

# Energy conversion, mixing energy and neutral surfaces with a nonlinear equation of state

Jonas Nycander

Dept. of Meteorology, Stockholm University, Stockholm, Sweden  
email: jonas@misu.su.se

November 6, 2009

## Abstract

A local neutral plane is defined so that a water parcel that is displaced adiabatically a small distance along the plane continues to have the same density as the surrounding water. Since such a displacement does not change the density field or the gravitational potential energy, it is generally assumed that it does not produce a restoring buoyancy force. However, it is here shown that, because of the nonlinear character of the equation of state (in particular the thermobaric effect), such a ‘neutral’ displacement is accompanied by a conversion between internal energy  $E$  and gravitational potential energy  $U$ , and an equal transformation between  $U$  and kinetic energy  $K$ . While there is thus no net change of  $U$ ,  $K$  does change. This implies that a force is in fact required for the displacement.

It is further shown that displacements that are orthogonal to a vector  $\mathbf{P}$  do not induce conversion between  $U$  and  $K$ , and therefore do not require a force. Analogously to neutral surfaces, which are defined to be approximately orthogonal to the dianeutral vector  $\mathbf{N}$ , one may define ‘ $\mathbf{P}$ -surfaces’ to be approximately orthogonal to  $\mathbf{P}$ . These  $\mathbf{P}$ -surfaces are intermediate between neutral surfaces and surfaces of constant  $\sigma_0$  (potential density reference to the surface).

If the equation of state is linear, there exists a well-known expression for the mixing energy in terms of the diapycnal flow. This expression is here generalised for a general nonlinear equation of state. The generalised expression involves the velocity component along  $\mathbf{P}$ , which is therefore the appropriate generalisation of the diapycnal velocity component.

# 1 Introduction

The nonlinear nature of the equation of state causes a number of difficulties for the analysis and modelling of ocean circulation. In order to deal with some of these, so-called ‘neutral surfaces’ have been introduced. They are defined so that a water parcel that is displaced adiabatically along a neutral surface always has the same density as the surrounding water (McDougall & Jackett, 1988; Eden & Willebrand, 1999). (This property cannot be satisfied exactly because of the helical nature of neutral particle trajectories, but by using a global optimization procedure surfaces can be found where it is satisfied to a very good approximation, as shown by Jackett & McDougall (1997).) Since a neutral displacement does not change the density field or the gravitational potential energy, it is generally assumed that it does not produce a restoring buoyancy force. As a consequence, it is thought that water masses primarily spread and mix along neutral surfaces.

However, this argument neglects possible conversion between internal energy  $E$  and gravitational potential energy  $U$ . In energy terms, a ‘restoring buoyancy force’ is equivalent to a conversion between kinetic energy  $K$  and  $U$ , but this conversion cannot be determined solely from the change of  $U$  if there is also a conversion between  $E$  and  $U$ .

We will here study the conversion between  $E$  and  $U$  in detail, by comparing the energetics of two different fluid models. The first model consists of the fluid equations in the Boussinesq approximation with a nonlinear equation of state, as given by Vallis (2006). In this model the buoyancy is defined as a function of potential temperature, salinity and depth. In the true equation of state, buoyancy is a function of pressure rather than depth, but by this slight modification a model that conserves energy in the adiabatic limit is obtained. (In his book, Vallis acknowledges this device as a personal communication from W.R. Young.) However, the potential energy  $H$  that appears in the energy budget of the Boussinesq model is not the same as the gravitational potential energy  $U$ ; we will call  $H$  the ‘effective potential energy’.

Since the Boussinesq model is incompressible, it does not explicitly describe the conversion between internal energy and mechanical energy. Instead, there is an external source of  $H$  which is proportional to the diabatic effects (mixing and diffusion). If  $H$  and  $U$  were equal, this external source would in reality be equal to the conversion of internal energy into potential energy.

The second model consists of the compressible form of the fluid equations, and this model explicitly describes the conversion between  $E$  and  $U$ . We will

particularly study this conversion in the ‘Boussinesq limit’ (i.e. the limit of small relative density perturbation). It turns out that this conversion is in general different from the external source of effective potential energy  $H$  in the Boussinesq model. Thus, while adiabatic particle displacements cannot change  $H$ , they may change  $U$ . Therefore,  $H$  is the relevant quantity if one wants to study irreversible mixing processes, rather than  $U$ .

A closely related problem, which was the original motivaton of this work, is how to compute the mixing energy, given the diapycnal flow. In the Boussinesq approximation with a linear and  $z$ -independent equation of state this is quite straightforward. The energy equation in this approximation is

$$\frac{\partial \rho_1}{\partial t} = -\mathbf{v} \cdot \nabla \rho_1 + \nabla \cdot \left( \kappa_h \nabla_h \rho_1 + \hat{z} \kappa_v \frac{\partial \rho_1}{\partial z} \right), \quad (1)$$

where  $\mathbf{v} = (u, v, w)$  is the three-dimensional velocity field,  $\rho_1$  is the deviation of the density from a constant background value,  $\kappa_h$  and  $\kappa_v$  are horizontal and vertical turbulent mixing coefficients, respectively, and  $\nabla_h$  is the horizontal gradient operator. Multiplying by  $gz$  and integrating over the entire volume, we obtain

$$\frac{d}{dt} \int gz \rho_1 dV = G - C. \quad (2)$$

Here, setting  $z = 0$  at the surface, the generation  $G$  of potential energy by mixing can be written

$$G = - \int g \kappa_v \frac{\partial \rho_1}{\partial z} dV, \quad (3)$$

and  $C$  is the conversion from potential energy to kinetic energy,

$$C = \int gz \mathbf{v} \cdot \nabla \rho_1 dV. \quad (4)$$

In a forced and dissipative system in steady state we have  $G = C$ , and therefore

$$G = \int gz \mathbf{v} \cdot \nabla \rho_1 dV, \quad (5)$$

which expresses the mixing energy in terms of the diapycnal flow.

A different form of the expression for the mixing energy can be obtained by introducing the buoyancy frequency  $N$ ,

$$N^2 = -\frac{g}{\rho_0} \frac{\partial \rho_1}{\partial z}. \quad (6)$$

Equation (3) can then be written

$$G = \int \rho_0 \kappa_v N^2 dV. \quad (7)$$

The expressions (7) and (5) were used by Munk and Wunsch (1998) to estimate the mixing energy necessary to sustain the overturning circulation of the deep ocean. However, these equations are based on the linear version of the equation of state, although it is known that the nonlinearity of this equation has important effects in the deep ocean.

One might think that eq. (5) could be generalized by interpreting the diapycnal velocity component as the component orthogonal to surfaces of constant potential density  $\sigma$ , i.e. to replace  $\mathbf{v} \cdot \nabla \rho_1$  by  $\mathbf{v} \cdot \nabla \sigma$ . However, there is no unique definition of the potential density, since it must be referenced to a particular depth, and it is well known that, for example, some parts of the deep ocean even appear as unstably stratified if the potential density is referenced to the surface. This is of course unphysical.

Here, eqs (7) and (5) will be generalized by using the fluid equations in the Boussinesq approximation as given by Vallis (2006). From the energy budget equations derived from this model we will identify the mixing energy.

After the submission of the present manuscript, I became aware of independent and simultaneous work on the same subject by W.R. Young (2009). Some of the present results, for example the realization that the effective potential energy of the Boussinesq model in fact includes a part of the internal energy, are also obtained by Young. However, he also obtains a deeper understanding of the connection between the Boussinesq model and the underlying thermodynamics. Appendix A summarizes some important results by Young that were missing in the original version of the present paper, written in a style adapted to the present paper. Furthermore, the notation of the present paper has partly been adapted to that used by Young (2009).

## 2 Energy conversion in a Boussinesq model with a nonlinear equation of state

We start from the fluid equations in the Boussinesq approximation as given by Vallis (2006, p. 72):

$$\frac{d\mathbf{v}}{dt} + f\hat{z} \times \mathbf{v} = -\frac{1}{\rho_0} \nabla p + b\hat{z} + \frac{\mathbf{F}}{\rho_0}, \quad (8)$$

$$\nabla \cdot \mathbf{v} = 0, \quad (9)$$

$$\frac{d\Theta}{dt} = \dot{\Theta}, \quad (10)$$

$$\frac{dS}{dt} = \dot{S}, \quad (11)$$

$$b = b(\Theta, S, z), \quad (12)$$

where

$$\frac{d}{dt} \equiv \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla.$$

Here  $\rho_0$  is a constant reference density and  $\rho_1$  the deviation from this density,  $p$  is the pressure deviation from the background pressure profile  $p_0 = -g\rho_0 z$ ,  $b = -g\rho_1/\rho_0$  is buoyancy,  $\mathbf{F}$  friction and  $S$  salinity.  $\Theta$  is the conservative temperature, as defined by McDougall (2003). He showed that it is considerably more accurate to use conservative temperature than the traditional potential temperature  $\theta$  (see also Appendix A). Diabatic effects (i.e. diffusion or mixing and external sources of heat and salt) are represented by  $\dot{\Theta}$  and  $\dot{S}$ .

