CHAPTER

4

Identifying Wave Processes Associated With Predictability Across Time Scales: An Empirical Normal Mode Approach

Gilbert Brunet[*], *John Methven*[†]

[*]Meteorological Research Division, Environment and Climate Change Canada, Dorval, QC, Canada [†]Department of Meteorology, University of Reading, Reading, United Kingdom

OUTLINE

1 Introduction 66

2 Partitioning Atmospheric Behavior Using Its Conservation Properties 68
- 2.1 *Partitioning Variability: Background State and Wave Activity* 69
- 2.2 *Wave Activity Conservation Laws* 74
- 2.3 *The Implications of Wave-Activity Conservation for Modes of Variability* 77

3 The ENM Approach to Observed Data and Models and Its Relevance to S2S Dynamics and Predictability 78
- 3.1 *ENMs: Bridging Principal Component, Normal Modes, and Conservation Laws* 79
- 3.2 *ENM in Applications Relevant to Predictability Across Time Scales* 83
- 3.3 *ENM Application to the Atmospheric S2S Variability* 86

4 Conclusion 89

Acknowledgments 90

Sub-seasonal to Seasonal Prediction
https://doi.org/10.1016/B978-0-12-811714-9.00004-8

1 INTRODUCTION

It can be demonstrated that predictive skill ranges from weeks out to seasons when the forecast metric involves the statistical treatment of weather variables. Typically, the range of regional forecast skill increases with the scale of the region considered, as well as the length of the time window used for verification (see Chapter 2). Comparisons can be made in terms of averages over forecast lead times, over ensemble members, or, in a probabilistic sense, using ensemble forecasts. These extended-range forecasts go beyond the limit of predictability for point forecasts (the value of an atmospheric variable at a particular time and location). Such forecasts depend on the nature of the variable being forecast and the phenomena dominating its fluctuations. For example, geopotential on a pressure surface is a smooth field dominated by synoptic-scale (or larger) weather systems and is predictable out to 7–15 days, depending upon the flow configuration. Finer-scale fields, like vorticity, have shorter predictability limits. For example, Frame et al. (2015) showed that the predictive skill for the strike probability of cyclonic vorticity centers, within a given radius of locations in the Euro-Atlantic sector, increases with feature intensity and scale. Precipitation typically has a much shorter limit to its predictability, owing to its dependence on vertical motion and convective-scale features in the flow.

There are many plausible sources for sub-seasonal to seasonal (S2S) predictability, including slowly varying boundary conditions for tropospheric weather systems: coupling with the ocean (see Chapters 5 and 9), land surface processes (see Chapter 8), the cryosphere (see Chapter 10), and the stratosphere (see Chapter 11). However, sub-seasonal regimes are characterized by large-scale patterns of variability that exhibit internal variability over long time scales, and they can be oscillatory in nature or characterized by long-range teleconnections. Consider three contrasting examples:

1. The weather in the tropics and the extratropics is influenced by the phase of the Madden-Julian Oscillation (MJO). The MJO is the dominant tropical mode of sub-seasonal variability, which propagates eastward along the equator but has strong remote influences (see Chapters 5 and 7).
2. Unprecedented extreme summer precipitation events in western Europe have occurred in the last decade, and they have been associated with quasi-stationary Rossby wave patterns on the midlatitude jet stream (Blackburn et al., 2008), prompting new theories of resonant excitation of Rossby modes (Petoukhov et al., 2013; Coumou et al., 2014).
3. The Russian heatwave that occurred in 2010 was associated with a persistent midlatitude blocking regime (Dole et al., 2011).

In these examples, a distinct dynamical phenomenon is involved, and the properties of that phenomenon influence the weather in a predictable manner. Here, we focus on isolating oscillatory phenomena or slowly propagating modes responsible for enhanced predictability in the S2S range. In addition to decoupling the phenomena from the data, it is important to gain insight into their intrinsic dynamical properties and interactions in order to anticipate how they will contribute to predictability. Also, in the context of a changing climate, if we understand how modes of variability relate to the climate state, we will be able to anticipate how variability and predictability may vary with climate change.

Energy spectra in the wave-number space (calculated by transform of spatial distributions and averaging over many realizations) show a smooth continuum from the planetary to the kilometer scale, indicating that there is no spectral gap distinguishing large-scale phenomena from smaller-scale ones. Similarly, frequency spectra exhibit a smooth continuum from seasons to hours. However, statistical techniques that utilize both spatial and temporal information, such as the popular empirical orthogonal function (EOF) technique, show that covariance in time is typically dominated by large-scale patterns of variability. The EOF technique for a given variable maximizes the variance, which is explained by a series truncation of fixed length.

A remarkable property of an EOF mathematical construction is that EOF patterns are orthogonal to one another (i.e., the integral over the domain of two different EOF patterns multiplied together equals zero) and their corresponding time series are orthogonal as well, making them a complete basis that can be used to project variability from a discrete data set. Hence, the first EOF is the spatial pattern that explains most temporal variance; the first and second EOFs form the two-dimensional (2D) orthogonal basis that explains the most variance, and so on for higher terms in the series. The disadvantage of using a purely statistical approach is that the spatial structures obtained (the EOFs) and their corresponding principal component time series do not have distinct physical properties; therefore, it is difficult to anticipate their behavior beyond the limits of the time series examined.

A relatively unexploited approach is to combine conservation laws derived from consideration of atmospheric dynamics with the orthogonality approach to identify distinct patterns of variability that can be linked to characteristic physical properties (e.g., an intrinsic frequency or a phase speed). This approach can be called *empirical normal mode (ENM) analysis.* Just as an analysis of the shape and physical construction of a bell can be used to anticipate the frequency at which it will ring when struck, the spatial structure of an ENM can be used to predict its intrinsic frequency or phase speed from conservation properties. If it can be shown in general that a small number of such modes dominate S2S variability, as was demonstrated by Brunet (1994) for the 315 K isentropic surface, a huge reduction in the dimensional size of the system to be solved will result. Hence, the ENM basis could be the natural one to use to study waves and weather regimes in low-order dynamical systems, as discussed in Chapter 6.

Other disturbances may perturb the modes over a wide range of frequencies or stochastically, which causes the observed frequency spectrum to resemble a continuum. Therefore, knowledge of their intrinsic frequencies provides potentially useful information. The well-known fluctuation-dissipation theorem (FDT) describes how the time-average response of a dynamic system to random perturbations will resemble the structure of the slowest (longest-time-scale), unforced mode of variability. For example, Ring and Plumb (2008) studied the response of the Southern Annular Mode to forcing by drawing on the FDT. They used the principal oscillation pattern (POP) analysis developed by Hasselman (1988) and Penland (1989), which relies on the calculation of lag covariances to obtain the temporal behavior of spatial patterns. An advantage of the ENM technique is that lag covariances are not required because the frequency information stems from the dynamical properties.

A popular approach used in both weather and climate sciences is *composite sampling* to study long geophysical time series. With this method, statistical properties (e.g., average and standard deviation) of similar segments of a time series are examined, and a given segment is included in the composite if a characteristic event occurs within it. There is a vast body

of literature describing such approaches; these studies use atmospheric reanalysis time series to link weather-climate events (e.g., Molteni et al., 1988; Robertson and Metz, 1989; Cotton et al., 1989; Vautard, 1990; Ferranti et al., 1989; Lin and Brunet, 2009). For example, Asaadi et al. (2016a, 2017) used this approach to identify fundamental dynamical and physical processes related to hurricane genesis. They showed that the coexistence of an African easterly wave nonlinear critical layer and a region of a weak meridional potential vorticity (PV) gradient over several days might be a major factor determining whether tropical disturbances develop into hurricanes. This finding answered the long-standing question of why only a small fraction of African easterly waves contribute to hurricane genesis. The study showed a way how S2S variability can modulate the hurricane season.

In general, the utilization of models of varying complexity is also needed for understanding and identifying sources of predictability. For example, with regard to sub-seasonal variability, this approach made it possible to demonstrate for the first time a two-way linkage between the Madden Julian Oscillation (MJO) and the Arctic Oscillation (AO) in a simplified general circulation model (GCM), numerical weather prediction (NWP) systems, and observations. These studies have pointed the way toward improving S2S predictive skill and provided examples of how global teleconnections are influencing regional weather in more complex modeling systems (Lin et al., 2007, 2009, 2010a; Lin and Brunet, 2011); also see Chapter 7.

A key message arising from this work is that even though the atmosphere is chaotic and stirring by large-scale waves and eddies generate fine-scale structures in air masses and conserved properties such as PV, the large-scale dynamics may be closer to linear behavior than expected. Characteristic spatial patterns that vary slowly but are not periodic in nature are often treated as oscillations. Teleconnections typically fall in this category. There are many situations in which interactions between wave phenomena can be identified and explained mechanistically using linear dynamics. The purpose of this chapter is to provide a theoretical and statistical framework for studying the sub-seasonal predictability in observational and model data using these concepts in an integrated manner.

Section 2 introduces the components of the framework, including the notional partition between a background state and perturbations to it, the concept of wave activity conservation, and the implications of conservation for modes of variability. Section 3 outlines the ENM technique and presents some implications for the behavior of perturbations that can be deduced from the approach. Several applications of the approach to global atmospheric data are presented to illustrate the technique and its potential. Conclusions are presented in Section 4.

2 PARTITIONING ATMOSPHERIC BEHAVIOR USING ITS CONSERVATION PROPERTIES

There are two major aims of the approach to understanding atmospheric variability presented here:

- Development of a theoretical framework that is capable of isolating a slowly varying component of the atmosphere that is influenced by slow processes, such as radiative forcing, and described by well-known equations. This phenomenon may be considered to be linked with climate.

