to a longitudinally confined region of enhanced baroclinicity, which in turn is caused by zonally asymmetric forcing from thermal and topographic anomalies. However, the storm tracks do not coincide with the region of enhanced baroclinicity, as can be seen both from Fig. 2 and from some idealized simulations with the dynamical core of a primitive equation GCM. In the first simulation, denoted T1, a zonally asymmetric thermal anomaly is introduced, as illustrated in Fig. 12. The anomaly consists of a region of high-latitude cooling adjacent to a region of high-latitude heating, so producing a region of enhanced (diminished) baroclinicity equatorwards of the region of cooling (heating), as measured by the Easy growth rate of the ensuing time-averaged flow. (The Eady growth rate may be defined by $\sigma_E \equiv 0.31 \, f N^{-1} |\partial \bar{u}/\partial z|$.) Such a set up might crudely correspond to the North America–Atlantic sector in winter: the cold region corresponds to the continent, and the warm region the North Atlantic ocean.

The enhanced baroclinicity (i.e., in the cooler region) does *not* coincide with the storm track; the latter, as measured by the eddy kinetic energy, is further downstream of the former. The eddy heat fluxes are just a little downstream of the baroclinic zone, and upstream of the region of maximum kinetic energy. The eddy momentum fluxes are downstream of the region of maximum kinetic energy, near the end of the storm-track. This pattern can be understood in part as a spatial mapping of the temporal life cycle of the baroclinic eddy life cycle. The region of enhanced baroclinicity is the region of cyclogenesis, and it is in the early part of the baroclinic life cycle that most of the eddy heat transport takes place. As a given baroclinic eddy grows, it is advected eastwards, and hence the maximum kinetic energy is further downstream. The eddy momentum transport occurs in the barotropic, decaying phase of the life cycle, and near the end of the storm-track. The first EOF of the surface pressure, which one might call the 'NAO' of this model, roughly co-incides with the region of the eddy momentum fluxes, where the eddy activity and the response is largely barotropic. The mapping of the classical temporal structure of the baroclinic life cycle on to the
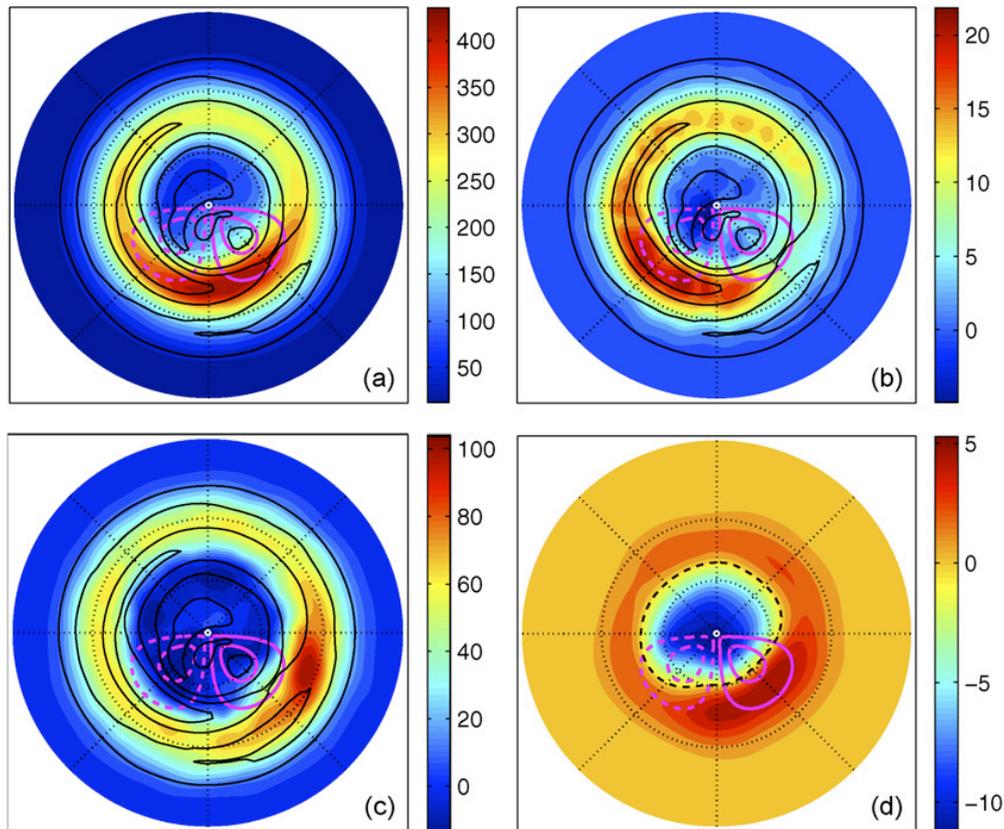
Fig. 12. Results from a numerical integration with a thermal anomaly (experiment T1), as indicated by the light contours: the dashed contours are a cold anomaly and the solid contours a warm anomaly. (a) The time average transient eddy kinetic energy at 250 hPa $(m/s)^2$, (b) the high pass eddy heat fluxes at 750 hPa (mK/s), (c) the eddy momentum fluxes at 250 hPa $(m/s)^2$, and (d), the first EOF of the 10 day averaged sea level pressure (hPa, corresponding to a one standard deviation anomaly). Black contours in (a), (b) and (c) mark levels of the Eady growth rate $(days^{-1})$. The eddy momentum fluxes and the center of the EOF pattern are both downstream of the region of maximum baroclinic instability.

spatial structure of the storm track is not perfect, because a baroclinic eddy always grows from the remnants of another eddy, and this is related to downstream development discussed more below.

A still more localized stormtrack, and a more localized EOF, may be produced by combining a thermal source with a suitably chosen topography, as illustrated in Fig. 13. (See also Cash et al. (2005) for a collection of related numerical simulations.) The topography is a ridge just west of the cold thermal anomaly, and might be thought of as a crude representation of the Rockies. The topography produces a trough downstream of the ridge and a wavetrain propagating and decaying eastwards. When combined with the thermal forcing, the ensuing eddy kinetic energy (i.e., the storm track) and the associated eddy kinetic energy are quite localized, and qualitatively resemble the Atlantic storm track in terms of its phase relative to North America and its zonal localization.