For simplicity, we assume that the surface is covered by a rigid lid at  $z = 0$ , and that there is no wind stress. Boundary fluxes of salinity and heat are allowed at the surface, but not through the bottom or the lateral boundaries. The bottom topography is arbitrary.

The rigid-lid condition is used to keep the calculations as simple as possible. However, free surface conditions are more realistic, and often used in numerical models. The differences caused by a free surface are therefore described in Appendix B, both for the Boussinesq model and for the compressible model to be considered in section 3. Essentially, a new barotropic contribution appears in the expressions for the potential energy and the conversion between kinetic and potential energy. However, the main conclusions of the present paper are not affected.

There are two forms of energy in the budget equations for the Boussinesq model: kinetic energy,

$$K = \int \rho_0 \frac{|\mathbf{v}|^2}{2} dV, \quad (13)$$

and effective potential energy,

$$H = \int \rho_0 h dV. \quad (14)$$

Here,  $h$  is defined by

$$h(\Theta, S, z) = \int_z^0 b(\Theta, S, z') dz'. \quad (15)$$

This is the same as the Boussinesq dynamic enthalpy defined by Young (2009), and can be interpreted as the energy required to move a fluid parcel of unit mass adiabatically (i.e. conserving  $\Theta$  and  $S$ ) from the surface to its actual position, taking into account the fact that the buoyancy force varies with depth. Equation (15) gives

$$b = - \left( \frac{\partial h}{\partial z} \right)_{\Theta, S} \quad (16)$$

and

$$\frac{dh}{dt} = \left( \frac{\partial h}{\partial \Theta} \right)_{S, z} \frac{d\Theta}{dt} + \left( \frac{\partial h}{\partial S} \right)_{\Theta, z} \frac{dS}{dt} - bw. \quad (17)$$

To derive the budget equation for kinetic energy, we take the scalar product between  $\rho_0 \mathbf{v}$  and eq. (8), and integrate over the entire fluid volume. This gives

$$\frac{dK}{dt} = C - D, \quad (18)$$

where the conversion between effective potential energy and kinetic energy is

$$C = \int \rho_0 bw \, dV, \quad (19)$$

and the energy dissipation is

$$D = - \int \mathbf{v} \cdot \mathbf{F} \, dV. \quad (20)$$

The budget equation for effective potential energy is obtained from eqs (14) and (17):

$$\frac{dH}{dt} = G_H - C, \quad (21)$$

where the generation of effective potential energy is given by

$$G_H = \int \rho_0 \left[ \left( \frac{\partial h}{\partial \Theta} \right)_{S, z} \dot{\Theta} + \left( \frac{\partial h}{\partial S} \right)_{\Theta, z} \dot{S} \right] dV. \quad (22)$$

We can obtain a different expression for  $C$  by noting that

$$\mathbf{v} \cdot \nabla h = \mathbf{v} \cdot \mathbf{P} - bw, \quad (23)$$

where

$$\mathbf{P} = \left( \frac{\partial h}{\partial \Theta} \right)_{S, z} \nabla \Theta + \left( \frac{\partial h}{\partial S} \right)_{\Theta, z} \nabla S. \quad (24)$$

Integrating eq. (23) over the entire fluid volume we obtain

$$C = \int \rho_0 \mathbf{v} \cdot \mathbf{P} dV. \quad (25)$$

This equation is analogous to eq. (4). In steady state  $G_H = C$ . Equation (25) therefore gives the desired generalisation of eq. (5) to a nonlinear equation of state.

With a linear and  $z$ -independent equation of state, the diapycnal flow component is unambiguously defined as the component parallel to  $\nabla \rho_1$ , as in eq. (4). Equation (25) shows that the analog of this when the equation of state is nonlinear is the component parallel to the vector  $\mathbf{P}$ , defined in eq. (24). Note that this vector is not parallel to the dianeutral vector  $\mathbf{N}$ , which is commonly used to define neutral surfaces. This issue will be discussed in detail below, in section 4.

From eq. (22) we will finally derive an expression for the mixing energy that is an analog of eq. (7). We assume that the diabatic terms in eqs (10) and (11) are given in terms of turbulent mixing coefficients  $\kappa_h$  and  $\kappa_v$ :

$$\dot{\Theta} = \nabla \cdot \left( \kappa_h \nabla_h \Theta + \hat{z} \kappa_v \frac{\partial \Theta}{\partial z} \right), \quad (26)$$

$$\dot{S} = \nabla \cdot \left( \kappa_h \nabla_h S + \hat{z} \kappa_v \frac{\partial S}{\partial z} \right), \quad (27)$$

where  $\nabla_h = \hat{x} \partial_x + \hat{y} \partial_y$  is the horizontal gradient operator. Inserting these expressions into eq. (22) we obtain

$$\begin{aligned} G_H = & - \int \rho_0 \left[ \left( \kappa_h \nabla_h \Theta + \hat{z} \kappa_v \frac{\partial \Theta}{\partial z} \right) \cdot \nabla \left( \frac{\partial h}{\partial \Theta} \right)_{S,z} \right. \\ & \left. + \left( \kappa_h \nabla_h S + \hat{z} \kappa_v \frac{\partial S}{\partial z} \right) \cdot \nabla \left( \frac{\partial h}{\partial S} \right)_{\Theta,z} \right] dV. \end{aligned}$$

We rewrite the integrand by using the relation

$$\nabla \left( \frac{\partial h}{\partial \Theta} \right)_{S,z} = \frac{\partial^2 h}{\partial \Theta^2} \nabla \Theta + \frac{\partial^2 h}{\partial \Theta \partial S} \nabla S + \frac{\partial^2 h}{\partial \Theta \partial z} \hat{z},$$

and similarly for the second term. Also setting

$$N^2 = \left( \frac{\partial b}{\partial \Theta} \right)_{S,z} \frac{\partial \Theta}{\partial z} + \left( \frac{\partial b}{\partial S} \right)_{\Theta,z} \frac{\partial S}{\partial z},$$

which is a generalization of eq. (6), we obtain

$$\begin{aligned}
G_H = & \int \rho_0 \left[ \kappa_v N^2 - \left( \kappa_h (\nabla_h \Theta)^2 + \kappa_v \left( \frac{\partial \Theta}{\partial z} \right)^2 \right) \frac{\partial^2 h}{\partial \Theta^2} \right. \\
& - 2 \left( \kappa_h \nabla_h \Theta \cdot \nabla_h S + \kappa_v \frac{\partial \Theta}{\partial z} \frac{\partial S}{\partial z} \right) \frac{\partial^2 h}{\partial \Theta \partial S} \\
& \left. - \left( \kappa_h (\nabla_h S)^2 + \kappa_v \left( \frac{\partial S}{\partial z} \right)^2 \right) \frac{\partial^2 h}{\partial S^2} \right] dV \quad (28)
\end{aligned}$$

This generalizes eq. (7). Notice that all terms in the integrand except the first one vanish if  $h$  is a linear function of  $\Theta$  and  $S$ .

### 3 Energy conversion in a compressible model

In this section we will use the fundamental, compressible form of the fluid equations. This permits us to study the conversion  $M$  between internal energy and mechanical energy, in contrast to when the Boussinesq approximation has been applied. After deriving the general expression for  $M$ , we will consider the limit of small relative density perturbations (the ‘Boussinesq limit’), and compare  $M$  to the external generation of effective potential energy derived directly for the Boussinesq model in section 2.

