
Clouds as Fluids

BJORN STEVENS & A. PIER SIEBESMA

From a fluid-dynamical point of view clouds are a dilute dispersion of condensate in a multi-phase and multi-component turbulent flow. Their description thus adopts the language of thermodynamics and fluid dynamics. The multi-phase (ice-liquid-vapor) and multi-component (dry-air and water) essence of the flow complicates the thermodynamic and fluid dynamical description of fluids, which is usually presented for single component, single phase flows, as for liquid water, or simply dry air. In this chapter an overview of the main equation systems and concepts used to describe clouds and cloudy flows is presented. It is assumed that the reader has a basic background in atmospheric thermodynamics and fluid mechanics, and this text attempts to build on this background to introduce the special elements related to both topics as applied to potentially cloudy systems and as required by subsequent chapters. For readers relatively new to particular topics supplementary reading is suggested at the end to help fill in what is otherwise a dense presentation of the material.

2.1 Thermodynamics

2.1.1 A multi-component multi-phase system

The atmosphere, or air, as we experience it, is a multi-component gas in which a great variety (if not great amount) of fine-scale particulate matter is suspended. The gas phase constituents include several major gases (Nitrogen, Oxygen, Argon) which through the current era have existed in a relatively fixed proportion to one another. To a large degree these determine the thermodynamic properties of "dry air", that is an ideal mixture composed of 78.08 % N₂, 20.95 % O₂ and 0.934 % Ar, by volume. Additionally, the atmosphere contains variable vapours such as carbon dioxide and water, along with a host of seemingly minor gases (e.g., Neon, Helium, Nitrous Oxide, Ozone, Methane, Sulphur compounds, organics) some of which can be important for determining the radiative properties of the atmosphere and air quality. Of the variable constituents, water is the most striking as it ranges from abundances that vary over many many orders of magnitude, from nearly zero in the coldest regions of the upper troposphere, to as much as 4 % by volume over very warm bodies of water. Because of its proclivity to change phase and the manner in which these phase changes affect the local temperature on the one hand, and foster diverse

interactions with radiant energy on the other, water indelibly marks motions in the lower atmosphere on all time and spatial scales. It is hard to think properly about atmospheric motions, let alone clouds, without considering how water is coupled to them. In this sense the simplest, accurate description of the dynamic atmosphere requires a description that admits at least two-components, dry air and water, with one component (water) admitting multiple phases.

The basic thermodynamic properties of the atmosphere thus depend on its component parts. Typically these are defined in terms of their mass, m such that for an equilibrium system four constituents of the moist atmosphere can be defined, dry air, vapour, liquid water, and solid (ice) water, denoted by the roman subscripts d, v, l and i respectively. The total mass of the system is thus given by sum of the constituent masses such that

$$\mathcal{M} = \mathcal{M}_d + \mathcal{M}_v + \mathcal{M}_l + \mathcal{M}_i. \quad (2.1)$$

The specific mass, rather than say the mixing ratio or a molar concentration, is used to describe the amount of matter (Table 2.1). The normalised, or specific mass of a component x is denoted by $q_x = \mathcal{M}_x/\mathcal{M}$. We distinguish between equilibrium condensed phases associated with clouds, which evolve with the thermodynamic state in a more or less reversible way, and larger hydrometeors which do not. Larger hydrometeors, like rain-drops and most forms of ice, develop through irreversible microphysical processes such as the collision and coalescence of water droplets or by rapid vapour deposition in conditions of very high-supersaturation. These are more difficult to approximate as an equilibrium phase. Because they are larger, non-equilibrium phases of water in the atmosphere are also more dilute and short-lived. Their presence, which we largely ignore, requires the introduction of a more expansive view of the thermodynamic constituents within a moist atmosphere, for instance by accounting for the mass (and perhaps temperature and velocity) of rain and most forms of ice. Non-equilibrium phases of water are discussed in terms of the microphysical processes in Chpt. 3.

Molar descriptions are useful for describing chemical reactions. The use of mass mixing ratios, where the constituent masses are described in terms of their mass fraction relative to dry-air, $\mathcal{M}_x/\mathcal{M}_d$, is often adopted by descriptions which introduce water as an additional constituent. The mixing ratio approach is also helpful when considering the possibility that different components of a mass element have their own velocity. In the case of the mixing ratio the normalising mass is invariant in so long as the basic flow describes the motion

Table 2.1. *Mass Variables*

\mathcal{M}	Mass
ρ	density
q_x	specific mass of some constituent x
p	thermodynamic pressure ($p_d + p_v$)
p_x	partial pressure of some constituent x

of the dry air. Because the focus in this chapter is on a thermodynamic description for an equilibrium system, we adopt the specific humidity description.