The baroclinic lifecyle is different in zonally varying storm tracks than in classical life cycles, primarily because a baroclinic eddy grows in the remnants of the last baroclinic eddy, while all the time being advected downstream (Chang, 1993; Orlanski and Chang, 2000). This leads to 'downstream development', in which a decaying eddy gives up its energy to neighboring eddies downstream, giving rise to coherent wave packets. In a zonally extended storm track, downstream development becomes increasingly important downstream of the entrance region. Now, the baroclinicity in the Atlantic storm track is largest over the eastern USA and western North Atlantic (Fig. 2), so forming the entrance of the Atlantic storm track. But it is also clear
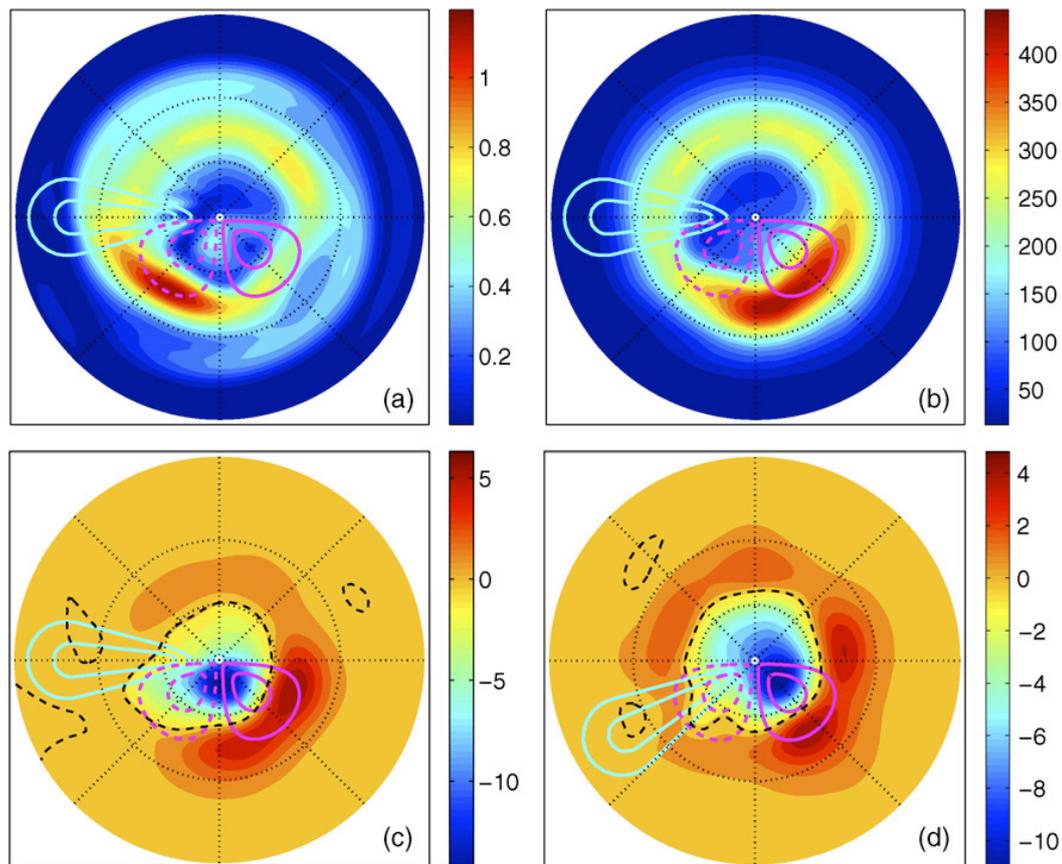
Fig. 13. Results from a numerical integration with a thermal anomaly and a mountain ridge (experiment TM1). The thermal anomaly is the same as in Fig. 12 and the ridge is denoted by light blue contours at 500 and 1500 m. In (a), (b) and (c) the ridge is at 9 O'clock and in (d) it is at 8 O'clock. (a) Eady growth rate; (b) time-averaged eddy kinetic energy; (c) First EOF of the sea-level pressure field; (d) As for (c) but different location of topographic ridge. Units and EOF normalization as in Fig. 12.

that in this region the Pacific storm track has not died, leading to a more intense Atlantic storm track than the local baroclinicity might indicate. Still, the signature of the classical life cycle is still apparent, and the eddy momentum fluxes are downstream of the eddy heat fluxes, as we saw in Fig. 3.

What is the conclusion to be drawn? Plainly, variability resembling NAO patterns is intimately linked to storm tracks and, rather roughly, corresponds to the eddy momentum fluxes of the storm tracks. The eddy momentum fluxes tend to be downstream of the region of maximum kinetic energy, which in turn is downstream of the region of maximum baroclinicity, which roughly coincides with the region of maximum eddy heat fluxes. These phase relationships are reasonably clean in the idealized numerical simulations of Figs. 12 and 13, but less so for the observations, probably, at least in part, because the entrance of the Atlantic storm track region and the exit of the Pacific storm track overlap. Numerical experiments covering a broad range of configurations suggest that a reasonably well-defined storm track termination region is needed to produce NAO-like patterns.

4.2. Teleconnections

A common way to look at the physical space characteristics of variability is via *teleconnections* (Namias, 1969; Wallace and Gutzler, 1981). Teleconnections are in general interdependencies
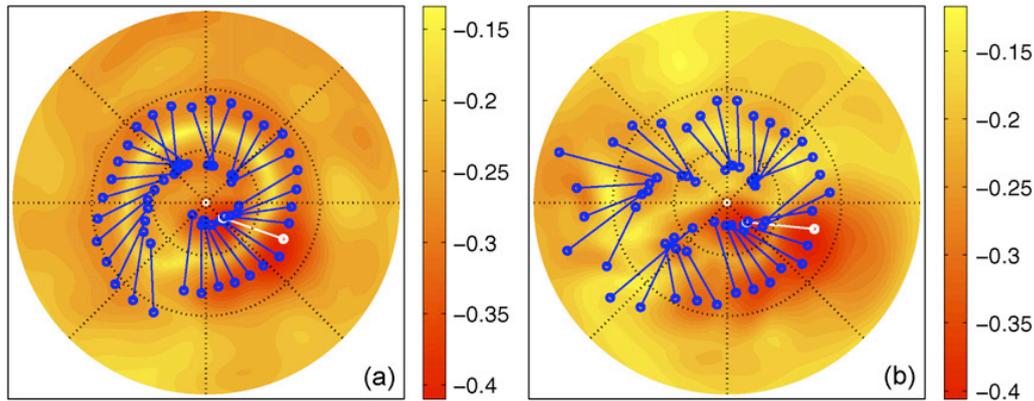
Fig. 14. Teleconnection lines between 32 pairs of points. The equatorward points are equally spaced with respect to longitude, and the points they each pair with are the points of maximum anti-correlation. The background color gives the maximum value of anti-correlation associated with each point; the darker the red, the stronger the anti-correlation. The data are taken from numerical simulations, left panel from T1 and right panel from TM1.

between spatially distant points, and as such are closely related to maps of the correlation between a 'base point' and all other locations on the globe. Obviously, there are a large number of such maps, many of them not very meaningful, and to identify the important ones we find those points that exhibit strong anticorrelations with other points. We may then construct a teleconnection map as half the difference of the correlation maps of two base points that have a strong anticorrelation.

