Tellus (1989), *41A*, 97–108

On exact and approximate energy equations in pressure coordinates

By G. J. BOER, *Canadian Climate Centre, Downsview, Ontario, Canada*

(Manuscript received 7 May 1987; in final form 6 May 1988)

ABSTRACT

Studies of the atmospheric energy cycle have been common in meteorology since the modern formulation of the concept by Lorenz. Most of these studies have been based on the approximate version of the energy budget equations using pressure as the vertical coordinate given by Lorenz in the same paper. Inexactitude in the formulation of the energy budget equations in general and the isobaric coordinate version in particular arises from several sources: the neglect of topography in the definition of the reference state for available potential energy, the simplification of the lower boundary condition, the definition of mean and eddy components and in particular in the use of the Lorenz approximate equations. An exact version of the energy cycle equations in isobaric coordinates is derived. The equations are exact in the sense that they are derived from the hydrostatic primitive equations without further approximation. It is pointed out that results already obtained indicate that there is a non-trivial difference between the exact and approximate versions of the energy budget equations.

1. Introduction

Data sets used for general circulation diagnostic calculations almost invariably have information given on standard pressure levels. In addition, the equations governing the behavior of the atmosphere generally have a simpler form when they are written using pressure as the vertical coordinate. Since pressure surfaces are almost horizontal, isobaric (i.e., pressure) coordinates are similar to height coordinates and their use does not introduce geometrical difficulties in interpretation as is sometimes the case when coordinate surfaces are not horizontal.

The difficulty with the use of pressure as the vertical coordinate resides in the treatment of the lower boundary which is not a coordinate surface. The surface pressure is a function of the two horizontal coordinates and of time. The same difficulty with the lower boundary applies to isentropic coordinates (for which potential temperature is the vertical coordinate), while for geometric height coordinates the surface height is a function of the two spatial coordinates but not of time. The lower boundary in scaled pressure or "sigma" coordinates is a coordinate surface so that the boundary conditions are particularly simple. However coordinate surfaces "follow the topography" in this case and so introduce difficulties in the interpretation of the usual diagnostic quantities.

Despite the difficulties with the lower boundary condition in isobaric coordinates, diagnostic calculations are almost exclusively performed in that coordinate system. The lower boundary condition is often replaced by a much simpler "flat earth" approximation which ignores topography and variations of surface pressure. For diagnostic calculations which deal with quantities which have a minimum near the surface, such as wind velocities and related quantities, this may be adequate. For diagnostic calculations involving quantities such as temperature and moisture, which reach their maximum values near the surface, the approximation may be less acceptable. A formalism which leads to "exact" diagnostic equations in pressure coordinates is given in Boer (1982).

One of the most ubiquitous general circulation diagnostic calculations is that of the energy cycle following Lorenz (1955). The calculations lead to highly integrated (formally globally integrated) values of terms representing the flow of energy in the atmospheric system from its generation by diabatic heating to its dissipation by friction. Despite the extensive application of the concept of the Lorenz energy cycle, most calculations of the quantities involved have been performed using approximate versions of the equations.

Approximations to the energy budget equations arise from four basic considerations. The first of these is the neglect of topography in the Lorenz (1955) and subsequent (e.g., Dutton and Johnson (1967), van Meighem (1973)) definitions of the reference state for available potential energy.

The second consideration arises from the neglect of topography in the calculation of the various terms in the energy equations in addition to the neglect of topography in the definition of the reference state. This approach is adopted because of the difficulties with the lower boundary conditions mentioned above.

The third consideration arises because it is natural to define available potential energy and to derive its governing equation using potential temperature as the vertical coordinate as was done originally by Lorenz. While it is possible to perform diagnostic calculations in this coordinate system (e.g., Dutton and Johnson (1967), Gall and Johnson (1971), Min and Horn (1982)), the data are generally available at standard pressure levels and it is more common to perform diagnostic calculations in the isobaric system. In his original paper, Lorenz therefore developed an approximate set of equations for available potential energy for use in isobaric coordinates and these approximate equations have been almost universally used for diagnostic calculations of energy cycle terms.

The fourth consideration involves the definition of the "mean" and "eddy" components of the terms in the energy cycle. In the Lorenz approximate version of the equations in isobaric coordinates, the decomposition is unambiguous and is performed by averaging quadratic terms with respect to time and/or space on pressure surfaces. This decomposition (or analysis of variance) depends on the quadratic nature of the quantities of interest in the approximate equations. The exact energy equations in either isobaric or isentropic coordinates have a different and non-quadratic form so that the decomposition must be defined in some different fashion. The decomposition of a term into "mean" and "eddy" quantities typically involves temporal and/or spatial averaging. It is immediately clear that averages and associated mean and eddy terms will differ when the averaging is performed in different coordinate systems.

Taylor (1979) provides a way or removing the first of these approximations by defining a suitable form of the reference state for available potential energy in the isentropic coordinate system which does not neglect topography.

The second difficulty may be overcome by using the formalism of Boer (1982). Holopainen (1970) uses a version of the energy cycle equations which takes account of topography but which is a modification of the approximate equations of Lorenz. In like manner, using the newly developed isobaric formalism, a version of the Lorenz approximate expressions for available potential energy which are modified to account for topography are presented in Boer (1982).

The third of the difficulties may be removed by providing exact equations for the energy cycle in pressure coordinates as is done below.

The fourth difficulty is not easily overcome since the exact equations for available potential energy in isobaric coordinates do not have the quadratic structure of the Lorenz approximate equations. If the decomposition into mean and eddy quantities is desired, some definition must be adopted which displays appropriate features. There is no reason to expect that such a decomposition for the equations written in isentropic coordinates would be the same as that for the equations written in isobaric coordinates, let alone that either would be the same as the standard decomposition used in the Lorenz approximate equations. This situation is discussed as is the development of various approximate versions of the equations.

The purpose of what follows is to develop the exact equations for the complete Lorenz energy cycle in isobaric coordinates and to contrast the form of these equations with those developed in isentropic coordinates as appropriate.

The concept of the energy cycle as applied to the global atmosphere by Lorenz has been a fruitful and widely used one. Nevertheless, the actual calculation of the energy budget terms is fraught with difficulty. Data difficulties for certain terms, the sensitivity of terms which are the differences between large positive and negative contributions and the approximations to the exact equations all contribute to uncertainty. Despite the considerable effort that has been expended on this topic, the terms in the energy cycle are only approximately known.