Thermodynamic systems in local equilibrium are thus identified with mass elements, sometimes referred to as fluid- or air-parcels. Formally the concept gains validity for mass elements small enough that the volume they occupy encompasses a scale much smaller than the scale over which thermodynamic properties vary, but much larger than the mean-free path. Diffusion rapidly homogenises the atmosphere on scales smaller than the Kolmogorov length scale, $l_K = (\nu^3/\varepsilon)^{1/4}$, where ν is the viscosity of the atmosphere and ε is the turbulent dissipation rate. In vigorous cumulus clouds the dissipation rate may approach $0.05 \text{ m}^2 \text{ s}^{-3}$ which given a kinematic viscosity, $\nu = 1.5 \times 10^{-5} \text{ m}^2 \text{ s}^{-1}$ implies that variations in thermal properties are not present on scales less than $l_K = 0.5 \text{ mm}$. This is several thousand times larger than the mean free path of an air molecule, making the concept of an air parcel a useful one.

In developing the present thermodynamic description of a cloudy atmosphere we make several assumptions. The atmosphere is assumed to be comprised of dry air and water. Both the dry air and water in its vapour form are assumed to be ideal gases (non-interacting point particles) in an ideal mixture. The condensate phase is taken to be an ideal liquid, so that it is incompressible, and its volume fraction is considered to be negligibly small, so that the total volume, \mathcal{V} , of the system is seen by both the gas and vapour constituents, i.e., $v_v = \mathcal{V}/\mathcal{M}_v$ denotes the specific volume of the vapour. The specific heats of the different constituents are considered to be small, surface effects, for instance at the condensate/vapour interface are neglected, as are electromagnetic interactions.

The assumption that the condensate volume fraction is vanishingly small arises from the diluteness of clouds. Typically the specific condensate mass is less than 1 g kg^{-1} , and given the approximately thousand-fold increase in density in the condensate versus vapour phases (for typical atmospheric pressures) this implies that the volume fraction of condensate in the atmosphere is of order 10^{-6} . The diluteness of condensate can challenge the concept of an equilibrium thermodynamic system, as on the Kolmogorov scale the condensate is not continuously distributed. To get around this difficulty one often imagines an air parcel as being on the scale of around 1 m^3 . Strictly speaking volumes of air this large cannot be thought of in terms of a single temperature, but the error of this approximation is typically much less than those associated with other approximations invoked in the description of such systems.

2.1.2 A notational challenge

A particular challenge of describing moist atmospheric systems is finding a suitable notation. Many symbols are overloaded. As an example, the symbol v is often used to denote the specific volume in thermodynamic systems, or the second component of the velocity vector in fluid dynamical systems. The roman form, “v” of the same letter is used to denote, in the form of a subscript “vapour”, and sometimes “virtual”. The symbol “s” may be used to denote static energy, entropy, or in subscript form a saturated state, or process, or simply a surface quantity.

This chapter is not the proper place to attempt to systematically overhaul an overloaded notation, but some slight deviations are introduced. For instance, density temperature terminology is favoured over the more old fashioned “virtual” temperature, and script fonts are introduced for many of the quantities of classical thermodynamics, e.g., \mathcal{M} and \mathcal{V} , for mass and volume. Following the usual convention, lower case denotes a mass specific, or intensive quantity, so that v denotes the specific volume, \mathcal{V}/\mathcal{M} . An exception to this rule is in the case of the specific mass, which following conventions in the atmospheric sciences is denoted by q , rather than m . δQ , however, denotes heating. Because no special symbol for the specific heating is introduced, no conflict arises with the use of q to denote specific mass.

For entropy, the text defers to the thermodynamic tradition and adopts the symbol \mathcal{S} , or s for the specific form. Clausius is said to have adopted the symbol \mathcal{S} for entropy in honour of Carnot, a tradition this chapter upholds. Likewise the classical notation is also used to denote enthalpy by \mathcal{H} , or h for the specific form. In some later chapters s and h may, for traditional reasons, be used to represent the dry and moist static energy respectively, what in this chapter is denoted by η_d and η_e . But in these chapters there should be no risk of confusion with the entropy and enthalpy. The Gibbs potential is given by G , whose specific form g is not to be confused with g the gravitational acceleration. Following these conventions a subtle distinction arises between the script font of particular symbol and the italic font, e.g., ‘ v ’ denotes specific volume, rather than v . The text also reserves the *italic* font for variables, i.e., measures of something, such as t for time. Abbreviations, even if only one letter long, are denoted by the roman font. Hence q_t is the total water specific humidity, and the subscript ‘t’ is written in the roman font because in this usage it abbreviates total water. In an attempt to minimise subsequent confusion, the common subscript abbreviations used in this chapter, and throughout the book, are presented in Table 2.2.