To get an overall sense of the teleconnection patterns, at 32 evenly spaced longitudes we identify a low-latitude 'center of action', meaning the latitude equatorwards of 45° that has the strongest anticorrelation of surface pressure with some other point on the hemisphere, and we connect the two points with lines, as illustrated in Fig. 14. The data set used to construct these figures is from numerical integrations T1 and TM1 (see Table 1). The background color corresponds to strength of the anti-correlation at each point. As can be seen, the teleconnected pairs do not necessarily fall along lines of constant longitude, although the two points that anti-correlate most strongly are usually at *similar* longitudes, this being close to the center of the first EOF of the corresponding simulation.

Given a teleconnection pair at locations x_1 and x_2, we may formally determine the corresponding teleconnection pattern as follows. We first construct a time series as a normalized difference of the time series of the pressure at the two points,

$$T_c(t) = A \left[\frac{p'(x_1, t)}{\sigma_{x_1}} - \frac{p'(x_2, t)}{\sigma_{x_2}} \right] \tag{4.1}$$

where $p'(x_i)$ denotes the pressure deviation from the mean and σ_{x_i} the standard deviation, and A is a normalization constant to give the time series unit variance. We then regress this time series

Table 1
Numerical experiments

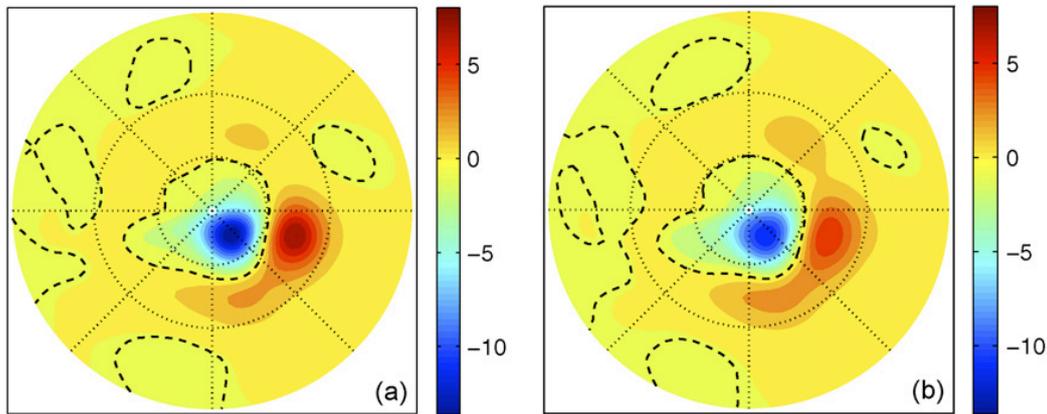| Experiment	Forcing	Appearing in Figures
Tl	Thermal anomaly only	Fig. 12, Fig. 14a
TM1	Thermal anomaly plus orographic (mountain)	Fig. 13a–c, Fig. 14b, Figs. 15, 17 and 18
TM2	Thermal anomaly plus orographic (mountain)	Fig. 13d
S1	Zonally symmetric	Figs. 16 and 19

Fig. 15. Teleconnection pattern (hPa) of the two points that have maximum anti-correlation (the two points connected by the white line in panel (b) of Fig. 14. The left pattern uses daily values of surface pressure, and the right panel uses pressure that has undergone a 30 day, low-pass filter. The data are from numerical simulations TM1, as in Figs. 13 and 14b.

on to the two-dimensional surface pressure field,

$$P_c(\boldsymbol{x}) = \frac{1}{n}\sum_{t=0}^{t=n} p'(x,t)T_c(t). \tag{4.2}$$

so producing the teleconnection pattern associated with the two points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, and this is shown in Fig. 15. From (4.1) and (4.2) it is clear that the teleconnection pattern is, in fact, the difference between the two single point correlation maps of points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.

There is a connection between the regression pattern shown in Fig. 15 and EOFs: the EOF pattern is a regression of the data set with the time series of the corresponding principal component. (The relationship between teleconnections and EOFs was also discussed by van den Dool et al. (2000).) Thus, the first EOF of surface pressure may be obtained using a formula similar to (4.2), but using the time series equal to that of the first principal component instead of the pressure difference, (4.1). However, there are important differences between EOFs and teleconnection patterns that are especially apparent when the *statistics* of the dynamics are zonally symmetric, but the dynamics themselves have limited correlations in the zonal direction. To illustrate this, let us construct a new time series, T_{c2} by spatially regressing the teleconnection pattern, (4.2), on to the pressure field so that

$$T_{c2}(t) = p'(\boldsymbol{x},t) \cdot P_c(\boldsymbol{x}), \tag{4.3}$$

where the dot-product indicates a summation over all spatial points. In general, $T_{c2}(t)$ is not the same as $T_c(t)$. We can iterate the procedure, generating a new spatial field given by

$$P_{c2}(\boldsymbol{x}) = \frac{1}{n}\sum_{t=0}^{t=n} p'(x,t)T_{c2}(t), \tag{4.4}$$

and then generate a new time series T_{c3}, and so on. [A similar procedure was used by Walker and Bliss (1932) and by Wallace (2000).] The time series given by (4.3) will be smoother than that given by (4.1), essentially because the time series is a weighted time series over many points, with $P_c(\boldsymbol{x})$ providing the weighting: note that the timeseries of (4.1) may be constructed by using a formula similar to (4.3), but with the weighting function being two delta functions at the points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.

At each step in the iteration procedure, weight is added to points correlated with the principal pattern of variability, and so the amount of variance the pattern represents increases. The amount of variance is bounded from above and the iteration procedure will eventually converge. At that point, the time series and the spatial patterns are related by

$$T_{c\infty}(t) = p'(\boldsymbol{x}, t) \cdot P_{c\infty}(\boldsymbol{x}) \tag{4.5}$$

$$P_{c\infty}(\boldsymbol{x}) = \frac{1}{n} \sum_{t=0}^{t=n} p'(\boldsymbol{x}, t) T_{c\infty}(t). \tag{4.6}$$

This is a property of EOFs and principal components: the EOF pattern, $P_{c\infty}(\boldsymbol{x})$ is the temporal regression of the data field $[p'(\boldsymbol{x},t)]$ on principal component time series, and the principal component, $T_{c\infty}(t)$, is the projection of the EOF pattern on to the data field, or equivalently the spatial regression of data field onto the EOF pattern. The property is a consequence of the fact that EOFs and PCs are orthogonal in space and time.