If such is the case, the "details" associated with more exact forms of the energy-cycle equations may not be of primary importance. However, it might be hoped that more exact forms of the equations together with better data will ultimately lead to stable estimates of these very basic terms in the global energy cycle of the atmosphere.

2. The exact available potential energy equations in isobaric coordinates

The isobaric coordinate formalism of Boer (1982), hereafter B, is used in the derivation. The approach handles the difficulty of the variable lower boundary at $p = p_s(\lambda, \phi, t)$ in pressure coordinates by extending the domain of the equations to $0 < p < p_0$ where p_0 is a *fixed* value which is greater than p_s at all times. Values of the variables are interpolated/extrapolated to those "subterranean" regions for which $p < p_s$, as is commonly done in practice, so that continuous fields on pressure surfaces may be analysed and mapped. The subterranean values do not enter the calculations since the equations are multiplied by the Heaviside or unit function

$$\beta(p - p_s) = \beta(\lambda, \phi, p, t) = \begin{cases} 1, & p < p_s \\ 0, & p > p_s. \end{cases} \tag{2.1}$$

The usual operations of calculus, extended to the behavior of this unit function and its derivative the Dirac delta function, $\partial\beta/\partial p = -\delta$, give slightly modified versions of the equations. The kinematic boundary conditions become $\omega = 0$ for $p = 0$ and $\beta = 0$ for $p = p_0$ in this case and $\mathrm{d}\beta/\mathrm{d}t = 0$.

2.1. *Total potential energy*

The total potential energy is defined as the sum of the internal plus potential energy. For the global atmosphere, the equation for total potential energy may be written alternatively as (B, eq. 21)

$$\begin{aligned} H &= \int_M C_p \beta T \,\mathrm{d}m + \int_S p_s \Phi_s \,\mathrm{d}\sigma/g, \\ &= \int_M C_p \beta \theta \left(\frac{p}{p_{\infty}}\right)^{\kappa} \mathrm{d}m + \int_S p_s \Phi_s \,\mathrm{d}\sigma/g, \end{aligned} \tag{2.2}$$

where $\mathrm{d}m$ is the element of mass and $\mathrm{d}\sigma$ that of area and the integrals are over the mass of the atmosphere and the area of the globe respectively. The usual relation $\theta = T(p_{\infty}/p)^{\kappa}$ connects temperature and potential temperature.

Using the thermodynamic equation in the form $C_p \,\mathrm{d}\beta T/\mathrm{d}t = \beta\omega\alpha + \beta Q$, the equation for total potential energy follows as

$$\frac{\partial H}{\partial t} = \int_M \beta\omega\alpha \,\mathrm{d}m + \int_S \frac{\partial p_s}{\partial t} \Phi_s \,\mathrm{d}\sigma/g + \int_M \beta Q \,\mathrm{d}m. \tag{2.3}$$

2.2. *The reference state*

The unavailable potential energy of the atmosphere is defined as the total potential energy of a particular "reference" state. This is the state of the atmosphere with the minimum total potential energy that may be obtained from the given state of the atmosphere by adiabatic rearrangement. It is shown (Lorenz, 1955; Dutton and Johnson, 1967; Van Mieghem, 1973) that the reference state is "horizontal" in that the reference state variables are functions only of height (and time). Since both pressure and potential temperature are monotonic functions of height for a hydrostatic atmosphere, it follows that the reference state variables are, similarly, functions only of pressure or potential temperature.

The equations for the reference state are most naturally developed using potential temperature as the vertical coordinate but the result may be subsequently expressed in pressure (or other) coordinates. The formalism of B, developed for pressure coordinates, passes over to isentropic coordinates, in particular, the existence and behavior of the β function. The reference state geopotential, temperature, beta and pressure, denoted respectively as Φ_r, T_r, β_r, π are functions only of θ and t.

In his discussion of available potential energy in the case of uneven topography, Taylor (1979) provides a set of equations to be solved to obtain the reference state variables. This set of equations may be rewritten in the form

$$\langle\beta_r\rangle\frac{\partial\pi}{\partial\theta}=\frac{\partial M}{\partial\theta}, \tag{2.4}$$

$$\beta_r(\lambda,\phi,\theta,t)=\beta(\Phi_r-\Phi_s)=\begin{cases}1, & \Phi_r>\Phi_s\\ 0, & \Phi_r<\Phi_s,\end{cases} \tag{2.5}$$

$$\frac{\partial\Phi_r}{\partial\theta}=-C_p\theta\frac{\partial}{\partial\theta}\left(\frac{\pi}{p_{00}}\right)^{\kappa}, \tag{2.6}$$

where $\langle\cdot\rangle$ represents the area weighted average of a quantity over the globe. There are three equations for the unknown reference state variables π, Φ_r, and β_r.

The first of these equations expresses the conservation of mass under the adiabatic redistribution to a horizontal state. The atmospheric mass above a particular isentropic surface for an observed state of the atmosphere is proportional to

$$M(\theta,t)=\left\langle-\int_{\theta}^{\theta_T}\beta\frac{\partial p}{\partial\theta}\,\mathrm{d}\theta\right\rangle, \tag{2.7}$$

where θ_T is some suitable value of θ chosen as that at the "top" of the atmosphere so that as $\theta\rightarrow\theta_T$, p becomes negligible as do terms involving p as a product. For the horizontal reference state, the mass above the same θ surface is expressed in terms of the reference state variables as

$$M_r(\theta,t)=\left\langle-\int_{\theta}^{\theta_T}\beta_r\frac{\partial\pi}{\partial\theta}\,\mathrm{d}\theta\right\rangle. \tag{2.8}$$

Since the reference state is obtained by an adiabatic rearrangement, each air parcel conserves its potential temperature and the mass above the isentropic surface must be the same in both the observed and reference states, i.e., $M(\theta,t)=M_r(\theta,t)$. Since this is true for arbitrary θ, it follows that

$$\left\langle\beta_r\frac{\partial\pi}{\partial\theta}\right\rangle=\langle\beta_r\rangle\frac{\partial\pi}{\partial\theta}=\left\langle\beta\frac{\partial p}{\partial\theta}\right\rangle, \tag{2.9}$$

which is equivalent to (2.4) using (2.7) and recalling that π is independent of the horizontal coordinates. Eq. (2.9) embodies the basic connection between the reference state and observed state thermodynamic variables which is subsequently used in the representation of the available potential energy.