We start from the fluid equations as given by Vallis (2006, p. 24):

$$\rho \left( \frac{d\mathbf{v}}{dt} + f \hat{z} \times \mathbf{v} \right) = -\nabla p - \hat{z} g \rho + \mathbf{F}, \quad (29)$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \rho \mathbf{v} = 0, \quad (30)$$

$$\frac{de}{dt} + \frac{p}{\rho} \nabla \cdot \mathbf{v} = \dot{Q}_T, \quad (31)$$

$$\frac{dS}{dt} = \dot{S}, \quad (32)$$

$$\frac{d\eta}{dt} = \frac{1}{T} \dot{Q}, \quad (33)$$

$$e = e(\rho, S, \eta), \quad (34)$$

$$\dot{Q}_T = \dot{Q} + \mu \dot{S}. \quad (35)$$

Here  $p$  is the full pressure,  $\rho$  density,  $\mathbf{F}$  friction force per unit volume,  $\eta$  entropy,  $e$  internal energy and  $\mu$  chemical potential (the latter three variables

defined per unit mass). Diabatic effects, i.e. diffusion and external sources, are represented by  $\dot{Q}$  (heating rate),  $\dot{Q}_T$  (total rate of energy input per unit mass) and  $\dot{S}$ . We use the same boundary conditions as for the Boussinesq model: a rigid lid at  $z = 0$ , zero wind stress, and boundary fluxes of heat and salinity only at the surface. (It would be more correct to use freshwater flux instead of salinity flux at the surface. However, this would be an unnecessary complication, since an exact representation of the surface fluxes is not crucial for the arguments to be made here.) The modifications caused by using a free surface instead of a rigid lid are summarized in Appendix B.

Taking the scalar product between  $\mathbf{v}$  and eq. (29) and multiplying eq. (30) by  $|\mathbf{v}|^2/2$ , adding the two resulting equations and integrating over the entire fluid volume, we obtain the budget equation for kinetic energy:

$$\frac{dK}{dt} = C - D, \quad (36)$$

where the kinetic energy is defined by

$$K = \int \rho \frac{|\mathbf{v}|^2}{2} dV, \quad (37)$$

the conversion between gravitational potential energy and kinetic energy by

$$C = - \int (\mathbf{v} \cdot \nabla p + g\rho w) dV, \quad (38)$$

and the energy dissipation by

$$D = - \int \mathbf{v} \cdot \mathbf{F} dV. \quad (39)$$

The budget equation for gravitational potential energy is obtained by multiplying eq. (30) by  $gz$  and integrating over the entire fluid volume:

$$\frac{dU}{dt} = M - C, \quad (40)$$

where the gravitational potential energy is

$$U = \int gz\rho dV, \quad (41)$$

and the thermodynamic work performed by the fluid is

$$M = \int p\nabla \cdot \mathbf{v} dV. \quad (42)$$

Finally, the budget equation for the internal energy  $E$  of the entire fluid is obtained by multiplying eq. (31) by  $\rho$ , eq. (30) by  $e$ , adding the resulting equations and integrating over the fluid volume:

$$\frac{dE}{dt} = G_E - M, \quad (43)$$

where the internal energy is defined by

$$E = \int e\rho dV, \quad (44)$$

and the external energy source is

$$G_E = \int \dot{Q}_T \rho dV. \quad (45)$$

We now decompose the density into a constant background value  $\rho_0$  and a deviation  $\rho_1$ ,

$$\rho = \rho_0 + \rho_1(x, y, z, t), \quad (46)$$

and assume that  $\rho_1 \ll \rho_0$ . We similarly decompose the pressure:

$$p = p_0(z) + p_1(x, y, z, t), \quad (47)$$

where  $p_0(z)$  is the pressure associated with the background density,

$$p_0 = -g\rho_0 z, \quad (48)$$

assuming that  $p_0 = 0$  at  $z = 0$ . We do not assume that the pressure perturbation  $p_1$  is hydrostatic, nor do we make the Boussinesq approximation, i.e.  $\nabla \cdot \mathbf{v}$  is nonzero (but small).

We first consider the conversion rate  $C$  in the limit  $\rho_1/\rho_0 \ll 1$ . Insert eqs (46) and (47) into eq. (38). Noting that  $\mathbf{v} \cdot \nabla p_0 + g\rho_0 w$  vanishes identically, we obtain after a partial integration:

$$C = \int (p_1 \nabla \cdot \mathbf{v} - g\rho_1 w) dV. \quad (49)$$

Estimate the magnitude of the first term. Inserting eq. (46) into eq. (30) we obtain

$$\nabla \cdot \mathbf{v} = -\frac{1}{\rho_0 + \rho_1} \left( \frac{\partial \rho_1}{\partial t} + \mathbf{v} \cdot \nabla \rho_1 \right).$$

The two terms on the right-hand side are in general of similar magnitude. Denote the horizontal and vertical length scales by  $L_h$  and  $L_v$ , and the scales

of the horizontal and vertical velocity components by  $U$  and  $W$ , respectively, and assume that  $U/L_h \sim W/L_v$ . The equation above then gives the estimate  $\nabla \cdot \mathbf{v} \sim (\rho_1/\rho_0)(W/L_v)$ . We use the fact that  $p_1$  is of the same magnitude as the hydrostatic pressure perturbation associated with  $\rho_1$ , i.e.  $p_1 \sim g\rho_1 L_v$ . (This does not mean that  $p_1$  is hydrostatic, merely that the hydrostatic contribution to  $\partial p_1/\partial z$  is not much smaller than the contribution from the vertical acceleration.) This gives the final estimate

$$p_1 \nabla \cdot \mathbf{v} \lesssim \frac{\rho_1}{\rho_0} g \rho_1 W.$$

Hence, this term can be neglected in eq. (49), and we obtain

$$C \simeq - \int g \rho_1 w \, dV.$$

This is the same as eq. (19). We conclude that the rate of conversion between  $U$  and  $K$  calculated from the compressible equations in the limit  $\rho_1/\rho_0 \rightarrow 0$  (the ‘Boussinesq limit’) is equal to the conversion rate between  $H$  and  $K$  calculated from the Boussinesq model.

We now consider the thermodynamic work  $M$  in the same limit. Using the equation of continuity and the relations (46) and (47) we obtain

$$p \nabla \cdot \mathbf{v} = -\frac{p}{\rho} \frac{d\rho}{dt} \simeq -\frac{p_0}{\rho_0} \frac{d\rho_1}{dt}.$$

Inserting this expression into eq. (42) and using eq. (48) we obtain

$$M \simeq \int g z \frac{d\rho_1}{dt} \, dV.$$

Thus, introducing the buoyancy  $b$  as in eqs (8)–(12), we conclude that the thermodynamic work in the Boussinesq limit can be written

$$M = - \int \rho_0 z \frac{db}{dt} \, dV. \quad (50)$$

Using the relation  $d(zb)/dt = wb + zdb/dt$ , this equation can be written

$$\frac{dU}{dt} = M - C,$$

where  $U$  is now defined by

$$U = - \int \rho_0 z b \, dV, \quad (51)$$

This only differs by the constant  $\int gz\rho_0 dV$  from the definition in eq. (41), and we thus have a consistent set of budget equations.

By using eqs (10) and (11), we can rewrite eq. (50):

$$M = - \int \rho_0 z \left[ \left( \frac{\partial b}{\partial \Theta} \right)_{S,z} \dot{\Theta} + \left( \frac{\partial b}{\partial S} \right)_{\Theta,z} \dot{S} + \left( \frac{\partial b}{\partial z} \right)_{\Theta,S} w \right] dV. \quad (52)$$

This relation, which gives the thermodynamic work  $M$  (i.e. the conversion between internal energy and gravitational potential energy  $U$ ) in the Boussinesq limit of the compressible model, is the main result of the present section. Note that  $M$  is in general nonzero even if the flow is adiabatic, i.e. if  $\dot{\Theta} = \dot{S} = 0$ , in contrast to the external generation  $G_H$  of effective potential energy in the Boussinesq model, given by eq. (22).

For  $M$  to vanish for adiabatic flow,  $(\partial b/\partial z)_{\Theta,S}$  must be a function of only  $z$ , i.e. we must have  $\partial^2 b/(\partial z \partial \Theta) = \partial^2 b/(\partial z \partial S) = 0$ , which means that the expansion coefficients  $(\partial b/\partial \Theta)_{S,z}$  and  $(\partial b/\partial S)_{\Theta,z}$  must be independent of  $z$ . This is not the case in seawater, because of the thermobaric effect.

## 4 Comparison between the Boussinesq model and the compressible model, and energetics of neutral flow

Above, we derived the expression (22) for the external generation  $G_H$  of effective potential energy  $H$  in the Boussinesq model, and the expression (52) for the conversion  $M$  of internal energy into gravitational potential energy  $U$  in the Boussinesq limit of the compressible model. If the equation of state does not explicitly depend on  $z$ , we have  $h = -zb$ , giving  $H = U$  and  $M = G_H$ . However, in general  $H$  and  $U$  differ, as do  $G_H$  and  $M$ . The implications of this fact will now be discussed in detail.