If the variance is fairly localized in space, as for example when there is a storm track like that over the Atlantic, the EOFs and the teleconnection patterns may be fairly similar. As noted above, the teleconnection pattern is essentially a correlation map, focusing on the regions where the anti-correlation is greatest, and it is localized because large correlations do not greatly extend beyond the synoptic scale. As the storm track becomes more zonally symmetric, the teleconnection pattern changes little, as can be seen by comparing Fig. 15 with Fig. 16. However, the first EOF becomes virtually zonally symmetric, because this is the pattern that captures the maximum possible variance as described in Section 2. As we demonstrated in Section 2.4, and as the construction with the teleconnections confirms, a zonally symmetric EOF does not necessarily correspond to zonally symmetric dynamics. Note, though, that the correlation pattern does become more extended zonally if the dynamical fields are low-pass filtered in time, as can be seen by comparing the two panels in Fig. 16.

4.3. Jet structure and a physical picture of the NAO

As noted in Section 3, there are two mechanisms that give rise to eastward flow in the Earth's atmosphere. One is that the meridional temperature gradient produces a thermal wind shear
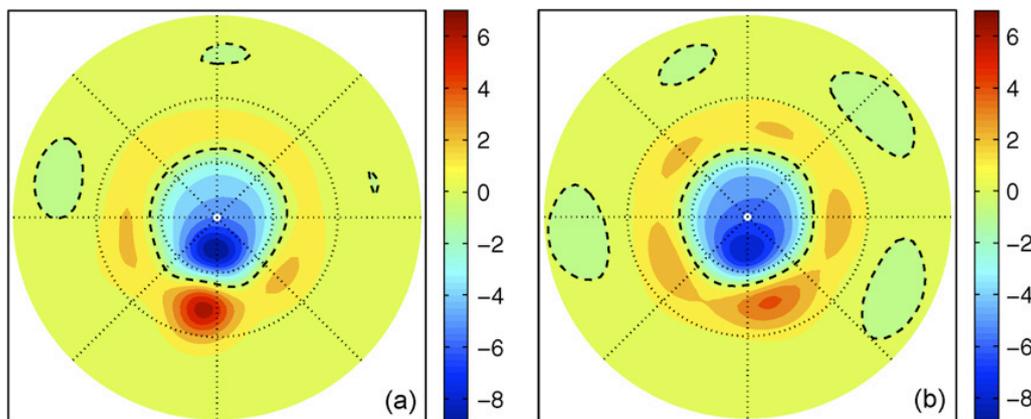

Fig. 16. Teleconnection pattern (hPa) of two points that have maximum anti-correlation, at an arbitrary longitude in a statistically zonally symmetric simulation (experiment S1). The left pattern, (a), uses daily values of surface pressure, and the right panel, (b), uses pressure that has undergone a 30 day, low-pass filter.
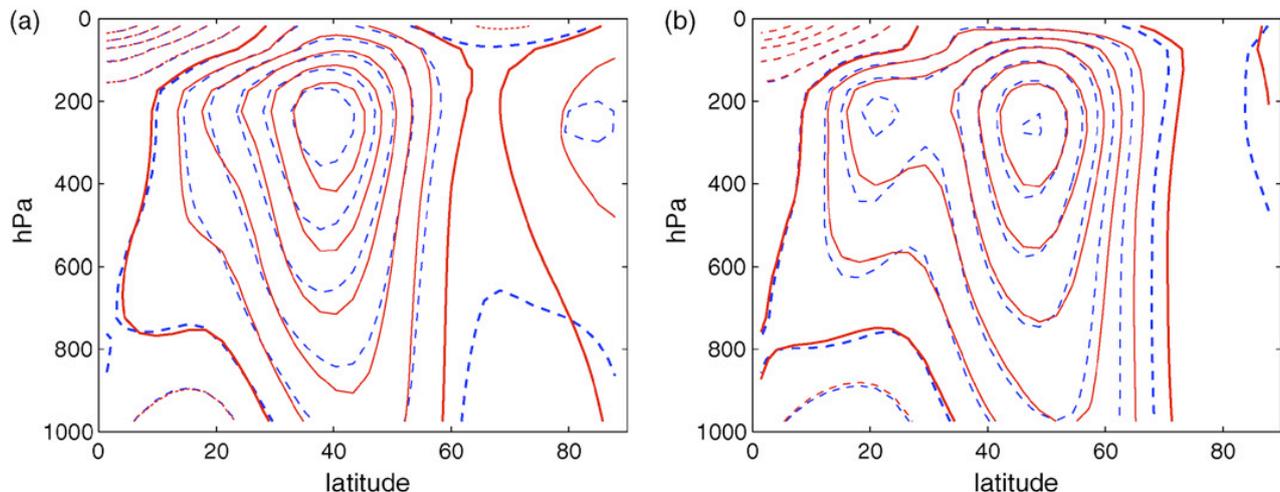
Fig. 17. Zonal winds, longitudinally averaged over the quadrant in which the eddy momentum fluxes are strongest in the simulation of Fig. 13. (a) Average over times when there is one jet (dashed lines) and average over times when the annular mode index is low (solid lines). (b) Average over times when there are two jets (dashed lines) and average over times when annular mode index is high (solid lines). Contour interval is 5 m/s and the slightly thicker line is zero.

and a baroclinic, subtropical jet. The second is the convergence of eddy momentum fluxes in baroclinic eddies. The momentum transfer occurs mainly in the barotropic, decaying phase of the baroclinic life cycle, and leads to the formation of a fairly barotropic jet. As the region of baroclinic instability is generally somewhat poleward of the region of maximum shear (the Eady growth rate is proportional to f times the shear, and so to the meridional temperature gradient) this tends to produce a jet polewards of the subtropical jet. (Also, the stabilizing effects of beta are weaker at higher latitudes, moving the region of maximum baroclinic growth polewards for a given shear.) In practice, the distinction between the thermal and eddy-driven jets is not large and over most of the year and over most longitudes the two jets merge into one, and we commonly observe one broad jet extending from the subtropics through the mid-latitudes. Certainly, after zonal and time averaging often only one jet may be discerned, although in equinoctal conditions two jets may often be discerned, especially in the Southern Hemisphere, even after zonal averaging (S. Lee, pers. comm.). Because the baroclinic subtropical jet tends to be somewhat stronger than the mid-latitude jet, but the momentum fluxes (and hence the surface winds) are co-located with the mid-latitude jet, the maximum in the surface winds tends to be somewhat polewards of the jet maximum aloft.

