



# Theory of the norm-induced metric in atmospheric dynamics

T.-Y. Koh<sup>1,2</sup> and F. Wan<sup>1</sup>

<sup>1</sup>School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore

<sup>2</sup>Earth Observatory of Singapore, Nanyang Technological University, Singapore

Correspondence to: T.-Y. Koh (kohty@ntu.edu.sg)

Received: 6 January 2014 – Published in Atmos. Chem. Phys. Discuss.: 11 February 2014

Revised: 3 February 2015 – Accepted: 21 February 2015 – Published: 9 March 2015

**Abstract.** We suggest that some metrics for quantifying distances in phase space are based on linearized flows about unrealistic reference states and hence may not be applicable to atmospheric flows. A new approach of defining a norm-induced metric based on the total energy norm is proposed. The approach is based on the rigorous mathematics of normed vector spaces and the law of energy conservation in physics. It involves the innovative construction of the phase space so that energy (or a certain physical invariant) takes the form of a Euclidean norm. The metric can be applied to both linear and nonlinear flows and for small and large separations in phase space. The new metric is derived for models of various levels of sophistication: the 2-D barotropic model, the shallow-water model and the 3-D dry, compressible atmosphere in different vertical coordinates. Numerical calculations of the new metric are illustrated with analytic dynamical systems as well as with global reanalysis data. The differences from a commonly used metric and the potential for application in ensemble prediction, error growth analysis and predictability studies are discussed.

dictability, of a chaotic system (Yoden, 1987). In practice, prediction also involves assimilating data to bring the first-guess modelled state into the “neighbourhood” of the observed state, putting an extrinsic constraint on predictability. In ensemble prediction methods, a cluster of close initial model states may be generated around an analysed state to yield “optimally growing” error structures so as to cover most efficiently the range of forecast uncertainty and guide targeted observations (Palmer et al., 1998; Mu et al., 2003).

In the preceding notions of “separation”, “neighbourhood” and “optimally growing”, the definition of a metric that measures the distance between two states is fundamental. There are a number of metrics used in the literature and many authors may hold the view that the definition of a metric is somewhat arbitrary a priori, especially in the weights given to the differently dimensioned state variables. The particular choice is often taken to depend on the application in mind, whether for investigating the theoretical predictability in a model, estimating optimally growing perturbations, or minimizing model departures from observations. For example, if temperature is rather constant in a region, more emphasis may be given to wind in the metric used to evaluate the theoretical predictability in that region; if temperature forecast is particularly bad, more emphasis may be put on temperature in the metric used to generate optimally growing perturbations or to minimize initial model errors.

In principle, one can adopt any expression to measure the distance between two points in the phase space of a dynamical system as long as the expression satisfies the properties of a metric. But only some expressions may have associated physical significance. For example, geopotential height is often taken to represent well the wind and temperature in mid-latitude regions through geostrophic and hydrostatic balance respectively. So in these regions, the phase space is

## 1 Introduction

### 1.1 The context

In predictability studies, the sensitivity of numerical models to initial conditions is an important topic. It has been demonstrated in Lorenz’s (1963) pioneering work that slightly different initial states diverge exponentially over time. Thus, theoretical predictability is often measured by the Lyapunov exponent, which is roughly speaking the long-term growth rate of the “separation” between neighbouring states (Lorenz, 1965). This characterizes only one aspect, the intrinsic pre-

single-variate and the metric may be defined simply from the  $p_1$  norm, i.e. the domain integral of the absolute difference in geopotential between two atmospheric states (Ding and Li, 2007). In multi-variate phase space, the situation is more complex as there are various ways of combining state variables into a single metric (Lorenz, 1969; Molteni and Palmer, 1993; Mu et al., 2003). No doubt each of these definitions have its merits for the purposes they serve. While the value of a metric for achieving a practical purpose is important in applications, our current work is mainly concerned with the fundamental theoretical question: is there a distinguished mathematical formulation of the metric that is consistent with the intrinsic dynamics of a physical system? Without jumping too much ahead, the answer lies in having a well-reasoned methodology to formulate such a metric rather than in a particular form of the metric itself.

Energy-like metrics are most commonly used to measure the distance between two states of the atmosphere. Some definitions look similar to wave energy (Bannon, 1995; Zou et al., 1997; Kim et al., 2011), while others use quadratic expressions that resemble kinetic and available potential energy (Buizza et al., 1993; Zhang et al., 2003; Leutbecher and Palmer, 2008; Rivière et al., 2009). But none of these metrics are truly energy or energy differences, as already noted by some authors (e.g. Ehrendorfer and Errico, 1995). Palmer et al. (1998) summarized and compared a number of metrics inspired by expressions of kinetic energy and total energy. Like energy, enstrophy is another dynamical invariant under certain conditions and Palmer et al. (1998) also investigated an enstrophy-like metric. But the commonality of such approaches lies in (1) the identification of a dynamical invariant, and (2) the formulation of a metric. The former is rather well-established in atmospheric dynamical theory; it is the latter formulation that needs clarification. We shall first review an often used metric as a concrete illustration of the problem.

## 1.2 An example of a metric

Talagrand (1981) considered dry, compressible flows linearized about a reference atmosphere at rest with temperature  $T_0$  and surface pressure  $p_0$  in the absence of surface topography, where  $T_0$  and  $p_0$  are constant in time and uniform in space. The following integral of quadratic forms over a horizontal domain  $A$  is conserved by the linearized flow when it is adiabatic and inviscid:

$$E_{T81} = \frac{1}{2} \int_A \int_0^{p_0} \left( u^2 + v^2 + \frac{c_p}{T_0} T'^2 \right) dp dA + \frac{1}{2} \int_A \frac{RT_0}{p_0} p_s'^2 dA, \quad (1)$$

where  $c_p$  and  $R$  are specific heat capacity at constant pressure and specific gas constant of dry air. The state variables

$u$ ,  $v$ ,  $T$  and  $p_s$  are zonal wind, meridional wind, temperature and surface pressure respectively while primes denote perturbations from the reference state. The tendency of  $E_{T81}$  is a small, time-varying fraction of the true energy tendency. A derivation of  $E_{T81}$  is given in Sect. A1 of the Appendix.

As stated in Sect. 4 of Ehrendorfer and Errico (1995), the temperature perturbation term in  $E_{T81}$  is the available potential energy (APE) of the system linearized about an isothermal atmosphere. But here we note that it is not the APE of the real atmosphere which has a non-trivial lapse rate (Lorenz, 1955, 1960):

$$\text{APE} = \frac{1}{2} \int_A \int_0^{p_0} \frac{1}{\Gamma_d - \Gamma} \frac{(T - \bar{T})^2}{\bar{T}} dp dA, \quad (2)$$

where  $\Gamma$  is the atmospheric lapse rate,  $\Gamma_d \equiv g/c_p$  is the adiabatic lapse rate,  $\bar{T}$  is the global isobaric mean temperature and  $p_0 = 1000$  mb.

The associated metric,  $M_{T81}$ , proposed by Talagrand (1981) is given by

$$M_{T81}^2 = \frac{1}{2} \int_A \int_0^{p_0} \left( (\delta u)^2 + (\delta v)^2 + \frac{c_p}{T_0} (\delta T)^2 \right) dp dA + \frac{1}{2} \int_A \frac{RT_0}{p_0} (\delta p_s)^2 dA, \quad (3)$$

where  $\delta$  denotes the difference between two evolving atmospheric states.  $M_{T81}$  is likewise invariant under the linearized dynamics.

The expression defined in Eq. (3) was originally formulated to study the convergence of the modelled state to the observed state with repeated data assimilation cycles in Talagrand (1981). It was used later by a number of authors (Ehrendorfer and Errico, 1995; Errico, 2000; Mu et al., 2009; Qin and Mu, 2012) to measure the evolving difference between two atmospheric states. Unfortunately, in the latter applications, the uniform  $T_0$  in Eq. (3) has lost its physical meaning as a reference state about which linearization takes place due to large realistic values of lapse rates. As Talagrand (1981) noted,  $E_{T81}$  and  $M_{T81}$  are not conserved due to non-linearity even if realistic flows were adiabatic and inviscid. We note additionally that significant surface topography like the Tibetan Plateau, the Rockies and the Andes would also invalidate the conservation of  $E_{T81}$  in realistic flows and renders questionable the use of  $M_{T81}$  as a metric. Many authors are probably aware of these shortcomings but for the lack of a better choice, continue to employ Eq. (3).

At a more fundamental level, while a dynamical invariant is a good metric to diagnose the change in a system due to data assimilation (which disrupts model dynamics and hence does not conserve that invariant), it is not a suitable metric to investigate sensitivity to initial conditions or to search for optimally growing initial perturbations, precisely because

it does not change during dynamical evolution. The former objective was the subject of Talagrand (1981) and he succeeded in finding  $M_{T81}$  as such an invariant metric in linearized flows. The latter two objectives were the interests of many other authors (Ehrendorfer and Errico, 1995; Errico, 2000; Qin and Mu, 2012) who used  $M_{T81}$  or  $M_{T81}$ -like metrics, sometimes appealing to the conditioned conservation of  $E_{T81}$  as a motivation. But for  $E_{T81}$  to be conserved, the flow has to be linear which also means that  $M_{T81}$  is invariant and useless for detecting growing perturbations. This inherent contradiction was not realized in the literature, no doubt because realistic flows manifest significant nonlinearity, thus never revealing the otherwise invariant property of  $M_{T81}$ , but also never conserving  $E_{T81}$ . This makes  $M_{T81}$  no more or less justifiable than other metrics, e.g. the total difference energy of Zhang et al. (2007) which is  $M_{T81}$  less the surface pressure contribution. Talagrand (1981) clearly did not intend his metric to be employed for those latter purposes while the community continues to use  $M_{T81}$  without a firm theoretical basis.

### 1.3 The essential problem

The essential question that this paper addresses is this: can a metric be theoretically determined a priori, other than being designed to fit a particular practical purpose a posteriori? In this theoretical work, we aim to develop a methodology to construct new *non-invariant* metrics based rigorously and consistently on invariant norms. These metrics should overcome the limitations of having unrealistic reference states and the need to linearize the flow about those states.

The organization of the paper is as follows. Section 2 illustrates the methodology by constructing energy-based metrics for the 2-D barotropic model and for the shallow-water model. Section 3 follows the same methodology and derives energy-based metrics for the dry, compressible model in different vertical coordinates. In Sects. 4–6, the new metrics are applied to analytic dynamical systems and reanalysis data of the atmosphere. Finally in Sect. 7, we discuss the theoretical and practical advantages of using an invariant norm-induced metric.

## 2 Basic methodology for simple fluid systems

### 2.1 Mathematical foundation

A norm on  $S$  can be any function  $\|\bullet\|: S \rightarrow [0, +\infty)$  which satisfies the following properties (Davidson and Donsig, 2010): non-negativity, absolute homogeneity, triangle inequality, and is zero only for the zero vector. Although given the flexibility of defining the norm, some norms may be interpreted with physical meanings while others may not when the vector space represents the state of a physical system. For atmospheric dynamical systems, a natural candidate for the norm is the square-root of energy which is invariant in the un-

forced flow. (“Forcing” here refers generally to diabatic heating, dissipation, mechanical forcing or gain/loss through domain boundaries.) The importance of the energy-norm is its conservation property so that any change in the norm means there is a net forcing or energy flux in or out of the system.

In a normed vector space, the metric between two vectors can be defined as the norm of the difference between them, and is called the “norm-induced metric”. But there are other ways of constructing a metric without first defining a norm, because a metric only needs to satisfy the following properties: non-negativity, identity of indiscernibles, symmetry and triangle inequality (Davidson and Donsig, 2010). When an inner product is defined for a vector space, the inner product of the difference between two vectors with itself yields the square of the norm-induced metric. In the literature, both the norm (Buizza et al., 1993; Ehrendorfer and Errico, 1995; Leutbecher and Palmer, 2008) and the inner product (Palmer et al., 1998) have been used to define a metric. We have the following hierarchy (Davidson and Donsig, 2010):

$$\{\text{metricspace}\} \supseteq \{\text{normed space}\} \supseteq \{\text{innerproductspace}\}.$$

In this section, the energy norm and the norm-induced metric are constructed on the phase space of two simple fluid systems, namely the 2-D barotropic model and the shallow-water model as an illustration of the basic methodology.

The two models are assumed to cover a horizontal domain  $A$  with periodic lateral boundary conditions.

### 2.2 2-D barotropic model

For the 2-D barotropic model in Cartesian coordinates  $(x, y)$ , the kinetic energy

$$E = \frac{1}{2} \int_A (u^2 + v^2) dA, \quad (4)$$

is conserved, where  $\mathbf{u} = (u, v)$  is the velocity vector. The barotropic flow is fully described by the phase vector  $\mathbf{x} = (u, v)$ , where  $u$  and  $v$  are functions on the domain  $A$  and all the possible states form the vector space  $\{\mathbf{x} \in S\}$ . It is easy to verify that energy could be used to define a norm such that  $\|\mathbf{x}\|^2 = E$ . This is the familiar Euclidean norm on the vector space.

Let  $\mathbf{x}_1$  and  $\mathbf{x}_2$  be two vectors in  $S$ . The norm-induced metric is defined by

$$\|\mathbf{x}_1 - \mathbf{x}_2\|^2 = \frac{1}{2} \int_A \left( (u_1 - u_2)^2 + (v_1 - v_2)^2 \right) dA, \quad (5)$$

which is similar to the “error kinetic energy” defined by Lorenz (1969). For ease of reference, the norm-induced metric can also be called the “separation” in this work.

### 2.3 Shallow-water model

For the shallow-water model, the sum of kinetic and geopotential energy,

$$E = \frac{1}{2} \int_A (h\mathbf{u}^2 + gh^2) dA, \quad (6)$$

is conserved (Vallis, 2006), where  $h$  is the height of the water surface,  $g$  is the gravitational acceleration, and the other symbols have the same meaning as in the barotropic model. Before making use of this energy expression as a norm, the subtle question is how the phase vector should be constructed.

The easiest way of constructing the phase space is adding another “dimension” to the phase space of the barotropic model, which results in a three-dimensional phase vector  $(\mathbf{u}, \mathbf{v}, h)$ .

However,  $\sqrt{E}$  is not a norm in this vector space because it does not have the property of absolute homogeneity, i.e.  $\sqrt{E[\mu u, \mu v, \mu h]} \neq \|\mu\| \sqrt{E[u, v, h]}$ , where  $\mu$  is a real number.

Moreover, since the integrand in Eq. (6) is not quadratic, despite being non-negative for a single state, it is not so for the corresponding difference vector between two states. So we also cannot use Eq. (6) to define a metric for this phase space.

The phase vector is constructed instead as  $(\sqrt{hu}, \sqrt{hv}, h)$  so that the energy is the Euclidean norm in this vector space. Let  $\mathbf{x}_1 = (\sqrt{h_1}u_1, \sqrt{h_1}v_1, h_1)$  and  $\mathbf{x}_2 = (\sqrt{h_2}u_2, \sqrt{h_2}v_2, h_2)$  be two phase vectors. The norm-induced metric or separation,  $M$ , is given by

$$M^2 = \|\mathbf{x}_1 - \mathbf{x}_2\|^2 = \frac{1}{2} \int_A \left( (\delta\sqrt{hu})^2 + g(\delta h)^2 \right) dA, \quad (7)$$

where  $\delta$  denotes taking the difference between the two phase vectors, e.g.  $\delta\sqrt{hu} = (\sqrt{h_1}u_1 - \sqrt{h_2}u_2)$ .

In mathematics, the axiomatic approach usually defines a norm after the construction of a vector space. But in physics, we suggest that it is more useful to first identify the physical quantity which we desire to be the norm and then try to construct the vector space such that this quantity is indeed a norm on that vector space. In the above example, we have adopted the latter approach and constructed the vector space such that energy is the familiar Euclidean norm again. The norm-induced metric follows naturally thereafter.