Eq. (2.5) is just a straightforward definition of β_r for the reference state. Note that it is a function of the horizontal variables since topography is *not* neglected in this development. Eq. (2.6) is the hydrostatic equation applied to the reference state. The three eqs. (2.4)–(2.6) serve to define the reference state in terms of the observed state of the atmosphere. The distribution of the reference state pressure may be easily obtained in the case where topography is neglected, i.e., as originally obtained by Lorenz, by setting $\beta_r=1$. A further approximation is introduced if $\beta=1$ is assumed for the observed state.

Presumably the solution of eqs. (2.4)–(2.6) for the distribution of the reference state variables in presence of topography is a matter of technique. Koehler (1986) reports difficulties in solving for the reference state variables by iteration in the fashion suggested by Taylor (1979) and proposes an alternative method.

2.3. *Unavailable potential energy*

The unavailable potential energy of the atmosphere is the total potential energy of the reference state. It is obtained by substituting reference state values of the variables into the isentropic version of (2.2) and using (2.9) to express the result. The integral of a quantity over the mass of the atmosphere in the reference state may be rewritten as the integral of that reference state variable in the framework of the observed state using

$$\begin{aligned}\int_M\beta_rX_r\,\mathrm{d}m_r&=\iint X_r\beta_r\frac{\partial\pi}{\partial\theta}\,\mathrm{d}\sigma\,\mathrm{d}\theta/g\\&=\iint X_r\beta\frac{\partial p}{\partial\theta}\,\mathrm{d}\sigma\,\mathrm{d}\theta/g=\int_M\beta X_r\,\mathrm{d}m,\end{aligned}$$

which follows from the independence of reference state variables X_r, π (but not β_r) to the horizontal coordinates and the relationship (2.9).

The unavailable potential energy is then

$$\begin{aligned}H_r&=\int_M C_p\beta T_r\,\mathrm{d}m+\int_S\pi_s\Phi_s\,\mathrm{d}\sigma/g,\\&=\int_M C_p\beta\theta\left(\frac{\pi}{p_{00}}\right)^{\kappa}\mathrm{d}m+\int_S\pi_s\Phi_s\,\mathrm{d}\sigma/g,\end{aligned} \tag{2.10}$$

in either θ- or p-coordinates. It follows from the thermodynamic equation that the rate of change of the unavailable potential energy may be written as

$$\frac{\partial H_r}{\partial t} = \int_M \beta \left(\frac{\pi}{p_{00}}\right)^\kappa \left(\frac{\theta}{T}\right) Q \, dm, \tag{2.11}$$

provided that the term

$$R = \int_M C_p \beta \theta \frac{d}{dt} \left(\frac{\pi}{p_{00}}\right)^\kappa dm + \int_S \frac{\partial \pi_s}{\partial t} \Phi_s \, d\sigma/g \tag{2.12}$$

is zero. This is shown to be the case in the Appendix.

2.4. *Available potential energy*

The available potential energy is the difference between the total potential energy H and the unavailable potential energy H_r which is the total potential energy of the reference state. It is obtained from (2.2) and (2.11) and may be expressed in a number of forms as

$$\begin{aligned} A &= \int_M C_p \beta (T - T_r) \, dm + \int_S (p_s - \pi_s) \Phi_s \, d\sigma/g \\ &= \int_M C_p \beta N T \, dm + \int_S (p_s - \pi_s) \Phi_s \, d\sigma/g \qquad (2.13) \\ &= \int_M C_p \beta M \theta \, dm + \int_S (p_s - \pi_s) \Phi_s \, d\sigma/g, \end{aligned}$$

where

$$N = 1 - (\pi/p)^\kappa$$
$$M = (p^\kappa - \pi^\kappa)/p_{00}^\kappa.$$

N has been termed an "efficiency factor" by Lorenz (1967) because of its role in the expression for the generation of available potential energy. Apparently it is a measure of the difference between the observed and reference states. This may be seen explicitly by writing it as

$$N = (T - T_r)/T$$

which shows that it is a local measure of the difference between the observed and reference state temperatures.

The first expression in (2.13) for available potential energy clearly illustrates its definition as the difference in the total potential energy between the observed and reference state. The second expression measures this difference in terms of the efficiency factor N while the third expression rewrites this in terms of potential temperature.

The equation for the rate of change of A follows from (2.3) and (2.11) as

$$\frac{\partial A}{\partial t} = \int_M \beta \omega \alpha \, dm + \int_S \frac{\partial p_s}{\partial t} \Phi_s \, d\sigma/g + \int_M \beta N Q \, dm, \tag{2.14}$$

or alternatively (using B, eq. 17) as

$$\frac{\partial A}{\partial t} = \int_M \beta V \cdot \nabla \Phi \, dm + \int_M \beta N Q \, dm. \tag{2.15}$$

3. The exact energy budget equations

The kinetic energy equation, where effects of topography have not been neglected, is given in B, Subsection 3.4. Combining this with the equation for available potential energy gives the exact equations for the atmospheric energy cycle:

$$\begin{aligned} \frac{\partial K}{\partial t} &= C - D, \\ \frac{\partial A}{\partial t} &= -C + G, \end{aligned} \tag{3.1}$$

where

$$K = \int_M \frac{1}{2} \beta V \cdot V \, dm,$$

$$A = \int_M C_p \beta N T \, dm + \int_S (p_s - \pi_s) \Phi_s \, d\sigma/g,$$

$$\begin{aligned} C &= -\int_M \beta V \cdot \nabla \Phi \, dm \\ &= -\left\{ \int_M \beta \omega \alpha \, dm + \int_S \frac{\partial p_s}{\partial t} \Phi_s \, d\sigma/g \right\}, \end{aligned}$$

$$D = -\int_M \beta V \cdot F \, dm,$$

$$G = \int_M \beta N Q \, dm,$$

and $N = 1 - (\pi/p)^\kappa$ is the efficiency factor.