If there is a zonally asymmetric storm track, as in Figs. 12 and 13, and in the real world, we can expect to see a greater jet separation in the region of strong eddy momentum fluxes, and towards the end of the storm track. Now, the eddy momentum fluxes will of course fluctuate, at times being stronger than at other times. When they are sufficiently strong and sufficiently polewards of the sub-tropical jet we can expect to see a mid-latitude jet that is *distinct* from the sub-tropical jet. Such times will also correspond to periods of a high index of the annular mode or NAO index, and the close correspondence between the two is illustrated in Fig. 17.

To obtain this figure, we first identify the quadrant over which the eddy momentum fluxes are, on average, strongest. We then examine the time evolution of the flow, and objectively partition the flow into three bins: times when there is just one jet, and times when there are two distinct jets, and third, intermediate group where there are two barely distinct jets. We also calculate the first EOF of the surface pressure and its corresponding time-varying principal component, and similarly divide the evolution into three bins. The state in which the annular mode index is high
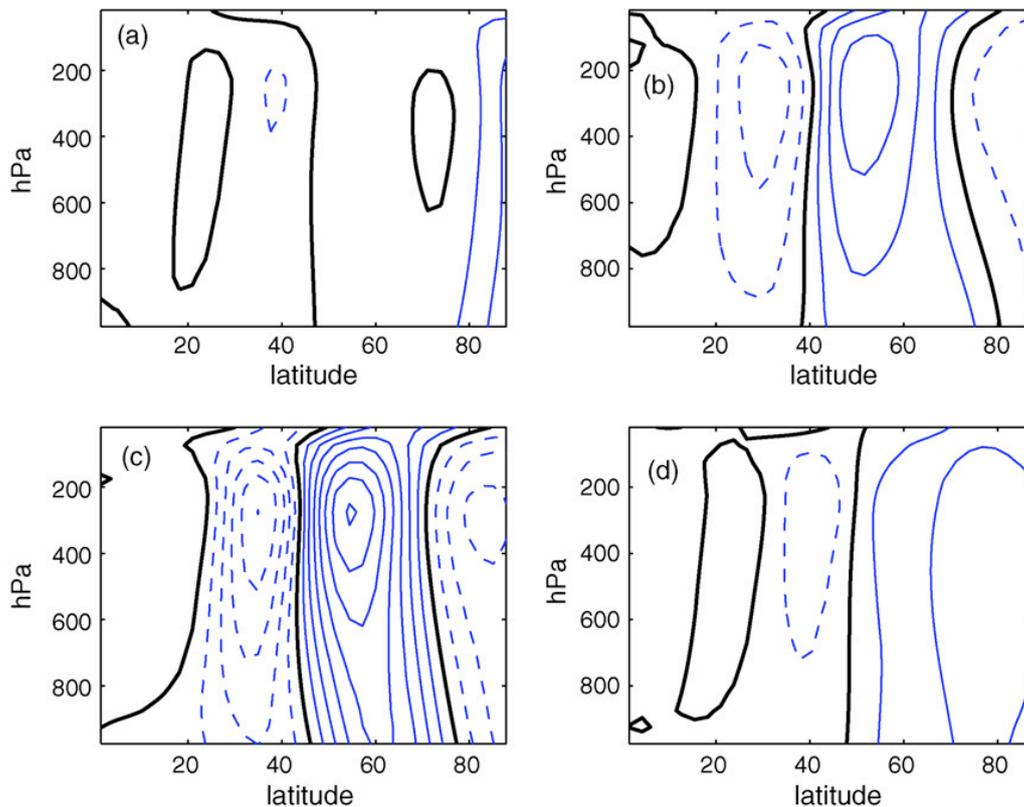
Fig. 18. Regression of the annular mode index onto the quadrant averaged flows (m/s). The third quadrant (c) was the quadrant used for the averaging in Fig. 17, which is the quadrant containing the end of the storm track.

corresponds very closely to the state in which there is a double jet. Indeed, one might go so far as to state that the annular mode index *is* an index of the jet structure, indicating when the jet bifurcates into two. A similar correspondence, although less marked, generally occurs in the other quadrants. [Hartmann and Lo (1998) also noted that transitions between one jet and two jet structures were related to the southern hemisphere 'high latitude mode', a precursor of the what is sometimes now called the Southern Annular Mode.]

We can see the vertical structure of the variability by regressing the annular mode index on to the quadrant averaged flows, as illustrated in Fig. 18. The pattern, most clearly illustrated in panel (c) of the figure, has little vertical phase shift—it is equivalent barotropic, consistent with the largely barotropic nature of the eddy momentum fluxes occurring in the decaying part of the baroclinic life cycle. (Similar patterns arise if we calculate and plot the EOF of the quadrant averaged zonal wind.)

4.3.1. Wave breaking and the phase of the NAO/annular mode

We mentioned above that the location and magnitude of the eddy momentum fluxes determine the location of the eddy-driven jet and so, effectively, the phase of the NAO. The eddy momentum fluxes occur in the later stages of a classical baroclinic life cycle, and are a manifestation of wave breaking (which in turn is a manifestation of the irreversible enstrophy cascade to small scales). It is therefore to be expected that the nature of the wave breaking will influence the phase of the NAO [see Benedict et al. (2004) and, with reference to the southern hemisphere, Hartmann and Zuercher (1998)].

Wavebreaking is essentially of two kinds: anti-cyclonic wavebreaking and cyclonic wave break-ing (AWB and CWB) Thorncroft et al. (1993). The former is characterized by wave breaking
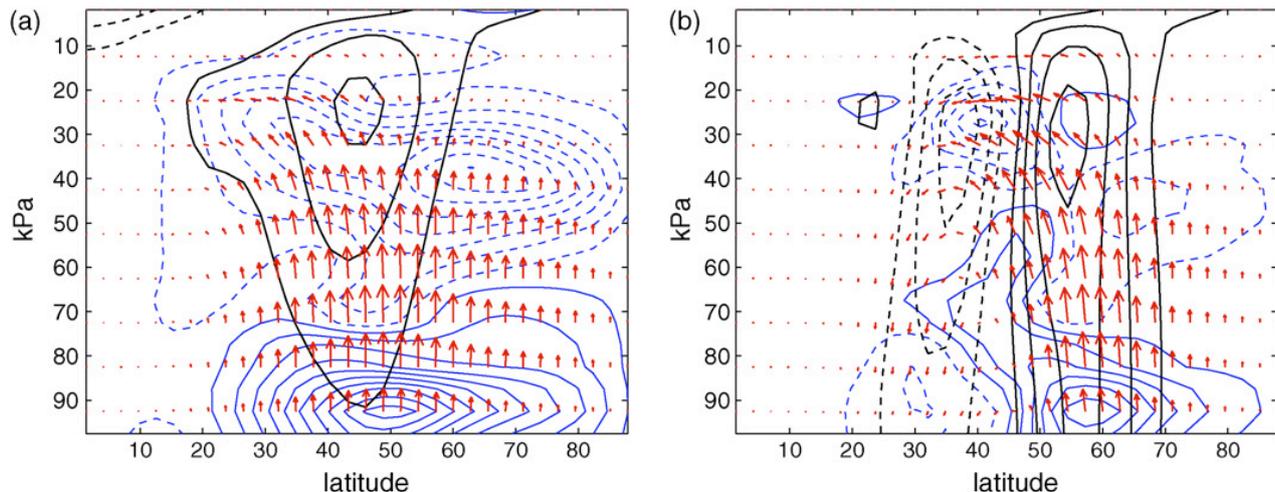
Fig. 19. EP fluxes (arrows), its divergence (blue contours, negative dashed, units of m/(s day)) and zonal wind (black contours) in a statistically zonally symmetric simulation. (a) Time average. (b) The EP fluxes regressed against the annular mode time series, with the fluxes leading the mode by 4 days. The black contours now show the zonal wind anomaly associated with the annular mode. From experiment S1.