The budget equations for the Boussinesq model, eqs (18) and (21), and those for the compressible model, eqs (36), (40) and (43), are illustrated in the box diagrams in Fig. 1. (As explained in Appendix C, this is not the only possible box diagram for the energy transformations in the compressible model, and a different version is sometimes used.) We may also combine the two box diagrams in Fig. 1 into a single diagram, by defining

$$U_{out} = U - H.$$

The combined diagram is shown in Fig. 2. We see that  $U_{out}$  can be regarded as that part of the gravitational potential energy  $U$  which is not included

in  $H$ , and therefore lies outside the Boussinesq model. An alternative box diagram is shown in Fig. 3, where we instead defined

$$E_{in} = H - U.$$

Here  $E_{in}$  can be regarded as that part of the internal energy  $E$  which is included in  $H$ , and therefore lies inside the Boussinesq model. In the figure we also introduced  $E_{out} = E - E_{in}$ , which is that part of  $E$  which is outside the Boussinesq model. In both diagrams the Boussinesq model contains everything to the right of the flux  $G_H$ .

The box diagrams in Fig. 2 and Fig. 3 are equivalent, since they correspond to the same set of budget equations. If the sign of  $E_{in}$  were known *a priori*, this would be a reason to prefer one of them, but that does not seem to be the case. We will here use the interpretation in Fig. 3.

The effective potential energy  $H$  of the Boussinesq model represents that part of the combined internal energy and gravitational potential energy which can be converted reversibly (i.e. adiabatically) to kinetic energy. It is conceptually similar to the ‘total potential energy’ of the primitive equations for the atmosphere (Holton, 2004; Peixoto & Oort, 1992). (The primitive equations are a compressible hydrostatic model for an ideal gas.) The ‘total potential energy’ is the sum of the internal energy and the gravitational potential energy, and it can in principle all be converted reversibly to kinetic energy. The difference between ‘effective’ and ‘total’ potential energy is that the former includes only a small part of the internal energy, while the latter includes all of it. Essentially, a thermodynamic constraint prevents the conversion of most of the internal energy when the relative density perturbation is small.

We now consider the vector  $\mathbf{P}$  defined in eq. (24). The Eulerian perturbation of the temperature and salinity fields in a fixed point in space caused by an infinitesimal adiabatic displacement  $\delta\mathbf{r}$  of the fluid elements is given by

$$\delta\Theta = -\delta\mathbf{r} \cdot \nabla\Theta, \quad (53)$$

$$\delta S = -\delta\mathbf{r} \cdot \nabla S. \quad (54)$$

The resulting Eulerian perturbation of the effective potential energy field  $h$  is

$$\delta h = \left( \frac{\partial h}{\partial \Theta} \right)_{S,z} \delta\Theta + \left( \frac{\partial h}{\partial S} \right)_{\Theta,z} \delta S = -\delta\mathbf{r} \cdot \mathbf{P} \quad (55)$$

Thus, if a fluid element is displaced adiabatically (i.e. preserving its values of  $\Theta$  and  $S$ ) in a direction perpendicular to  $\mathbf{P}$ , its effective potential energy  $h$

will be the same as that of the surrounding fluid at its new location (but not the same as its own value at the original location). Also, if all displacements are perpendicular to  $\mathbf{P}$ , the perturbation of the effective potential energy of the whole fluid is zero, since, using eqs (14) and (55),

$$\delta H = - \int \rho_0 \delta \mathbf{r} \cdot \mathbf{P} dV. \quad (56)$$

Analogously, the Eulerian perturbation of the buoyancy field caused by an adiabatic displacement is

$$\delta b = -\delta \mathbf{r} \cdot \mathbf{N} \quad (57)$$

where the dianeutral vector  $\mathbf{N}$  is defined by

$$\mathbf{N} = \left( \frac{\partial b}{\partial \Theta} \right)_{S,z} \nabla \Theta + \left( \frac{\partial b}{\partial S} \right)_{\Theta,z} \nabla S. \quad (58)$$

Thus, if a fluid parcel is displaced adiabatically in a direction perpendicular to  $\mathbf{N}$ , its buoyancy (and hence density) will be the same as that of the surrounding fluid at its new location. Also, the perturbation of the gravitational potential energy of the whole fluid is zero, since

$$\delta U = \int \rho_0 z \delta \mathbf{r} \cdot \mathbf{N} dV. \quad (59)$$

A local plane that is perpendicular to  $\mathbf{N}$  is called a ‘neutral plane’ (McDougall & Jackett, 1988; 2007; Eden & Willebrand, 1999). Since displacements along such a plane do not change the buoyancy field or the gravitational potential energy  $U$ , it has been argued that they do not produce a ‘buoyant restoring force’ (McDougall, 1987a; 1987b), and that turbulent property fluxes across neutral planes are several orders of magnitude less than the lateral fluxes along such planes (McDougall, 1987a; Davis, 1994).

However, this argument ignores the conversion between internal and potential energy via thermodynamic work. It is true that displacements along neutral planes do not change the gravitational potential energy  $U$ , but that is because the thermodynamic work  $M$  exactly equals the conversion  $C$  between  $U$  and  $K$ , as seen from eq. (40) or Fig. 3. The crucial fact is that  $C$  is nonzero, which implies that a force is necessary for such ‘neutral’ displacements. (Whether this force is ‘restoring’ is another question.)

A possible conclusion is that the properties usually associated with neutral planes, which are orthogonal to  $\mathbf{N}$ , should instead be attributed to planes that are orthogonal to  $\mathbf{P}$  ( $\mathbf{P}$ -planes). This would mean that the

‘lateral’ planes, where the turbulent fluxes are maximal, coincide with  $\mathbf{P}$ -planes rather than with neutral planes. (Note that if the buoyancy only depends on  $\Theta$  and  $S$ , i.e. if the equation of state is  $z$ -independent, we have  $h = -zb$ , and hence  $\mathbf{P} = -z\mathbf{N}$ , so that the two vectors are parallel. However, in general they are not parallel, as will be seen in section 5.) The support for this conclusion comes from eq. (56), which shows that if all displacements are along  $\mathbf{P}$ -planes,  $H$  does not change, and there is therefore no conversion between  $H$  and  $K$ . No force is therefore necessary. It is true that the gravitational potential energy  $U$  is changed, but that is done by the thermodynamic work  $M$ , which converts internal energy to gravitational potential energy, without affecting the motion. Similarly, eq. (25) shows that a flow along  $\mathbf{P}$ -planes does not involve any conversion between kinetic energy and gravitational potential energy; thus, if such a flow could somehow be excited, it could continue indefinitely by inertia (if it were not for friction).

However, a serious problem with this argument is that the gradient of an arbitrary scalar field may be added to  $\mathbf{P}$  in eq. (56) without changing the result, since  $\delta\mathbf{r}$  must be divergence-free. (Changing the reference level in the definition of  $h$  is a special case of this.) Thus, it appears that the locally preferred direction of mixing cannot be determined uniquely on the basis of the integral relation (56).

Therefore, the main conclusion is that a flow along the traditional neutral surfaces involves conversion between  $H$  and  $K$ , as seen from eq. (25), and therefore requires (or generates) a force. If such a flow is to be steady, this energy conversion must be compensated by generation of effective potential energy by diapycnal mixing, as seen from eq. (22). In other words, neutral flow requires a source of mixing energy.

According to McDougall & Jackett (2005b), the best empirical evidence that the preferred direction of mixing is not simply horizontal comes from the smallness of the dissipation of mechanical energy in the ocean, together with the large observed value of lateral mesoscale mixing. To be compatible with the observed dissipation, the lateral mixing must occur along a slope that minimizes the energy dissipation. However, it is doubtful whether this argument is enough to distinguish between neutral surfaces and  $\mathbf{P}$ -surfaces.

Both the Boussinesq model and the compressible model used above are non-hydrostatic, while most general circulation models are hydrostatic. I have also derived the budget equations for the hydrostatic versions of both models. The calculations are not shown here, but the results are the same as those obtained in sections 2 and 3 for the non-hydrostatic models.

It is well known that the local neutral planes cannot be connected to global surfaces. This follows from the fact that the helicity  $H_N = \mathbf{N} \cdot \nabla \times \mathbf{N}$

of  $\mathbf{N}$  is in general nonzero (Jackett & McDougall, 1997; Eden & Willebrand, 1999; McDougall & Jackett, 2007).