### 2.4 Linearization of separation

Let the norm be  $E = f(a, b) + g(a, b)$ , where  $f$  and  $g$  are positive definite functions of variables  $a$  and  $b$ , so that  $E$  is the Euclidean norm on the vector space  $\{\mathbf{x} \in S | \mathbf{x} = (\sqrt{f}, \sqrt{g})\}$ . If variables  $a$  and  $b$  are observed and recorded in practice rather than  $f$  and  $g$ , it is more convenient to transform  $(\sqrt{f}, \sqrt{g})$  coordinates to the more conventional  $(a, b)$  coordinates. Although the relation between

the two coordinate systems may involve nonlinear transformations, the increments  $(\delta\sqrt{f}, \delta\sqrt{g})$  can be approximated by linear combinations of  $(\delta a, \delta b)$  assuming the increments are small (Fleisch, 2011). One example is the transformation from Cartesian coordinates  $(x, y, z)$  to spherical polar coordinates  $(\lambda, \phi, r)$  where  $\lambda$  is longitude,  $\phi$  is latitude and  $r$  is the distance from origin (Fleisch, 2011).

The norm-induced metric in  $(\sqrt{f}, \sqrt{g})$  coordinates is given by

$$M^2 = (\delta\sqrt{f}, \delta\sqrt{g}) (\delta\sqrt{f}, \delta\sqrt{g})^T. \quad (8)$$

Using the total increment theorem,  $M$  can be approximated in the tangent space  $\{\mathbf{y} \in T | \mathbf{y} = (\delta a, \delta b)\}$  by

$$M^2 \approx (\delta a, \delta b) \mathbf{G} (\delta a, \delta b)^T, \quad (9)$$

where the metric tensor

$$\mathbf{G} = \frac{1}{4} \begin{pmatrix} f_a^2/f + g_a^2/g & f_a f_b/f + g_a g_b/g \\ f_a f_b/f + g_a g_b/g & f_b^2/f + g_b^2/g \end{pmatrix}, \quad (10)$$

where the subscripts denote derivatives with respect to that variable. Notice that when  $f = a^2$  and  $g = b^2$ , the metric tensor is simply the identity matrix and Eq. (9) reproduces the Euclidean norm.

In the shallow-water model, we may transform  $(\sqrt{hu}, \sqrt{hv}, h)$  coordinates to the more conventional  $(u, v, h)$  coordinates. If two vectors  $\mathbf{x}_1$  and  $\mathbf{x}_2$  are close to each other, the total increment  $\delta\sqrt{hu}$  can be linearized about a reference state which could be either one of the two vectors. Hence, Eq. (7) can be rewritten as

$$M^2 \approx \frac{1}{2} \int_A (\delta u, \delta v, \delta h) \begin{pmatrix} h & 0 & u/2 \\ 0 & h & v/2 \\ u/2 & v/2 & (u^2 + v^2)/4h + g \end{pmatrix} \begin{pmatrix} \delta u \\ \delta v \\ \delta h \end{pmatrix} dA, \quad (11)$$

which is non-Euclidean as the metric tensor matrix is not diagonal. The separation-squared between two neighbouring states is linearized about one of them but importantly, the dynamics governing the evolution of both states remain nonlinear.

## 3 Dry compressible atmosphere

In this section, the definition of the separation metric is extended using the same methodology as above to fully compressible equations of a dry, adiabatic and inviscid atmosphere in different vertical coordinates. The horizontal domain  $A$  is assumed to be closed or periodic.

### 3.1 Pressure coordinate

The formulation of a metric in pressure coordinate  $p$  is useful because much observation and reanalysis data are presented on pressure levels. In  $p$ -coordinate, it can be proven that the following quantity is conserved (Trenberth, 1997):

$$E = \frac{1}{g} \int_A \int_0^{p_H} \left( \frac{1}{2} \mathbf{u}^2 + c_p T + \Phi_H \right) dp dA, \quad (12)$$

where  $T$  is temperature,  $\Phi_H$  is surface geopotential,  $p_H$  is surface pressure,  $c_p$  is the specific heat capacity of dry air at constant pressure, and the other symbols are as before. The energy in Eq. (12) is not a norm in the vector space  $(u, v, \sqrt{T})$  as  $p_H$  also appears as a variable in the upper limit of the first integral. This means that Eq. (12) cannot be used directly to define the norm-induced metric.

To make further progress, consider a constant reference pressure  $p_r$  close to but smaller than  $p_H$  such that the vertical integration of the first two terms in Eq. (12) could be separated into a main contribution  $[0, p_r]$  and a boundary-layer contribution  $[p_r, p_H]$ . In the boundary layer, the kinetic energy  $\mathbf{u}^2/2$  is always much less than  $c_p T$  and the temperature does not deviate much from a reference temperature  $T_r(x, y)$ , which could be conveniently defined by the vertical gradient of a hydrostatically balanced geopotential field  $\Phi_{\text{ref}}(x, y, p)$  at pressure  $p_r$ :

$$T_r = -\frac{p_r}{R} \left( \frac{\partial \Phi_{\text{ref}}}{\partial p} \right)_{p=p_r}. \quad (13)$$

So Eq. (12) can be approximated by

$$E \approx \frac{1}{g} \int_A \int_0^{p_r} \left( \frac{1}{2} \mathbf{u}^2 + c_p T \right) dp dA + \frac{1}{g} \int_A (c_p T_r (p_H - p_r) + \Phi_H p_H) dA, \quad (14)$$

where  $(p_H - p_r)$  represents the boundary-layer mass. Note that the atmosphere stays close to the reference state  $\Phi_{\text{ref}}(x, y, p)$  because hydrostatic balance must be dominant for the pressure coordinate to be reasonably employed. This implies that surface pressure and boundary-layer temperature must always stay close to their reference values. So the above approximation is as good as the hydrostatic balance implicitly assumed in pressure coordinate and a modified energy expression differing by a constant from Eq. (14) can be de-

finied:

$$E_{\text{mod}} = \frac{1}{g} \int_A \int_0^{p_r} \left( \frac{1}{2} \mathbf{u}^2 + c_p T \right) dp dA + \frac{1}{g} \int_A (c_p T_r + \Phi_H) p_H dA. \quad (15)$$

The energy expression Eq. (15) is used to define the norm with the phase vector defined as  $\mathbf{x} = (u, v, \sqrt{T}, \sqrt{p_H})$  since  $\Phi_H$  and  $T_r$  are time independent. The separation metric of compressible flows in pressure coordinate is given by

$$M^2 = \frac{1}{g} \int_A \int_0^{p_r} \left( \frac{1}{2} (\delta \mathbf{u})^2 + c_p (\delta \sqrt{T})^2 \right) dp dA + \frac{1}{g} \int_A (c_p T_r + \Phi_H) (\delta \sqrt{p_H})^2 dA, \quad (16)$$

where the three contributions by differences in wind, temperature and surface pressure are henceforth called kinetic, enthalpy and surface pressure components of separation-squared respectively.

Equation (16) could be approximated in terms of perturbations of the more conventional variables  $u, v, T$  and  $p_H$  as

$$M^2 \approx \frac{1}{g} \int_A \int_0^{p_r} \left( \frac{1}{2} (\delta \mathbf{u})^2 + c_p \frac{(\delta T)^2}{4T} \right) dp dA + \frac{1}{g} \int_A (c_p T_r + \Phi_H) \frac{(\delta p_H)^2}{4p_H} dA. \quad (17)$$

The approximation in Eq. (16) only requires  $\delta T/T \ll 1$  because hydrostatic balance then implies  $\delta p_H/p_H \ll 1$ .

The separation metric in Eq. (17) is linearized in the sense that it has been transformed into the tangent linear space at  $(T, p_H)$ . It is different from  $M_{T81}$  in Eq. (3) in the coefficients of temperature difference  $(\delta T)^2$  and surface pressure difference  $(\delta p_H)^2$  by factors 2 and  $2R/c_p$  respectively. In our expression, the reference state is realistic and can evolve nonlinearly with time. We also account for the influence of surface topography. Moreover, no linearization is assumed in the flow dynamics in developing Eq. (17) and a more accurate expression for the separation metric is available in Eq. (16) for atmospheric states that are not close, i.e.  $\delta T/T \gtrsim 0.1$ , or less likely,  $\delta p_H/p_H \gtrsim 0.1$ .

### 3.2 Isentropic coordinate

The use of potential temperature as a vertical coordinate dates back half a century when for example, Lagrangian parcel trajectories were traced on isentropic surfaces (Green

et al., 1966). Hoskins (1991) further proposed a potential vorticity – potential temperature view of the general circulation which has advantages in understanding atmospheric dynamics and advanced mid-latitude weather forecasts. Thus, it is of both theoretical and practical interest to examine our separation metric in isentropic coordinate.

In isentropic coordinate  $\theta$ , the conserved energy in Eq. (12) takes the form (Trenberth, 1997; Staniforth and Wood, 2003)

$$E = \int_A \int_{\theta_H}^{\infty} \left( \frac{1}{2} \mathbf{u}^2 + \theta \Pi + \Phi_H \right) \sigma \, d\theta \, dA, \quad (18)$$

where  $\sigma = -g^{-1} \partial p / \partial \theta$  is isentropic density,  $\Pi = c_p (p/1000 \text{ mb})^{R/c_p}$  is Exner's function,  $\theta_H$  is surface potential temperature, and the other symbols are as before. Similar to the case of pressure coordinate, we define an appropriate reference potential temperature at the lower boundary  $\theta_r(x, y)$  to separate the main contribution and the boundary-layer contribution as

$$E \approx \int_A \int_{\theta_r}^{\infty} \left( \frac{1}{2} \sigma \mathbf{u}^2 + \theta \sigma \Pi + \sigma \Phi_H \right) d\theta \, dA + \int_A \int_{\theta_H}^{\theta_r} (\theta \Pi + \Phi_H) \sigma \, d\theta \, dA, \quad (19)$$

where we have ignored the kinetic energy in the boundary layer as before. Since  $\theta \Pi = c_p T$  and  $\sigma d\theta = \rho dz$ , where  $\rho$  is mass density and  $z$  is height, the boundary-layer term can be evaluated as

$$\int_{\theta_H}^{\theta_r} (\theta \Pi + \Phi_H) \sigma \, d\theta = \int_{z_H}^{z_{\theta_r}} \left( 1 + \frac{\Phi_H}{\theta \Pi} \right) \frac{c_p}{R} p \, dz \approx \left( 1 + \frac{\Phi_H}{\theta_r \Pi_r} \right) \frac{c_p}{R} p_r (z_{\theta_r} - z_H), \quad (20)$$

where  $z_H$  is the surface topography, and  $z_{\theta_r} = z(x, y, \theta_r, t)$  is the elevation of the  $\theta_r$ -surface as further elaborated in Sect. 3.5. The reference boundary-layer pressure  $p_r(x, y)$  and Exner's function  $\Pi_r(x, y)$  are defined from the vertical gradient of the hydrostatically balanced Montgomery potential field  $M_{\text{ref}}(x, y, \theta)$  at isentropic level  $\theta_r$ :

$$\Pi_r \equiv c_p (p_r/1000 \text{ mb})^{R/c_p} = \left( \frac{\partial M_{\text{ref}}}{\partial \theta} \right)_{\theta=\theta_r}. \quad (21)$$

This allows a modified energy expression differing from Eq. (19) by a constant to be defined:

$$E_{\text{mod}} = \int_A \int_{\theta_r}^{\infty} \left( \frac{1}{2} \sigma \mathbf{u}^2 + \theta \sigma \Pi + \sigma \Phi_H \right) d\theta \, dA + \frac{c_p}{R} \int_A \left( 1 + \frac{\Phi_H}{\theta_r \Pi_r} \right) p_r z_{\theta_r} \, dA. \quad (22)$$

We define the phase vector as  $\mathbf{x} = (\sqrt{\sigma} u, \sqrt{\sigma} v, \sqrt{\sigma} \Pi, \sqrt{\sigma}, \sqrt{z_{\theta_r}})$  so that the separation metric in isentropic coordinate induced by the Euclidean norm in Eq. (22) is given by

$$M^2 = \int_A \int_{\theta_r}^{\infty} \left( \frac{1}{2} (\delta \sqrt{\sigma} \mathbf{u})^2 + \theta (\delta \sqrt{\sigma} \Pi)^2 + \Phi_H (\delta \sqrt{\sigma})^2 \right) d\theta \, dA + \frac{c_p}{R} \int_A \left( 1 + \frac{\Phi_H}{\theta_r \Pi_r} \right) p_r (\delta \sqrt{z_{\theta_r}})^2 \, dA. \quad (23)$$

Compared to the separation metric in pressure coordinate, Eq. (23) depends on one more thermodynamic variable, the isentropic density  $\sigma$ , because the flow is compressible in isentropic coordinate whereas it is non-divergent in pressure coordinate. The linearized separation-squared in the tangent linear space  $(u, v, \Pi, \sigma, z_{\theta_r})$  is given by the non-Euclidean form

$$M^2 \approx \int_A \int_{\theta_r}^{\infty} \left( \frac{\sigma}{2} (\delta \mathbf{u})^2 + \theta \sigma \frac{(\delta \Pi)^2}{4 \Pi} + \left( \frac{u^2}{2} + \theta \Pi + \Phi_H \right) \frac{(\delta \sigma)^2}{4 \sigma} + (\mathbf{u} \cdot \delta \mathbf{u} + \theta \delta \Pi) \frac{\delta \sigma}{2} \right) d\theta \, dA + \frac{c_p}{R} \int_A \left( 1 + \frac{\Phi_H}{\theta_r \Pi_r} \right) p_r \frac{(\delta z_{\theta_r})^2}{4 z_{\theta_r}} \, dA. \quad (24)$$

The variation of the elevation of the isentropic surface  $\theta_r$  can be further related to the variation of the surface potential temperature at each location  $(x, y)$ :

$$\delta z_{\theta_r} \approx - \frac{\delta \theta_H}{\Theta_{zr}}, \quad (25)$$

$$\Theta_{zr} = - \frac{g}{\theta_r} \left( \frac{\partial^2 M_{\text{ref}}}{\partial \theta^2} \right)_{\theta=\theta_r}^{-1}, \quad (26)$$

where  $\Theta_{zr}(x, y)$  is the reference (positive) static stability in the boundary layer defined consistently above as

$$\frac{\partial \theta}{\partial z} \frac{\partial^2 M}{\partial \theta^2} = \frac{\partial \theta}{\partial z} \frac{\partial \Pi}{\partial \theta} = \frac{\partial p}{\partial z} \frac{d \Pi}{d p} = -g \rho \frac{R \Pi}{c_p p} = -g \frac{\Pi}{c_p T} = -\frac{g}{\theta}. \quad (27)$$

### 3.3 Geopotential height coordinate

Both pressure and isentropic coordinate formulations above are limited by their underlying assumption of hydrostatic balance. Current numerical weather prediction (NWP) models are able to model non-hydrostatic flows at mesoscale resolution, and many predictability studies are conducted based on NWP model results (Zhang et al., 2007; Hohenegger and Schar, 2007; Qin and Mu, 2012). Therefore, it is useful to derive the separation metric without making the hydrostatic assumption and here we make use of the geopotential height coordinate.

In geopotential height coordinate,  $z \equiv \Phi/g$  and total energy is the sum of kinetic energy, internal energy and geopotential energy (Vallis, 2006)

$$E = \int_A \int_{z_H}^{\infty} \left( \frac{1}{2} \rho \mathbf{v}^2 + \rho c_v T + \rho g z \right) dz dA, \quad (28)$$

where  $\mathbf{v} = (u, v, w)$  is the 3-D velocity,  $c_v$  is the specific heat capacity for dry air at constant volume, and the other symbols are as before. For an ideal gas, Eq. (28) can be rewritten as

$$E = \int_A \int_{z_H}^{\infty} \left( \frac{1}{2} \rho \mathbf{v}^2 + \frac{c_v}{R} p + \rho g z \right) dz dA. \quad (29)$$

By defining the phase vector as  $\mathbf{x} = (\sqrt{\rho}u, \sqrt{\rho}v, \sqrt{\rho}w, \sqrt{p}, \sqrt{\rho})$ , Eq. (29) specifies a Euclidean norm. Hence, the separation metric is given by

$$M^2 = \int_A \int_{z_H}^{\infty} \left( \frac{1}{2} (\delta \sqrt{\rho} \mathbf{v})^2 + \frac{c_v}{R} (\delta \sqrt{p})^2 + g z (\delta \sqrt{\rho})^2 \right) dz dA. \quad (30)$$

Actually,  $M$  is not dependent on the precise set of variables  $E$  is expressed in. The separation metric induced by Eq. (28) is the equivalent to that induced by Eq. (29). Compared to the separation metric in pressure and isentropic coordinate, Eq. (30) depends on  $w$  because the flow is non-hydrostatic. Note that the absence of a boundary-layer term in Eq. (30) is because the bottom boundary is rigid in geopotential height coordinate.