equatorwards of the main westerly jet, this being the more common, classical behavior. Cyclonic wave breaking corresponds to breaking polewards of the main westerly flow, and may lead to vortices wrapping up cyclonically. From a more linear perspective, AWB corresponds to an equator-ward flux of wave activity (Eliassen–Palm or E–P fluxes), followed by absorption and dissipation in a critical layer equatorwards of the jet. Similarly, CWB corresponds to a veering poleward flux of wave activity. The effect on the mean flow, or at least the zonal mean flow, can be transparently seen by writing the zonal mean momentum equation in the residual form, as in (3.1). A convergence of the wave activity ($\nabla \cdot \mathscr{F} < 0$) evidently leads to flow deceleration. AWB, which corresponds to a convergence of wave activity equatorwards of the main westerly flow, thus gives rise to flow deceleration in that region, effectively pushing the core of the eddy-driven jet polewards and into a positive phase of the NAO, or a positive annular mode or zonal index. On the other hand, CWB will lead to an equatorwards tendency of the jet. As the scale of the NAO is not so much larger than the scale of single baroclinic eddy, it can be expected that that the phase of the NAO in particular is influenced by wave breaking. Indeed, it seems that a single wave breaking event or single storm can determine the NAO index for period of several days (Franzke et al., 2004; Riviere and Orlanksi, 2007).

The EP fluxes produce a *tendency* in $\bar{u}$, and so can be expected to lead the actual zonal index (i.e., $\bar{u}$), and that the two quantities are connected is illustrated in Fig. 19 (see Limpasuvan and Hartmann (2000) for some related observational plots). The figure shows the EP fluxes, regressed onto the annular mode time series (with its mean removed, and normalized to have unit variance) with the time series offset such that fluxes lead the annular mode by 4 days. The time mean E–P fluxes (which are qualitatively quite similar to those observed in the atmosphere) show a broad region of upward propagation in mid-latitudes (much as is seen in the Eady problem), the fluxes then veering equatorwards aloft and converging, causing the mean eastward flow to decelerate. The regressed pattern, shows a tighter region of upward propagation, offset polewards from the mean, and a greater extent of equatorwards propagation aloft, and this pushes the eddy-driven jet polewards.

We might interpret the regressed pattern as the difference of the fluxes driving the positive index state and those driving the negative state. When the annular mode has a positive tendency,

the E–P fluxes tend to originate from higher latitudes, and travel more towards the equator. When the annular mode is negative, the E–P fluxes tend to originate from the lower latitudes, as the jet (and heat fluxes) have shifted equatorward. There is also more poleward movement of E–P fluxes aloft (or at least less equatorward flux) and less upward flux on the poleward flank of the jet: the opposite pattern as shown. Overall, the pattern suggests that strong E–P fluxes drive the growth of the positive phase of the annular mode, while the negative phase is associated with an equatorward shift and weakening of the eddy fluxes.

5. Discussion

At a fundamental level, the NAO is the pattern of variability associated with the eddy fluxes of the North Atlantic storm track. The eddy momentum fluxes produce a dipolar pattern of variability in the surface fields – such as zonal winds or surface pressure – that is robustly identified by diagnostics such as EOF or teleconnection analysis. The same eddy fluxes produce changes in the zonal index and, when they are statistically zonally symmetric, produce a zonally symmetric EOFs that are meridionally dipolar; these are the annular modes, although they may well have zonally localized teleconnections. This is very far from the whole story, and it may help to express our conclusions (or at least our point of view) as a series of questions and answers, as follows.

- *Why are the pattern of zonal wind variability dipolar?*
  This is ultimately a result of there being only one baroclinic zone. If the variability is small at two latitudes, such as the equator and the pole, then as shown in Section 2, under rather weak assumptions the dominant pattern of variability (meaning the first EOF) of the zonal wind will be dipolar in the meridional direction. Physically, we may think of this as a wobbling of the eddy-driven jet. (The other main modes of variability and their associated patterns are illustrated in Fig. 5.) Momentum conservation is a key aspect in this: if the meridional integral of the surface zonal wind is constrained to be zero, then its dominant pattern of variability (the first EOF) will be dipolar.
- *Why is the pattern equivalent-barotropic?*
  The eddy momentum fluxes occur in the decaying phase of the baroclinic life cycle, where the eddy fluxes are largely barotropic. In a zonally localized storm track, with a preferred location for cyclogenesis, the temporal progression of the baroclinic life cycle is mapped into an eastward spatial progression, and the eastward end of the storm track corresponds to decaying phase of the baroclinic life cycle.
- *What do the two phases of the NAO represent?*
  When the eddy momentum fluxes are strong and are distinctly polewards of the subtropics, then a mid-latitude jet may be generated of comparable magnitude to the sub-tropical jet, and the two jets become distinct. When eddy momentum fluxes are weak and/or close to the subtropics, the mid-latitude jet is weak and the sub-tropical jet dominates. Such differences may also reflect differences in the type of wavebreaking—a cyclonic wavebreaking leading to a more polewards jet, anticyclonic wavebreaking leading to a more equatorwards jet. Additional dynamics may arise because of the jet structure: if the mid-latitude jet and the subtropical jet become too close they may merge; if the eddy momentum fluxes become strong enough and are poleward enough, an existing single jet may quickly break into two. Such dynamics may provide a degree of bimodality to the NAO structure, so that it either appears as one jet or as two well-separated jets, with not much middle ground. However, this remark is conjectural and, to our knowledge, unsupported by observations.