We will now compare  $H_N$  with the helicity  $H_P$  of  $\mathbf{P}$ . Begin with the relation

$$\mathbf{P} = \nabla h + b\hat{z}, \quad (60)$$

and the corresponding relation for  $\mathbf{N}$ ,

$$\mathbf{N} = \nabla b - \hat{z} \left( \frac{\partial b}{\partial z} \right)_{\Theta, S}. \quad (61)$$

Taking the curl of eq. (60) and using eq. (61) we obtain

$$\nabla \times \mathbf{P} = \mathbf{N} \times \hat{z}.$$

The helicity of  $\mathbf{P}$  is then

$$H_P = \mathbf{P} \cdot \nabla \times \mathbf{P} = \hat{z} \cdot (\mathbf{P} \times \mathbf{N}). \quad (62)$$

Thus, if  $\mathbf{P}$  and  $\mathbf{N}$  are parallel, the helicity of  $\mathbf{P}$  vanishes.

We can also express  $H_P$  in the gradients  $\nabla\Theta$  and  $\nabla S$ . Inserting eqs (24) and (58) into eq. (62) we obtain

$$H_P = \left[ \left( \frac{\partial h}{\partial \Theta} \right)_{S, z} \left( \frac{\partial b}{\partial S} \right)_{\Theta, z} - \left( \frac{\partial h}{\partial S} \right)_{\Theta, z} \left( \frac{\partial b}{\partial \Theta} \right)_{S, z} \right] \hat{z} \cdot \nabla \Theta \times \nabla S. \quad (63)$$

A similar expression for  $H_N$  may also be derived. First take the curl of eq. (61):

$$\nabla \times \mathbf{N} = \hat{z} \times \nabla \left( \frac{\partial b}{\partial z} \right)_{\Theta, S}.$$

Using this relation together with eq. (61) we obtain

$$H_N = -\hat{z} \cdot \nabla b \times \nabla \left( \frac{\partial b}{\partial z} \right)_{\Theta, S}. \quad (64)$$

Express this in the gradients  $\nabla\Theta$  and  $\nabla S$ :

$$H_N = - \left[ \left( \frac{\partial b}{\partial \Theta} \right)_{S, z} \left( \frac{\partial^2 b}{\partial S \partial z} \right)_{\Theta} - \left( \frac{\partial b}{\partial S} \right)_{\Theta, z} \left( \frac{\partial^2 b}{\partial \Theta \partial z} \right)_S \right] \hat{z} \cdot \nabla \Theta \times \nabla S. \quad (65)$$

The expressions in square brackets in both eq. (63) and eq. (65) are state variables, i.e. functions of  $\Theta$ ,  $S$  and  $z$ , and they are in general nonzero.

(In section 5 it will be seen that, for a simplified equation of state, these expressions are always nonzero.) Thus,  $H_P$  and  $H_N$  vanish under the same condition: that isolines of  $\Theta$  and  $S$  coincide on all depth surfaces. From eq. (62) we conclude that if  $\mathbf{N}$  and  $\mathbf{P}$  are parallel, both helicities vanish. However, the converse is not true: even if the helicities vanish,  $\mathbf{N}$  and  $\mathbf{P}$  are in general not parallel.

We conclude that the helicity of  $\mathbf{P}$  is in general nonzero. Hence, the local  $\mathbf{P}$ -planes (as the local neutral planes) cannot be connected to global surfaces.

Although the helicity of  $\mathbf{N}$  is nonzero, it is in practice quite small, and approximate global 'neutral surfaces' can therefore be computed by minimizing a global measure of their mismatch with the local neutral planes (Jackett & McDougall, 1997). Each surface is then labeled by a neutral density  $\gamma^n$ , which is approximately equal to the potential density  $\sigma_0$  on that surface in the middle of the Pacific (Jackett & McDougall, 1997). It has been shown that neutral trajectories (which are everywhere exactly orthogonal to  $\mathbf{N}$ ) with global extent deviate from these neutral surfaces by at most a few tens of meters in the vertical direction (McDougall & Jackett, 2005a). Neutral surfaces have, for example, been used to define water masses in inverse models of the ocean circulation (Lumpkin & Speer, 2007).

Similarly, global ' $\mathbf{P}$ -surfaces' could be computed by minimizing their mismatch with the local  $\mathbf{P}$ -planes. Essentially, this would give a redefined 'neutral density'. If these surfaces and the corresponding redefined neutral density were used in an inverse model, the 'diapycnal' flow obtained could readily be used for computing the global mixing energy by using eq. (25). That is not the case with the existing inverse models, in which the neutral surfaces are approximately orthogonal to  $\mathbf{N}$ .

## 5 Application to a simple nonlinear equation of state

We will here evaluate the main expressions obtained in the previous sections using a version of the equation of state proposed by Vallis (2006). This version has a simple analytic form, while still retaining the most essential nonlinearities:

$$b = g \left[ \frac{gz}{c^2} + \beta_T(1 - \gamma z)\Theta + \frac{\beta_T^*}{2}\Theta^2 - \beta_S S \right]. \quad (66)$$

Here  $c$  is the speed of sound,  $\beta_T$  the reference value of the thermal expansion coefficient,  $\beta_S$  the haline contraction coefficient, and  $\Theta$  is the deviation

from the reference temperature 10°C. The term proportional to  $\gamma$  describes the thermobaric effect, i.e. the fact that the thermal expansion coefficient increases with increasing pressure, and the term proportional to  $\beta_T^*$  means that the thermal expansion coefficient increases with increasing temperature, which gives rise to cabbeling. These are the two main nonlinear effects of the equation of state, and they are both retained in the equation above. Numerical values of the coefficients in eq. (66) are given by Vallis (2006).

From eqs (15) and (66) we obtain the effective potential energy:

$$h = -g \left[ \frac{gz^2}{2c^2} + \beta_T \left( z - \frac{\gamma z^2}{2} \right) \Theta + \frac{\beta_T^*}{2} z \Theta^2 - z \beta_S S \right]. \quad (67)$$

This permits us to evaluate the vector  $\mathbf{P}$  in eq. (24):

$$\mathbf{P} = -g \left[ \beta_T \left( z - \frac{\gamma z^2}{2} \right) + \beta_T^* z \Theta \right] \nabla \Theta + gz \beta_S \nabla S. \quad (68)$$

The dianeutral vector  $\mathbf{N}$  in eq. (58) becomes

$$\mathbf{N} = g [\beta_T (1 - \gamma z) + \beta_T^* \Theta] \nabla \Theta - g \beta_S \nabla S. \quad (69)$$

Thus, because of the thermobaric effect (which is proportional to  $\gamma$ ),  $\mathbf{P}$  and  $\mathbf{N}$  are not parallel. The potential buoyancy referred to the sea surface is defined by

$$b_0(\Theta, S) \equiv b(\Theta, S, 0) = g \left[ \beta_T \Theta + \frac{\beta_T^*}{2} \Theta^2 - \beta_S S \right].$$

Surfaces of constant  $b_0$  are the same as the traditional  $\sigma_0$ -surfaces, i.e. surfaces of constant potential density referenced to the sea surface, and they are orthogonal to  $\nabla b_0$ :

$$\nabla b_0 = g (\beta_T + \beta_T^* \Theta) \nabla \Theta - g \beta_S \nabla S. \quad (70)$$

We find

$$\mathbf{N} = \nabla b_0 - g \beta_T \gamma z \nabla \Theta,$$

and

$$\mathbf{P} = -z \left( \nabla b_0 - \frac{1}{2} g \beta_T \gamma z \nabla \Theta \right),$$

giving

$$-\frac{\mathbf{P}}{z} = \frac{\nabla b_0 + \mathbf{N}}{2}. \quad (71)$$

Thus, within the framework of the simple equation of state (66), the angle between  $\mathbf{P}$  and  $\nabla b_0$  is exactly half the angle between  $\mathbf{N}$  and  $\nabla b_0$ . In other words, the  $\mathbf{P}$ -surfaces are intermediate between the potential density surfaces and the neutral surfaces.

McDougall & Jackett (2005a) have shown that a neutral surface and a potential density surface that are at the same depth in the North Atlantic may be separated vertically by more than 1000 m in the Southern Ocean. We conclude that a  $\mathbf{P}$ -surface which is at the same depth as those surfaces in the North Atlantic will be situated approximately mid-way between the neutral surface and the potential density surface in the Southern Ocean, i.e. it may be separated by more than 500 m from both these surfaces.

We then calculate the helicity. Evaluating eqs (63) and (65) using eqs (66) and (67) we obtain

$$H_P = \frac{1}{2}g^2\gamma\beta_T\beta_S z^2 \hat{z} \cdot \nabla\Theta \times \nabla S, \quad (72)$$

and

$$H_N = g^2\gamma\beta_T\beta_S \hat{z} \cdot \nabla\Theta \times \nabla S. \quad (73)$$

Thus,

$$H_P = \frac{z^2}{2}H_N. \quad (74)$$

We may normalize the helicities by dividing  $H_P$  by  $|\mathbf{P}|^2$  and  $H_N$  by  $|\mathbf{N}|^2$ , respectively. Noting that  $|\mathbf{P}| \approx |z\mathbf{N}|$ , we find that  $\mathbf{P}$  is half as helical as  $\mathbf{N}$ . Since the helicity of  $\nabla b_0$  is zero, this agrees with the fact that the angle between  $\mathbf{P}$  and  $\nabla b_0$  is half the angle between  $\mathbf{N}$  and  $\nabla b_0$ .