The eqs. (3.1) express the rate of change of kinetic and available potential energy for the atmosphere in terms of the conversion between the two forms, the dissipation of kinetic energy by frictional effects and the generation of available potential energy by diabatic heating. The conversion is expressed alternatively as the work done by down gradient ageostrophic motions or by the rising of warm and sinking of cool air. The efficiency factor N arises both in the expression for the available potential energy and its generation. In the generation term it measures how "efficient" the heating is in generating available potential energy.

Note that these equations are generally valid for various vertical coordinate systems but, because of the form of the expressions, may be thought of as being particularly appropriate in isobaric coordinates. Beginning with the hydrostatic primitive equations and the kinematic boundary conditions, the development of the energy budget equations has been without further approximation so that they are exact in that sense. In particular, the presence of topography has not been neglected in the determination of the reference state for available potential energy nor in the subsequent derivation of the equations for available potential energy or for kinetic energy.

4. Mean and eddy components

The usual "four-box" energy budget diagram of Lorenz represents a decomposition of the basic energy equations into "mean" and "eddy" components. In the case of the Lorenz (1955) approximate equations in isobaric coordinates, the available potential energy and kinetic energy associated with the zonal average of the flow and the deviations therefrom, termed the eddies, are considered separately. Equations expressing the amounts of the four components of energy, their sources and sinks and the conversions between them are derived. This approach represents a relatively straightforward decomposition of variance of the quadratic quantities kinetic energy and squared temperature deviation which arise in the Lorenz approximate equations under the averaging operation. Various decompositions are possible and are frequently used. The most common involve time and/or zonal averaging.

Unfortunately, the decomposition of the energy budget equations into various mean and eddy components introduces a number of difficulties. The most commonly used decompositions may be represented as $X=\overline{X}+X'$, $X=[X]+X^*$, and $X=[X]+\overline{X}^*+X'$ where the overbar represents a time mean and the square brackets a zonal average. The decompositions are into the time mean and transient eddies, the zonal mean and eddies, and the time and zonal mean and standing and transient eddies respectively.

It is immediately apparent that the value of mean and eddy components will not be the same when the time and zonal averages are carried out with different vertical coordinate values kept fixed. Hence there cannot be a single set of "exact" component equations for the energy cycle since the various components will have different values depending on the vertical coordinate system in which the averaging is performed. It makes little sense to compare the mean and eddy quantities calculated in different coordinate systems as to their "correctness". At best the component equations which are "exact for a particular coordinate system" can be obtained. The actual values of the various terms will differ depending on the coordinate system in which the calculation is performed.

Another consideration rests with the development of the equations linking the various components with each other and with their sources and sinks. A definition of "mean" and "eddy" components of available potential energy, for instance, which is not accompanied by the associated budget equations for these components is incomplete.

Finally, the available potential energy in its exact form is not a quadratic expression so that, even in isobaric coordinates, the decomposition into mean and eddy components is not straightforward as is the case for kinetic energy. The form of the mean and eddy components of available potential energy is not, therefore, without ambiguity. It will depend on the choice of a suitable definition of the mean and eddy components.

For all these reasons it is clear that while (3.1) are the exact equations for the energy budget, there can be no universal "exact" equations for the mean and eddy components of the energy budget. Because of the various advantageous properties of the isobaric coordinate system, the component equations in that coordinate system are investigated in what follows.

4.1. Exact component equations for kinetic energy

In isobaric coordinates, the exact component equations for kinetic energy may be obtained following the approach in B. These equations are unique in that the presence of topography is not ignored. The approach makes use of so-called representative values which take into account the fact that the fields on a pressure surface are not everywhere defined due to the intersection of the pressure surface and the topography.

The decomposition of the kinetic energy equation into components is illustrated for what might be called the "traditional" case of the zonally averaged and eddy components. This is the decomposition discussed in Lorenz (1955) for instance. Similar results for cases involving time and both time and zonal averaging may be developed similarly.

The kinetic energy may be decomposed into its zonal and eddy components as $\frac{1}{2}[\beta \boldsymbol{V}\cdot\boldsymbol{V}] = \frac{1}{2}[\beta][\boldsymbol{V}]_{\mathrm{R}}\cdot[\boldsymbol{V}]_{\mathrm{R}} + \frac{1}{2}[\beta \boldsymbol{V}^*\cdot\boldsymbol{V}^*]$. The equation for the zonal component is obtained by taking the dot product of $[\boldsymbol{V}]_{\mathrm{R}}$ with the averaged equation of motion while the eddy component equation is obtained by subtracting the zonal component equation from the full equation.

The resulting component equations for kinetic energy are

$$\begin{aligned}\frac{\partial K_{\mathrm{Z}}}{\partial t} &= C_{\mathrm{Z}} - C_{\mathrm{K}} - D_{\mathrm{Z}},\\ \frac{\partial K_{\mathrm{E}}}{\partial t} &= C_{\mathrm{E}} + C_{\mathrm{K}} - D_{\mathrm{E}},\end{aligned} \tag{4.1}$$

where

$$K_{\mathrm{Z}} = \int_{\mathrm{M}} \frac{1}{2}[\beta][\boldsymbol{V}]_{\mathrm{R}}\cdot[\boldsymbol{V}]_{\mathrm{R}}\,\mathrm{d}m,$$

$$K_{\mathrm{E}} = \int_{\mathrm{M}} \frac{1}{2}[\beta \boldsymbol{V}^*\cdot\boldsymbol{V}^*]\,\mathrm{d}m,$$

$$\begin{aligned}C_{\mathrm{K}} = -\int_{\mathrm{M}} a\cos\phi \bigg\{&\bigg([\beta u^* \boldsymbol{V}^*]\cdot\nabla\\ &+ [\beta u^*\omega^*]\frac{\partial}{\partial p}\bigg)\left(\frac{[u]_{\mathrm{R}}}{a\cos\phi}\right)\\ &+ \bigg([\beta v^* \boldsymbol{V}^*]\cdot\nabla + [\beta v^*\omega^*]\frac{\partial}{\partial p}\\ &- \frac{\tan\phi}{a}[\beta \boldsymbol{V}^*\cdot\boldsymbol{V}^*]\bigg)\left(\frac{[v]_{\mathrm{R}}}{a\cos\phi}\right)\bigg\}\,\mathrm{d}m\\ &+ \int_{\mathrm{M}} \bigg\{\left[\beta\frac{\partial\Phi^*}{\partial t}\right] + [\boldsymbol{V}]_{\mathrm{R}}\cdot[\beta\nabla\Phi]\\ &+ [\omega]_{\mathrm{R}}\left[\beta\frac{\partial\Phi^*}{\partial p}\right]\bigg\}\,\mathrm{d}m,\end{aligned}$$