- *What are the timescales of the NAO and zonal index, and what mechanisms are responsible for producing them?*

  The atmosphere is, *inter alia*, a chaotic dynamical system and almost any dynamical field will consequently fluctuate, including the zonal index and the NAO index. An important timescale of these fields is the memory that is associated with them. The decorrelation timescale of the NAO is measured in days, primarily reflecting the variations in wavebreaking that occur on the timescale of baroclinic eddies and the associated variability of the storm track. Variations in the stationary wave pattern may provide a longer (i.e., weeks) timescale to variations in the NAO index, and still slower forcing from either the ocean or stratosphere may lead to a tendency for the NAO to exist in a preferred state over the course of a season, or longer, although we should remember that the synoptic variability is stronger than the interannual variability.

  Changes in the zonal index do tend to occur on a somewhat longer timescale than that of baroclinic eddies. Various models and mechanisms have been put forward, and some are summarized in this paper (in Section 3.1 we discuss the tendency of a zonal flow to remain at its current latitude, possibly providing some persistence to the NAO index). However, in our view none of these makes a strong enough case for us to regard the issue as wholly settled, and the relative importance of all these potential mechanisms remains to be elucidated.

- *Should we consider the NAO to be a local or a regional or hemispheric phenomenon? What is its relation to annular modes?*

  We have seen that a zonally symmetric first EOF will result from dynamics that are statistically zonally symmetric, but in which the zonal correlations are quite local. That is, the presence of a nearly zonally symmetric EOF does not imply dynamics on a hemispheric scale. Indeed, in the atmosphere, there is little correlation between locations in the Pacific and in the Atlantic except on long timescales, and although other possibilities exist the simplest explanation of this is that the dynamics themselves do not have robust structures on such large scales.

  However, this is not to say that the dynamics of the NAO are local, because it is the stationary wave pattern that organizes the baroclinicity and allows there to be a storm track in the first instance. At leading order, the Atlantic storm track may be thought of as a consequence of the Rockies and the thermal contrasts provided by the North American continent and the North Atlantic ocean, and these features have a near-hemispheric scale. Further, the Pacific storm track never really dies over North America, and its lingering presence (combined with the enhanced baroclinicity over eastern North America and the western North Atlantic) provides the seed – a finite amplitude seed – of the Atlantic storm track. The state of Pacific storm track may thus provide a partial determinant of the index of the NAO.

  In this picture the annular mode itself should not be considered a dynamical mode of variability, like the eigenmode of a linear system. The annular mode is not a zonally symmetric linear response to a stimulus. But nor are the dynamics wholly local, for they involve the near-hemispheric stationary wave pattern.

- *What is the role of the stratosphere?*

  As we ascend into the stratosphere, the dynamics themselves become more zonally symmetric and the annular mode gains more dynamical significance and its time becomes longer. There is also some evidence (e.g., Baldwin et al., 2003; Polvani and Kushner, 2002; Scaife et al., 2005) that the stratosphere can influence the NAO, and given the somewhat longer timescales in the stratosphere it may provide a source of predictability. However, the extent of this remains an open question.

- *What is the role of the ocean?*

The NAO, and annular modes, are certainly primarily modes of atmospheric variability that would exist in the absence of an ocean. The ocean is nevertheless important in two regards. First, it provides zonally asymmetric thermal forcing that is one of the reasons that a localized storm track, and hence the NAO, exists in the first instance. Second, given that the timescales of the ocean are much longer than those of the atmosphere, variability in the ocean may be the source of variability in the NAO on interannual and longer timescales. However, the importance of this remains an open question.

- *Why is the NAO more obvious than an equivalent pattern in the Pacific?*

There may be two reasons for this. First, an EOF analysis over the Pacific will also identify other sources of variability, for example that associated with El Niño and the Southern Oscillation. Second, the Pacific storm track, although as strong as that over the Atlantic, does not completely die out over North America (Fig. 2), whereas the Atlantic storm track dies rather abruptly over western Europe, and there is little of it remaining east of the U.K. and Scandinavia, so concentrating the eddy momentum fluxes. However, observations of the eddy momentum fluxes do not unambiguously support this hypothesis, and we do not consider the relation between the NAO and eddy momentum fluxes, or the zonal pattern of eddy momentum fluxes in the Pacific and Atlantic storm tracks, to have been completely explained.

- *Are our current appellations appropriate?*

The 'North Atlantic Oscillation' is not a simple oscillation like the pendulum. Nevertheless, there is an average state to the zonal flow across the Atlantic, and if the NAO index is highly positive or negative there will be a general tendency for it to revert to that average value. Regarding annular patterns, referring to more zonally symmetric patterns of variability as annular *modes* seems appropriate if one uses the word mode is the common, non-technical sense of 'a particular form or variety of something' or 'a manifestation, form or arrangement' (Mirriam Webster). However, there is less reason to suppose that an annular mode, at least in the troposphere, represents a linear mode of response, such as that which occurs when striking a bell with a gong, and annular 'patterns' may be a more neutral, albeit less memorable, moniker.

Acknowledgements

References

Ambaum, M.H.P., Hoskins, B.J., Stephenson, D.B., 2001. Arctic oscillation or North Atlantic oscillation? J. Climate 14, 3495–3507.

Andrews, D.G., McIntyre, M.E., 1976. Planetary waves in horizontal and vertical shear: the generalized Eliassen–Palm relation and the mean zonal acceleration. J. Atmos. Sci. 33, 2031–2048.

Ashmead, H.G., 1884. History of Delaware County, Pennsylvania. L. H. Everts and Co., Philadelphia.

Baldwin, M.P., Stephenson, D.B., Thompson, D.W.J., Dunkerton, T.J., et al., 2003. Stratospheric memory and skill of extended-range weather forecasts. Science 301, 636–640.

Benedict, J.J., Lee, S., Feldstein, S.B., 2004. Synoptic view of the North Atlantic oscillation. J. Atmos. Sci. 61, 121–144.

Cash, B.A., Kushner, P., Vallis, G.K., 2002. The structure and composition on the annular modes in an aquaplanet General Circulation Model. J. Atmos. Sci. 59, 3399–3414.

Cash, B.A., Kushner, P., Vallis, G.K., 2005. Zonal asymmetries, teleconnections, and annular patterns in a GCM. J. Atmos. Sci. 62, 207–219.

Chang, E.K.M., 1993. Downstream development of baroclinic waves as inferred from regression analysis. J. Atmos. Sci. 50, 2038–2053.