As noted in section 4, eqs (72) and (73) show that the helicities are nonzero unless the isolines of  $\Theta$  and  $S$  on constant-depth surfaces coincide. Again, the thermobaric term is the crucial one, while the cabbeling term plays no role.

We also evaluate the difference between  $U$  and  $H$ . From eqs (66) and (67) we obtain

$$zb + h = -\frac{1}{2}\gamma g\beta_T\Theta z^2.$$

Hence, eqs (14) and (51) give

$$E_{in} = H - U = -\frac{1}{2} \int \rho_0 \gamma g \beta_T \Theta z^2 dV.$$

Thus,  $E_{in}$  is nonzero (i.e.  $U$  and  $H$  differ) because of the thermobaric effect.

We finally evaluate the expression (28) for the mixing energy. Inserting eq. (67) we obtain

$$G_H = \int \rho_0 \left[ \kappa_v N^2 + g z \beta_T^* \left( \kappa_h (\nabla_h \Theta)^2 + \kappa_v \left( \frac{\partial \Theta}{\partial z} \right)^2 \right) \right] dV. \quad (75)$$

In this case the cabelling term is the important one, and it decreases the mixing energy.

## 6 Conclusion

Conversion between gravitational potential energy and internal energy is achieved through the thermodynamic work  $M$ , resulting from expansion or contraction of fluid parcels. A Boussinesq model does not describe this conversion directly, since the fluid is formally assumed to be incompressible. Instead, such a model has an external source  $G_H$  of effective potential energy  $H$  that depends on vertical diffusion. It is tempting to guess that this external energy source is in fact conversion from internal energy by thermodynamic work, i.e. that  $G_H$  and  $M$  are equal. This issue has here been investigated in detail, by calculating the energy conversion in a general compressible fluid model, and then considering the limit of small relative density perturbations (the Boussinesq limit).

It was found that if the equation of state is independent of pressure,  $M$  is indeed equal to  $G_H$  in the Boussinesq limit. However, for a general equation of state, which includes the thermobaric effect, this is no longer true: there is a mismatch between  $G_H$  and  $M$ . The explanation is that the effective potential energy  $H$  of the Boussinesq model is not the same as the gravitational potential energy  $U$ . One way to put this is to say that  $H$  also includes a part of the internal energy, as illustrated in Fig. 3. The significance of  $H$  is that it is the part of the combined internal energy and gravitational potential energy that can be converted to kinetic energy through reversible adiabatic processes. The thermodynamic origin of  $H$  has been clarified by Young (2009), who calls it ‘Boussinesq dynamic enthalpy’ (see also Appendix A).

We also derived the expression (25) for the energy conversion  $C$  between effective potential energy and kinetic energy in terms of the diapycnal flow, using a Boussinesq model with a nonlinear equation of state. In steady state,  $C$  must equal the mixing energy  $G_H$ , given by eq. (22). The conversion  $C$  is found to depend on the flow component along the vector  $\mathbf{P}$  defined in eq. (24).  $\mathbf{P}$  is not parallel to the dianeutral vector  $\mathbf{N}$  traditionally used to define

neutral surfaces, and a flow that is perpendicular to  $\mathbf{N}$  is therefore in general not perpendicular to  $\mathbf{P}$ . Thus, steady flow along neutral surfaces in general has a component along  $\mathbf{P}$ , and requires diapycnal mixing, defined as mixing that requires an external energy source. If we instead take ‘diapycnal’ to mean ‘across neutral surfaces’, the conclusion is that isopycnal mixing, i.e. mixing along neutral surfaces, in general requires an external energy source.

This surprising result is explained by the reversible conversion between internal energy and gravitational potential energy, which is ‘hidden’ in the Boussinesq model. Adiabatic displacements along neutral surfaces do not change the density field, and therefore also not the gravitational potential energy  $U$ . However, the internal energy does change, and therefore also the effective potential energy  $H$ . Since the external source  $G_H$  vanishes for adiabatic displacements, the conversion  $C$  to kinetic energy is then nonzero, as seen from eq. (21) or Fig. 1. In steady state, this conversion must be balanced by mixing energy. This mixing energy can be calculated from eq. (25), or, if the turbulent mixing coefficients are assumed to be known, from eq. (28).

These results have been obtained from non-hydrostatic model equations, but also hold for the hydrostatic versions of the models.

It is well known that it is impossible to connect the local neutral planes (which are orthogonal to  $\mathbf{N}$ ) to global surfaces, since  $\mathbf{N}$  has nonzero helicity. Neutral surfaces are therefore defined by a global optimization procedure (Jackett & McDougall, 1997), and they are only approximately orthogonal to  $\mathbf{N}$ . It was here shown that the helicity of  $\mathbf{P}$  is also nonzero, and it is therefore likewise impossible to find global surfaces that are everywhere orthogonal to  $\mathbf{P}$ . However, a similar optimization procedure as the one used to define neutral surfaces could also be used to define global ‘ $\mathbf{P}$ -surfaces’, which are approximately orthogonal to  $\mathbf{P}$ . It was shown in section 5 that such  $\mathbf{P}$ -surfaces are approximately midway between neutral surfaces and  $\sigma_0$ -surfaces (i.e. surfaces of constant potential density referenced to the sea surface). A flow along  $\mathbf{P}$ -surfaces does not cause any conversion of kinetic energy, and therefore does not require any mixing energy in steady state.

## Acknowledgements

This work was inspired by discussions with Trevor McDougall. I am also grateful to William Young and two anonymous referees for useful comments.

## Appendix A: connection between effective potential energy and dynamic enthalpy, following Young (2009)

In independent and simultaneous work, Young (2009) has clarified the thermodynamic origin of the effective potential energy  $h$ , which he calls the ‘Boussinesq dynamic enthalpy’. Some of his main results, and the connections to the present work, will here be summarized.

We start from the compressible model used in section 3, eqs (29)-(35). These equations are now rewritten by introducing the enthalpy  $\tilde{h}$ :

$$\tilde{h} = e + \alpha p, \quad (76)$$

where  $\alpha = \rho^{-1}$ . Equation (31) then gives

$$\frac{d\tilde{h}}{dt} = \alpha \frac{dp}{dt} + \dot{Q}_T. \quad (77)$$

One of the classic thermodynamic Maxwell relations reads

$$\left( \frac{\partial \tilde{h}}{\partial p} \right)_{\eta, S} = \alpha. \quad (78)$$

Integrating this, we obtain

$$\tilde{h}(p, \eta, S) = \tilde{h}(p_s, \eta, S) + \int_{p_s}^p \alpha(\eta, S, p') dp', \quad (79)$$

where  $p_s$  is the surface pressure. McDougall (2003) defined the potential enthalpy  $\tilde{h}_0$  as

$$\tilde{h}_0(\eta, S) = \tilde{h}(p_s, \eta, S), \quad (80)$$

which is the enthalpy of a fluid parcel after it has been brought adiabatically to the surface. He also defined the conservative temperature  $\Theta$  as

$$\Theta = \frac{\tilde{h}_0(\eta, S)}{C_{p_s}}, \quad (81)$$

where  $C_{p_s}$  is the heat capacity at surface pressure. Following Young (2009), the dynamic enthalpy  $\tilde{h}_d$  is defined as

$$\tilde{h}_d(p, \eta, S) = \int_{p_s}^p \alpha(\eta, S, p') dp'. \quad (82)$$

Thus, eq. (79) can be written

$$\tilde{h}(p, \eta, S) = C_{p_s} \Theta(\eta, S) + \tilde{h}_d(p, \eta, S). \quad (83)$$

Using the Maxwell relation (78) and eqs (32) and (33), we obtain

$$\frac{d\tilde{h}_d}{dt} = \alpha \frac{dp}{dt} + \frac{\partial \tilde{h}_d}{\partial \eta} \frac{\dot{Q}}{T} + \frac{\partial \tilde{h}_d}{\partial S} \dot{S} \quad (84)$$

Subtracting this equation from eq. (77) we obtain

$$C_{p_s} \frac{d\Theta}{dt} = \dot{Q}_T - \frac{\partial \tilde{h}_d}{\partial \eta} \frac{\dot{Q}}{T} - \frac{\partial \tilde{h}_d}{\partial S} \dot{S}. \quad (85)$$

Using eq. (35), this can be written

$$C_{p_s} \frac{d\Theta}{dt} = \left(1 - \frac{1}{T} \frac{\partial \tilde{h}_d}{\partial \eta}\right) \dot{Q} + \left(\mu - \frac{\partial \tilde{h}_d}{\partial S}\right) \dot{S} \approx \dot{Q}. \quad (86)$$

It was shown by McDougall (2003) that the last approximation above is highly accurate, and that the error is much larger in the corresponding approximate equation for potential temperature  $\theta$ . This is why conservative temperature  $\Theta$  should be used in the Boussinesq model in section 2.