$$C_{\mathrm{Z}} = -\int_{\mathrm{M}} [\beta][\omega]_{\mathrm{R}}[\alpha]_{\mathrm{R}}\,\mathrm{d}m - \int_{\mathrm{S}}\left[\frac{\partial p_{\mathrm{s}}}{\partial t}\Phi_{\mathrm{s}}\right]\mathrm{d}\sigma/g,$$

$$C_{\mathrm{E}} = -\int_{\mathrm{M}} [\beta\omega^*\alpha^*]\,\mathrm{d}m,$$

$$D_{\mathrm{Z}} = -\int_{\mathrm{M}} [\beta][\boldsymbol{V}]_{\mathrm{R}}\cdot[\boldsymbol{F}]\,\mathrm{d}m,$$

$$D_{\mathrm{E}} = -\int_{\mathrm{M}} [\beta \boldsymbol{V}^*\cdot\boldsymbol{F}^*]\,\mathrm{d}m.$$

Component equations under time or time and zonal averagings are of the same general form.

These equations are similar to the usual form except for the presence of β in the various terms and the additional terms which arise in the expressions for C_{K} and C_{Z}. These latter terms are a consequence of retaining topographic information and go to zero in the "flat earth" approximation ($\Phi_{\mathrm{s}} = 0$, $\beta = 1$). They should not, however, be interpreted as reflecting the net effect of the presence of topography on the energy budget.

4.2. *Component equations for available potential energy in isobaric coordinates*

As noted above, there is some ambiguity in the definition of "mean" and "eddy" components for available potential energy. If the decomposition involves averaging, the results will necessarily be different in the different vertical coordinate systems. In addition, for non-quadratic quantities, the form of the decomposition into components is a matter of definition.

The following definition is proposed for the decomposition of available potential energy into its components in pressure coordinates:

> If $X = \{X\}_{\mathrm{R}} + X^{\#}$ represents the decomposition of the variable $X(\lambda, \phi, p, t)$ into a representative average part and the deviation therefrom, and if the available potential energy is written as the integral over the mass of the atmosphere of some function of X (and possibly other variables), as
>
> $$A = \int_{\mathrm{M}} \{\beta f(X)\}\,\mathrm{d}m,$$
>
> then the averaged or "mean" component of available potential energy is defined as
>
> $$A_{\mathrm{mean}} = \int_{\mathrm{M}} \{\beta f(\{X\}_{\mathrm{R}})\}\,\mathrm{d}m \tag{4.2}$$

and the corresponding "eddy" component as

$$A_{\mathrm{eddy}} = A - A_{\mathrm{mean}} = \int_{\mathrm{M}} \{\beta(f(X) - f(\{X\}_{\mathrm{R}}))\}\,\mathrm{d}m. \tag{4.3}$$

In other words, the "mean" component of the available potential energy is obtained by replacing the observed atmospheric quantity by its averaged value. The "eddy" component is obtained by subtraction. The same definition applied to a quadratic quantity like kinetic energy gives the expected result as in Section 4.1. The definition clearly makes sense in the case that the observed state of the atmosphere has no deviations from the averaged state (i.e., when $X^* = 0$).

The definition may be applied to the expression for available potential energy in isobaric coordinates for the various decompositions. The method is illustrated using the "traditional" decomposition into zonal and eddy components as $X = [X] + X^*$. Of course other decompositions give analogous results.

For this decomposition, the zonal available potential energy in isobaric coordinates following the definition (4.2) can be expressed in a variety of forms as:

$$\begin{aligned} A_{\mathrm{Z}} &= \int_{\mathrm{M}} C_{\mathrm{p}}\beta([T]_{\mathrm{R}} - T_{\mathrm{rZ}})\,\mathrm{d}m + \int_{\mathrm{S}} (p_{\mathrm{s}} - \pi_{\mathrm{sZ}})\Phi_{\mathrm{s}}\,\mathrm{d}\sigma/g \\ &= \int_{\mathrm{M}} C_{\mathrm{p}}\beta N_{\mathrm{Z}}[T]_{\mathrm{R}}\,\mathrm{d}m + \int_{\mathrm{S}} (p_{\mathrm{s}} - \pi_{\mathrm{sZ}})\Phi_{\mathrm{s}}\,\mathrm{d}\sigma/g \\ &= \int_{\mathrm{M}} C_{\mathrm{p}}\beta M_{\mathrm{Z}}[\theta]_{\mathrm{R}}\,\mathrm{d}m + \int_{\mathrm{S}} (p_{\mathrm{s}} - \pi_{\mathrm{sZ}})\Phi_{\mathrm{s}}\,\mathrm{d}\sigma/g, \end{aligned} \tag{4.3}$$

where

$$N_{\mathrm{Z}} = 1 - (\pi_{\mathrm{Z}}/p)^{\kappa}$$

$$M_{\mathrm{Z}} = (p^{\kappa} - \pi_{\mathrm{Z}}^{\kappa})/p_{\infty}^{\kappa}$$

$$\pi_{\mathrm{Z}} = \pi([\theta]_{\mathrm{R}}, t)$$

$$[\theta]_{\mathrm{R}} = [T]_{\mathrm{R}}\left(\frac{p_{\infty}}{p}\right)^{\kappa} = T_{\mathrm{rZ}}\left(\frac{p_{\infty}}{\pi_{\mathrm{Z}}}\right)^{\kappa}.$$

These expressions are obtained by replacing θ by its averaged value on pressure surfaces, $[\theta]_{\mathrm{R}}$, in eqs. (2.13).