Chang, E.K.M., Lee, S., Swanson, K.L., 2002. Storm track dynamics. J. Climate 15, 2163–2183.

Crantz, D., 1767. *The History of Greenland.* London.

Edmon, H.J., Hoskins, B.J., McIntyre, M.E., 1980. Eliassen–Palm cross sections for the troposphere. J. Atmos. Sci. 37, 2600–2616.

Feldstein, S.B., Franzke, C., 2006. Are the North Atlantic Oscillation and the Northern Annular Mode distinguishable? J. Atmos. Sci. 129, 2915–2930.

Feldstein, S.B., Lee, S., 1998. Is the atmospheric zonal index driven by an eddy feedback? J. Atmos. Sci. 55, 3077–3086.

Franzke, C., Lee, S., Feldstein, S.B., 2004. Is the North Atlantic Oscillation a breaking wave? J. Atmos. Sci. 61, 145–160.

Gerber, E.P., Vallis, G.K., 2005. A stochastic model for the spatial structure of annular patterns of variability and the NAO. J. Climate 18, 2102–2118.

Gerber, E.P., Vallis, G.K., 2007. Eddy-zonal flow interactions and the persistence of the zonal index. J. Atmos. Sci. 64, 3296–3311.

Hartmann, D.L., Lo, F., 1998. Wave-driven zonal flow vacillation in the Southern Hemisphere. J. Atmos. Sci. 55, 1303–1315.

Hartmann, D.L., Zuercher, P., 1998. Response of baroclinic life cycles to barotropic shear. J. Atmos. Sci. 55, 297–313.

Hurrell, J.W., 1995. Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. Science 269, 676–679.

James, P.M., Fraedrich, K., James, I.N., 1994. Wave-zonal flow interaction and ultra-low-frequency variability of the mid-latitude troposphere. Quart. J. R. Meteor. Soc. 120, 1045–1067.

Lee, S., Feldstein, S., 1996. Mechanism of zonal index evolution in a two-layer model. J. Atmos. Sci. 53, 2232–2246.

Limpasuvan, V., Hartmann, D.L., 2000. Wave-maintained annular modes of climate variability. J. Climate 13, 4414–4429.

Lorenz, D.J., Hartmann, D.L., 2001. Eddy-zonal flow feedback in the Southern Hemisphere. J. Atmos. Sci. 58, 3312–3327.

Namias, J., 1950. The index cycle and its role in the general circulation. J. Meteor. 7, 130–139.

Namias, J., 1969. Seasonal interactions between the North Pacific Ocean and the atmosphere during the 1960's. Mon. Wea. Rev. 97, 173–192.

Orlanski, I., Chang, E.K.M., 2000. The life cycle of baroclinic eddies in a storm track environment. J. Atmos. Sci. 57, 3498–3513.

Polvani, L.M., Kushner, P.J., 2002. Tropospheric response to stratospheric perturbations in a relatively simple general circulation model. Geophys. Res. Lett. 29 (7), 1114.

Randel, W.J., Held, I.M., 1991. Phase speed spectra of transient eddy fluxes and critical layer absorption. J. Atmos. Sci. 48, 688–697.

Riviere, G., Orlanksi, I., 2007. Characteristics of the Atlantic storm-track activity and its relation with the North Atlantic Oscillation. J. Atmos. Sci. 64, 241–266.

Robinson, W.A., 1996. Does eddy feedback sustain variability in the zonal index? J. Atmos. Sci. 53, 3556–3569.

Robinson, W.A., 2000. A baroclinic mechanism for eddy feedback on the zonal index. J. Atmos. Sci. 57, 415–422.

Rossby, C.-G., Collaborators, 1939. Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semi-permanent centers of action. J. Mar. Res. 2, 38–55.

Scaife, A., Knight, J., Vallis, G., Folland, C., 2005. A stratospheric influence on winter climate in the North Atlantic region. Geophys. Res. Lett. 32, L18715, doi:10.1029/2005GL023226.

Simmons, A., Hoskins, B., 1978. The life-cycles of some nonlinear baroclinic waves. J. Atmos. Sci. 35, 414–432.

Stephenson, D.B., Wanner, H., Brönniman, S., Luterbacher, J., 2003. The history of scientific research on the North Atlantic Oscillation. In: Hurrell, J.W., Kushnir, Y., Ottersen, G., Vis-beck, M. (Eds.), The North Atlantic Oscillation: Climate Significance and Environmental Impact. American Geophysical Union, pp. 37–50.

Thompson, D.W.J., Lee, S., Baldwin, M.P., 2003. Atmospheric processes governing the northern hemisphere annular mode/North Atlantic oscillation. In: Hurrell, J.W., Kushnir, Y., Ottersen, G., Visbeck, M. (Eds.), The North Atlantic Oscillation: Climate Significance and Environmental Impact. American Geophysical Union, pp. 81–112.

Thorncroft, C.D., Hoskins, B.J., McIntyre, M.E., 1993. Two paradigms of baroclinic-wave life-cycle behaviour. Quart. J. R. Meteor. Soc. 119, 17–55.

Vallis, G.K., 2006. Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation. Cambridge University Press, 745 pp.

Vallis, G.K., Gerber, E.P., Kushner, P.J., Cash, B.A., 2004. A mechanism and simple dynamical model of the North Atlantic Oscillation and annular modes. J. Atmos. Sci. 61, 264–280.

van den Dool, H.M., Saha, S., Johansson, A., 2000. Empirical orthogonal teleconnections. J. Climate 13, 1421–1435.

van Loon, H., Rogers, J.C., 1978. The seesaw in winter temperatures between Greenland and Northern Europe. Part I: General description. Mon. Wea. Rev. 106, 296–310.

Walker, G.T., Bliss, E.W., 1932. World weather V. Memoirs R.M.S. 4, 53–83.

Wallace, J.M., 2000. North Atlantic Oscillation/Annular Mode: two paradigms—one phenomenon. Quart. J. Roy. Meteor. Soc. 126, 791–805.

Wallace, J.M., Gutzler, D.S., 1981. Teleconnections in the geopotential height field during the northern hemisphere winter. Mon. Wea. Rev. 109, 784–812.

Wanner, H., Bronnimann, S., Casty, C., Gyalistras, D., et al., 2001. North Atlantic Oscillation—concepts and studies. Surv. Geophys. 22, 321–381.

Watterson, I.G., 2002. Wave-mean flow feedback and the persistence of simulated zonal flow vacillation. J. Atmos. Sci. 59, 1274–1288.

Wittman, M.A., Charlton, A.J., Polvani, L.M., 2005. On the meridional structure of annular modes. J. Atmos. Sci. 18, 2119–2122.