Equation (30) can be linearized and approximated in  $(u, v, w, p, \rho)$ -space as

$$M^2 \approx \int_A \int_{z_H}^{\infty} \left( \frac{\rho}{2} (\delta \mathbf{v})^2 + \frac{c_v}{R} \frac{(\delta p)^2}{4p} + (w^2/2 + g z) \frac{(\delta \rho)^2}{4\rho} + \frac{\mathbf{v}}{2} \cdot \delta \mathbf{v} \delta \rho \right) dz dA. \quad (31)$$

The linear approximation requires the fractional difference  $\delta p/p$  and  $\delta \rho/\rho$  to be much smaller than one. The terms

above are physical analogues to those in Eq. (11) for the shallow-water model, apart from the additional internal energy term. Equation (31) can be further simplified to

$$M^2 \approx \int_A \int_{z_H}^{\infty} \left( \frac{1}{2} \rho (\delta \mathbf{v})^2 + \frac{c_v}{R} \frac{(\delta p)^2}{4p} + g z \frac{(\delta \rho)^2}{4\rho} \right) dz dA, \quad (32)$$

if  $|\delta \mathbf{v}|/|\mathbf{v}| \gg |\delta \rho|/\rho$  holds true over most of the integration domain.

### 3.4 Generalized coordinate and finite upper boundary

In the preceding sections, the upper boundary of the atmosphere is always assumed to be at zero or infinity. But it is impossible to span the whole atmosphere in a numerical model and a finite upper boundary is prescribed. In this section, we treat the case of a generalized vertical coordinate with finite upper and lower boundaries. The two boundaries are assumed to be material surfaces to conserve the mass between them.

It has been shown (Kasahara, 1974; Staniforth and Wood, 2003) that energy for a dry, compressible atmosphere in generalized vertical coordinate  $s$  takes the form

$$E = \int_A \int_{s_H}^{s_T} \left( (u^2 + v^2 + \epsilon_v w^2)/2 + c_v T + \Phi \right) \sigma_s ds dA, \quad (33)$$

where  $\sigma_s = \rho \partial z / \partial s$ ,  $\Phi$  is the geopotential and  $\delta_v$  is the switch between non-hydrostatic ( $\epsilon_v = 1$ ) and hydrostatic ( $\epsilon_v = 0$ ) flows. The subscripts  $H$  and  $T$  denote values at the lower and upper boundaries respectively and the other symbols are as before. The integrand is similar to that in geopotential height coordinate except that the density is multiplied by the Jacobian of the vertical coordinate transformation.

$E$  is conserved only if the upper boundary is a rigid lid, i.e.  $z_T = z_T(x, y)$ , so that no work is done there. But this boundary condition is not realistic for the atmosphere. Instead, we consider the case of an ‘‘elastic lid’’, i.e.  $p_T = \text{constant}$ , where an energy-like invariant exists (Staniforth et al., 2003). For a non-hydrostatic atmosphere, this invariant is

$$E = \int_A \int_{s_H}^{s_T} \left( \sigma_s (u^2 + v^2 + w^2)/2 + \sigma_s c_v T + \sigma_s \Phi + p_T J \right) ds dA, \quad (34)$$

where  $J = \partial z / \partial s$  is the Jacobian of the vertical coordinate transformation. The last term in the integrand arises from work done at the upper boundary. For a hydrostatic atmosphere, the energy-like invariant in Eq. (34) is simplified by combining pressure work, internal energy and gain in geopotential above the surface into enthalpy and dropping away the vertical velocity contribution (Staniforth et al., 2003) to get

$$E = \int_A \int_{s_H}^{s_T} \left( (u^2 + v^2)/2 + c_p T + \Phi_H \right) \sigma_s ds dA. \quad (35)$$

Note that the energy density in Eq. (35) reproduces the energy density in pressure and isentropic coordinates with zero and infinite upper boundary respectively (see Eqs. 12 and 18).

As before, Eqs. (34) and (35) can be approximated by decomposing  $E$  into an integral with constant integration limits  $[s_L, s_U]$  over the main atmospheric body and boundary-layer contributions over  $[s_H, s_L]$  and  $[s_U, s_T]$ . To make use of the rigid lower boundary condition, we integrate with respect to  $z$  over the lower boundary layer (except for  $s \equiv p$ , see Sect. 3.1). Likewise, to make use of the elastic upper boundary condition, we integrate with respect to  $p$  over the upper boundary layer. We make the hydrostatic approximation in both boundary-layer integrations because hydrostatic balance is still dominant in the atmosphere even when the flow is non-hydrostatic. Thus, modified energy expressions for the non-hydrostatic and hydrostatic atmosphere,  $E_{\text{mod}}^{\text{nh}}$  and  $E_{\text{mod}}^{\text{h}}$  respectively, can be defined after dropping away constant contributions:

$$\begin{aligned} E_{\text{mod}}^{\text{nh}} &= \int_A \int_{s_L}^{s_U} \left( \sigma_s (u^2 + v^2 + w^2) / 2 + c_v \sigma_s T + \sigma_s \Phi + p_T J \right) ds dA \\ &+ \int_A \left( c_v T_L + \Phi_H + \frac{p_T}{\rho_L} \right) \rho_L z_L dA \\ &+ \frac{1}{g} \int_A \left( \frac{c_p}{R} \frac{p_T}{\rho_U} + \Phi_U \right) p_U dA, \end{aligned} \quad (36)$$

$$\begin{aligned} E_{\text{mod}}^{\text{h}} &= \int_A \int_{s_L}^{s_U} \left( \sigma_s (u^2 + v^2) / 2 + c_p \sigma_s T + \sigma_s \Phi_H \right) ds dA \\ &+ \int_A (c_p T_L + \Phi_H) \rho_L z_L dA \\ &+ \frac{1}{g} \int_A \left( \frac{c_p}{R} \frac{p_T}{\rho_U} + \Phi_H \right) p_U dA, \end{aligned} \quad (37)$$

where  $\rho_L$  and  $T_L$  are reference functions at  $s_L$ , while  $\rho_U$  and  $\Phi_U$  are reference functions at  $s_U$ , all of which are functions of  $(x, y)$  only. The  $z_L$  is the elevation at  $s_L$  and  $p_U$  is the pressure at  $s_U$ . So the phase vectors  $\mathbf{x}^{\text{nh}}$  and  $\mathbf{x}^{\text{h}}$  respectively for the non-hydrostatic and hydrostatic atmosphere are defined as

$$\mathbf{x}^{\text{nh}} = \left( \sqrt{|\sigma_s|} u, \sqrt{|\sigma_s|} v, \sqrt{|\sigma_s|} w, \sqrt{|\sigma_s|} T, \sqrt{|\sigma_s|} \Phi, \sqrt{|J|}, \sqrt{z_L}, \sqrt{p_U} \right), \quad (38)$$

$$\mathbf{x}^{\text{h}} = \left( \sqrt{|\sigma_s|} u, \sqrt{|\sigma_s|} v, \sqrt{|\sigma_s|} T, \sqrt{|\sigma_s|}, \sqrt{z_L}, \sqrt{p_U} \right), \quad (39)$$

and the norm-induced metrics can be defined as before. For a non-hydrostatic atmosphere, the degree of freedom  $|J|$  in Eq. (34) compensates for the loss in internal energy due to work done by the atmosphere at the upper boundary.

But  $|J|$  is not a degree of freedom for a hydrostatic atmosphere. The reason is that when pressure  $p$  is kept fixed,  $c_p(dT)_p \equiv (\delta Q)_p$  so that enthalpy in Eq. (35) can only be changed by heat transfer and is invariant to work done at the upper boundary.

In geopotential height coordinate ( $s \equiv z$ ), the Jacobian  $J$  is identical to one and so drops out from the phase vector of a non-hydrostatic atmosphere, while  $\Phi$  is a function of the coordinate only and so  $\sqrt{\rho}$  and not  $\sqrt{\rho}\Phi$  is the phase coordinate. The lower-boundary coordinate is time-independent ( $z_L \equiv z_H$ ) and so the lower-boundary integrals are constant and can be dropped from Eqs. (36) and (37). So our generalization is consistent with the results of Sect. 3.3.

In pressure coordinate ( $s \equiv p$ ), the ‘‘density’’  $|\sigma_s| = 1/g$  is a constant and so drops out from the phase vector. The upper-boundary coordinate is constant ( $p_U \equiv p_T$ ) and so the upper-boundary integrals are also constant and can be dropped from Eq. (37). So the separation metrics in Eqs. (16) and (17) in pressure coordinate are still valid for an atmosphere with an elastic lid, although the vertical integrals start from  $p_T$  instead of zero.

When the atmosphere has an elastic lid, the metrics for geopotential height and isentropic coordinates have additional upper-boundary terms while the vertical integrals over the main atmospheric body have finite constant upper limits, unlike Eqs. (23) and (30). The pressure change  $\delta p_U$  on the constant upper-limit coordinate surface is directly related to the movement of the elastic lid as follows:

$$\delta p_U(z_U) \approx g \rho_U \delta z_T, \quad (40)$$

where  $\rho_U$  is the reference mass density at  $z_U$  in geopotential height coordinate:

$$\delta p_U(\theta_U) \approx \frac{g \rho_U}{\Theta_{zU}} \delta = \frac{g c_p}{R} \frac{p_T}{\Pi_T} \frac{\delta \theta_T}{\theta_U \Theta_{zU}} \quad (41)$$

where  $\Theta_{zU}$  is the reference static stability at  $\theta_U$  defined similarly to Eq. (26) and  $\Pi_T$  is the constant Exner function on the elastic lid in isentropic coordinate.

### 3.5 Elevation at the top of the lower boundary layer

Geopotential is assigned to be zero at the lowest point on Earth’s surface for energy to satisfy the non-negative requirement of a norm. So the null vector corresponds to a state where Earth’s surface is flat. This zero-point is not arbitrary as it is not possible to extract any more geopotential energy from this point by moving air around. So by definition, *geopotential height*  $z$  is also zero at the lowest point on Earth’s surface.

The  $z_L$  in Eqs. (36) and (37) (or  $z_{\theta r}$  in Eq. (30) for  $\theta$ -coordinate) is the *elevation* of the coordinate surface  $s_L$  at the top of the lower boundary layer (or  $\theta_r$  for  $\theta$ -coordinate). Elevation is really just a geometric coordinate that can have a zero-point at any level. Thus an arbitrary constant can be



added to  $E_{\text{mod}}^{\text{nh}}$  or  $E_{\text{mod}}^{\text{h}}$  so that  $z_L$  is arbitrary to a constant value. In other words,  $z_L$  need not be identical to the geopotential height  $z$  at  $s_L$  as is implicitly assumed so far. This means that our theoretical formulation of the norm and norm-induced metric is not yet complete at this juncture. (There is no corresponding problem at the upper boundary because  $p_U$  is a thermodynamic coordinate of which the zero-point is well-defined physically.)

In pressure coordinate, we treat the lower boundary differently, integrating with respect to  $p$  instead of  $z$ . The basic reason is that the formulation is simpler in pressure coordinate: the phase space has one less state variable (as “density” is a constant in pressure coordinate), the top of the boundary layer,  $p_r$ , is constant and the phase coordinate arising from the lower-boundary contribution is a well-defined thermodynamic function,  $\sqrt{p_H}$ . Moreover, our formulation in Eq. (17) has the advantage of sharing essentially the same form as the already widely used  $M_{T81}$  in Eq. (3) (apart from factors of 2 and  $2R/c_p$  in the temperature and surface pressure terms and the inclusion of surface topography). In contrast, integrating with respect to  $z$  leads to a phase coordinate  $\sqrt{z_L}$  where  $z_L$  is an arbitrary constant and in general, one must assume reference values for two thermodynamic functions ( $T_L$  and  $\rho_L$ ) instead of one.

The uniqueness of the pressure coordinate allows us to calibrate  $z_L$  by implementing Eq. (37) for  $s \equiv p$  and requiring that the lower boundary-layer contribution to the energy norm be equal to that in Eq. (15). Hence,

$$\frac{p_H}{g} = \rho_L (z_L + z_o), \quad (42)$$

where  $z_L$  is the geopotential height at  $s_L$  as before and  $z_o$  is the arbitrary constant to be calibrated. From hydrostatic balance, to first order,

$$p_H - p_r = g\rho_L (z_L - z_H), \quad (43)$$

where  $p_r$  is the *time-independent* reference pressure at the top of the boundary layer in pressure coordinate as defined in Sect. 3.1. Equations (42) and (43) imply

$$z_o = -z_H + \frac{p_r}{g\rho_L} = -z_H + \frac{RT_r}{g}, \quad (44)$$

where  $T_r$  is the reference boundary-layer temperature in pressure coordinate as defined in Sect. 3.1.

Now we define a generalized “local elevation”

$$Z_L \stackrel{\text{def}}{=} z_L - z_H + \frac{RT_L}{g}, \quad (45)$$

which is the elevation from a zero-point located locally at a distance  $RT_L/g$  below the surface.  $RT_L/g$  is the density scale height derived from the reference lower boundary-layer temperature  $T_L(x, y)$  in the generalized coordinate. The zero-point of local elevation is shallower underground

in regions of high terrain. When we add the relevant constants to  $E_{\text{mod}}^{\text{nh}}$  and  $E_{\text{mod}}^{\text{h}}$ , the phase coordinate arising from the lower boundary condition becomes the locally calibrated  $\sqrt{Z_L}$  instead of the globally calibrated, geopotential-based  $\sqrt{z_L}$ . Note that the definitions of geopotential and geopotential height are not affected by this local calibration.

Application of the calibration in isentropic coordinate leads to  $z_{\theta r}$  being replaced by

$$Z_{\theta r} \stackrel{\text{def}}{=} z_{\theta r} - z_H + \frac{R}{gc_p} \theta_r \Pi_r, \quad (46)$$

which is the local elevation of the  $\theta_r$ -surface in Sect. 3.2. This fixes the hidden problem in Eq. (24), and hence in Eq. (23), where the boundary-layer contribution could be arbitrarily small because  $z_{\theta r}$  in the denominator of the integrand is arbitrary to a constant. In pressure coordinate, the local elevation actually measures surface pressure as it can be shown that  $Z_L = (p_H/p_r)(RT_r/g)$ . Our theoretical formulation is now complete and consistent among all coordinates.

#### 4 Example I: geostrophic balanced flow of shallow-water model

In this section, the separation metric of the shallow-water model is applied to an axisymmetric geostrophically balanced flow in polar coordinates in a rotating frame. Let  $r$  and  $\lambda$  be the radial and angular coordinates respectively, and  $u$  and  $v$  be the radial and azimuthal velocity components respectively, and the rest of the symbols follow the same notation as in previous sections. The flow is initially at rest with height given by

$$h_o = \begin{cases} -h'_o + H_o & r < a_1 \\ H_o & r > a_1, \end{cases} \quad (47)$$

where  $H_o$  is the basic height and  $h'_o$  is the initial disturbance. The initial potential vorticity profile is

$$\xi = \frac{f + \zeta}{h} = \begin{cases} \frac{f}{-h'_o + H_o} & r < a_1 \\ \frac{f}{H_o} & r > a_1, \end{cases} \quad (48)$$

where  $\xi$  is potential vorticity (PV),  $f$  is the Coriolis parameter (or “planetary” vorticity), and  $\zeta = r^{-1} \partial(rv)/\partial r$  is the relative vorticity. The geostrophic balanced state can be solved analytically by PV conservation without assumption of  $h'_o \ll H_o$  as pointed by Mak (2011), where he gave the non-dimensional solution of geostrophic adjustment in Cartesian coordinates. The boundary conditions are:  $\partial h/\partial r = 0$  at  $r = 0$  and  $h = 0$  as  $r \rightarrow \infty$ , which means the azimuthal velocity at the origin is zero and the perturbation dies away at infinity.