The corresponding eddy available potential energy is

$$\begin{aligned} A_{\mathrm{E}} &= A - A_{\mathrm{Z}} \\ &= \int_{\mathrm{M}} C_{\mathrm{p}}\beta(T_{\mathrm{rZ}} - T_{\mathrm{r}})\,\mathrm{d}m + \int_{\mathrm{S}} (\pi_{\mathrm{sZ}} - \pi_{\mathrm{s}})\,\mathrm{d}\sigma/g \\ &= \int_{\mathrm{M}} C_{\mathrm{p}}\beta(NT - N_{\mathrm{Z}}[T]_{\mathrm{R}})\,\mathrm{d}m \\ &\quad + \int_{\mathrm{S}} (\pi_{\mathrm{sZ}} - \pi_{\mathrm{s}})\,\mathrm{d}\sigma/g \\ &= \int_{\mathrm{M}} C_{\mathrm{p}}\beta(M\theta - M_{\mathrm{Z}}[\theta]_{\mathrm{R}})\,\mathrm{d}m \\ &\quad + \int_{\mathrm{S}} (\pi_{\mathrm{sZ}} - \pi_{\mathrm{s}})\,\mathrm{d}\sigma/g. \end{aligned} \tag{4.4}$$

While it is not particularly obvious from the first of these expressions, the eddy available potential energy does depend on the deviation of temperature from the zonal mean. The second and third forms of the expressions may be used to illustrate this. For example

$$\begin{aligned} A_{\mathrm{E}} &= \int_{\mathrm{M}} C_{\mathrm{p}}\beta(NT^* + (N - N_{\mathrm{Z}})[T]_{\mathrm{R}})\,\mathrm{d}m \\ &\quad + \int_{\mathrm{S}} (\pi_{\mathrm{sZ}} - \pi_{\mathrm{s}})\,\mathrm{d}\sigma/g. \end{aligned}$$

Note, however, that this expression does not obviously involve the variance of temperature on pressure surfaces as do the approximate expressions of Lorenz (1955) or Boer (1975).

4.3. Equations for the mean and eddy components of available potential energy in isobaric coordinates

Differentiating the expression (4.3) for M and multiplying by $C_{\mathrm{p}}\beta[\theta]_{\mathrm{R}}$ adding it to the product of M_{Z} with the thermodynamic equation and integrating over the mass of the atmosphere gives

$$\begin{aligned} \frac{\partial}{\partial t}\int_{\mathrm{M}} [\beta] C_{\mathrm{p}} M_{\mathrm{Z}}[\theta]_{\mathrm{R}}\,\mathrm{d}m &= \int_{\mathrm{M}} [\beta][\omega]_{\mathrm{R}}[\alpha]_{\mathrm{R}}\,\mathrm{d}m \\ &\quad - \int_{\mathrm{M}} M_{\mathrm{Z}}\left(\nabla\cdot C_{\mathrm{p}}[\beta\theta^* V^*] + \frac{\partial}{\partial p} C_{\mathrm{p}}[\beta\theta^*\omega^*]\right)\mathrm{d}m \\ &\quad + \int_{\mathrm{M}} M_{\mathrm{Z}}\left(\frac{\theta}{T}\right)[\beta][Q]_{\mathrm{R}}\,\mathrm{d}m + R_{\mathrm{Z}}, \end{aligned}$$

where R_{Z} is of the same form as previously in (2.12) but with the variables replaced with their averaged values in the same sense as for the definition of the decomposition into mean and eddy terms.

Since $R = 0$ for an arbitrary atmospheric state, as shown in the Appendix, it follows that it is zero for the state which is obtained by replacing the observed distribution of the variables by their zonal averages. Thus $R_Z = 0$ and the equations for zonal and eddy available potential energy follow as

$$\begin{aligned}\frac{\partial A_Z}{\partial t} &= G_Z - C_Z - C_A,\\ \frac{\partial A_E}{\partial t} &= G_E - C_E + C_A,\end{aligned} \tag{4.5}$$

for

$$A_Z = \int_M C_p \beta N_Z [T]_R \, dm + \int_S (p_s - \pi_{sZ}) \Phi_s \, d\sigma/g,$$

$$A_E = \int_M C_p \beta (N - N_Z) T \, dm + \int_S (\pi_{sZ} - \pi_s) \Phi_s \, d\sigma/g,$$

$$G_Z = \int_M \beta N_Z [Q]_R \, dm,$$

$$G_E = \int_M \beta (N - N_Z) Q \, dm,$$

$$\begin{aligned} C_A &= \int_M C_p N_Z \left(\frac{T}{\theta}\right) \left(\nabla \cdot \left(\frac{\theta}{T}\right) [\beta T^* \boldsymbol{V}^*] \right.\\ &\qquad \left. + \frac{\partial}{\partial p} \left(\frac{\theta}{T}\right) [\beta T^* \omega^*] \right) dm \\ &= - \int_M C_p \left(\frac{\theta}{T}\right) \left([\beta T^* \boldsymbol{V}^*] \cdot \nabla \right.\\ &\qquad \left. + [\beta T^* \omega^*] \frac{\partial}{\partial p} \right) \left(\frac{T}{\theta} N_Z\right) dm. \end{aligned}$$

The terms representing the generation of zonal and eddy available potential energy involve versions of the efficiency factor while the terms representing conversions with kinetic energy have already been given in (4.1).

The terms representing the interaction between zonal and eddy forms of the available potential energy differs from the corresponding approximate Lorenz form although some parallelism with that form can be seen. The first form is perhaps the most straightforward. N_Z is a measure of the difference between the observed and the zonally averaged reference states. It is typically positive in tropical regions and negative in extratropical regions. Thus, divergence of energy from the tropical region and convergence into the extratropics due to the eddies acts to reduce the difference between the observed and zonal reference states and represents a conversion from zonal to eddy forms of available potential energy. Alternatively, the term may be represented by the transport of energy by the eddies down the gradient, which is basically from warm to cold temperatures. Note that the variance terms may be alternatively written in terms of θ and M.

Eqs. (4.5) for available potential energy together with eqs. (4.1) for kinetic energy represent an "exact" set of energy budget equations in isobaric coordinates.

4.4. Component equations in isentropic coordinates

In his book, Lorenz (1967, p. 1087) states that "It is not at all certain how the zonal average of the mass field is best defined" when referring to the proposed decomposition of the available potential energy into its zonal and eddy components using the so-called "exact" formula. This difficulty is manifest in the several definitions that have been used to obtain zonal and eddy components of the available potential energy and kinetic energy in isentropic coordinates. Several such expressions have been proposed and evaluated (e.g., Min and Horn, 1982; Zautner et al., 1986). Budget equations for the components in isentropic coordinates, corresponding to eqs. (4.1) and (4.5) in isobaric coordinates, do not seem to have been derived.