- Development of a technique to isolate coherent, dynamical modes of variability from observed global data and models. These modes have intrinsic properties that can be deduced from theory.

In spite of the fact that there is a continuum of complex behavior (including nonlinear interactions) across scales, our goal is to take the underpinning theory as far as possible in terms of isolating processes from observations and representing them in forecasts. The robustness of the approach will be tested with global reanalysis data. Possible applications that result from the deductions about forcing of variability, dynamical processes and wave resonance also will be discussed.

Areas where this approach could be useful include the following:

- Anticipating how variability might change with changing climate by increasing our understanding of the properties of dynamical modes of variability and their dependence on the background state
- Identifying the physical links between large-scale modes of variability and high-impact weather that typically occurs on smaller scales
- Using these links to forecast the likelihood of high-impact weather, even though representing high-impact weather itself in models may be very challenging
- Diagnostics to identify model errors in the representation of dynamics

Examples include the risk of extreme precipitation that is conditional on the phase of the MJO and persistent midlatitude weather extremes associated with a particular phase of quasi-stationary Rossby waves, such as the extremely wet summers that occurred in western Europe in 2007 and 2012 (Blackburn et al., 2008; de Leeuw et al., 2016), or the Russian heatwave in 2010 (Dole et al., 2011).

2.1 Partitioning Variability: Background State and Wave Activity

Typically, atmospheric variability is identified through statistical analysis that does not explicitly use the properties of atmospheric dynamics. A starting point is to identify an anomaly: it could be some form of readily recognized coherent structure in the atmosphere (such as a tropical cyclone), but more often it is obtained by subtraction of some form of mean state to define perturbation fields at every point in the model:

$$q' = q - q_0 \tag{1}$$

These perturbations are then analyzed statistically. In such an approach, the definition of perturbations and their properties clearly depends on the mean state, q_0, that is chosen.

The three most popular approaches for defining mean state are as follows:

1. *Global average*. In this case, the system can be described by the global integral of the evolution equations, but all the dynamics are contained entirely within the perturbations that dominate the atmospheric response to forcing (e.g., the global temperature response to greenhouse gas forcing or volcanic eruption).
2. *Eulerian time average (at fixed locations)*. The mean state can readily be calculated from data, but it is not a complete solution of the governing equations. Forcing from eddy fluxes

must be added. All the time dependence rests within the perturbations. One example of this approach is forecasting a regional seasonal temperature anomaly in contrast to numerical weather prediction (NWP) of the total temperature field.

3. *Eulerian zonal average (average around latitude circles).* Although it is often used in dynamic meteorology, as in the wave-mean flow interaction problem, this approach has a disadvantage: The mean can vary as quickly as the perturbations because it may be changed by adiabatic eddy fluxes.

Research in the 1960s, 1970s, and 1980s showed that the evolution of the mean meridional circulation (a zonal mean, usually time-filtered) depends crucially on the variables chosen to describe the data (Andrews et al., 1987). In particular, there is a very marked difference in the mean state deduced by averaging pressure-level data (the easiest coordinate to use in assimilating observational profile data), compared with averaging along isentropic levels (surfaces of constant potential temperature). The key reason is that potential temperature is materially conserved following adiabatic motion, so isentropic-coordinate averages partition diabatic behavior from adiabatic behavior. In contrast, vertical motion across pressure surfaces can occur through both adiabatic and diabatic processes. A major example is the Ferrell cell, which appears in the midlatitudes as a thermally indirect circulation in the pressure-coordinate average, but is absent in the isentropic-coordinate average (Townsend and Johnson, 1985). The origin of this structure is predominantly adiabatic motion along sloping isentropic surfaces within midlatitude weather systems.

An extension of the isentropic-coordinate approach involves using coordinates where two of the variables are properties that are approximately materially conserved. The most common example is to use PV in conjunction with potential temperature (e.g., Nakamura, 1995). If motions are adiabatic and frictionless, then the PV and potential temperature cannot be modified along trajectories. If surfaces of constant potential temperature and PV intersect, those intersections must be transported around by the fluid as material lines. For example, the midlatitude tropopause is often described as a particular PV surface (2 PVU), and the position of the tropopause on each isentropic surface that intersects it is then stirred by the fluid motion along that surface (Hoskins et al., 1985). However, for conservative motion, mass cannot be transported across isentropic surfaces or PV surfaces, nor can it be transported across the intersections between the surfaces. This is a strong constraint on fluid behavior.

In a conceptual sense, we can describe the entire atmosphere in conserved variable coordinates (i.e., a 2D plane with potential temperature and PV as the axes), where motion of atmospheric mass is possible only through the action of diabatic or frictional processes. This framework is described as a *modified Lagrangian mean (MLM)* (McIntyre, 1980). A true Lagrangian mean would require calculation of the trajectories of all air parcels and some form of time-averaging along trajectories and over trajectories from similar initial conditions. In contrast, the MLM state is obtained by using the approximately conserved variables as markers of air masses, and therefore tracers of fluid motion. This has a distinct advantage. The time-dependent winds stir tracers so that fine-scale structures are developed through chaotic advection, and there is a cascade to ever-finer-scale structures in the absence of nonconservative processes. However, real tracers including chemical constituents, as well as PV and potential temperature, are subject to nonconservative processes that act to dissipate the smallest features (halting the scale cascade) and also act to maintain large-scale contrasts

(so that the entire atmosphere does not become well-mixed). The MLM approach takes advantage of this property.

Another crucial aspect of introducing a partition between a background state and perturbations about that state is that it is necessary to predict the evolution of both components using the equations of motion and thermodynamics. First, we will consider the background state (the evolution of the perturbations will be considered in Section 2.2). An important way to predict background state evolution is to use the integral conservation properties of the full state. The definition of the MLM state uses two material conservation properties. From this, it is possible to deduce that mass cannot cross surfaces of constant potential temperature (θ) or constant Ertel PV (q) if the flow is conservative (adiabatic and frictionless). Therefore, the mass enclosed in a volume bounded above and below by two neighboring isentropic surfaces ($\Delta\theta$) and laterally by a PV contour (with value Q) must be conserved:

$$M(Q,\theta) = \frac{1}{\Delta\theta}\iiint r\, dA\, d\theta \tag{2}$$

where r is the density in isentropic coordinates and A is the area enclosed.

Kelvin's circulation theorem also states that the circulation within any closed material contour on an isentropic surface is invariant if the flow is adiabatic and frictionless. A set of circulation integrals can be defined by the integral of the tangential absolute velocity (in an inertial frame) around all closed PV contours (i.e., varying Q):

$$C(Q,\theta) = \oint_{q=Q} \underline{u}.d\underline{l} = \langle u_t \rangle L_Q \tag{3}$$

where the average tangential speed of the flow around the circuit, $\langle u_t \rangle$, must depend inversely upon the length of the circuit, L_Q. Although the mass enclosed by the circuit is invariant, the length of its boundary depends upon the degree of contortion of the PV contour by the flow. Therefore, the average speed and local velocity must depend on the shape of PV contours. To specify the background state and perturbations to it completely, it is necessary to make an assumption about the shape of the PV contours that define the background state.

One option is to define the MLM background state as zonally symmetric and to require it to obey the same equations of motion as the full flow. This can be achieved by an adiabatic rearrangement, in which the mass and circulation enclosed by every PV contour in isentropic layers are the same as in the full state. The geometry of the contours is made to be concentric around latitude circles (McIntyre, 1980; Methven and Berrisford, 2015). Using Stokes's theorem, the circulation also can be expressed as a volume integral of PV:

$$C(Q,\theta) = \frac{1}{\Delta\theta}\iiint rq\, dA\, d\theta \tag{4}$$

because the absolute vorticity on an isentropic surface is given by rq.

Consider a thought experiment where there is a distorted polar vortex in an isentropic layer characterized by uniform high PV (value Q) inside the vortex and zero PV outside. For illustration purposes, perturbations in isentropic density along the layer are assumed to be small compared with the mean density, R. In this case, the circulation within the wavy

contour is approximately RQA, where A is the area enclosed. Therefore, the background state zonal flow around the vortex edge is

$$u_0 = \frac{RQA - C_p}{L_0} \tag{5}$$

where L_0 is the length of the latitude circle encompassing a vortex of area A that is concentric about the pole in the background state. It is usual to define an equivalent latitude for the contour such that $L_0 = 2\pi a(\pi/2 - \phi_e)$ and $A = 2\pi a^2(1 - \sin\phi_e)$, where a denotes the Earth's radius. The area integral of the vertical component of the planetary vorticity ($2\Omega \sin\phi$) within the background state contour is $C_p = \Omega.\, 2\pi a^2(1 - (\sin\phi_e)^2)$. It is subtracted to obtain the zonal flow in the rotating frame of the Earth. The zonal average of the flow around latitude ϕ_e is then given by

$$[u] = \frac{RQ(A - B) - C_p}{L_0} \tag{6}$$

where B is the area within the latitude circle that is external to the disturbed vortex where the PV is zero. Two properties are immediately apparent. If there is any disturbance, we expect $[u] < u_0$ because $0 < B < A$, and we also can anticipate that the Eulerian zonal average $[u]$ will fluctuate with the disturbance amplitude, as characterized by the area occupied by low PV ridges (B).

The flow and density can be obtained by inversion of the PV distribution (Methven and Berrisford, 2015). In this way, we can define the distribution of dynamical atmospheric variables and their evolution. Because the PV and θ distributions of the MLM state are zonally symmetric, the zonal flow obtained from PV inversion must be symmetric as well. Furthermore, because the zonal flow is parallel to the PV contours along isentropic surfaces, there can be no change associated with advection. Without additional approximations, the solution to the primitive equations on the sphere in this situation is a state that is in hydrostatic and gradient wind balance. Because the zonal integral of the full state conserves zonal angular momentum, so must the zonally symmetric background state due to translational invariance in the zonal direction. The perturbations also must obey a pseudoangular momentum conservation law, as described in Section 2.2.