For simplicity, the following non-dimensional variables are introduced:

$$\tilde{h}' = \frac{h}{H_o} - 1, \quad \tilde{r} = \frac{r}{L_d}, \quad \tilde{v} = \frac{v}{fL_d}, \quad \tilde{\xi} = \frac{\xi}{f/H_o}, \quad (49)$$

where  $\tilde{h}'$  is the non-dimensional perturbation height and  $L_d = \sqrt{gH_0}/f$  is the Rossby radius of deformation. As finally axisymmetry and geostrophic balance is restored,  $\tilde{u} = 0$  and  $\tilde{v} = \partial\tilde{h}'/\partial\tilde{r}$ . Since the initial disturbance could be strong, it is necessary to consider the advection of fluid columns from initial positions (Mak, 2011). Let  $a_2$  be the new PV discontinuity point, i.e.  $a_2 - a_1$  is the displacement of the PV boundary. Then the non-dimensional conservation of PV becomes

$$\frac{\frac{\partial^2 \tilde{h}'}{\partial \tilde{r}^2} + \frac{1}{\tilde{r}} \frac{\partial \tilde{h}'}{\partial \tilde{r}} + 1}{\tilde{h}' + 1} = \begin{cases} 1/\eta & \tilde{r} < a_2/L_d \\ 1 & \tilde{r} > a_2/L_d, \end{cases} \quad (50)$$

where  $\eta = -h'_0/H_0 + 1$ . The solution is determined by parameters  $h'_0/H_0$  and  $a_2/L_d$ .

Solving Eq. (50) separately for  $\tilde{r} < a_2/L_d$  and  $\tilde{r} > a_2/L_d$  and matching the solutions at  $\tilde{r} = a_2/L_d$  gives the balanced perturbation height

$$\tilde{h}' = \begin{cases} -\frac{h'_0}{H_0} \left(1 - \frac{I_0(\tilde{r}/\sqrt{\eta})}{M(a_2/L_d)}\right) & \tilde{r} \leq a_2/L_d \\ -\frac{h'_0}{H_0} \frac{K_0(\tilde{r})}{N(a_2/L_d)} & \tilde{r} > a_2/L_d, \end{cases} \quad (51)$$

where  $I_\alpha$  and  $K_\alpha$  are modified Bessel functions of the first and second kind, and

$$M(x) = I_0(x/\sqrt{\eta}) + \frac{1}{\sqrt{\eta}} \frac{K_0(x)}{K_1(x)} I_1(x/\sqrt{\eta}), \quad (52)$$

$$N(x) = K_0(x) + \sqrt{\eta} \frac{I_0(x/\sqrt{\eta})}{I_1(x/\sqrt{\eta})} K_1(x). \quad (53)$$

The balanced velocity can be obtained as

$$\tilde{v} = \begin{cases} \frac{h'_0}{H_0} \frac{1}{\sqrt{\eta}} \frac{I_1(\tilde{r}/\sqrt{\eta})}{M(a_2/L_d)} & \tilde{r} \leq a_2/L_d \\ \frac{h'_0}{H_0} \frac{K_1(\tilde{r})}{N(a_2/L_d)} & \tilde{r} > a_2/L_d, \end{cases} \quad (54)$$

where  $a_2$  is determined by the mass conservation equation

$$\int_0^\infty \tilde{h}' \tilde{r} d\tilde{r} = \int_0^{a_1/L_d} (-h'_0/H_0 \tilde{r}) d\tilde{r}. \quad (55)$$

The left-hand side of Eq. (55) is a monotonic decreasing function of  $a_2$ , hence the solution of  $a_2$  is unique and increases as  $a_1$  increases.

The non-dimensional separation metric in polar coordinates is given by

$$M^2 = \frac{1}{2} \int_0^\infty \left( (\delta\sqrt{\tilde{h}\tilde{v}})^2 + (\delta\tilde{h})^2 \right) \tilde{r} d\tilde{r}, \quad (56)$$

which linearizes as

$$M^2 \approx \frac{1}{2} \int_0^\infty \left( \tilde{v}\delta\tilde{v}\delta\tilde{h} + \tilde{h}(\delta\tilde{v})^2 + \left(\frac{\tilde{v}^2}{4\tilde{h}} + 1\right) (\delta\tilde{h})^2 \right) \tilde{r} d\tilde{r}, \quad (57)$$

where the first three quadratic terms involving  $\delta\tilde{v}$  and  $\tilde{v}$  sum up to approximate the kinetic separation-squared.

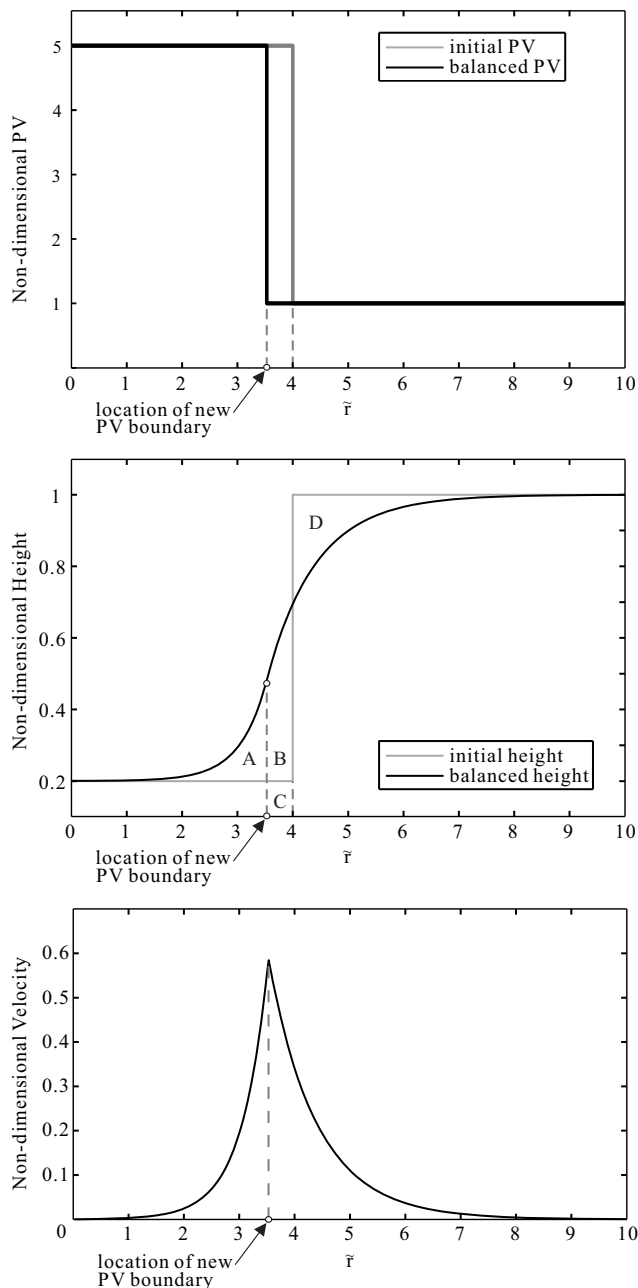
The non-dimensional PV profile is shown in Fig. 1a with  $a_1/L_d = 4$  and  $h'_0/H_0 = 0.8$ . Figures 1b and 1c show the non-dimensional solutions of height and tangential velocity. The PV boundary is displaced from  $\tilde{r} = 4$  to  $\tilde{r} = 3.53$ . The low-PV water mass originally at region D now moves to B and C and pushes the high-PV water mass originally at C to A. The tangential velocity maximizes at the new PV boundary  $a_2$ . The Rossby number  $R_o = V/fL = \tilde{v}L_d/a_2$  is about 0.17 which is small so that the geostrophic approximation is good.

In order to investigate the importance of the mixed term in separation-squared, two sets of balanced solutions with different initial height discontinuity but the same initial radius of high PV ( $a_1/L_d = 4$ ) are investigated. The first case is  $h'_0/H_0 = 0.1$ , where the flow is more like a linear system. The second case is  $h'_0/H_0 = 0.8$ , where the flow is nonlinear. Separation metrics are calculated by adding perturbations to the two control parameters  $a_1/L_d$  and  $h'_0/H_0$  for both cases. All perturbations are sufficiently small that linearization of separation-squared is good. The ratio between the mixed term  $\frac{1}{2} \int_0^\infty (\tilde{v}\delta\tilde{v}\delta\tilde{h}) \tilde{r} d\tilde{r}$  and the other two quadratic terms of the linearized kinetic separation-squared for both cases are shown in Fig. 2.

In the first case of almost linear flow, the mixed term's contribution is always less than 5% of the  $\delta\tilde{v}^2$  term's contribution. Hence, the non-Euclidean separation metric can be approximated by ignoring the mixed term. But in doing so, one must also ignore the kinetic enhancement of the  $\delta\tilde{h}^2$  term as it is generally even smaller. This is because the flow has only small PV differences which induces small velocities and so all terms involving  $\tilde{v}$  must be consistently ignored.

In the second case of nonlinear flow, the contribution of the mixed term could be comparable to that of the  $\delta\tilde{v}^2$  term unless the perturbations are almost entirely in the extent  $a_1/L_d$  rather than the magnitude  $h'_0/H_0$  of the initial low-PV fluid. Here, it is also generally inconsistent to keep the kinetic enhancement of the  $\delta\tilde{h}^2$  term without keeping the mixed term. Therefore, in nonlinear flows, the non-Euclidean characteristic of the linearized metric cannot be neglected because large PV differences lead to large velocities  $\tilde{v}$ .

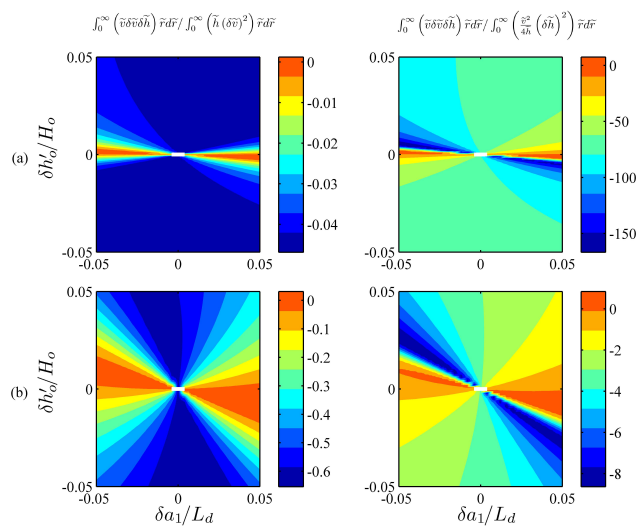
The kinetic separation-squared (the first term of Eq. 56) for the nonlinear flow where  $h'_0/H_0 = 0.8$  and the fractional error made in using the linearized expression (the first three terms in Eq. 57) are shown in Fig. 3. Note that the size of perturbations on  $a_1/L_d$  and  $h'_0/H_0$  is comparable to the parameters themselves. The linearized separation-squared is only valid for very small perturbations, or for a special subset of the dynamical parameters.



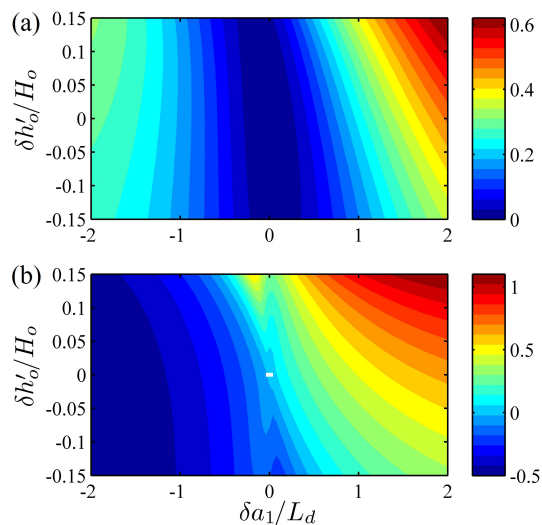
**Figure 1.** Non-dimensional solution of PV, height and tangential velocity with  $a_1/L_d = 4$  and  $h'_o/H_o = 0.8$  in the geostrophically balanced shallow-water model.

### 5 Example II: 2-D thermal wind model in pressure coordinate

In this section, the separation metric of a dry compressible atmosphere in pressure coordinate is applied to a 2-D thermal wind flow in the Northern Hemisphere. The zonal wind  $u$  at the surface is assumed to be zero. The potential temperature



**Figure 2.** The ratio between the mixed term and the other two quadratic terms of the linearized kinetic separation-squared when  $a_1/L_d = 4$  and (a)  $h'_o/H_o = 0.1$  and (b)  $h'_o/H_o = 0.8$ . The value at the origin is not defined.

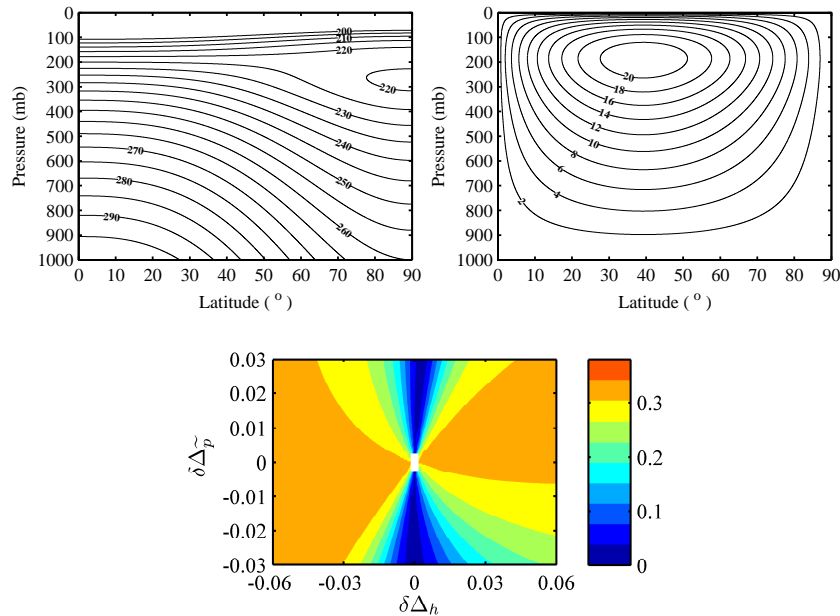


**Figure 3.** (a) The non-dimensional kinetic separation-squared when  $a_1/L_d = 4$  and  $h'_o/H_o = 0.8$ . (b) Fractional error of linearized kinetic separation-squared from (a). The white dot at the centre marks the origin.

$\theta$  under radiative equilibrium is assumed to be

$$\frac{\theta}{\theta_o} = 1 - \Delta_h \sin^{\frac{8}{3}} \phi \frac{2 \exp\{(p - p_{ch})/\Delta_p\} - 1}{2 \exp\{(p - p_{ch})/\Delta_p\} + 1} + \Delta_v \ln \frac{p_o}{p}, \quad (58)$$

where  $\phi$  is latitude,  $p_o$  is the constant surface pressure,  $\theta_o$  is the constant surface temperature at the equator,  $\Delta_h$  and  $\Delta_v$  control the fractional change of potential temperature from equator to pole and from the surface to the tropopause respectively,  $p_{ch}$  controls the pressure where the equator–pole temperature difference changes sign,  $\Delta_p$  is a factor control-



**Figure 4.** Balanced (a) temperature (K) and (b) zonal wind ( $\text{m s}^{-1}$ ) in the 2-D thermal wind model. (c) Ratio of the non-dimensional kinetic to enthalpy components of separation-squared when  $\Delta_h$  and  $\Delta_{\tilde{p}}$  are perturbed.

ling the vertical extent of the balanced jet and the other symbols follow the same notation as in previous sections. The form of Eq. (58) is inspired by the work of Held and Hou (1980).