Equation (86) readily gives a conservation law for the total heat content of the fluid:

$$\frac{d}{dt} \int C_{p_s} \Theta \rho dV = \int \dot{Q} \rho dV. \quad (87)$$

We also have the conservation law for the total energy in the compressible model, the sum of eqs. (36), (40) and (43):

$$\frac{d}{dt} \int \left( e + gz + \frac{|\mathbf{v}|^2}{2} \right) \rho dV = \int (\dot{Q}_T - \mathbf{v} \cdot \mathbf{F}) \rho dV. \quad (88)$$

Thus, there are two independent conservation laws. Instead of using the total energy, we may use the difference between total energy and heat as the second conserved quantity. However, since the internal energy dominates the total energy, the two quantities are almost equal. We therefore cannot use the approximate relation (86) when calculating the difference. Using eqs (76) and (83) we obtain the exact relation

$$e + gz + \frac{|\mathbf{v}|^2}{2} - C_{p_s} \Theta = -p\alpha + \tilde{h}_d + gz + \frac{|\mathbf{v}|^2}{2}. \quad (89)$$

Differentiating the right-hand side, and using eqs (29), (30) and (84), we obtain after some algebra:

$$\frac{d}{dt} \left( -p\alpha + \tilde{h}_d + gz + \frac{|\mathbf{v}|^2}{2} \right) = -\frac{\nabla \cdot (\mathbf{v}p)}{\rho} + \frac{\partial \tilde{h}_d}{\partial \eta} \frac{\dot{Q}}{T} + \frac{\partial \tilde{h}_d}{\partial S} \dot{S} + \frac{\mathbf{v} \cdot \mathbf{F}}{\rho}.$$

Multiplying by  $\rho$  and integrating over the fluid volume, we get the following integral relation:

$$\frac{d}{dt} \int \left[ -p + \left( \tilde{h}_d + gz + \frac{|\mathbf{v}|^2}{2} \right) \rho \right] dV = \int \left[ \left( \frac{\partial \tilde{h}_d}{\partial \eta} \frac{\dot{Q}}{T} + \frac{\partial \tilde{h}_d}{\partial S} \dot{S} \right) \rho + \mathbf{v} \cdot \mathbf{F} \right] dV \quad (90)$$

It remains to show that this relation reduces to the conservation law for total energy of the Boussinesq model in the limit of small relative density perturbation. We set

$$\alpha = \alpha_0 + \alpha_1(x, y, z, t),$$

where  $\alpha_0 = \rho_0^{-1}$  is a constant, and assume that  $\alpha_1/\alpha_0 \ll 1$ . Also set

$$p = p_0(z) + p_1(x, y, z, t),$$

where

$$p_0(z) = p_s - \rho_0 gz.$$

Inserting these expressions into eq. (82) we obtain to first order

$$\tilde{h}_d = -gz + \alpha_0 p_1 + \tilde{h}_{d1}, \quad (91)$$

where

$$\tilde{h}_{d1} = \int_{p_s}^{p_0(z)} \alpha_1(\Theta, S, p') dp'.$$

Inserting eq. (91) into eq. (90), and dropping the time-independent term  $p_0$  in the integrand, we finally obtain

$$\frac{d}{dt} \int \left( \tilde{h}_{d1} + \frac{|\mathbf{v}|^2}{2} \right) \rho_0 dV = \int \left[ \left( \frac{\partial \tilde{h}_{d1}}{\partial \eta} \frac{\dot{Q}}{T} + \frac{\partial \tilde{h}_{d1}}{\partial S} \dot{S} \right) \rho_0 + \mathbf{v} \cdot \mathbf{F} \right] dV. \quad (92)$$

This is the conservation law for total energy in the Boussinesq model. The ‘effective potential energy’  $\tilde{h}_{d1}$  is simply denoted  $h$  in the main text. It is the same as the ‘Boussinesq dynamic enthalpy’ introduced by Young (2009).

## Appendix B: energy budgets with a free surface

A free surface is more realistic than the rigid-lid condition used in sections 2 and 3. We will here indicate the modification of the energy budget equations that are incurred by the free surface condition.

For the Boussinesq model, we again use eqs (8)-(12), but now explicitly denoting the pressure perturbation  $p_1$  instead of  $p$ . The background pressure is

$$p_0 = -\rho_0 g z.$$

We define the surface to be at  $z = \zeta(x, y, t)$ . The kinematic boundary condition at the free surface is

$$\frac{d\zeta}{dt} = w \quad \text{at} \quad z = \zeta. \quad (93)$$

We also take the total pressure  $p_0 + p_1$  to be zero at the surface, which gives

$$p_1 = \rho_0 g \zeta \quad \text{at} \quad z = \zeta. \quad (94)$$

In deriving the budget equations, the following general integral relation is used:

$$\frac{d}{dt} \int_{z < \zeta} a \, dV = \int_{z < \zeta} \frac{da}{dt} \, dV,$$

where  $a(x, y, z, t)$  is an arbitrary field. On the right-hand side of this relation,  $d/dt$  as previously denotes  $\partial/\partial t + \mathbf{v} \cdot \nabla$ , while on the left-hand side  $d/dt$  is unambiguous, since the integral depends only on  $t$ . The resulting budget equations are still written as in eqs (18) and (21), and the kinetic energy  $K$  is still given by eq. (13), but with the integral taken over the region below the free surface. The effective potential energy  $H$ , on the other hand, has a new contribution from the free surface:

$$H = \int_{z < \zeta} \rho_0 h \, dV + \int \rho_0 g \frac{\zeta^2}{2} \, dx dy, \quad (95)$$

where the surface integral is taken over the free surface.

The generation of effective potential energy,  $G_H$ , is still given by eq. (22), with the integral taken over the region below the free surface. The conversion  $C$  between  $H$  and  $K$ , on the other hand, has one term associated with the interior, and one with the free surface:

$$C = C_i + C_s. \quad (96)$$

The interior term  $C_i$  is given by eq. (19). By using eq. (94) the surface term can be written as follows:

$$C_s = \int_{z < \zeta} \mathbf{v} \cdot \nabla p_1 dV = \oint \rho_0 g \zeta \mathbf{v} \cdot d\mathbf{S} = \int_{z < \zeta} \rho_0 g w dV. \quad (97)$$

Notice that the last integral above is nonzero only because of the free surface. Finally,  $C_i$  may be rewritten in terms of the diapycnal velocity by inserting eq. (23) into eq. (19):

$$C_i = \int_{z < \zeta} \rho_0 \mathbf{v} \cdot \mathbf{P} dV - \oint \rho_0 h \mathbf{v} \cdot d\mathbf{S}. \quad (98)$$

In the second term on the right-hand side we can set  $h \approx -bz$ . This term is thus a factor  $\rho_1/\rho_0$  smaller than  $C_s$ , and may be neglected. The expression for  $C_i$  then reduces to eq. (25).

For the compressible model we use eqs (29)-(35) together with the boundary conditions (93) and

$$p = 0 \quad \text{at} \quad z = \zeta. \quad (99)$$

In deriving the budget equations, the following general integral relation is used:

$$\frac{d}{dt} \int_{z < \zeta} a \rho dV = \int_{z < \zeta} \frac{da}{dt} \rho dV,$$

where  $a(x, y, z, t)$  is an arbitrary field. The resulting budget equations are written as in eqs (36), (40) and (43), and the expressions for  $K$ ,  $U$ ,  $E$ ,  $G_E$ ,  $M$ ,  $C$  and  $D$  remain the same as in section 3, with the integrals taken over the volume below the free surface.