An alternative would be to define the background as a strictly steady state. In this case, the time invariance implies energy conservation of the background, as well as global energy conservation of the full state, and the perturbations obey a pseudoenergy conservation law. However, there are several disadvantages to this approach. The background state is not exactly in balance if the flow is zonally asymmetric, so the PV distribution cannot, in general, be inverted to obtain the flow and density. Furthermore, except in a special situation where the flow happens to be parallel to PV contours everywhere, the background could not be steady without the continuous action of forcing introduced into the evolution equations.

Therefore, a compromise must be made. Either we identify the background state with zonal symmetry, in which case even stationary waves and zonal variations in climate must be regarded as part of the perturbation field, or we identify the background with a steady state (time symmetry), in which case the evolution of the background is not considered (by definition), and the maintenance of the background also would require a forcing term

in the equations. Most people would define the term *climate* using some notional component that is slowly varying and inherently large-scale, but this approach loses the advantages of precise symmetry in space or time.

In the analysis of perturbations that follows in this chapter, we will use a zonally symmetric background state, but also assume that it evolves much more slowly than the perturbations. As explained previously, this is approximately true for the MLM state because it can evolve only through nonconservative processes, and these modify global circulation only slowly. Fig. 1 illustrates the evolution of the MLM state for the Northern Hemisphere in June 2007, obtained using the equivalent latitude iteration by PV inversion (ELIPVI) method of Methven and Berrisford (2015).

It is immediately apparent that $u_0 > [u]$, and also that $[u]$ varies more rapidly, as predicted previously. The solid black contour on both plots marks the tropopause (PV = 2 PVU;

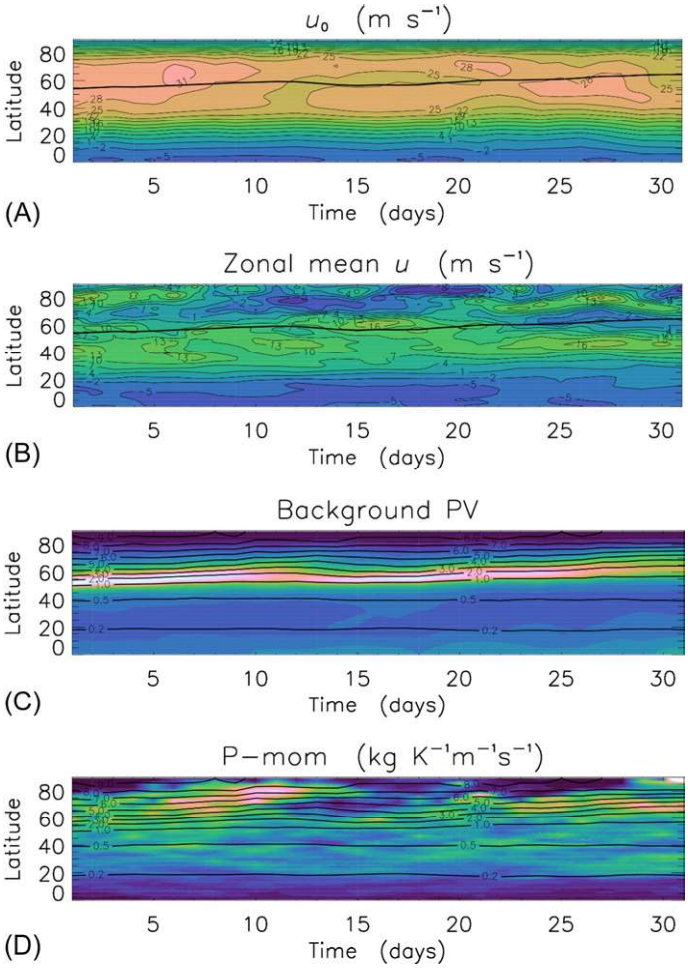

FIG. 1 Evolution of the atmosphere on the 320K isentropic surface calculated from daily ERA-Interim data for June 2007. (A) MLM background state zonal flow. (B) Eulerian zonal mean flow. (C) PV (PVU) in contours; color shading indicates the meridional PV gradient (the largest values are pink). (D) Wave activity (pseudomomentum density) in color shading overlain on background state PV contours (CI = 5000 up to 153,000 kg K^{-1} m^{-1} s^{-1} shown in lightest *pink*).

PVU $= 10^{-6}$ kg^{-1} s^{-1} K m^2) on the 320K isentropic surface. It is obvious that it migrates poleward over the month, aside from a brief period in the middle. Fig. 1C shows the MLM state PV distribution on the 320K surface. All the midlatitude PV contours are migrating poleward, which can occur only through nonconservative (diabatic or frictional) processes. The meridional gradient of PV (color shading) also migrates with the tropopause on this surface. This is the seasonal march of the tropopause.

The mechanism is "vortex erosion" (Legras and Dritschel, 1993), where continuous filamentation of PV by the breaking of Rossby waves on the polar vortex edge transports mass within PV filaments away from the edge region into the surf zone, where the PV is mixed (McIntyre and Palmer, 1984). The net result is less mass within the vortex, and thus a poleward displacement. The high PV in the lower stratosphere (polar regions) is maintained by radiative cooling (Haynes, 2005), but this is weakest at the summer solstice; therefore, the high PV is not maintained. This occurs until late August, when the cooling begins to strengthen again and the high PV reservoir builds. The tropopause progresses slowly equatorward during the autumn. The time scale of delay behind the cycle in solar insolation, therefore, is related to the radiative equilibrium time scale of the troposphere (30 days, James, 1994). Fig. 1D shows a measure of wave activity on the 320K surface, which will be explained in the next section. It can be seen that the marked variations in the zonal mean are related to variations in wave activity, as argued earlier in this chapter.

2.2 Wave Activity Conservation Laws

Once we have isolated the background state, the partition is useful only if we also can anticipate the properties for evolution of the perturbations from it. The aim is to find a definition for wave activity that satisfies a conservation law of the following form:

$$\frac{\partial A}{\partial t} + \nabla . \underline{F} = D \tag{7}$$

which stems from the conservation laws obeyed by the full system combined with those satisfied by the background state due to its symmetry. In Eq. (7), A is a wave activity density, $_F$ is the wave activity flux, and D stands for the effects of nonconservative processes only. McIntyre and Shepherd (1987) set out a systematic approach to find the conservation laws for perturbations by combining globally conserved properties (such as angular momentum or energy) with properties, called *Casimirs*, which depend only on materially conserved properties. Two key examples are the mass and circulation enclosed by PV contours within isentropic layers, which can be described as functions of the PV and potential temperature (θ) coordinates.

The method proceeds by defining the pseudo(angular)momentum density as

$$P = -r(Z + S) + r_0(Z_0 + S_0) \tag{8}$$

where Z is the specific zonal angular momentum, $S(q,\theta)$ is the Casimir density (as yet unspecified), and the subscript zero refers to the same quantities in the background state. A central aspect is that there is a whole continuum of conserved properties, but the approach is to identify the property where the first-order contribution (in wave amplitude) is zero by construction, therefore ensuring that the resulting wave activity is second order (or higher).

Haynes (1988) showed the full nonlinear result for the primitive equations on the sphere which, in the limit of small wave slope, reduces to the more familiar form of wave activity density (see, e.g., Vallis, 2006, Section 7, for a simple derivation of the first term):

$$P = \frac{1}{2} r_0 Q_y \eta^2 - r' u' \cos\phi + \left(\frac{1}{2} {r_0}^2 q_0 {\eta_b}^2 - r_0 u' \eta_b \right)_b \cos\phi \frac{\partial \theta_{0b}}{\partial y} \tag{9}$$

where $Q_y = r_0 \cos\phi \partial q_0 / \partial y$ is the appropriate mass-weighted meridional PV gradient on the sphere, $y = a\phi$ is the meridional coordinate, and $\eta = -q'/(\partial q_0/\partial y)$ is defined as the meridional displacement of a PV contour relative to its latitude in the background state. Because the interior PV gradient of the background state is positive, the first term in Eq. (9) is positive definite; therefore, it is a useful measure of the amplitude of Rossby wave activity. Also, southward (negative) displacements give rise to positive PV anomalies through the advection of PV. The second pseudomomentum term, $-r'u' \cos\phi$, is often described as the *gravity wave term* because it is absent under quasi-geostrophic balanced dynamics and does not involve meridional PV fluxes. However, it can be an important player in some large-scale motions. For example, it is the dominant term in equatorial Kelvin wave activity.

The last term is proportional to the meridional gradient of potential temperature along the lower boundary. Because $\frac{1}{2} r_0 {\eta_b}^2$ is positive definite, the boundary term takes the sign of $r_0 q_0 \partial \theta_{0b} / \partial y$, which is typically negative (in both hemispheres) and opposes the interior term (although the terms involving u' are not sign definite). Note that η_b represents the meridional displacement of θ contours along the lower boundary in the full state relative to their position in the background state.

Fig. 2 shows the PV anomaly pattern (color shading) for a particular snapshot of the atmosphere on an isentropic surface (320K) that intersects the tropopause. The Ertel PV anomalies are weighted by the background-state isentropic density because this quantity, $r_0(q - q_0)$, reduces to the quasi-geostrophic PV anomaly field under the approximations of QG theory (see Section 12.4 of Hoskins and James, 2014). Because the density is much smaller at high latitudes in the stratosphere, this downweights the high-latitude negative anomalies; consequently, the positive anomalies are much more prominent. The positive anomalies are in tropopause troughs where the air has been displaced far equatorward and the tropopause is lower than its surroundings. Note that although each trough is distorted differently by advection, it is clear that there are seven centers of action around the midlatitudes. However, they are not equally spaced, with the strong anomalies over the eastern United States and Western Europe being separated the most, and the anomalies over Alaska and the west coast of North America being closest together.