It is convenient to introduce the following non-dimensional variables:

$$\tilde{T} = \frac{T}{\theta_o}, \quad \tilde{u} = \frac{u}{R\theta_o/\Omega_e a}, \quad \tilde{p} = \frac{p}{p_o}, \quad \tilde{E} = \frac{E}{c_p \theta_o p_o a/g}, \quad (59)$$

where  $\Omega_e$  is the angular speed of rotation of Earth,  $a$  is the radius of Earth, and the rest of the symbols follow the same notation as Sect. 3. So the non-dimensional thermal wind equation can be written as

$$2\sin\phi \frac{\partial \tilde{u}}{\partial \tilde{p}} = \frac{1}{\tilde{p}} \frac{\partial \tilde{T}}{\partial \phi}, \quad (60)$$

Given the equilibrium potential temperature, the non-dimensional temperature is

$$\tilde{T} = \tilde{p}^{R/c_p} \left( 1 - \Delta_h \sin^{\frac{8}{3}} \phi \frac{2 \exp\{(\tilde{p} - \tilde{p}_{\text{ch}})/\Delta_{\tilde{p}}\} - 1}{2 \exp\{(\tilde{p} - \tilde{p}_{\text{ch}})/\Delta_{\tilde{p}}\} + 1} - \Delta_v \ln \tilde{p} \right), \quad (61)$$

where  $\Delta_{\tilde{p}}$  is the non-dimensionalized  $\Delta_p$ . The solution for zonal wind is obtained by integrating the right-hand side of Eq. (60) from the surface to  $\tilde{p}$ . The non-dimensional separation metric is given by

$$M^2 = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_0^1 \left( \frac{R^2 \theta_o}{c_p \Omega_e^2 a^2} \frac{(\delta \tilde{u})^2}{2} + (\delta \sqrt{\tilde{T}})^2 \right) \cos \phi \, d\tilde{p} \, d\phi, \quad (62)$$

where we have used the fact that the depth of the atmosphere is much smaller than the radius of Earth. Notice that  $\tilde{T}$  and  $\tilde{u}$  are both well-defined finite functions of  $\tilde{p}$  over  $[0, 1]$  so the integral is finite.

The following parameters are specified:  $\theta_o = 300$  K,  $p_o = 1000$  mb,  $p_{\text{ch}} = 220$  mb,  $\Delta_v = 35/300$ . We consider a reference solution with  $\Delta_h = 40/300$  and  $\Delta_{\tilde{p}} = 50/1000$ . Figure 4a and b show the balanced temperature and zonal wind profiles of the reference solution. The westerly jet is formed around 40° N with a maximum velocity of 21.5  $\text{m s}^{-1}$  at about 200 mb. The thermal wind balance model is more valid in the mid-latitudes and hence the solution in the tropics is not a good approximation of the real atmosphere. Another deficiency is the vertical temperature profile in this model does not describe the temperature inversion above tropopause. Therefore we shall only make use of the data below 100 mb and between 35–65° N.

Separation metrics are calculated when  $\Delta_h$  and  $\Delta_{\tilde{p}}$  are perturbed. From the results in Fig. 4c, the kinetic component is always less than the enthalpy component though the ratio between them varies with perturbations. The order of magnitude of the ratio of kinetic to enthalpy components is set fundamentally by the ratio of temperature and specific angular momentum parameters on Earth,  $\theta_o/(\Omega_e a)^2$ . In the next section, the relative importance of kinetic and enthalpy components of separation-squared is further investigated with re-analysis data.

## 6 Example III: reanalysis data of the atmosphere

The separation metric of a dry compressible atmosphere in pressure coordinate is also applied to the reanalysis data. The data used in this study are the NCEP Climate Forecast System Reanalysis (CFSR) monthly mean of 6-hourly forecasts (CFSR, Accessed 15 May 2013, 2010). The reanalysis monthly mean data cover 31 years from January 1979 to December 2009 with  $0.5^\circ \times 0.5^\circ$  spatial resolution and 37 vertical levels from 1000 mb to 1 mb. Temperature, zonal and meridional wind at 37 pressure levels as well as geopotential height, pressure and temperature at the surface are used for the calculation. The separation metric Eq. (17) is transformed from Cartesian coordinates to spherical coordinates with the Jacobian  $r^2 \cos \phi$ . Since the depth of the atmosphere is much smaller than the radius of Earth, the Jacobian can be approximated by  $a^2 \cos \phi$ .

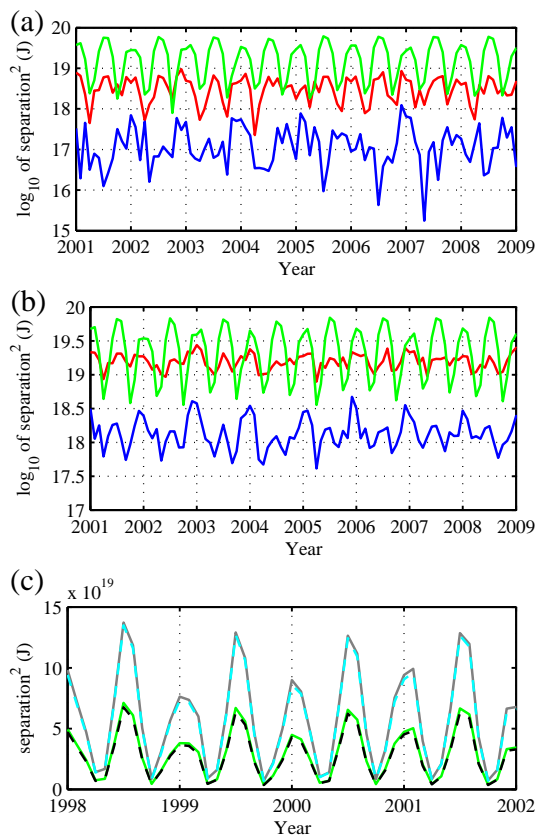
For the work here, at each grid point, we chose  $p_r$  to be the smallest surface pressure ever attained and linearly interpolate for  $T_r$  at  $p_r$  between the mean temperature at the lowest pressure-level above the surface and the mean surface temperature (assumed to be at the mean value of surface pressure) in the data set. A quick check with the data set shows that temperature within 100 mb from the surface never deviates by more than 2.5 % and so the approximation  $T \approx T_r$  in the boundary layer in Eq. (14) is valid. We also confirmed that linearization of the separation-squared in Eq. (17) is justified.

We investigate the separation between the monthly mean state of the atmosphere represented by CFSR data and its annual mean climatology. The annual mean climatology is defined as the mean over all months in 31 years of CFSR data and so is time independent. It provides the values for  $T$  and  $p_H$  in the denominators of the terms in Eq. (17).

### 6.1 Mid-latitude zonal mean and eddies

The separation-squared between zonal mean CFSR monthly mean data and its annual mean climatology in mid-latitudes ( $35\text{--}65^\circ\text{N}$ ) up to 100 mb between 2001 and 2009 is shown in Fig. 5a. The averaging interval for zonal mean is confined to isobars above the surface. Kinetic and enthalpy components show a synchronous semi-annual oscillation, which maximizes in January or February and in July or August. This is consistent with the seasonal cycle: the atmosphere moves furthest from the annual mean state during winter and summer. The surface pressure component is noisier and the semi-annual oscillation is not obvious. The reason is that surface pressure has strong zonal asymmetry due to the distribution of continents and oceans and the seasonal cycles of surface pressure are out of phase between continents and oceans.

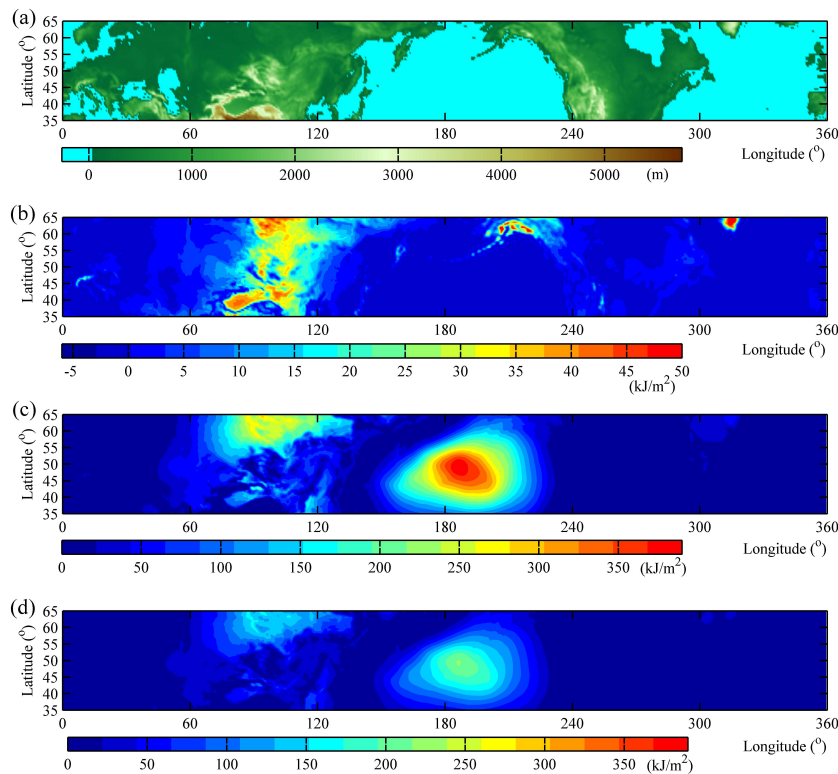
The kinetic component is smaller than the enthalpy component, which agrees with the results from the analytical thermal wind model in Sect. 5. The ratio of kinetic to enthalpy separation-squared is about 0.37 on average.



**Figure 5.** (a) Kinetic (red), enthalpy (green) and surface pressure (blue) separation-squared between CFSR zonal mean monthly mean data and its annual mean climatology in mid-latitudes ( $35\text{--}65^\circ\text{N}$ ) up to 100 mb. (b) As (a), but using full 3-D data including the eddies. (c) Enthalpy separation-squared calculated with different formulae: as the author proposed (green solid line), using constant temperature  $T_0 = 270\text{ K}$  in the integrand  $c_p(\delta T)^2/(4T_0)$  (black dashed line), further multiplying by a factor of 2 to get  $c_p(\delta T)^2/(2T_0)$  (cyan dashed line), and using constant  $p_0 = 1000\text{ mb}$  instead of  $p_r(x, y)$  for the integration upper limit but excluding isobars below the surface (grey solid line). Note that the vertical scale is logarithmic in (a, b) but is linear in (c).

Insight from Eq. (62) shows that it is because Earth is a rapidly rotating planet (i.e.  $\Omega_e$  is large) resulting in smaller geostrophically balanced flow for the same equator–pole temperature difference. The possibility to reveal this reason stems from the separation metric being induced from the energy norm which respects the fundamental dynamical ratios in the system. Such dynamical reasoning would not be possible if the metric was arbitrarily constructed with user-prescribed ratios.

It has been shown in the Lorenz energy cycle (Holton, 2004) that baroclinic eddies are the primary driving force for the energy exchange in mid-latitudes. However, the eddies are neglected in the calculation above with zonal mean data. To reveal the contribution to the separation metric from eddies, the same separation-squared is calculated from the



**Figure 6.** (a) Surface elevation  $z_H$  in the northern mid-latitudes ( $35\text{--}65^\circ\text{N}$ ). Surface pressure separation-squared in Dec 2005 contributed by (b) surface topography,  $z_H(\delta p_H)^2/(4p_H)$ , and (c) enthalpy,  $\frac{1}{g}c_p T_r(\delta p_H)^2/(4p_H)$ . (d) The enthalpy contribution using Talagrand's formula,  $\frac{1}{g}RT_0(\delta p_H)^2/(2p_0)$ , where  $T_0 = 270\text{ K}$  and  $p_0 = 1000\text{ mb}$ , for comparison with (c).

full 3-D data as shown in Fig. 5b. The kinetic separation-squared is found to be comparable to the enthalpy separation-squared when the mid-latitude eddies are included. We attribute the difference between Fig. 5a and b to the existence of mid-latitude eddies. So on average, eddies contribute 78 % to the total kinetic separation-squared and 28 % to the total enthalpy separation-squared.

These percentages are consistent with the contribution from eddies in the Lorenz energy cycle, where eddies contribute 71 % to the total kinetic energy and 31.8 % to the total APE (Oort and Peixóto, 1974).

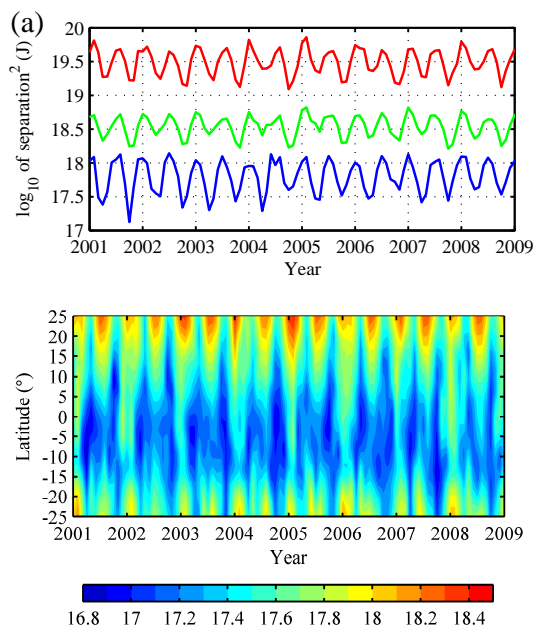
The enthalpy separation-squared is not conceptually related to Lorenz's definition of APE in Eq. (2). There are superficial resemblances because of the quadratic form, but in Eq. (17), no fixed reference state is assumed and the atmospheric lapse rate plays no role. Nonetheless, the partition between zonal mean and eddy contributions in enthalpy separation-squared and in APE are comparable because  $\Gamma$  is nearly constant in the troposphere while climatological temperature  $T_{\text{clim}} \sim \bar{T}$ , leading to enthalpy separation-squared from the annual mean climatology being roughly proportional to APE.

The coefficient of temperature difference  $(\delta T)^2$  is different in our linearized separation-squared in Eq. (17) from

that in the often used metric  $M_{T81}$  in Eq. (3) (Ehrendorfer and Errico, 1995; Errico, 2000; Qin and Mu, 2012). The enthalpy separation-squared is recalculated using progressively modified formulae in Fig. 5c (see the details in the caption). It is found that using constant reference temperature  $T_0 = 270\text{ K}$  and constant reference pressure  $p_0 = 1000\text{ mb}$  does not change the enthalpy separation-squared much.

But the increase of the coefficient by a factor of 2 from our metric to  $M_{T81}$  (green to grey line in 5c) makes the enthalpy contribution to the total separation-squared twice as significant!

Although the surface pressure separation-squared is 1 order of magnitude smaller than the enthalpy separation-squared (Fig. 5b), the coefficient of  $(\delta p_H)^2$  also differs between our metric and  $M_{T81}$ . Figure 6a and b compare the northern mid-latitude terrain and the topographic contribution to the surface pressure separation-squared. However, the boundary-layer enthalpy contribution to the surface pressure separation-squared is nearly an order of magnitude larger and maximizes over the central Pacific Ocean (Fig. 6c). If  $M_{T81}$  was used instead, the surface pressure separation-squared would be considerably smaller (see Fig. 6c and d).



**Figure 7.** (a) Kinetic (red), enthalpy (green) and surface pressure (blue) separation-squared between CFSR monthly mean data and its annual mean climatology in the tropical region ( $25^{\circ}\text{S}$ – $25^{\circ}\text{N}$ ) up to 100 mb. (b) Time–latitude cross-section of the logarithm to base 10 of the total separation-squared integrated zonally and vertically up to 100 mb in the tropical region.