In isentropic coordinates kinetic energy is formally a triple product under averaging, i.e., $[\boldsymbol{V} \cdot \boldsymbol{V} \, \partial p/\partial \theta]$, although it may be conveniently decomposed into mean and eddy components using a mass-weighted average as was apparently done by Zautner et al. (1986). The situation is less clear for available potential energy since its equation is normally written as

$$\begin{aligned} A = \frac{C_p}{g p_{\infty}^{\kappa} (1 + \kappa)} &\int_\theta \int_S (p^{\kappa+1} - \pi^{\kappa+1}) \, d\theta \, d\sigma \\ &+ \int_S (p_s^{\kappa+1} - \pi_s^{\kappa+1}) \theta_s \, d\sigma \\ &+ \int_S (p_s - \pi_s) \Phi_s \, d\sigma/g \end{aligned}$$

where the factor $\partial p/\partial \theta$ is notable by its absence.

One possible method of decomposition for the available potential energy is to use the same definition for the zonal and eddy decomposition as proposed for the isobaric case, the mass weighted averagings in isentropic coordinates defined above and the formula (2.13) for available potential energy.

This would apparently not correspond to the proposed decomposition given by Lorenz in his book nor would it correspond to that used by Min and Horn (1982). The latter definition would seem to be flawed in any case since an observed state of the atmosphere which happened to be zonally symmetric when viewed in isentropic coordinates apparently gives $A_Z = 0$ and $A_E = A$.

5. Approximate forms of the available potential energy equations

While it may seem somewhat obtuse to consider approximate versions of the equations when so much effort has been expended to obtain the "exact" forms, it is nevertheless instructive to do so. One set of approximations has to do with terms involving β and hence the treatment of the boundary conditions and the presence of topography. A second set of approximations arises when the reference state variables, which are functions only of θ and time, are expanded in power series about various averaged values of θ.

Several approximations which are obtained by replacing β by various of its averages and ultimately by making the "flat earth" approximation for which $\beta = 1$ are available. In particular, a very good approximation, and one which simplifies calculations considerably is obtained by replacing β everywhere by its time average $\overline{\beta}$ when considering time averaged versions of the equations. Experience suggests that this approximation is quite acceptable. Other approximations obtained by replacing β by its zonal or time and zonal average are not particularly acceptable nor is replacing β by 1.

One of the main differences between the components eqs. (4.5) and the Lorenz approximate equations is in the appearance in the Lorenz equations of several terms involving the variance of temperature and the covariance of temperature and adiabatic heating rates. Instead, the eqs. (4.5) have terms involving the efficiency factor.

It is possible to derive the Lorenz equations from the exact equations in pressure coordinates and there is also a set of approximate equations which is intermediate between the exact equations and the Lorenz approximate equations. This intermediate approximate set of equations is derived in Boer (1975) and are developed from the exact equations by expanding the efficiency factor $M(\theta, t)$ and hence the reference pressure π, in Taylor series about the representative mean $[\theta]_R$ and retaining terms of second order in θ^*. The resulting budget equations have some aspects of the exact equations and some aspects of the Lorenz approximate equations. The stability parameter γ arising in this intermediate approximation is a function of latitude and time as opposed to the stability parameter of Lorenz which is a function of pressure only. This intermediate approximation gives results which differ appreciably from those obtained with the same climatological data using the Lorenz approximation.

The Lorenz approximation is obtained by expanding in Taylor series about the global average $\langle\theta\rangle_R$ and retaining terms of second order in the eddy components $[\theta]^+ + \theta^*$ (where $[\theta]^+ = [\theta] - \langle\theta\rangle$). In addition the assumption that

$$\pi(\langle\theta\rangle_R \sim p,$$

and

$$\left.\frac{\partial\pi}{\partial\theta}\right|_{\theta=\langle\theta\rangle_R} \sim \left(\frac{\partial\langle\theta\rangle_R}{\partial p}\right)^{-1}$$

is made to obtain the Lorenz approximate equations. Since the neglected terms in the expression depend, in this case, on quantities of the order of $([\theta]^+ + \theta^*)^3$ compared to θ^{*3} in the previous case, the approximation will generally be less accurate.

6. Concluding remarks

The "exact" equations for the available potential and kinetic energy budgets of the energy cycle have been obtained in isobaric coordinates. The equations are exact in the sense that they are obtained from the hydrostatic primitive equations together with the kinematic boundary condition without further approximation. Information concerning the distribution of topography is retained throughout.

A decomposition of the equations into mean and eddy components is proposed and budget equations are obtained for the components. These budget terms include expressions for the generation and dissipation of mean and eddy components of available potential and kinetic energy and for the conversions between components. These "exact" component equations in isobaric coordinates depend on averaging on pressure surfaces so they necessarily differ from analogous component equations in isentropic or other coordinates, if they were to be developed. There is little point in comparing, as to their "correctness", terms in the component energy budget equations as calculated for data in different vertical coordinate systems.

These exact energy cycle equations in pressure coordinates are quite different in appearance from the Lorenz approximate equations which have formed the basis of most energy cycle calculations to date. In particular, the "efficiency" factor which is a direct measure of the difference between the observed and reference state is seen in the formulae for the amount of available potential energy, the generation of available potential energy and the conversion between mean and eddy forms of available potential energy. In addition, the retention of topographic information leads to a modification of the usual expression for the conversion between mean and eddy components in the kinetic energy equations.

There are a number of possible approximations to the exact energy budget equations. Various approximate versions of the equations can be obtained by "neglecting the topography" or by neglecting certain features of it in the calculation of the reference pressure and the various energy terms. An approximation "intermediate" between the exact equations and the Lorenz approximate equations may be derived by expanding the reference state parameters in Taylor series about the zonal mean and retaining terms to second order in the "eddies". The Lorenz approximate equations are obtained in a similar fashion where the expansion is about the global surface mean.

Considering the difficulties with the observation of the global structure of the atmosphere and the difficulty in calculating the various terms in the energy budget equations in a stable way, it may be that the details of "exact" versus the various approximate versions of the energy budget are unimportant. Nevertheless it is already clear (e.g., Boer, 1975) that the exact forms of the equations give different results from the Lorenz approximate form. It might be hoped that improved observational coverage and analysis methods together with the exact form of the equations will lead to stable estimates of these very basic atmospheric quantities.