We then decompose the density and the pressure as in eqs (46)-(48), and investigate the Boussinesq limit,  $\rho_1 \ll \rho_0$ . Using  $\int_{z < \zeta} z dV = \text{const.} + \int (\zeta^2/2) dx dy$ , the gravitational potential energy in eq. (41) can be written

$$U = \int_{z < \zeta} g z \rho_1 dV + \int \rho_0 \frac{\zeta^2}{2} dx dy,$$

where we dropped a time-independent constant. Proceeding as in section 3, it can be shown that the conversion  $C$  approaches the value given by eqs (96), (19) and (97) above. The thermodynamic work  $M$ , finally, can be shown to approach the expression given in eq. (52) in the Boussinesq limit.

To summarize: using a free surface condition results in new contributions to the gravitational potential energy  $U$ , the effective potential energy  $H$ , and to the conversion  $C$  between these energy forms and kinetic energy. However,

the internal energy  $E$ , the thermodynamic work  $M$  and the generation of effective potential energy,  $G_H$ , are not affected, and all the conclusions in the main text therefore remain the same.

## Appendix C: alternative box diagrams

The lower box diagram in Fig. 1 illustrates the budget equations (36), (40) and (43) of the compressible model. We may rewrite these equations by a simple redefinition of the fluxes:

$$F = C - M \quad (100)$$

Substituting  $F$  for  $C$  we obtain the following modified set of budget equations:

$$\frac{dK}{dt} = M + F - D, \quad (101)$$

$$\frac{dU}{dt} = -F, \quad (102)$$

$$\frac{dE}{dt} = G_E - M, \quad (103)$$

These equations are illustrated in Fig. 4, which is similar to the box diagram shown by Peixoto & Oort (1992) in their Fig. 13.1.

Although the box diagrams in Fig. 1 and Fig. 4 look very different, they are in fact equivalent. A box diagram is merely a graphical representation of a set of budget equations. It is useful because it aids the thinking about the system; for example, one can immediately write down budget equations for various subsets of the system, or for the system as a whole, by looking at the diagram.

When studying a system there is a choice between different but equivalent diagrams, obtained by redefining the fluxes as above, and one is free to choose anyone. But one should not attach any deep physical significance to the way the various boxes are connected, since that is a matter of choice. For example, the box diagram in Fig. 1 does not prove that there is no direct conversion between  $E$  and  $K$ , and Fig. 4 does not prove that there is no direct conversion between  $E$  and  $U$ .

## References

- [1] Davis, R.E., 1994: Diapycnal mixing in the ocean: equations for large-scale budgets. *J. Phys. Oceanogr.*, **24**, 777–800.
- [2] Eden, C., and J. Willebrand, 1999: Neutral density revisited. *Deep-Sea Res. II*, **46**, 33–54.
- [3] Holton, J.R., 2004: *An introduction to dynamic meteorology*. Elsevier Academic Press, 535 pp.
- [4] Jackett, D.R., and T.J. McDougall, 1997: A neutral density variable for the world’s oceans. *J. Phys. Oceanogr.*, **27**, 237–263.
- [5] Lumpkin, R., and K. Speer, 2007: Global ocean meridional overturning. *J. Phys. Oceanogr.*, **37**, 2550–2562.
- [6] McDougall, T.J., 1987a: Neutral surfaces. *J. Phys. Oceanogr.*, **17**, 1950–1964.
- [7] McDougall, T.J., 1987b: Thermobaricity, cabbeling, and water-mass conversion. *J. Geophys. Res.*, **92**, 5448–5464.
- [8] McDougall, T.J., 2003: Potential enthalpy: a conservative ocean variable for evaluating heat content and heat fluxes. *J. Phys. Oceanogr.*, **33**, 945–963.
- [9] McDougall, T.J., and D.R. Jackett, 1988: On the helical nature of neutral trajectories in the ocean. *Progress in Oceanography*, Vol. 20, Pergamon, 153–183.
- [10] McDougall, T.J., and D.R. Jackett, 2005a: An assessment of orthobaric density in the global ocean. *J. Phys. Oceanogr.*, **35**, 2054–2075.
- [11] McDougall, T.J., and D.R. Jackett, 2005b: The material derivative of neutral density. *J. Mar. Res.*, **63**, 159–185.
- [12] McDougall, T.J., and D.R. Jackett, 2007: The thinness of the ocean in  $S - \Theta - p$ -space and the implications for mean diapycnal advection. *J. Phys. Oceanogr.*, **37**, 1714–1732.
- [13] Munk, W.H., and C.I. Wunsch, 1998: Abyssal recipes. II: Energetics of tidal and wind mixing. *Deep-Sea Res. I*, **45**, 1977–2010.

- [14] Peixoto, J.P., and A.H. Oort, 1992: *Physics of Climate*. Springer, New York, 624 pp.
- [15] Vallis, G.K., 2006: *Atmospheric and oceanic fluid dynamics*. Cambridge University Press, Cambridge, 745 pp.
- [16] Young, W.R., 2009: Dynamic enthalpy, conservative temperature, and the seawater Boussinesq approximation. *J. Phys. Oceanogr.*, (submitted).

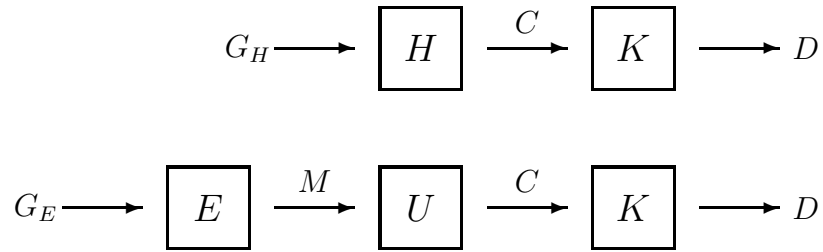


Figure 1: Illustration of the budget equations for energy. The upper diagram illustrates the incompressible Boussinesq model, eqs (18) and (21), while the lower diagram illustrates the compressible model, eqs (36), (40) and (43).  $E$  is the internal energy,  $U$  the gravitational potential energy,  $H$  the effective potential energy,  $K$  the kinetic energy,  $M$  the thermodynamic work,  $C$  the conversion between potential and kinetic energy,  $D$  the viscous dissipation, and  $G_H$  and  $G_E$  external sources of effective potential energy and internal energy, respectively.

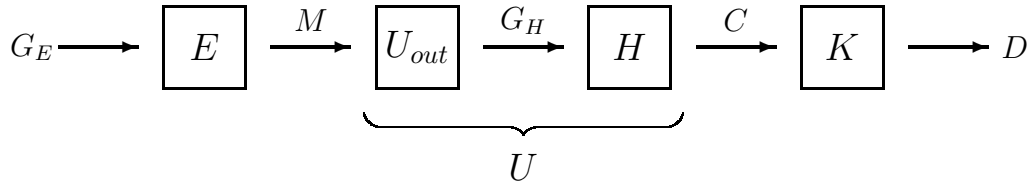


Figure 2: Combination of the two box diagrams in Fig. 1 into one diagram. Here  $U_{out} = U - H$  can be thought of as the part of the gravitational potential energy  $U$  that lies outside of the Boussinesq model.

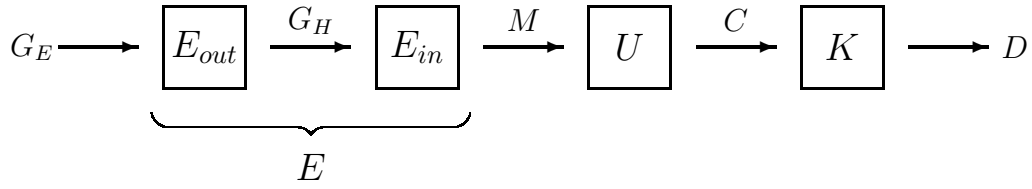


Figure 3: Alternative combination of the two box diagrams in Fig. 1 into one diagram. Here  $E_{in} = H - U$  can be thought of as that part of the internal energy which lies inside the Boussinesq model, and  $E_{out} = E - E_{in}$  as that part which lies outside it.

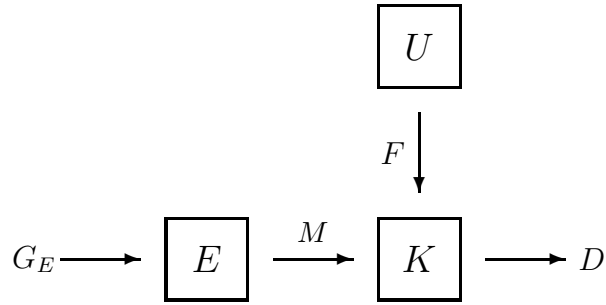


Figure 4: Box diagram for the compressible model, illustrating the redefined energy budget equations (101), (102) and (103). This diagram is equivalent to the lower diagram in Fig. 1, since it illustrates the same set of budget equations, with a trivial redefinition of the fluxes.