## 6.2 Tropical oscillations

The separation-squared between CFSR monthly mean 3-D data and its annual mean climatology in the tropical region ( $25^{\circ}\text{S}$ – $25^{\circ}\text{N}$ ) up to 100 mb is calculated and the results between 2001 and 2009 are shown in Fig. 7a. Surface pressure separation-squared, like kinetic and enthalpy separations-squared, shows a synchronous semi-annual oscillation as oceans cover more area than land in the tropics.

The kinetic separation-squared is about 1 order of magnitude larger than the enthalpy separation-squared, quite unlike in the mid-latitudes (see Figs. 7a and 5b). This can be explained by the near constancy of temperature and surface pressure as opposed to large seasonality of wind in the tropics, e.g. due to the monsoons, whereas geostrophic and thermal wind balance necessitate surface pressure and temperature to have accompanying large variation to wind variation.

The semi-annual oscillation at different latitudes is further investigated in Fig. 7b. Contributions from the higher tropical latitudes are an order of magnitude larger than from the equatorial latitudes, which attests to the constant climate near the equator. The seasonality is stronger in the Northern Hemisphere compared to the same latitude in the Southern Hemisphere.

We next focus on the equatorial atmosphere to see what the separation metric can reveal about tropical dynamics. The separation-squared between CFSR monthly mean data and

its annual mean climatology in the equatorial tropics ( $5^{\circ}\text{S}$ – $5^{\circ}\text{N}$ ) integrated from the surface to the stratopause (1 mb) is shown in Fig. 8a. Further investigation of the separation-squared level by level shows that on average, the kinetic and enthalpy separations-squared in the stratosphere (70 to 1 mb) contribute 39.3 and 47.3 % respectively to the kinetic and enthalpy separations-squared in the whole column.

This is noteworthy because the stratosphere only makes up about 10 % of the atmospheric mass but it accounts for up to about 40 % of the combined monthly variance of the equatorial troposphere and stratosphere as measured by the energy norm-induced metric.

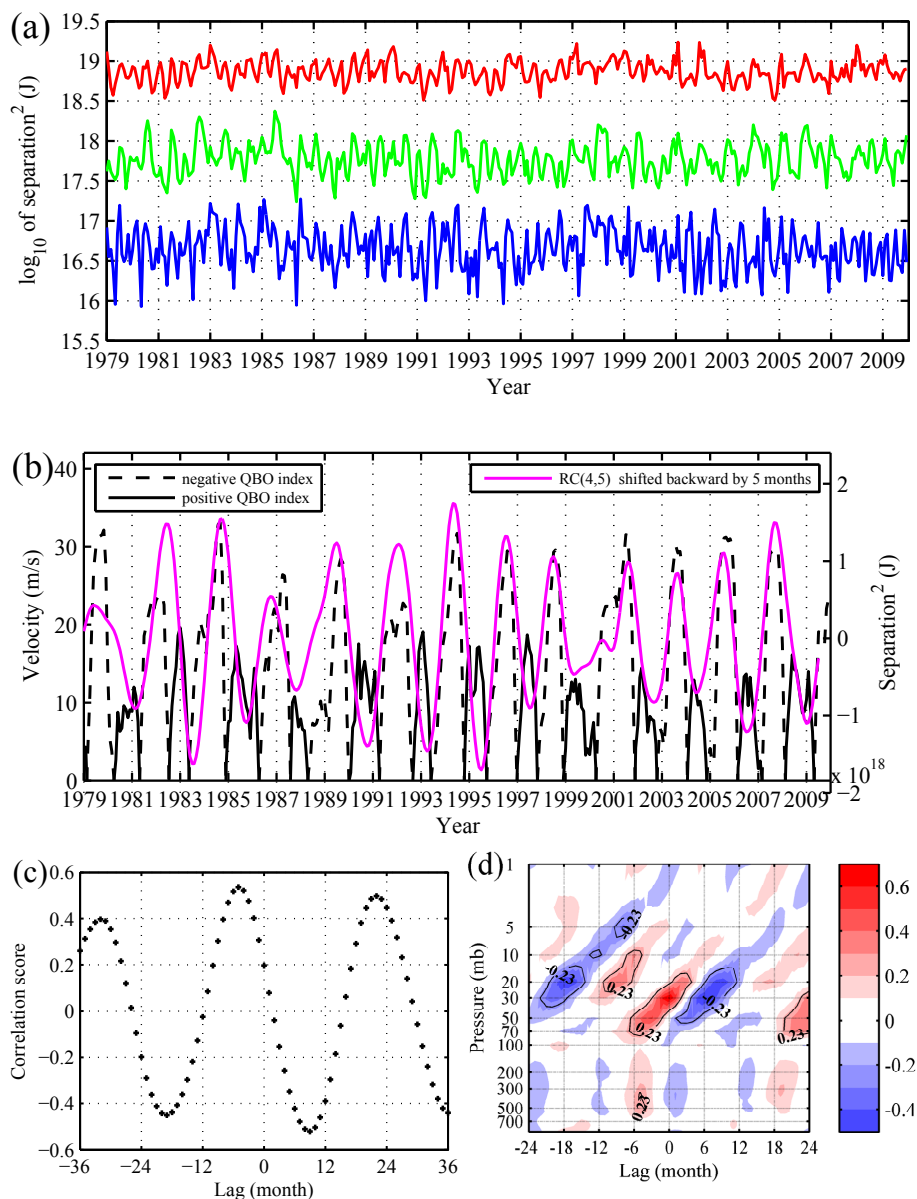
The quasi-biennial oscillation (QBO) is a quasi-periodic reversal of the mean zonal wind in the equatorial stratosphere, and is well-known to influence the global stratosphere through modulation of zonal wind, temperature, humidity and the meridional circulation (Baldwin et al., 2001). The meridional distribution of QBO amplitude is approximately Gaussian, centred at the equator with a  $12^{\circ}$  half-width (Wallace, 1973).

Although the QBO has a signature in temperature, it is weak because geostrophic balance is not dominant near the equator and thus the QBO signal is not identifiable in the enthalpy separation-squared against large signals arising from seasonal variation in insolation. We present only the analysis of the kinetic separation-squared here, which is a very good approximation to the total separation-squared because of its overwhelmingly large contribution.

Using singular spectrum analysis (SSA), we first decompose the kinetic separation-squared into a trend, seasonal oscillation and interannual oscillation, leaving a residue. Between 1979 and 2009, there are 372 sample points in time. The window length for SSA used is 36 sample points, i.e. 3 years. The first reconstructed component, RC(1), explains 90.9 % of the total variance. It traces the decadal variation which is in anti-phase to the 10.7 cm solar flux (Patat, 1998) and has a secular rising trend.

The semi-annual oscillation due to the seasonality of hemispheric insolation is captured by RC(2, 3), explaining 3.6 % of the total variance. RC(4, 5) explains 1.4 % of the total variance and when shifted backward by 5 months matches the QBO index very well (Fig. 8b), including years when the QBO period is longer than average, e.g. 1999 to 2002. (The QBO index is defined as the zonal wind at 30 mb over Singapore.) RC(4, 5) is lag correlated with the absolute value of the QBO index in Fig. 8c. The maximum correlation score of 0.53 is attained when the RC(4, 5) lags behind the QBO index by 5 months. The correlation score peaks every 28 months which is the average period of the QBO.

The lag correlation of the absolute value of the QBO index with the (full) kinetic separation-squared at different pressure levels is shown in Fig. 8d. The correlation score maximizes at 30 mb with zero lag simply because the QBO index is defined at this level. Since the QBO phase propagates downward, the kinetic separation-squared leads the QBO in-



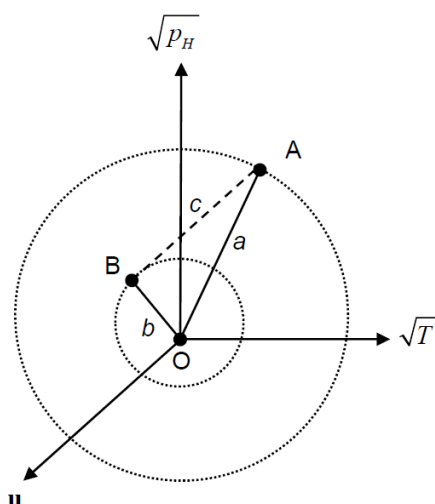
**Figure 8.** (a) Kinetic (red), enthalpy (green) and surface pressure (blue) components of separation-squared between CFSR monthly mean data and its annual mean climatology in the equatorial tropics ( $5^{\circ}\text{S}$ – $5^{\circ}\text{N}$ ) integrated over the whole atmosphere column. (b)  $\|\text{QBO index}\|$  (black) and RC(4, 5) of kinetic separation-squared (magenta) shifted backward by 5 months. Black solid (dashed) lines represent westerly (easterly) phase of the QBO. (c) Lag correlation of  $\|\text{QBO index}\|$  with the RC(4, 5) of kinetic separation-squared. (d) Lag correlation of  $\|\text{QBO index}\|$  with kinetic separation-squared at different pressure levels. The black contours denote the 99%-confidence level. In (c, d), a positive lag denotes the signal leading the QBO index.

dex at higher pressure levels than 30 mb and lags behind at pressures lower than 30 mb. Only the tilted positive correlation band centred at zero lag denotes a real physical connection. The other tilted correlation bands located at about multiples of 7 months away are just mirrored images produced by the quasi-periodicity of the QBO. The correlation score drops rapidly in the troposphere and becomes insignificant at the 99 % confidence level (except around 500 mb). To minimize the effect of auto-correlation in time, we assume

that the number of independent samples is 124, which is the number of seasons in our time series, giving the number of degrees of freedom as 122.

There is some indication that QBO has a significant but weak influence in the mid-troposphere around 500 mb. Such an influence may not be as unreasonable as it first sounds because of the disproportionately large stratospheric contribution to equatorial atmospheric variance mentioned earlier. A plausible dynamical reason could be that the zonal mean





**Figure 9.** Multi-variate phase space for a hydrostatic flow in pressure coordinate. For ease of visualization, wind  $\mathbf{u}$  is shown as a 1-D axis instead of a 2-D plane and each axis represents one among an infinite number of degrees of freedom on the surface (for  $\sqrt{p_H}$ ) or in the volume (for  $\sqrt{T}$  and  $\mathbf{u}$ ) of a fluid. The length of phase vectors A and B are  $a$  and  $b$  respectively while their separation is  $c$  in this subspace.  $\mathbf{O}$  is the null vector. The spheres represent hyper-surfaces of constant energy.

wind in the lowermost stratosphere ( $\sim 70$  mb) modifies the vertical propagation of equatorial waves, reflecting certain waves downwards so that their trapped energy maximizes in the mid-troposphere at the peaks of QBO easterly or westerly phases. Because we use a metric induced by the energy norm, when the energy of the atmospheric state is enhanced, the separation of that state from the annual mean climatology is correspondingly enhanced.

## 7 Discussion and summary

To date, much of the literature's rationale to the definition of a metric runs along two lines of thinking which are not mutually exclusive:

- to employ the quadratic form of a norm and its induced metric beyond the restricted dynamical regime for which the norm is proven to obey a conservation law (Ehrendorfer and Errico, 1995), with the confidence that the form is at least valid in that regime;
- to justify the quadratic form of a metric based on its simplicity (Palmer et al., 1998) and on dimensional consistency among the contributions by different state variables: the weighing coefficient on each variable depends on the suitable choice of a convenient reference state, certain physical constants and dimensionless numbers, as well as the practical importance of emphasizing that variable.

Neither line of thinking is without its merits and both arguments are substantial enough if practical application demands utility more than theoretical rigour. For instance, to find a singular vector or conditional nonlinear optimal perturbation (CNOP) in a model forecast for adaptive observation (Buizza et al., 1993; Mu et al., 2009), knowing that temperature has larger normalized error variance than wind (Koh and Ng, 2009) would favour a metric definition that emphasizes temperature deviations more, such as  $M_{T81}$  instead of Eq. (17). In that case, extending the use of a metric beyond the regime for which it was originally designed –  $M_{T81}$  was formulated by Talagrand (1981) for linearized, adiabatic, inviscid flows and is constant for the forecast of such flows in between consecutive data assimilations – is justifiable at least because the practical use of the metric in nonlinear, forced-dissipative regimes enables important advancement in NWP.

There are other practical considerations: Sect. 4a of Ehrendorfer and Errico (1995) mentions the relevance of numerical discretization schemes in determining whether a norm is practically invariant or not. Sections 4 and 6 of Palmer et al. (1998) distinguish the analysis error covariance metric for practical predictions, which depends on the observation network and the data assimilation scheme, against the geophysical fluid dynamics (GFD) covariance metric for GFD studies, which depends on dynamics only and is defined from the invariant measure associated with the system's attractor. The results of ensemble predictions, error growth analysis and predictability studies will be sensitive to the choice of the metric. Section 5 of Palmer et al. (1998) showed that the  $M_{T81}$  metric may be more suitable for practical prediction studies than in pure GFD problems on predictability.

Having recognized the merits of the above approaches, it is instructive to examine the comparative advantages of our theoretical approach. For that, we need to elucidate the underlying physical basis of the norm-induced metric.

Figure 9 is a Cartesian representation of the phase space  $(\mathbf{u}, \sqrt{T}, \sqrt{p_H})$  constructed for a hydrostatic flow in pressure coordinate. To our knowledge, this is the first time that quantities like  $\sqrt{T}$  and  $\sqrt{p_H}$  are constructed to serve as phase coordinates and we emphasize that only in these constructed coordinates does the square-root of true energy satisfy all the axioms of a norm.  $\sqrt{E_{\text{mod}}}$  of Eq. (15) is the Euclidean norm on the vector space of  $(\mathbf{u}, \sqrt{T}, \sqrt{p_H})$ ; it is not even a norm in the conventional vector space of  $(\mathbf{u}, T, p_H)$  because it does not have absolute homogeneity there.

With reference to Fig. 9,  $\sqrt{E_{\text{mod}}}$  is used to measure  $a$  and  $b$ , the lengths of phase vectors A and B respectively. The  $M$  of Eq. (16) is the metric used to measure the separation  $c$  between A and B. As the metric is induced by the norm, the separation of a phase vector from the null vector  $\mathbf{O}$  is identical to the phase vector's length. This means that  $a$ ,  $b$  and  $c$  are measured with the same "ruler". By using any other metric in a normed vector space, we are measuring  $c$  on a different ruler from  $a$  and  $b$ , which is admissible mathematically but goes against physical sense.

For adiabatic, inviscid flows, it is a law of physics that energy  $E$  is conserved and hence  $a$  and  $b$  are invariant. The vectors  $\mathbf{A}$  and  $\mathbf{B}$  move on constant energy hyper-surfaces which take the form of hyper-spheres when the phase coordinates  $(\mathbf{u}, \sqrt{T}, \sqrt{p_H})$  are scaled by factors of  $\sqrt{dp/2g}$ ,  $\sqrt{c_p dp/g}$ ,  $\sqrt{(c_p T_r + \Phi_H)/g}$  respectively and have common units of  $m^{-1}$ . Unless the flow is linear (like the case in Talagrand (1981)),  $c$  will generally not be invariant. Then the norm induced metric can indeed be used to detect changing separations between  $\mathbf{A}$  and  $\mathbf{B}$  while the invariant lengths of  $\mathbf{A}$  and  $\mathbf{B}$  provide physical justification for using the same “ruler” to measure separation. In this way, we avoid the inherent contradiction that  $M_{T81}$  faces: it is only useful when the norm is not conserved (see the end of Sect. 1.2). Moreover, linearization of the flow is never required for the conservation of energy. Without detracting from the last statement, where mathematically valid, the separation metric can be transformed into the tangent linear space of conventional but non-Cartesian coordinates  $(\mathbf{u}, T, p_H)$  where either  $\mathbf{A}$  or  $\mathbf{B}$  provides a realistic nonlinearly evolving reference state for the linearization of the metric. In these coordinates, the norm and metric take on a non-Euclidean form.