Overall, the pattern projects most strongly onto zonal wave number 6 and the correspondence is highlighted by overplotting the wave number 6 Fourier component of the PV field. This serves as an indication that despite the nonlinearities introduced by advection and Rossby wave–breaking, the dynamics of the large-scale pattern may be interpreted in terms of wave propagation and interaction. This hypothesis will be tested using the ENM approach, which is derived by combining statistical analysis of data with wave-activity-conservation properties.

A similar, but less frequently used, conservation law exists for pseudoenergy. The large-amplitude derivation (Haynes, 1988) begins with a definition similar to Eq. (8), but using the

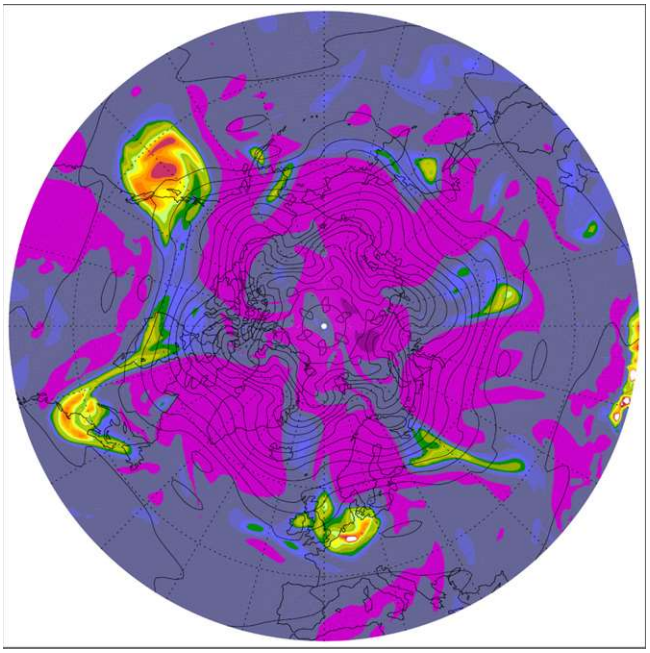

FIG. 2 Snapshot of the atmosphere (00UT June 23, 2007) on the 320K isentropic surface. The contour lines show PV, Fourier-filtered to zonal wave number 6 (interval 0.5 PVU), and the shading shows the PV anomaly, $r_0(q - q_0)$, from the unfiltered data overlain (negative anomalies in pink; the most positive anomalies in *red*; interval 10^{-5} s^{-1}). This pattern recurred throughout June and July 2007, and the trough (positive PV anomaly) over Western Europe gave rise to extreme monthly rainfall (Blackburn et al., 2008).

energy density in place of the zonal angular momentum. The pseudoenergy conservation law is then obtained using the time symmetry (i.e., steadiness) of the background state. Methven (2013) derived the small-amplitude expression for pseudoenergy, including perturbations near the lower boundary:

$$H = \frac{1}{2} r_0 \left(u'^2 + v'^2 \right) + \frac{1}{2} \frac{h_0}{g p_0 \theta} p'^2 - \frac{u_0}{\cos\phi} P \tag{10}$$

where the first term is the kinetic energy of the perturbations, the second is the available potential energy, and the third is called the *Doppler term* (for reasons which will become apparent), and is proportional to P (including the boundary terms). Note that perturbations are defined relative to the background state by using Eq. (1), where position is identified in isentropic coordinates (λ, ϕ, θ). For example, background pressure can differ from the full state at the same position in θ-coordinates.

Even at large amplitudes, the conservation laws imply certain properties for the perturbations. Consider a coherent disturbance that is neither growing nor decaying, but rather translating primarily along latitude circles. If the background can be defined as both zonally symmetric and steady, then the disturbance must have both a conserved pseudomomentum and pseudoenergy. The ratio of pseudoenergy to pseudomomentum gives the translation speed of the reference frame from which the disturbance appears steady to the observer. In other words, it defines the phase speed of the disturbance (Held, 1985; Zadra, 2000). This is the one of the central properties that we will use to characterize modes of variability in data.

2.3 The Implications of Wave-Activity Conservation for Modes of Variability

The small-amplitude limit of wave activity in Eq. (9) is precisely quadratic in disturbance amplitude, and this property has been exploited to make deductions about the disturbances in general and the properties of normal modes of atmospheric dynamics in particular (Held, 1985):

1. Because the wave activity is globally conserved, its rate of change is zero. Therefore, if the background state is also steady, disturbance amplitude can grow everywhere only if the global wave activity is identically zero. From Eq. (9), this leads to the celebrated Charney-Stern necessary condition for shear instability: The PV gradient must change sign somewhere within the domain. This argument includes the possibility that the negative pseudomomentum could be associated with the boundary wave activity in Eq. (9), as first described for baroclinic instability by Bretherton (1966).
2. Consequently, growing normal modes must have zero pseudomomentum, while neutral modes could have nonzero values.
3. Because normal modes evolve independently, each conserves pseudomomentum on its own. Therefore, if the global pseudomomentum of a superposition of modes is to be conserved, then the normal modes must be orthogonal with respect to pseudomomentum.
4. If the background state is also steady, pseudoenergy will be conserved and conclusions 2 and 3 also pertain to pseudoenergy. Also, because the disturbance energy is positive definite, from Eq. (10), we can obtain the Fjortoft necessary condition for shear instability: the zonal flow and meridional PV gradient must be positively correlated on average across the domain.
5. As argued in Section 2.2, the phase speed of neutral modes is given by the ratio of pseudoenergy to pseudomomentum. Considering the small-amplitude quadratic forms, we see that there are two distinct influences on phase speed:

$$c_p = -\frac{\langle H \rangle}{\langle P \rangle} = \left\langle \frac{u_0}{\cos\phi} P \right\rangle / \langle P \rangle - \langle E \rangle / \langle P \rangle \tag{11}$$

Here, the angle brackets denote an integral over the entire domain. The Doppler term in pseudoenergy gives the rate of advection of the disturbance by the zonal flow, even in the presence of shear. Because the wave frequency can be defined as $\omega = c_p k$, the Doppler term represents the shift in frequency associated with zonal advection by the background flow. Its sign depends only on the sign of the background zonal flow weighted by locations where the wave activity is largest. The second term describes propagation relative to the zonal flow and is proportional to the disturbance energy E. Because the energy is positive definite, the direction of propagation depends on the sign of the mode's pseudomomentum.

Note that it is not immediately obvious how to predict the phase speed of growing normal modes because they must have zero pseudomomentum and pseudoenergy. However, a solution to this problem was obtained by Heifetz et al. (2004), who used wave activity orthogonality to recast the growing and decaying normal modes obtained from complex conjugate normal mode solutions in terms of the linear superposition of a pair of counterpropagating Rossby waves (CRWs). By construction, CRWs are orthogonal with respect to

pseudomomentum, and therefore, one CRW has positive and the other CRV negative pseudomomentum such that their sum is zero when only the growing normal mode is present. Therefore, they describe disturbances that propagate in opposite directions (relative to the flow where wave activity is large). However, they are not orthogonal with respect to energy that can grow or decay as a CRW pair evolves.

They are used to give a mechanistic explanation for baroclinic or barotropic instability (a generalization of the Eady model for any unstable parallel zonal flow). The growth rate of normal modes can be expressed in terms of the energy of interaction between the CRWs. Perhaps most important, an expression for the phase speed of a growing normal mode is obtained as the average of the intrinsic phase speeds of the two CRWs:

$$c_{NM} = -\frac{1}{2}\left(\frac{\langle H_1\rangle}{\langle P_1\rangle} + \frac{\langle H_2\rangle}{\langle P_2\rangle}\right)$$

where $\langle P_1\rangle = -\langle P_2\rangle$ in the CRW construction.

3 THE ENM APPROACH TO OBSERVED DATA AND MODELS AND ITS RELEVANCE TO S2S DYNAMICS AND PREDICTABILITY

One way of diagnosing and characterizing the atmospheric S2S variability is to use a *phase space approach,* which has been shown to be very valuable in mathematics, physics, and atmospheric dynamics. The phase space of a geophysical fluid is a space where the state of the flow at a given time corresponds to one unique point. Usually, the phase space for a geophysical fluid can be represented by a steady basic state and superposed wave disturbances that are each represented individually as an oscillation in a 2D-phase plane (with characteristic amplitude and phase). In a nonlinear flow, this decomposition is nonunique (as discussed in Section 2.1), and the phase space trajectory can be complicated.

Here, we will focus on the evolution of waves and the insights that come from the small-amplitude limit for disturbances where linear wave theory applies. This is motivated by the prevalence of wave propagation on larger scales (e.g., Fig. 3), even though stirring by the large scales results in a continuous cascade of PV to smaller scales and the wavelike patterns associated with teleconnections. Even with such a drastic assumption, the proposed ENM diagnostic framework is very insightful and can be applied often to various type of flows with success, including nonlinear flow. Evidence from examples will be discussed next.

In general, waves transfer energy and momentum in flows through dynamical processes that obey the conservation laws introduced in Section 2. In Section 3.1, we will show that conservation laws constrain fundamentally the space-time characteristics of waves and are central to normal mode theory. For a given dissipative and stochastically forced flow, the conservation laws can be used to augment the physical relevance of the statistical principal component analysis (PCA) and its associated EOFs. In the context of waves relative to a steady basic state, if PCA is performed using wave-activity conservation laws, we demonstrate that the EOFs are the normal modes obtained from the linear wave theory.