7. Acknowledgements

I would like to thank L. Smith and J. Cantin for their expert processing of the manuscript.

8. Appendix

The equation for unavailable potential energy

As indicated in Subsection 2.3, the equation for the rate of change of the unavailable potential energy is given by (2.11) provided that the term R of (2.12) is zero. This can be shown to be the case as follows.

Since π is a function only of θ and t,

$$\frac{\mathrm{d}\pi}{\mathrm{d}t}=\frac{\partial\pi}{\partial t}+\dot{\theta}\,\frac{\partial\pi}{\partial\theta}$$

and

$$\left\langle\beta\,\frac{\partial p}{\partial\theta}\,\frac{\partial\pi}{\partial t}\right\rangle=\frac{\partial\pi}{\partial\theta}\left\{\langle\beta_{\mathrm{r}}\rangle\,\frac{\partial\pi}{\partial t}+\left\langle\beta\,\frac{\partial p}{\partial\theta}\,\dot{\theta}\right\rangle\right\}.$$

Averaging and integrating the continuity equation in isentropic coordinates in the vertical gives

$$\frac{\partial M}{\partial t}=\frac{\partial}{\partial t}\left\langle-\int_{\theta}^{\theta_{\mathrm{T}}}\beta\,\frac{\partial p}{\partial\theta}\,\mathrm{d}\theta\right\rangle=-\left\langle\beta\,\frac{\partial p}{\partial\theta}\,\dot{\theta}\right\rangle.$$

The conservation of mass for the atmosphere expressed as $\partial\langle\pi_{\mathrm{s}}\rangle/\partial t=0$ together with the integrated form of (2.8) written as $M_{\mathrm{r}}=\langle\beta_{\mathrm{r}}\rangle\pi+\langle(1-\beta_{\mathrm{r}})\pi_{\mathrm{s}}\rangle$ gives

$$\frac{\partial M}{\partial t}=\frac{\partial M_{\mathrm{r}}}{\partial t}=\langle\beta_{\mathrm{r}}\rangle\,\frac{\partial\pi}{\partial t}-\left\langle\beta_{\mathrm{r}}\,\frac{\partial\pi_{\mathrm{s}}}{\partial t}\right\rangle=-\left\langle\beta\,\frac{\partial p}{\partial\theta}\,\dot{\theta}\right\rangle,$$

so that R may be written as

$$R=\frac{S}{g}\left\langle\frac{\partial\pi_{\mathrm{s}}}{\partial t}\left\{\Phi_{\mathrm{s}}+\int_{0}^{\theta_{\mathrm{T}}}C_{\mathrm{p}}\theta\,\frac{\kappa}{\pi}\left(\frac{\pi}{p_{00}}\right)^{\kappa}\beta_{\mathrm{r}}\,\frac{\partial\pi}{\partial\theta}\,\mathrm{d}\theta\right\}\right\rangle,$$

where S is the earth's surface area. Using the hydrostatic equation for the reference state (2.6) gives

$$R = \frac{4\pi a^2}{g} \left\langle \frac{\partial \pi_s}{\partial t} \left\{ \Phi_s - \int_0^{\theta_T} \beta_r \frac{\partial \Phi}{\partial \theta} \, d\theta \right\} \right\rangle$$

$$= \frac{4\pi a^2}{g} \left\langle \frac{\partial \pi_s}{\partial t} \right\rangle \Phi_r(\theta_T) = 0,$$

since $\partial \langle \pi_s \rangle / \partial t = 0$.

An equation for the total derivative of the reference pressure may also be obtained from the expressions above as

$$\langle \beta_r \rangle \frac{d\pi}{dt} = -\left\langle \beta \frac{\partial p}{\partial \theta} \dot{\theta} \right\rangle + \langle \beta_r \rangle \dot{\theta} \frac{\partial \pi}{\partial \theta} + \left\langle \beta_r \frac{\partial \pi_s}{\partial t} \right\rangle.$$

In the absence of topography this takes the form

$$\frac{d\pi}{dt} = -\left\langle \dot{\theta} \frac{\partial p}{\partial \theta} \right\rangle + \dot{\theta} \frac{\partial \pi}{\partial \theta},$$

as obtained by Egger (1976) and the demonstration that $R = 0$ follows much more directly as in Boer (1976).

REFERENCES

Boer, G. J. 1975. Zonal and eddy forms of the available potential energy equations in pressure coordinates. *Tellus 27*, 433–442.

Boer, G. J. 1976. Reply to J. Egger. *Tellus 28*, 379.

Boer, G. J. 1982. Diagnostic Equations in Isobaric Coordinates. *Mon. Wea. Rev. 110*, 1801–1820.

Dutton, J. A. and Johnson, D. R. 1967. The theory of available potential energy and a variational approach to atmospheric energetics. *Advances in Geophysics 12*, 333–436.

Egger, J. 1976. Comments on "Zonal and eddy forms of available potential energy equations in pressure coordinates" by G. J. Boer. *Tellus 28*, 377–378.

Gall, R. L. and Johnson, D. R. 1971. The generation of available potential energy by sensible heating: A case study. *Tellus 23*, 466–482.

Holopainen, E. O. 1970. An observational study of the energy balance of the stationary disturbances in the atmosphere. *Quart. J. Roy. Meteor. Soc. 96*, 626–644.

Koehler, T. L. 1986. A terrain-dependent reference atmosphere determination method for available potential energy calculations. *Tellus 38A*, 42–48.

Lorenz, E. N. 1955. Available potential energy and the maintenance of the general circulation. *Tellus 7*, 157–167.

Lorenz, E. N. 1967. *The nature and theory of the general circulation of the atmosphere. WMO—No. 218 TP.115.* Geneva: World Meteorological Organization, 161 pp.

Min, K. D. and Horn, L. H. 1982. Available potential energy in the northern hemisphere during the FGGE year. *Tellus 34*, 526–539.

Taylor, K. E. 1979. Formulas for calculating available potential energy over uneven topography. *Tellus 31*, 236–245.

van Meighem, J. 1973. *Atmospheric energetics.* Oxford: Clarendon Press.

Zautner, J. H., Whittaker, L. M. and Horn, L. H. 1986. Northern Hemisphere Kinetic Energy during the First GARP Global Experiment Year. *J. of Geophys. Res. 91*, 7809–7815.