There is another advantage of measuring  $a$ ,  $b$  and  $c$  on the same “ruler”: the angle  $\psi$  between vectors  $\mathbf{A}$  and  $\mathbf{B}$  in Cartesian coordinates  $(\mathbf{u}, \sqrt{T}, \sqrt{p_H})$  can be consistently defined by the cosine rule:

$$\cos \psi = \frac{a^2 + b^2 - c^2}{2ab}, \quad (63)$$

which is equivalent to the definition of the inner product:

$$\langle \mathbf{A}, \mathbf{B} \rangle = (a^2 + b^2 - c^2)/2. \quad (64)$$

For the metric in Eq. (17), the inner product defined in this way is

$$\begin{aligned} \langle \mathbf{A}, \mathbf{B} \rangle &= \frac{1}{g} \int_A \int_0^{p_r} \left( \frac{1}{2} \mathbf{u}^A \cdot \mathbf{u}^B + c_p \sqrt{T^A T^B} \right) dp \, dA \\ &+ \frac{1}{g} \int_A (c_p T_r + \Phi_H) \sqrt{p_H^A p_H^B} \, dA, \end{aligned} \quad (65)$$

where the superscripts on the variables refer correspondingly to states  $A$  and  $B$ . Equation (65) may be contrasted against Eq. (11) of Palmer et al. (1998) which is the inner product in  $(\mathbf{u}, T', p'_s)$ -space related to the  $M_{T81}$  metric, reproduced in the notation of this paper as

$$\begin{aligned} \langle \mathbf{A}, \mathbf{B} \rangle_{p98} &= \frac{1}{2g} \int_A \int_0^{p_o} \left( \mathbf{u}^A \cdot \mathbf{u}^B + \frac{c_p}{T_o} T'^A T'^B \right) dp \, dA \\ &+ \frac{1}{2g} \int_A R T_o p_o \ln \left( \frac{p_s^A}{p_o} \right) \ln \left( \frac{p_s^B}{p_o} \right) \, dA. \end{aligned} \quad (66)$$

The set of angles a phase vector makes with the Cartesian axes fixes the direction of the phase vector. The notions of di-

rection and inner product are fundamental to many concepts and applications in predictability (e.g. Lyapunov exponents) and optimization of error growth (e.g. CNOP). Like our definition of separation, our definitions of direction and inner product ultimately rest upon the physical principle of energy conservation as the basis for the invariance of the norm and Euclidean geometry is manifest in  $(\mathbf{u}, \sqrt{T}, \sqrt{p_H})$ -space. In contrast, the set of metric  $M_{T81}$ , norm  $\sqrt{E_{T81}}$  and inner product  $\langle \mathbf{A}, \mathbf{B} \rangle_{p98}$  respects Euclidean geometry in  $(\mathbf{u}, T', p'_s)$ -space, but  $E_{T81}$  is not energy, causing the norm to vary in nonlinear, adiabatic, inviscid flows. In such flows, quantifying the separation and angle between two state vectors by  $M_{T81}$  and  $\langle \mathbf{A}, \mathbf{B} \rangle_{p98}$  would be like measuring distance and angle with elastic rulers and protractors. Appendix A delves further into the origin of the difference between  $E_{T81}$  and energy  $E$ .

Placed in the context of applications like ensemble forecast, the above theoretical development provides a physically based metric that can be used to, for instance, measure the spread of member states about the observation in multi-variate phase space where no single variable can summarize the model performance, such as at the surface (Scherrer et al., 2004) or in the tropics (Koh et al., 2012). The development of such multi-variate spread diagnostics would complement existing univariate spread measures (Buizza, 1997). Another use can be in error growth analysis to define the norm of a CNOP used to identify area targets for observation (Mu et al., 2009). A major practical advantage in the above applications would be that even large separations can be rigorously quantified using the Cartesian coordinates in which the energy norm is Euclidean, such as illustrated in Fig. 3a of Sect. 4. This would be essential, for example, in shallow-water simulations of tsunamis (Wang and Liu, 2007).

The importance of the non-Euclidean form of the metric for nonlinear flows illustrated in Sect. 4 (Fig. 2) is an important advancement in the definition of separation. For highly nonlinear, non-hydrostatic atmospheric flows at mesoscales, especially those manifesting strong convection such as around the core region of a tropical cyclone (TC), the separation between atmospheric states involves a kinetic–buoyant energy inter-conversion term,  $\delta w \delta \rho$  in Eq. (31). The practical implementation of such a metric for computing CNOP may result in a discernible impact on the areas targeted for observation to improve TC intensity forecasts.

By agreeing on the invariant norm relevant to the dynamics of a system, for instance the energy norm in pressure coordinate, the relative ratios of contribution among the conventional state variables,  $\delta \mathbf{u}$ ,  $\delta T$  and  $\delta p_H$  in this case, to the total separation-squared are no longer arbitrary to some dimensionless constants or dependent on the user’s choice of the reference state, as the case would be for a metric constructed from dimensional consistency arguments alone.

For example, the small ratio of enthalpy to kinetic components of separation-squared in the tropics (Fig. 7a) compared with the mid-latitudes (Fig. 5a, b) cannot be increased in an

ad hoc manner, e.g. by replacing the nonlinear reference-state  $T$  by the climatological amplitude of diurnal temperature fluctuations in Eq (17). The reason is that this ratio is reflective of the lack of geostrophic balance in the monthly mean tropical climate (i.e. contrary to the case in Eq. (62) and Fig. 4). Likewise, in the mid-latitudes, the use of  $M_{T81}$ , where the enthalpy component is roughly doubled (Fig. 5c), would not be recommended if the dimensionless ratio of system constants governing the dynamics of thermal wind balance in Eq. (62) is to be respected. Nonetheless, we do not believe having doubled the enthalpy contribution in  $M_{T81}$  would detract from the qualitative conclusions of much previous work even if details might have been altered, e.g. the consistency of the “energy” norm to the “analysis error covariance metric” in Palmer et al. (1998).

We have given a firm theoretical basis for the contribution of surface topography on the metric. Previous theoretical literature (e.g. Talagrand, 1981) did not allow for the presence of topography. It would not be possible to guess the form of topographic influence by dimensional analysis alone. For example, by dimensional analysis, topography could well modify the enthalpy contribution as  $(c_p T_o + \Phi_H) (\delta T / T_o)^2$  in Eq. (3) instead of modifying the surface pressure contribution in Eq. (17), especially looking at the form of Eq. (12). With our approach, the topographic term is negligible in pressure coordinate (Fig. 6b, c) because  $\Phi_H / (c_p T) \lesssim 1\%$ , and this is also true in isentropic coordinate, see Eq. (24). But one would not be able to consistently neglect the topographic effect if one used the expression  $(c_p T_o + \Phi_H) (\delta T / T_o)^2$  because surface pressure separation-squared can be about 1% of enthalpy separation-squared (e.g. in Fig. 5a) and is retained within the metric expression. In height coordinate, topography does not even appear except as the fixed lower limit of vertical integration for the semi-infinite atmosphere.

By using the energy norm-induced metric, we detected a weak but statistically significant teleconnection between the QBO phase in the lower stratosphere and the monthly variability at mid-tropospheric levels (Fig. 8d) which may be worth further investigation in future. At this early juncture, it is understandable if a teleconnection is selectively picked out by our metric for atmospheric variation because of the principle of energy conservation on which the metric is based: a longer state vector due to accumulation of tropospheric energy by equatorial wave reflection from the lower stratosphere would manifest larger variation from the climatic mean state since we use the same “ruler” to measure energy and separation in phase space (see Fig. 9). If the enthalpy contribution  $c_p T (\delta T / T)^2$  was artificially exaggerated roughly a hundredfold by normalizing the temperature difference by its variance  $\Delta T$  instead of by  $T$ , i.e.  $c_p T (\delta T / \Delta T)^2$ , the metric would now be dominated by temperature variability in which the QBO signal is swamped by seasonal signals. The above teleconnection between the mid-troposphere and the QBO index would be lost when using such an ad hoc metric.

While useful, the energy norm is not the only invariant norm from which a metric can be induced. Other dynamical invariants, e.g. enstrophy (Vallis, 2006) and wave activity (Haynes, 1988), could be used. One only needs to construct the phase space judiciously following the approaches demonstrated in Sects. 2 and 3 so that the invariant quantity takes the form of a Euclidean norm. Hence, the above theoretical advantages may potentially be relevant to most problems with a conserved physical quantity. For instance, in homogeneous, isotropic turbulence of a 2D incompressible fluid, the metric induced by the enstrophy norm may be useful in investigating chaotic dynamics of turbulence: e.g. defining  $\zeta$  as absolute vorticity, one might consider the spectral power of  $\int_A (\delta \zeta)^2 dA$  in the enstrophy cascading inertial sub-range as a measure of the separation between two mature turbulent flows.

In summary, we propose a new two-step approach in tackling the problem of metric definition: (1) constructing the phase space specifically so that an invariant based on a physical conservation law is the Euclidean norm on this space; and (2) defining the norm-induced metric to quantify the separation in phase space between two states. This methodology is mathematically rigorous and physically meaningful. The norm can be invariant even for nonlinear flows and the norm-induced metric is valid even for large separations. We have applied this approach to examine analytical examples and realistic reanalysis data, and discussed its potential applications in ensemble prediction, error growth analysis and predictability studies. But we note that practical and other theoretical considerations may favour alternative approaches to defining the metric. Finally, the separation metric in this study is developed for dry atmospheres. We are working next on the separation metric including moisture.

## Appendix A: Difference between $E_{T81}$ and energy $E$

### A1 Derivation of $E_{T81}$

We first re-derive  $E_{T81}$ , following Talagrand (1981) but using the notation of this paper. The inviscid, adiabatic flow is first linearized about the reference state,  $\mathbf{u} = 0$ ,  $T = T_o$  and  $p_s = p_o$ , where  $T_o$  and  $p_o$  are constants. The reference-state geopotential  $\Phi_o$  is given by hydrostatic balance:

$$\begin{aligned} \frac{d\Phi_o}{dp} &= -\frac{RT_o}{p}, \\ \Rightarrow \Phi_o &= -RT_o \ln\left(\frac{p}{p_o}\right), \end{aligned} \quad (\text{A1})$$

where we have chosen  $\Phi_o(p_o) = 0$ .

In pressure coordinate, the equations of motion for the perturbation state are

$$\frac{\partial \mathbf{u}}{\partial t} + f\hat{k} \times \mathbf{u} + \nabla_p \Phi' = 0, \quad (\text{A2})$$

$$c_p \frac{\partial T'}{\partial t} - \omega \frac{RT_o}{p} = 0, \quad (\text{A3})$$

$$\nabla_p \cdot \mathbf{u} + \frac{\partial \omega}{\partial p} = 0, \quad (\text{A4})$$

$$\frac{\partial \Phi'}{\partial p} + \frac{RT'}{p} = 0, \quad (\text{A5})$$

where  $f$  is the Coriolis parameter and  $\hat{k}$  is a unit vector pointing upwards.

Equation (A2) is dot-multiplied by  $\mathbf{u}$  to obtain the kinetic energy tendency, while Eq. (A3) is multiplied by  $T'/T_o$  to get a *varying fraction* of enthalpy tendency and of adiabatic heating:

$$\frac{\partial \mathbf{u}^2}{\partial t} + \mathbf{u} \cdot \nabla_p \Phi' = 0, \quad (\text{A6})$$

$$\frac{\partial c_p T'^2}{\partial t} - \omega \frac{RT'}{p} = 0. \quad (\text{A7})$$

Equation (A5) and (A7) can be combined as

$$\frac{\partial c_p T'^2}{\partial t} + \omega \frac{\partial \Phi'}{\partial p} = 0. \quad (\text{A8})$$

Equation (A8) is added to Eq. (A6) and with the help of Eq. (A4), we get

$$\frac{\partial}{\partial t} \left( \frac{\mathbf{u}^2}{2} + \frac{c_p T'^2}{2T_o} \right) + \nabla_p \cdot (\mathbf{u} \Phi') + \frac{\partial}{\partial p} (\omega \Phi') = 0. \quad (\text{A9})$$

Using the fact that  $\omega \rightarrow 0$  exponentially fast with height, Eq. (A9) is integrated vertically to the surface to obtain

$$\begin{aligned} \int_0^{p_s} \frac{\partial}{\partial t} \left( \frac{\mathbf{u}^2}{2} + \frac{c_p T'^2}{2T_o} \right) dp + \nabla_p \cdot \int_0^{p_s} \mathbf{u} \Phi' dp \\ + \frac{\partial p_s}{\partial t} \Phi'(p_s) = 0. \end{aligned} \quad (\text{A10})$$

Unlike Talagrand (1981), we allow a weak surface topography to exist without interfering with the linearization of the flow, so that  $\Phi(p_s) = \Phi'_s(x, y)$ . Thus, we may write

$$\Phi'(p_s) = \Phi(p_s) - \Phi_o(p_s) \Phi'_s + RT_o \ln\left(\frac{p_s}{p_o}\right), \quad (\text{A11})$$

where we have made use of Eq. (A1). The above is substituted into Eq. (A10) to yield

$$\begin{aligned} \int_0^{p_s} \frac{\partial}{\partial t} \left( \frac{\mathbf{u}^2}{2} + \frac{c_p T'^2}{2T_o} \right) dp + \left( RT_o \ln\left(\frac{p_s}{p_o}\right) + \Phi'_s \right) \frac{\partial p_s}{\partial t} \\ + \nabla_p \cdot \int_0^{p_s} \mathbf{u} \Phi' dp = 0. \end{aligned} \quad (\text{A12})$$

Integrating horizontally under periodic lateral boundary conditions causes the last term to vanish. Ignoring third-order terms, as is consistent with the effect of flow linearization on Eqs. (A6) and (A7), we then have

$$\begin{aligned} \frac{\partial E_{T81}^{\text{topo}}}{\partial t} &\equiv \frac{\partial}{\partial t} \left( \frac{1}{g} \int_A \int_0^{p_o} \left( \frac{\mathbf{u}^2}{2} + \frac{c_p}{2T_o} T'^2 \right) dp dA \right. \\ &\left. + \frac{1}{g} \int_A \left( \frac{RT_o}{2p_o} p_s'^2 + \Phi'_s p_s' \right) dA \right) = 0, \end{aligned} \quad (\text{A13})$$

In Eq. (A13), the influence of surface topography is exerted through a linear term in  $p_s'$ , which was fortuitously excluded from  $E_{T81}$  in Eq. (1) by Talagrand (1981); the other terms all have homogeneous quadratic forms. Thus, only when the surface is flat,  $E_{T81}$  is derived as the square of a norm on  $(\mathbf{u}, T', p_s')$ -space and is invariant for adiabatic, inviscid flows linearized about a reference isothermal atmosphere at rest. As only  $T'/T_o$  of the adiabatic heating and enthalpy tendency is included in the formulation of Eq. (A7),  $E_{T81}$  is not total energy. But its tendency is a time-varying fraction of the energy tendency, which explains why  $E_{T81}$  is not conserved in general adiabatic, inviscid flows.

### A2 Derivation of perturbation energy $E'$ in pressure coordinate

We now derive the conservation of total energy  $E$  in pressure coordinate, paying attention to the differences from the

derivation in the previous section. We only assume that the basic atmospheric state is hydrostatic and isothermal at temperature  $T_0$  so that Eq. (A1) applies, where the constant  $p_0$  is now defined as the reference pressure at the lowest point on Earth's surface. The full equations of motion for the perturbation state, which may not be close to the basic state, are

$$\left(\frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla_p + \omega \frac{\partial}{\partial p}\right) \mathbf{u} + f \hat{k} \times \mathbf{u} + \nabla_p \Phi' = 0, \quad (\text{A14})$$

$$c_p \left(\frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla_p + \omega \frac{\partial}{\partial p}\right) T' - \omega \frac{R}{p} (T' + T_0) = 0, \quad (\text{A15})$$

and Eqs. (A4) and (A5) as before.