The latter result permits the development of a statistical and empirical diagnostic framework with a built-in linear wave theory interpretation. It has been named ENM analysis by

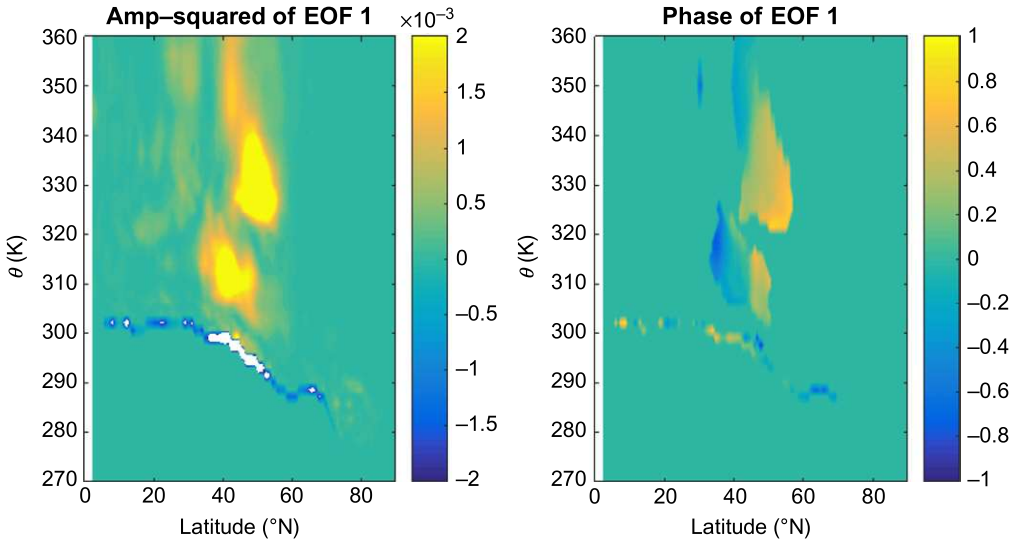

FIG. 3 The structure of the leading ENM pair for zonal wave number 6 during June and July 2007. (A) The amplitude squared of the ENM is its pseudomomentum structure, which is positive in the interior and negative in the boundary domain, which spans the space between the wavy and background state θ contours on the lower boundary (for each θ value). ENM is normalized such that the integral of amplitude squared over the latitude-θ section is unity (units $\text{rad}^{-1}\,\text{K}^{-1}$). (B) The phase of the meridional air parcel displacement associated with the ENM pair (radians/π).

Brunet (1994). In circumstances when the phase speeds obtained from ENM properties are consistent with the phase speed of waves observed by tracking (the ENM phase speed condition; see Eq. 19), we also will show in Section 3.1 that some aspects of ENM analysis are still conceptually and quantitatively relevant to stochastically forced and damped nonlinear flows. In Section 3.2, we will discuss the potential of ENM analysis for diverse applications, and in Section 3.3, we will focus on the ENM analysis of S2S variability.

3.1 ENMs: Bridging Principal Component, Normal Modes, and Conservation Laws

In Section 2.2, we discussed two well-known conservation laws for wave activity in geophysical fluid dynamics (Haynes, 1988). These wave activities are the total pseudoenergy for a steady basic state and the total pseudomomentum for a zonal symmetric basic state. When the basic-state wind is uniform, pseudoenergy and pseudomomentum reduce to total energy and total enstrophy, respectively, but shear in the basic state flow changes the dynamics fundamentally.

In the past, the quest for optimal bases has given rise to many independent rediscoveries of what is now known as *EOF space-time biorthogonal expansions.* A historical review of this topic is given by Sirovich and Everson (1992). The physical interpretation in terms of normal mode bases was explored by North (1984) for an atmospheric linear dynamical system with normal modes satisfying a self-adjoint equation. This work was extended by Brunet (1994) to normal

modes of the primitive atmospheric equation for sheared flows (in general not a self-adjoint problem) using the conserved pseudomomentum and pseudoenergy wave activities. In this more general case, the ENMs are not statistical eigenfunctions of a covariance matrix, but rather a solution of a generalized symmetric eigenvalue problem.

The statements about dynamical systems, normal modes, phase speed, and conservation laws discussed here can be found in Brunet (1994), Brunet and Vautard (1996), and Charron and Brunet (1999). For the rest of this discussion, we will assume that a zonal basic state exists, ensuring that pseudomomentum and pseudoenergy are conserved for an unforced and inviscid flow.

Consider a nonlinear dynamical system expressed in the following general form:

$$\frac{\partial X}{\partial t} = G(X) \tag{12}$$

where X is the state vector (e.g., the distribution of variables required to define the flow state), and $G(X)$ represents a dynamical operator that gives the rate of change of the state given the current state. If we linearize around a time-independent zonal basic state X_0 solution of Eq. (12), we obtain

$$\frac{\partial X'}{\partial t} = iG_0X' = iH_AAX' \tag{13}$$

where H_A and nonsingular A are time-independent Hermitian operators if and only if $W_A = \langle X', AX' \rangle$ is a conserved quantity (Charron and Brunet, 1999). This was demonstrated explicitly for the shallow-water model on the sphere for Rossby and gravity waves in Brunet and Vautard (1996). For the rest of this discussion, we will assume that W_A is the total pseudoenergy.

The bracketed term $\langle f, g \rangle = \int f^{\dagger} g dv$ represents an integral that can be over a line, area, or volume. The specific example is given by Eq. (10) for the dynamics described by the primitive equations on the sphere, where the state vector would need to be described by $X' = (u', v', p', r', \eta)$, including the perturbations in the interior and appearing in the lower-boundary terms, and the bracket represents the volume integral over the atmosphere.

Considering the normal mode expansion, in which each Z_n is a monochromatic wave solution of Eq. (13), then

$$X' = \sum_n a_n(t) Z_n = \sum_n a_{n,0} e^{i\omega_n t} Z_n$$

$$\omega_n Z_n = H_A A Z_n \text{ where } \quad \langle Z_n, AZ_m \rangle = \alpha_n \delta_{n,m} \text{ and } \quad \overline{a_n a_m^*} = \delta_{nm} \tag{14a}$$

in which α_n is the total pseudoenergy of the normal mode Z_n, a_n. ω_n is the natural frequency of the normal mode, and the overbar is the time average. This is a generalized eigenvalue problem, and because A and H_A are Hermitian operators and A is nonsingular, we can show for a space of finite dimension that the set of $\{Z_n\}$ forms a complete orthogonal basis under the pseudoenergy metric $M_A = \langle f, Ag \rangle$ (Bai et al., 2000). Hence, the normal mode expansion provides a full solution to the initial value problem associated with Eq. (13). The normal modes span a phase plane in phase space. They represent propagating waves except for stationary waves (i.e., $\omega_n = 0$).

If we linearize around a zonally symmetric basic-state X_0 solution of Eq. (12), then pseudomomentum $W_J = \langle X', JX' \rangle$ is conserved and normal mode solutions satisfy:

$$\omega_n Y_n = H_J J Y_n \text{ where } \langle Y_n, JY_m \rangle = \beta_n \delta_{n,m} \text{ and } \overline{a_n a_m^*} = \delta_{nm} \tag{14b}$$

where β_n is the total pseudomomentum of the normal mode Y_n. Note that the normal modes $\{Z_n\}$ and $\{Y_n\}$ are identical and form a unique basis if the eigenvalue pairs (α_n, β_n) are nondegenerate. We will assume for the rest of this discussion that this is the case. It follows from the framework of complete ensemble of commutating operators (CECO), where a set of many operators have the same eigenvectors if they all commute and their eigenvalues are determined uniquely for each eigenvector (Cohen-Tannoudji et al., 1973). For the primitive equations on the sphere with pseudoenergy and pseudomomentum conservation laws, we can write:

$W_A = cW_J$, and hence, for each normal mode, we have

$$c_n = \frac{\langle Z_n', AZ_n' \rangle}{\langle Z_n', JZ_n' \rangle} \tag{15}$$

where c is the mean phase speed of the flow (Held, 1985; Zadra, 2000), as discussed in Section 2.2 with respect to Eq. (11). This relationship also holds true for modified wave activities (by adding a divergent term that is conserved by construction) for individual isentropic surfaces that do not intersect with the surface (Zadra, 2000).

In the geophysical context, it is often more representative of observed flows to have a damped and stochastic forced model (for more on this, see Chapter 6). For the sake of simplicity, we will assume in the following that we have already decomposed Eq. (13) for each zonal wave number k.

Then the forced-dissipative version of Eq. (13) for each zonal wave number (ignore the subscript) can be expressed in the following form:

$$\frac{\partial X'}{\partial t} = iH_A A X' - \gamma X' + \varepsilon \tag{16}$$

where γ is a Raleigh-damping coefficient, and ε is a random forcing that is uncorrelated in time (e.g., Wiener process) and should be square-summable in space.

If we expand Eq. (16) in terms of normal modes $\{Z_n\}$, the complete time series is

$$X'(x,t) = \sum a_n(t) Z_n(x)$$

and the Fourier transform of the time domain in the equation yields the following coefficients:

$$\widetilde{a}_n = \frac{\widetilde{\varepsilon}_n}{i(\omega - \omega_n) + \gamma} \text{ where } \alpha_n \widetilde{\varepsilon}_n = \langle Z_n, A\widetilde{\varepsilon} \rangle \tag{17}$$

where the Fourier transformed variables, $\widetilde{g}(\omega) = (2\pi)^{-1} \frac{1}{T} \int_{-T/2}^{T/2} g(t) e^{-i\omega t} dt$ and α_n, are defined by Eq. (14a).

When $\gamma = 0$, a solution for Eq. (17) exists only if the Fredholm alternative (Riesz and Sz-Nagy, 1953) is satisfied; hence, $\widetilde{\varepsilon}_n|_{\omega=\omega_n} = 0$ for all n.

If the pair (α_n, β_n) as defined in Eqs. (14a), (14b) is nondegenerate, then in the limit $T \to \infty$ and using Eqs. (14a), (17), we can show that if A and B are nonsingular and we have a time series X' solution of the stochastically forced and damped dynamical system (Eq. 16), then for each zonal wave number k:

$$CJX_n = \beta_n X_n \text{ and } CAX_n = \alpha_n X_n \text{ with } \omega_n = k\frac{\alpha_n}{\beta_n} \tag{18}$$

where $\{X_n\}$ are the normal modes of Eq. (13) and the covariance matrix elements are defined by $C(x, x') = \overline{X'(x)X'(x')^*}$. A normal mode X_n obtained using a covariance matrix approach as in Eq. (18) is named an ENM.

The phase speed relationship (Eq. 15) implies that by knowing the X' time series, we have completely solved the initial value problem of Eq. (13), including its nonhomogeneous damped version with stochastic forcing (Eq. 16). This is possible because the unforced and nondissipative evolution equation (Eq. 13) is a completely integrable Hamiltonian system, which for a given truncation N has $2N$ constants of motion. Associated with the complete ENM basis $\{X_n\}$, we also have a complete orthonormal basis (i.e., principal components) in time $\{a_n\}$, where $\overline{a_n a_m^*} = \delta_{n,m}$. It means that each individual wave/ENM of a different type (e.g., gravity and Rossby waves) obtained through an ENM analysis will span biorthogonal subspaces in space and time with a clear dynamical interpretation.

ENM analysis permits, in practice, the ability to diagnose a specific wave spatial structure and its time evolution, which is not contaminated by other waves present in the flow. In particular, it will efficiently partition fast and slow modes without using any time-filtering technique. In Brunet and Vautard (1996), this has been shown to be very advantageous relative to a standard EOF analysis for simulated linear and nonlinear upper-tropospheric barotropic flows. It should be noted that other statistical techniques are available for studying and predicting atmospheric oscillations in the presence of damping and stochastic forcing. Two examples are the principal oscillation pattern (POP) method (Penland, 1989) and the constructed analog (CA) method (Van den Dool, 1994). They have been successfully used in long-range forecasting (e.g., Van den Dool and Barnston, 1995).

These two statistical methods rely fundamentally on temporal lag techniques for a given variable *time series* in which a linear *regression* equation is used to predict future values based on both the current variable values and the lagged (past period) values. It can be readily shown that, in general, these two are mathematically equivalent. The ENM analysis is fundamentally different because it is not based on time-lag correlation, but only on the existence of conserved wave activities. This makes ENM analysis very robust when dealing with noise. For example, the ENMs derived from Eq. (19) are not affected at all by a random reordering of the time series. This is not the case for POP and CA analyses because the covariance matrix depends only on time-averaging operations and hence is invariant under a reordering of the time series.

Of course, when performing an ENM analysis in practice, we need to assess the validity of the underlying normal mode assumptions for a given time series, such as the choice of basic state, conservation laws, and small-amplitude wave activities. The small-amplitude approximation can be relaxed by using finite-amplitude wave activities, but as discussed in Brunet (1994), the interpretation of the ENM analysis results is definitely more problematic. With the

exception of the research done by Brunet (1994), all ENM analyses to date have been performed using small-amplitude wave activities.

In general, we can objectively falsify the effectiveness of ENMs at representing the dynamic of atmospheric flows (whether simulated or observed). It can be done in the context of CECO theory and using temporal lag techniques, as in CA and POP analyses. One important aspect of such an evaluation methodology are the ENM phase speed conditions:

$$\overline{\Omega_n} = -i\overline{\frac{da}{dt}a^*} = k\frac{\langle X_n, AX_m\rangle}{\langle X_n, JX_m\rangle} = \omega_n \tag{19}$$

where $\overline{\Omega_n}$ are the observed mean natural frequencies. As discussed previously, these relations are necessary to demonstrate that an atmospheric flow is an integrable dynamical system. In practice, the principal components are generally not monochromatic, but the ENM phase speed conditions require that for each ENM, the observed mean natural frequency $\overline{\Omega_n}$ (derived from the principal component time series) is equal to its intrinsic natural frequency ω_n. These phase speed conditions have been verified within statistical estimation errors (e.g., dependence on the length of time series) for a wide variety of geophysical flows spanning mesoscale to planetary scales and have provided significant insights for many problems.

3.2 ENM in Applications Relevant to Predictability Across Time Scales

First, we will illustrate ENM analysis in application with a relative simple and explicit example taken from (Brunet and Vautard, 1996): the shallow-water model on the sphere in spherical coordinate. In many aspects, this model is relevant to the S2S prediction problem. It is a global barotropic model that supports Rossby, Rossby-gravity, Kelvin, and gravity waves typical of the midlatitude and tropical-upper-troposphere regions. Then for the ENM analysis of the evolution equation (Eq. 12), we have the following perturbation state vector for a given zonal wave number s:

$$X' = X - X_0 = \begin{pmatrix} u' \\ v' \\ \sigma' \\ P' \end{pmatrix} \text{ with basic state } X_0 = \begin{pmatrix} u_0 \\ 0 \\ \sigma_0 \\ P_0 \end{pmatrix} \tag{20}$$

where u', v', σ' and P' are the nondimensional zonal wind, meridional wind, height, and PV perturbations, respectively. The pseudoenergy A and pseudomomentum J operators are explicitly

$$A = \frac{1}{2}\begin{pmatrix} \sigma_0 & 0 & u_0 & 0 \\ 0 & \sigma_0 & 0 & 0 \\ u_0 & 0 & 1/F_R & 0 \\ 0 & 0 & 0 & -\dfrac{u_0\sigma_0^2}{\dfrac{d\mathrm{P}_0}{d\phi}} \end{pmatrix} \text{ and } J = \frac{\cos(\varphi)}{2}\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\dfrac{\sigma_0^2}{\dfrac{d\mathrm{P}_0}{d\phi}} \end{pmatrix} \tag{21}$$

where F_R is the Froude number and ϕ the latitude. Hence, from the considerations of the previous section, and with the sole knowledge of these operators, we can perform an ENM

analysis for a given shallow-water model time series X by solving the generalized eigenvalue problem (Eq. 18).

Note that the uniqueness and completeness of the ENM basis depend only on the rank of these two matrices (and of the covariance matrix). From their determinant we can show that this tantamount to have bounds a and j for which, for any latitude,

$$0 < \frac{\sigma_0^2}{\frac{d\mathrm{P}_0}{d\varphi}} < j \text{ and } 0 < -\frac{u_0\sigma_0^3}{\frac{d\mathrm{P}_0}{d\varphi}}\left(\frac{\sigma_0}{F_R} - u_0^2\right) < a \tag{22}$$

If they are not satisfied, the first and second conditions are related to the Charney-Stern and Fjortoft necessary conditions for shear instability, respectively, as discussed in Section 2.3. But the second condition also states that the wind should not match the local value of the gravity wave phase speed. The second stability criterion was first derived in a somewhat different manner by Ripa (1983), and it guarantees that there is no unstable normal mode. In the presence of unstable normal modes, the ENM analysis is still valid, but it needs a different approach, along the line explained for CRWs in Section 2.3. Of course, the unstable ENMs also can be obtained directly from the kernel of the pseudomomentum generalized eigenvalue problem (Eq. 18) (Martinez et al., 2010a).

Using the spectral shallow model on the sphere, Brunet and Vautard (1996), the ENM phase speed conditions (Eq. 19) were tested for linear and nonlinear regimes and various metrics (e.g., pseudomomentum versus the square of height) for typical Northern Hemisphere winter jets. The EOF diagnostics based on different metrics has clearly shown the advantage of using wave activities.

The time and zonal mean basic state of the time series was also the best choice. It minimizes the perturbation variance, and hence it is the closest to the small-amplitude limit. Of note, the phase speed conditions were satisfied for sheared, low-frequency Rossby waves in the linear regime only, for a relatively high spectral resolution of T100 compared to typical climate model resolutions in the 1990s (T32-64). The modes of variability of the shallow-water model with a realistic radius of deformation were well tuned only for sufficiently high spectral resolution. This example highlights the potential of ENM analysis as a suitable tool for assessing the numerical accuracy of dynamical processes in climate and prediction models. Zadra et al. (2002b) extended the method to multi-layer data from a primitive equation model and applied it to the Canadian Global Environmental Multiscale (GEM) NWP model dynamical core.

An important finding is that ENM phase speed conditions were verified even for nonlinear simulations with wave-breaking events (Brunet and Vautard, 1996). This indicates that over a sufficiently long period of time, the cumulative effect of nonlinear terms can be considered negligible when verifying the phase speed conditions, even if the ENMs are interacting nonlinearly. For example, although Rossby wave–breaking and the complexity of the flow results in the stretching and folding of individual troughs in different ways in Fig. 2, the propagation of the underlying wave pattern of PV anomalies is quantitatively predicted by Eq. (19). Therefore, ENMs are in some respects dynamically relevant to nonlinear problems. The framework for studying nonlinear interaction of normal modes in sheared flows and their relation to wave activities has been established in Vanneste and Vial (1994). This may

provide an avenue for studying the nonlinear interactions of ENMs with applications to the study of the predictability and dynamical processes of the S2S phase space and could pave the way for the empirical application of different chaos theory techniques to identify routes to chaos (e.g., KAM theory and period doubling).