Equation (A14) is dot-multiplied by  $\mathbf{u}$  to obtain the kinetic energy tendency while Eq. (A15) is combined with Eq. (A5) retaining the *full* enthalpy tendency:

$$\left(\frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla_p + \omega \frac{\partial}{\partial p}\right) \frac{\mathbf{u}^2}{2} + \mathbf{u} \cdot \nabla_p \Phi' = 0, \quad (\text{A16})$$

$$c_p \left(\frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla_p + \omega \frac{\partial}{\partial p}\right) T' + \omega \frac{d\Phi'}{dp} - \omega \frac{RT_0}{p} = 0. \quad (\text{A17})$$

Equations (A16) and (A17) are added with the help of Eq. (A4) so that

$$\begin{aligned} & \frac{\partial}{\partial t} \left( \frac{\mathbf{u}^2}{2} + c_p T' \right) + \nabla_p \cdot \left( \mathbf{u} \left( \frac{\mathbf{u}^2}{2} + c_p T' + \Phi' \right) \right) \\ & + \frac{\partial}{\partial p} \left( \omega \left( \frac{\mathbf{u}^2}{2} + c_p T' + \Phi' \right) \right) - \omega \frac{RT_0}{p} = 0. \end{aligned} \quad (\text{A18})$$

As  $\omega \rightarrow 0$  exponentially fast with height, Eq. (A18) is integrated vertically to the surface to obtain

$$\begin{aligned} & \frac{\partial}{\partial t} \int_0^{p_H} \left( \frac{\mathbf{u}^2}{2} + c_p T' \right) dp + \nabla_p \cdot \int_0^{p_H} \mathbf{u} \left( \frac{\mathbf{u}^2}{2} + c_p T' + \Phi' \right) dp \\ & + \frac{\partial p_H}{\partial t} \Phi'(p_H) - RT_0 \int_0^{p_H} \frac{\omega}{p} dp = 0. \end{aligned} \quad (\text{A19})$$

The penultimate term in Eq. (A19) is evaluated following Eq. (A11) to get

$$\frac{\partial p_H}{\partial t} \Phi'(p_H) = \frac{\partial p_H}{\partial t} \Phi_s + RT_0 \frac{\partial p_H}{\partial t} \ln \left( \frac{p_H}{p_0} \right), \quad (\text{A20})$$

where  $\Phi_s$  is the non-trivial surface geopotential due to topography. The last integral in Eq. (A19) can be integrated by parts, with the help of Eqs. (A1) and (A4), as follows:

$$\begin{aligned} & -RT_0 \int_0^{p_H} \frac{\omega}{p} dp = RT_0 \int_0^{p_H} \frac{\partial \omega}{\partial p} \ln \left( \frac{p}{p_0} \right) dp - RT_0 \left[ \omega \ln \left( \frac{p}{p_0} \right) \right]_0^{p_H} \\ & = \nabla_p \cdot \int_0^{p_H} \mathbf{u} \Phi_0 dp - RT_0 \frac{\partial p_H}{\partial t} \ln \left( \frac{p_H}{p_0} \right). \end{aligned} \quad (\text{A21})$$

The last term of Eq. (A20) is cancelled by the last term of Eq. (A21) and its contribution is replaced by the first term

of Eq. (A21) which is the flux divergence of reference-state geopotential. In contrast, the last term of Eq. (A11) survives and becomes the quadratic  $p'_s$  term in Eq. (A13). This is because the bulk of adiabatic heating,  $\omega RT_0/p$ , that is responsible for the terms in Eq. (A21) is dropped out when only a fraction  $T'/T_0$  of adiabatic heating is retained in Eq. (A7).

Substituting Eqs. (A20) and (A21) into Eq. (A19), we finally get

$$\begin{aligned} & \frac{\partial}{\partial t} \int_0^{p_H} \left( \frac{\mathbf{u}^2}{2} + c_p T' \right) dp + \Phi_s \frac{\partial p'_H}{\partial t} \\ & + \nabla_p \cdot \int_0^{p_H} \mathbf{u} \left( \frac{\mathbf{u}^2}{2} + c_p T' + \Phi \right) dp = 0, \end{aligned} \quad (\text{A22})$$

where  $p'_H$  is the perturbation of surface pressure from  $p_0$  and is not necessarily small. Integrating horizontally under periodic lateral boundary conditions, the flux divergence term vanishes leaving

$$\begin{aligned} & \frac{\partial E'}{\partial t} \equiv \frac{\partial}{\partial t} \left( \frac{1}{g} \int_A \int_0^{p_H} \left( \frac{\mathbf{u}^2}{2} + c_p T' \right) dp dA \right. \\ & \left. + \frac{1}{g} \int_A \Phi_s p'_H dA \right) = 0. \end{aligned} \quad (\text{A23})$$

The above is the basis for the conservation of perturbation energy  $E'$  for fully nonlinear, adiabatic, inviscid flows.  $E'$  is different from total energy  $E$  in Eq. (12) by a constant as mass is conserved, i.e.  $\int_A p_H/g dA = \text{constant}$ .

### A3 Comparison between $E_{T81}$ and linearized perturbation energy $E'_{\text{lin}}$

To compare with Eq. (A13) on equal footing, we linearize Eq. (A23) about the same reference state as in Sect. A1 where surface topography is weak, i.e.  $\Phi_s = \Phi'_s$ . First, consider the following integral over the lower boundary layer [ $p_H, p_0$ ]:

$$\int_{p_H}^{p_0} \left( \frac{\mathbf{u}^2}{2} + c_p T' \right) dp = c_p T'_H (p_0 - p_H) = -c_p T'_H p'_H, \quad (\text{A24})$$

where  $T'_H$  and  $p'_H$  are the perturbations on surface temperature and surface pressure respectively and we have ignored third-order terms. Thus, Eq. (A23) becomes

$$\begin{aligned} & \frac{\partial E'_{\text{lin}}}{\partial t} \equiv \frac{\partial}{\partial t} \left( \frac{1}{g} \int_A \int_0^{p_0} \left( \frac{\mathbf{u}^2}{2} + c_p T' \right) dp dA \right. \\ & \left. + \frac{1}{g} \int_A (c_p T'_H + \Phi'_s) p'_H dA \right) = 0. \end{aligned} \quad (\text{A25})$$

As the full enthalpy tendency and adiabatic heating are kept in Eq. (A15), the enthalpy contribution in Eq. (A25) is

linear in  $T'$  and there are no quadratic terms in  $p'_H$ . Instead, the topographic influence clearly shows a linear dependence on  $p'_H$ , which is already seen in Eq. (A13) and would have been similarly present in  $E_{T81}$  if surface topography had not been ignored in Talagrand (1981)'s approach. There is additionally a quadratic interaction term  $T'_H p'_H$  in Eq. (A25) whose counterpart in Eq. (A13) is a third-order term that has been ignored.

The above comparison shows clearly that  $E_{T81}^{\text{topo}}$  or  $E_{T81}$  is different from  $E'_{\text{lin}}$ . Both linearized formulations are inherently problematic because the atmosphere is significantly different from the isothermal reference state. The relative ratio of the terms in Eqs. (A13) or (A25) also depends on the arbitrary choice of  $T_o$  and  $p_o$ . Adopting the full (non-linearly conserved) energy in Eq. (12) requires us to innovate on the construction of the phase space  $(\sqrt{T}, \sqrt{p_H})$  in Sect. 3.1, in order to solve the problem of having a Euclidean energy norm  $\sqrt{E}$  when energy is a linear function of temperature and surface pressure. Subsequent transformation of our norm-induced metric into the tangent linear space  $(T, p_H)$  yields the quadratic dependence on  $\delta T$  and  $\delta p_H$  but with different weighing coefficients than in the metric  $M_{T81}$  induced by the norm  $\sqrt{E_{T81}}$  on the phase space  $(T', p'_H)$ . This difference arises because we use a different norm on a different vector space from Talagrand (1981).

*Acknowledgements.* The authors would like to thank the helpful comments of the anonymous reviewers and examiners of F. Wan's PhD thesis. This work comprises Earth Observatory of Singapore contribution no. 81. It is supported in part by the National Research Foundation Singapore and the Singapore Ministry of Education under the Research Centres of Excellence initiative.

Edited by: P. Haynes

## References

- Baldwin, M. P., Gray, L. J., Dunkerton, T. J., Hamilton, K., Haynes, P. H., Randel, W. J., Holton, J. R., Alexander, M. J., Hirota, I., Horinouchi, T., Jones, D. B. A., Kinnersley, J. S., Marquardt, C., Sato, K., and Takahashi, M.: The quasi-biennial oscillation, *Rev. Geophys.*, 39, 179–229, 2001.
- Bannon, P. R.: Hydrostatic adjustment: Lamb's problem, *J. Atmos. Sci.*, 52, 1743–1752, 1995.
- Buizza, R.: Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system, *J. Atmos. Sci.*, 45, 99–119, 1997.
- Buizza, R., Tribbia, J., Molteni, F., and Palmer, T.: Computation of optimal unstable structures for a numerical weather prediction model, *Tellus A*, 45, 388–407, 1993.
- CFSR: Environmental Modeling Center/National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce. NCEP Climate Forecast System Reanalysis (CFSR) Monthly Products, January 1979 to December 2010. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, <http://rda.ucar.edu/datasets/ds093.2> (last access: 15 May 2013) 2010.
- Davidson, K. R. and Donsig, A. P.: Real analysis and applications theory in practice, Undergraduate Texts in Mathematics, Springer, 2010.
- Ding, R. and Li, J.: Nonlinear finite-time Lyapunov exponent and predictability, *Phys. Lett. A*, 364, 396–400, 2007.
- Ehrendorfer, M. and Errico, R. M.: Mesoscale predictability and the spectrum of optimal perturbations, *J. Atmos. Sci.*, 52, 3475–3500, 1995.
- Errico, R. M.: Interpretations of the total energy and rotational energy norms applied to determination of singular vectors, *Q. J. Roy. Meteorol. Soc.*, 126, 1581–1599, 2000.
- Fleisch, D.: A student's guide to vectors and tensors, Cambridge University Press, 1 edn., 2011.
- Green, J. S. A., Ludlam, F. H., and McIlveen, J. F. R.: Isentropic relative-flow analysis and parcel theory, *Q. J. Roy. Meteorol. Soc.*, 92, 210–219, 1966.
- Haynes, P. H.: Forced dissipative generalizations of finite-amplitude wave-activity conservation relations for zonal and nonzonal basic flows, *J. Atmos. Sci.*, 45, 2352–2362, 1988.
- Held, I. M. and Hou, A. Y.: Nonlinear axially symmetric circulations in a nearly inviscid atmosphere, *J. Atmos. Sci.*, 37, 515–533, 1980.
- Hohenegger, C. and Schar, C.: Atmospheric predictability at synoptic versus cloud-resolving scales, *B. Am. Meteor. Soc.*, 88, 1783–1794, 2007.
- Holton, J. R.: An introduction to dynamic meteorology, International Geophysics Series, Academic press, 4 edn., 2004.
- Hoskins, B. J.: Towards a PV-theta view of the general circulation, *Tellus*, 43AB, 27–35, 1991.
- Kasahara, A.: Various vertical coordinate systems used for numerical weather prediction, *Mon. Weather Rev.*, 102, 509–522, 1974.
- Kim, H., Kim, S., and Jung, B.: Real-time adaptive observation guidance using singular vectors for typhoon Jangmi (200815) in T-PARC 2008, *Weather Forecast.*, 26, 634–649, 2011.
- Koh, T.-Y. and Ng, J. S.: Improved diagnostics for NWP verification in the tropics, *J. Geophys. Res.*, 114, D12102, doi:10.1029/2008JD011179, 2009.
- Koh, T.-Y., Wang, S., and Bhatt, B. C.: A diagnostic suite to assess NWP performance, *J. Geophys. Res.*, 117, D13109, doi:10.1029/2011JD017103, 2012.
- Leutbecher, M. and Palmer, T. N.: Ensemble forecasting, *J. Comput. Phys.*, 227, 3515–3539, 2008.
- Lorenz, E. N.: Available potential energy and the maintenance of the general circulation, *Tellus*, 7, 157–167, 1955.
- Lorenz, E. N.: Energy and numerical weather prediction, *Tellus*, 12, 364–373, 1960.
- Lorenz, E. N.: Deterministic nonperiodic flow, *J. Atmos. Sci.*, 20, 130–141, 1963.
- Lorenz, E. N.: A study of the predictability of a 28-variable atmospheric model, *Tellus*, 17, 321–333, 1965.
- Lorenz, E. N.: Predictability of a flow which possesses many scales of motion, *Tellus*, 21, 289–307, 1969.
- Mak, M.: Atmospheric dynamics, Cambridge University Press, 1 edn., 2011.
- Molteni, F. and Palmer, T. N.: Predictability and finite-time instability of the northern winter circulation, *Q. J. Roy. Meteorol. Soc.*, 119, 269–298, 1993.
- Mu, M., Duan, W., Wang, B., et al.: Conditional nonlinear optimal perturbation and its applications, *Nonlin. Processes in Geophys.*, 10, 493–501, 2003.
- Mu, M., Zhou, F., and Wang, H.: A method for identifying the sensitive areas in targeted observations for tropical cyclone prediction: conditional nonlinear optimal perturbation, *Mon. Weather Rev.*, 137, 1623–1639, 2009.
- Oort, A. H. and Peixóto, J.: The annual cycle of the energetics of the atmosphere on a planetary scale, *J. Geophys. Res.*, 79, 2705–2719, 1974.
- Palmer, T. N., Gelaro, R., Barkmeijer, J., and Buizza, R.: Singular vectors, metrics, and adaptive observations, *J. Atmos. Sci.*, 55, 633–653, 1998.
- Patat, F.: The dancing sky: 6 years of night-sky observations at Cerro Paranal, *Astron. Astrophys.*, 481, 575–591, 1998.
- Qin, X. and Mu, M.: Influence of conditional nonlinear optimal perturbations sensitivity on typhoon track forecasts, *Q. J. Roy. Meteorol. Soc.*, 138, 185–197, 2012.
- Rivière, O., Lapeyre, G., and Talagrand, O.: A novel technique for nonlinear sensitivity analysis: Application to moist predictability, *Q. J. Roy. Meteorol. Soc.*, 135, 1520–1537, 2009.
- Scherrer, S. C., Appenzeller, C., Eckert, P., and Cattani, D.: Analysis of the spread-skill relations using the ECMWF ensemble prediction system over Europe, *Weather Forecast.*, 19, 552–565, 2004.
- Staniforth, A. and Wood, N.: The deep-atmosphere Euler equations in a generalized vertical coordinate, *Mon. Weather Rev.*, 131, 1931–1938, 2003.

- Staniforth, A., Wood, N., and Girard, C.: Energy and energy-like invariants for deep non-hydrostatic atmospheres, *Q. J. Roy. Meteorol. Soc.*, 129, 3495–3499, 2003.
- Talagrand, O.: A study of the dynamics of four-dimensional data assimilation, *Tellus*, 33, 43–60, 1981.
- Trenberth, K. E.: Using atmospheric budgets as a constraint on surface fluxes, *J. Climate*, 10, 2796–2809, 1997.
- Vallis, G. K.: Atmospheric and oceanic fluid dynamics: fundamentals and large-scale circulation, Cambridge University Press, 2006.
- Wallace, J. M.: General circulation of the tropical lower stratosphere, *Rev. Geophys.*, 11, 191–222, 1973.
- Wang, X. and Liu, P. L.-F.: Numerical simulations of the 2004 Indian Ocean tsunamis – coastal effects, *J. Earthq. Tsunami*, 1, 273–297, 2007.
- Yoden, S.: Dynamical aspects of stratospheric vacillations in a highly truncated model, *J. Atmos. Sci.*, 44, 3683–3695, 1987.
- Zhang, F., Snyder, C., and Rotunno, R.: Effects of moist convection on mesoscale predictability, *J. Atmos. Sci.*, 60, 1173–1185, 2003.
- Zhang, F., Bei, N., Rotunno, R., Snyder, C., and Epifanio, C. C.: Mesoscale predictability of moist baroclinic waves: Convection-permitting experiments and multistage error growth dynamics, *J. Atmos. Sci.*, 64, 3579–3594, 2007.
- Zou, X., Vandenberghe, F., Pondecà, M., and Kuo, Y. H.: Introduction to adjoint techniques and the MM5 adjoint modeling system, NCAR Technical Note, 1997.