In the literature so far, ENM phase speed conditions have been verified when analyzing global atmospheric North Hemisphere atmospheric variability on the 315K isentropic surface (Brunet, 1994), shallow-water-model Rossby waves (Brunet and Vautard, 1996), multi-layer NCEP winter reanalyses (Zadra et al., 2002a), simulated gravity waves in Charron and Brunet (1999), and for hurricane vortex, Rossby wave dynamics (Chen et al., 2003; Martinez et al., 2010a,b, 2011). In these studies, the ENM phase speed conditions were demonstrated within reasonable margins for a large portion of the studied wave activity variances.

Most recently, Methven et al. (2018) have extended the technique to the stratified problem, including the lower boundary, which introduces considerable complexity in using reanalysis data; but it is important due to the negative term introduced in the pseudomomentum and the corresponding term in pseudoenergy. In other words, the propagation of the potential temperature wave along the lower boundary modifies the phase speed of baroclinic waves, as is most familiar in the Eady and Charney models of baroclinic instability (Heifetz et al., 2004). Fig. 3 illustrates the structure of the largest amplitude ENM at zonal wave number 6 obtained from ERA-I data for June–July 2007. Note that in the calculation, the perturbations are defined relative to the MLM background state calculated by Methven and Berrisford (2015). The mode is propagating, and thus described by a pair of ENM structures in quadrature.

It has the largest amplitude in the upper troposphere and the lower stratosphere (straddling the tropopause) and a distinct negative pseudomomentum contribution associated with the boundary terms (here seen sloping along the lower boundary of the background state). Therefore, it is a baroclinic wave, although it has much stronger interior wave activity than the boundary term, and so it does not have zero pseudomomentum, as would be expected for a baroclinic growing normal mode. Its phase changes across the midlatitude jet in the troposphere, which is a signature of the different wave-breaking directions on either side of the jet. Evidence for this can be seen in Fig. 2 in the Fourier-filtered PV field. This particular structure meets the ENM phase speed conditions within a few meters per second (Methven et al., 2018). These studies clearly show the relevance of combining PCA, wave activities (including their associated Elliasen-Palm flux), and normal mode theory to diagnose atmospheric dynamics.

The advantage of the biorthogonal ENMs is evident: It permits a systematic approach for examining the statistics of data and exploring whether a modal perspective is dynamically relevant or not. The ENM phase speed conditions provide important nontrivial information on the dynamic processes (e.g., gravity versus Rossby waves) found in each of the subspaces spanned by the ENMs, as in hurricanes, where there is no separation of time scales between gravity and vortex Rossby waves due to a finite Rossby number (Chen et al., 2003). Similarly, the technique can be used to distinguish among types of wave modes when they are not separated, in terms of spatial scale. In this regard, it holds promise for addressing tropical dynamics in particular.

Falsification of the ENM phase speed conditions can happen for the following reasons: (1) nonlinear effects cannot be neglected; (2) the damping is not Raleigh or stochastic forcing is not Wiener; (3) the wave variability is not well simulated in the model; and (4) some dynamical and physical processes are not represented properly in the wave activities. The latter can

be sensitive to the basic state choice, as demonstrated in Brunet and Vautard (1996) with a shallow-water model. In some situations (e.g., baroclinic development), the important contributions of the boundary terms have been neglected (see Zadra et al., 2002a), as pointed out by Methven (2013).

The ENM approach can be used to study the response of a conservative dynamical system to arbitrary forcing. Nowadays, this is a key issue in climate science. According to Cooper and Haynes (2011), a study about the FDT, the response of a conservative dynamical system to steady forcing applied for $t > 0$, and in the limit, $t \to \infty$ would be equal in the ENM framework to $\delta x = -T^{-1}\delta f$, where T is simply the diagonal matrix composed of the intrinsic natural frequency of the ENMs. The advantage relative to the POP approach is that the response can be studied individually for each biorthogonal ENM with a direct physical interpretation. Of course, one of the focuses of climate studies should be on low-frequency dynamical processes like S2S variability.

3.3 ENM Application to the Atmospheric S2S Variability

In Brunet (1994) and Zadra et al. (2002a), large-scale atmospheric variability was shown to be spanned by ENMs, with intrinsic natural oscillation periods ranging from days to months. The ENM phase speed conditions were verified for almost all the wave activity except for some specific ENMs. For example, Brunet (1994) showed that the ENMs associated with Atlantic blocking did not satisfy the ENM phase speed conditions, possibly due to the nonlinear transient feedback and omission of boundary terms in the wave activities.

In Brunet (1994), the S2S North Hemisphere variability on the 315K isentropic surface for 24 winters was characterized quantitatively and empirically within an ENM framework. Fig. 4 shows the distribution of the total observed wave activity per day as a function of the ENM intrinsic natural oscillation period ($\omega_n = c_n k$) for various truncation thresholds. Eight discretelike ENMs were identified, which show a finite contribution to the total wave activity (over 1% individually), with distinct intrinsic periods from 14 to 200 days spanning the S2S time range. They represent around 20%–30% of the total wave activity and are closely linked to large-scale patterns like the AO and Atlantic blocking.

The rest of the wave activity (70%–80%) is represented by a continuous spectrum of ENMs with a peak around 3–5 days associated with transients, storms, and baroclinic waves (diagnosed as very distinct propagating ENM pairs, as shown in Fig. 3). The discretelike ENM pairs were shown to be relatively predictable, with a large e-folding time of 3–5 days. Here, the e-folding time is defined for each ENM pairs as the time average of its amplitude tendency (growth or decay) in the phase plane (Brunet, 1994), which is a good measure of predictability. The continuous spectrum has e-folding times of less than 3 days, which is consistent with predictability theory for baroclinic wave activity (Leith, 1978). The eight discretelike ENM pairs partly control the evolution of the continuous spectrum and the distribution of high-impact weather because they are large-scale features dominating the advection of PV. They are good candidates to span the phase space of a low-order model of S2S variability (see Chapter 6).

For example, Fig. 5 shows the phase-space-probability density of the zonal wave number 2 ENM pair, with an intrinsic period of 35 days for the winters from 1963 to 1987. An asymmetric bimodal signature with one localized small peak and one wide peak can be clearly

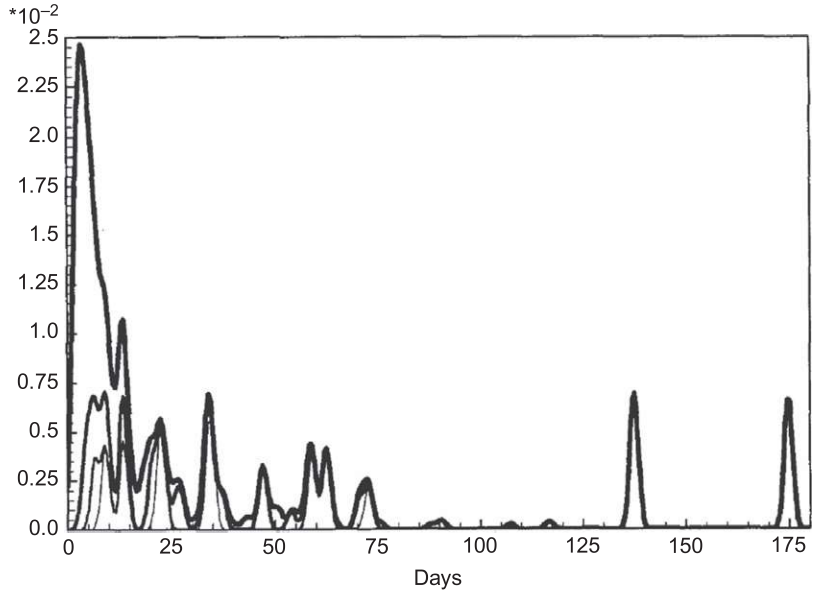

FIG. 4 Distribution of the observed wave-activity spectral density as a function of the ENM oscillation period for the NH 315K isentropic surface. The solid curves, from the thicker to the thinner, correspond to the total wave activity in percentage with individual ENM contribution higher than 0.1%, 0.4%, 0.7% and 1%, respectively. *From Brunet, G., 1994. Empirical normal mode analysis of atmospheric data. J. Atmos. Sci., 51, 932–952.*

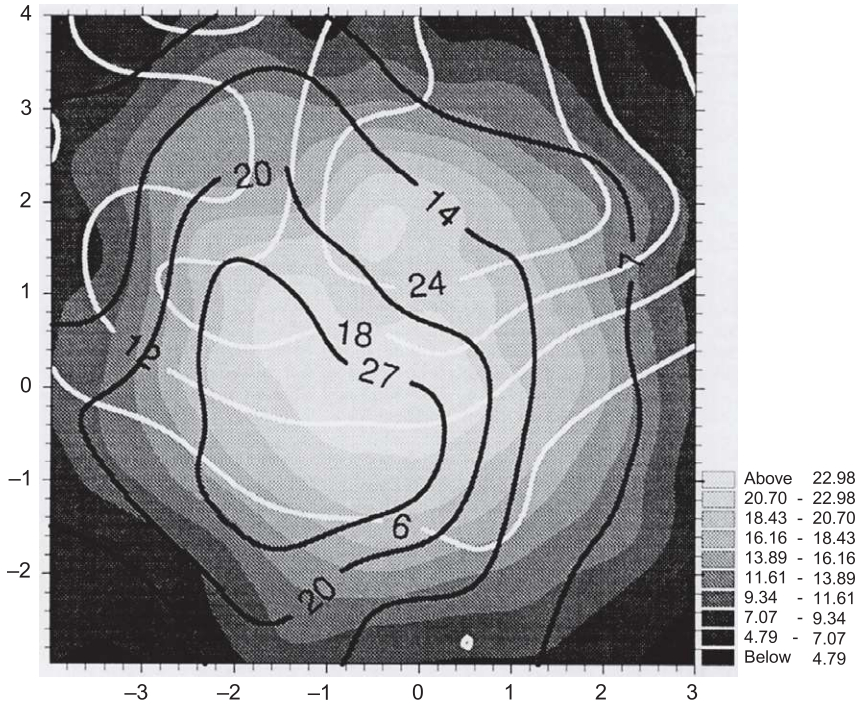

FIG. 5 Phase-space-probability density of the zonal wave number 2 ENM, with an intrinsic period of 35 days. The black and white contour lines correspond to the ZO and BL weather regime events, respectively. The probability density is shaded (units are 10^{-3}). *From Brunet, G., 1994. Empirical normal mode analysis of atmospheric data. J. Atmos. Sci., 51, 932–952.*

observed. A composite of wave activity maps shows that the small peak is associated with Atlantic blocking. This is confirmed in the same image by the probability densities of the zonal (ZO) and Atlantic blocking (BL) weather regime events obtained by Vautard (1990), where the ZO event density has a pattern similar to the main structure of the ENM bimodality.

It is noteworthy that the ZO and BL densities in Fig. 5 show a strong amplification of the ENM pair wave activity for the BL relative to ZO regimes, with an approximate π phase change in the amplitude that can be readily observed by noting that the BL maximum density is farther from the phase-plane origin than the ZO maximum density. This is observed for two other discretelike ENM pairs, and is typical of a resonant process. The presence of weakly unstable hemispherical normal modes in the slow variability is predicted by the wave-mean flow interaction theory of Charney and DeVore (1979); also see Chapter 6. They proposed orographically induced linear and nonlinear resonance mechanisms to explain phase locking and multiple equilibria of weather regimes.

It is recognized that the North Atlantic Oscillation (NAO) temporal variability spans many time scales and is subject to a wide range of atmospheric and oceanic forcing. For example, Molteni et al. (2015) has shown extratropical teleconnections with the Indo-Pacific region that have common atmospheric responses associated with different forcing spanning weeks to interdecadal time scales, which is consistent with resonant behavior. More in-depth studies of discretelike ENM phase space, weather regimes, and resonance mechanisms are needed to make progress on these S2S issues. One step toward this objective was achieved by the three-dimensional (3D) ENM analyses of the observed and simulated global atmosphere by Zadra et al. (2002a,b). The focus of Zadra et al. (2002a) was not S2S variability per se, but the upper-troposphere and lower-stratosphere observed variability along the tropopause. It has been shown that a large part of the wave activity around the tropopause was spanned by ENMs with intrinsic oscillation periods of less than 14 days, and this can be explained using the theory of quasi-modes (Rivest and Farrell, 1992).

Quasi-modes are defined as superpositions of singular modes that are sharply peaked in the phase speed domain but have large-scale (delocalized) structures, as opposed to singular modes, which represent sheared disturbances advected by the flow. They are often reminiscent of monochromatic discrete modes (neutral or unstable) that have been displaced into the continuous spectrum by the modification of the wind basic state (e.g., addition of wind shear) or dynamical processes like f-plane versus β-plane (Zadra, 2000). Quasi-modes are often weakly damped by critical layer stirring (Briggs et al., 1970; Schecter et al., 2000, 2002; Schecter and Montgomery, 2006; Martinez et al., 2010a) and maintain their energy for relatively long periods. They are easily excited by external forcing due to their large spatial structure, which is typical of discrete modes. They are not easily computed or identified by numerical or analytical methods (e.g., identification of relevant Landau poles), but they have been readily identified by ENM analyses in hurricane simulation (Martinez et al., 2010a) and atmospheric reanalysis (Zadra et al., 2002a) diagnostics. The latter study was able to identify leading modes with dipolar pressure patterns along the summer hemisphere tropopause that have well-defined phase speeds and decay rates of a few days, which can be explained by the theory of quasi-modes (e.g., wave number 5 with phase speed of 12 m/s and 3-day decay rate).

It is quite possible that the eight discretelike ENMs spanning the S2S variability in the 2D barotropic study of Brunet (1994) are quasi-modes. Further studies are required to confirm

this hypothesis, and these will probably need to be 3D ENM analyses because Zadra et al. (2002a,b) have identified a relatively larger number of discretelike ENMs with intrinsic periods of over 14 days spanning the S2S variability (e.g., more than a half-dozen for the zonal wave number 1 alone).

The ENM also could be used in the atmosphere-ocean S2S context where, in general, pseudoenergy is expected to be conserved, but not pseudomomentum. For example, the latter is not conserved when an oceanic basin has irregular boundaries because the zonal symmetry is broken. The ocean ENM diagnostic could be used to look at the MJO atmosphere-ocean coupled problem, with the ocean limited to the mixing layer. This would be the first step toward looking at ENMs of the ocean-atmosphere with sub-seasonal to multidecadal time scales. It is noteworthy that wave-activity study of the ocean is almost nonexistent except for the work of P. Ripa (Shepherd, 2003).

4 CONCLUSION

The first important objective of this chapter was to tackle the problem of diagnosing S2S atmospheric variability by splitting the diabatic and adiabatic flow components using fundamental principles from geophysical fluid dynamics. This was shown to be possible by taking advantage of the MLM theory based on the conservation of PV and potential temperature. The MLM partitioning of S2S variability in terms of slow diabatic processes, such as radiative forcing, and large adiabatic dynamical processes leads to the second important goal of our proposed methodology, which is to be able to extract dynamical modes of S2S variability from observed global data and model simulations with coherent space-time characteristics using fundamental properties that can be deduced from theory.

We demonstrate in this chapter that ENM analysis, with its built-in characteristics based on conservation laws, PCAs, and normal mode theory, provides an appropriate theoretical framework. ENM analysis is able to frame S2S scientific studies suitably and bring new perspectives (e.g., partitioning the S2S variability in fast and slow modes based on the ENM intrinsic phase speed). For example, the use of the ENM technique to date has revealed that a small number of structures dominate the observed variability at lower phase speeds than baroclinic waves. However, there are many unanswered questions regarding the nature of the modes. Are the ENM structures consistent with the structure of normal modes? Can we confirm or rule out their quasi-modal interpretation? Are they robust in a very long time series? The background state has a strong seasonal variation, but the calculation assumes a steady background, so should the modes be obtained for each season separately? What is the role of boundary wave activity?

Once the discretelike ENM phase space spanning the S2S variability has been characterized properly, we will be well positioned to address the S2S predictability problem. To address the predictability challenges, we need to understand the following points better:

- The extent to which a low number of distinct modes describe variability on S2S time scales and the robustness of those structures from year to year.
- How these discretelike modes interact with one another, with faster disturbances, and with the background state through nonlinear interactions. The role of physical mechanisms,

such as wave resonance, multiple equilibria, attractor sets, and stable and unstable limit cycles (see Chapter 6 for more on this topic), is also important.
- The role of the slow modes in predictability on S2S time scales.
- How the ENM approach can be used to examine tropical-extratropical interactions and the teleconnections that result in longer-term predictability in the extratropics.
- The degree to which ENMs present a reduction of atmospheric dynamics in terms of the average response of the system to stochastic and other forcings and use the phase space of ENMs to understand changes in S2S variability with climate change, including weather-regime responses to climate change (structure and occurrence), by performing intercomparisons of global and regional climate models as recommended by Palmer (1999).

Using statistical and theoretical research programs together will improve our knowledge of the S2S forecast problem and point the way toward exploiting new sources of predictability. For example, Brunet (1994) and the subsequent applications of ENM analysis to the observed and simulated global atmosphere (Zadra et al., 2002a,b) provided the guidance needed to look at the problem of the MJO and the NAO two-way interaction problem through teleconnections (Lin et al., 2009) and its impact on NWP skill (Lin et al., 2010a; Lin and Wu, 2011). As in the previous example, to make future advances in the S2S forecast problem, we also will need to use a hierarchy of GCMs of increasing complexity to gain the necessary dynamical and physical insights (e.g., Derome et al., 2005; Lin et al., 2007).

We believe that the S2S forecast problem is at the forefront of the weather and climate predictability continuum, where S2S variability can be represented by a finite number of relatively large-scale discretelike modes. These discretelike modes evolve in a complex manner through nonlinear interactions with themselves and transient eddies and weak dissipative processes. The sources of predictability are a mixture of fast adiabatic and slow diabatic processes that can be differentiated and diagnosed properly with a phase space approach based on ENM and MLM theories. Although the approach described here is not unique, the key to better prediction of S2S variability and weather regimes in a changing climate lies in improved understanding of the fundamental nature of S2S phase space structure and associated predictability arising from dynamical processes.

Acknowledgments

We thank Martin Charron, Yongsheng Chen, Yosvany Martinez, and Ayrton Zadra for their important contributions to ENM analyses. G. Brunet would like to express thanks to Hai Lin for his continuous leadership, discussions, and support throughout the years on the S2S diagnostic and prediction problem. Also, many thanks for the work of Tom Frame in cosupervision of the master's dissertation projects of Lina Boljka and Carlo Cafaro, which have advanced the ENM analysis, including the MLM background-state and lower-boundary-wave activity, and led to Fig. 3, and Paul Berrisford, for his contributions to the development of the background-state calculation. We thank Fréderic Vitart and Andrew Robertson for their helpful comments.