2.1.3 Equation of state

Taking an air-parcel to be comprised of an ideal mixture of ideal gases, perhaps in the presence of condensate, the equation of state is that for an ideal gas of variable composition, such that

$$p = p_d + p_v = \left(\frac{\mathcal{M}_d R_d}{\mathcal{V}} + \frac{\mathcal{M}_v R_v}{\mathcal{V}} \right) T, \quad (2.2)$$

Table 2.2. *Subscript notation*

d	dry air
v	water vapour
l	liquid water
i	solid water (ice)
c	condensate
t	total water
s	saturated state, or process
e	equivalent (all condensate) reference state
ℓ	liquid-free (all vapour) reference state
p	isobaric process
0	some specified reference state

where R_d and R_v are the specific gas constants of “dry air” and water vapour respectively. Using the subscript “c” and subscript “t” to denote the total amount of condensate, and total amount of water (irrespective of phase) respectively (Table 2.2), the specific mass of total water can be expressed as

$$q_t = q_v + q_l + q_i = q_v + q_c. \quad (2.3)$$

Defining the density of the gaseous/vapour mixture as $\rho = \mathcal{M}\mathcal{V}^{-1}$ allows one to formulate the equation of state as

$$p = \rho RT \quad (2.4)$$

where the specific gas constant depends on the amount and distribution of water,

$$R = R_d + q_v R_v - q_t R_d. \quad (2.5)$$

The ratio of the gas constants often arises, albeit in different forms, which makes it helpful to define two constants that depend only on this ratio:

$$\epsilon_1 = \frac{R_d}{R_v} \approx 0.622 \quad \text{and} \quad \epsilon_2 = \left(\frac{1}{\epsilon_1} - 1 \right) \approx 0.608. \quad (2.6)$$

Many derivations are aided by simply accepting that the gas constants and specific heats depend on the composition of the fluid. The density temperature, T_ρ

$$p = \rho R_d T_\rho \quad \text{where} \quad T_\rho = T(1 + \epsilon_2 q_v - q_c) \quad (2.7)$$

makes the sensitivity to composition explicit. In the absence of water the density temperature is the air temperature, otherwise it can be interpreted as the temperature of a dry air parcel having the same density and pressure as the given air parcel.

In the older literature the density temperature is often called the virtual temperature, although some authors use both terms, distinguishing between them based on whether or not condensate effects are included. Here the term density temperature is preferred as it makes physical reference to the specific quantity that this particular temperature is meant to help describe.

Given the pressure, the density temperature determines the density and thus is important to the concept of the buoyancy, or effective acceleration, of a fluid parcel in the presence of a gravitational field. From a scale analysis of the vertical momentum equation the buoyancy of a fluid parcel can be

measured by the extent to which its density differs from a background or reference density. For instance, assume that locally the density is given in terms of a deviation from such a reference state density, ρ_0 such that

$$\rho = \rho_0 + \rho' \quad (2.8)$$

where the \prime denotes a deviation. In this case the buoyancy, b has units of acceleration and can be defined as

$$b \equiv -g \frac{\rho'}{\rho_0} \approx g \left(\frac{T'}{T_0} + \frac{R'}{R_0} \right) = g \frac{T'_\rho}{T_{\rho,0}}. \quad (2.9)$$

The approximation of the buoyancy in terms of the density temperature follows from the assumption that the relative change in pressure is small compared to the relative change in density, i.e., $p'/p_0 \ll T'/T_0$, and thus Eq. (2.7) identifies the density temperature as the dynamically relevant variable for describing fluid motions.

2.1.4 The first law and its consequences

For an atmospheric system it proves useful to use temperature and pressure to describe the state of the system. The choice of pressure rather than volume, as is more customary in the description of laboratory systems, arises because the pressure is fixed externally by the weight of the surroundings.¹

With pressure as a thermodynamic coordinate the First Law becomes

$$d\mathcal{H} = \delta Q + \mathcal{V}dp \quad (2.10)$$

where δQ denotes an infinitesimal amount of heat imparted upon the system, and \mathcal{H} is the enthalpy, or heat function. Because \mathcal{H} is an extensive variable, its specific value, h , can be written as a linear combination of the mass weighted contribution of the constituent enthalpies:

$$h = q_d h_d + q_v h_v + q_l h_l + q_i h_i. \quad (2.11)$$

The laws of thermodynamics only constrain *changes* in enthalpy, which means that the actual value of enthalpy can only be given relative to a reference value, and can thus differ depending on what one adopts for this reference value. In Eq. (2.11) the reference state enthalpies are hidden in the specific enthalpies of the various constituents.

Different phases of matter, for instance liquid water versus water vapour, differ in their specific enthalpy, so that phase changes imply a change in the enthalpy of the system, which may be used to do work or as heat lost to the environment. Hence these are often referred to as latent heats, *i.e.*, an amount of heat that is realised only through a change

¹ As a result the isobaric specific heat of some substance x is written as c_{p_x} . Dropping the subscript ‘ p ’ simplifies the notation, but may cause confusion in a few places. For example, the isobaric specific heat for vapour, c_{p_v} , can be mistaken for c_v , which is commonly used to denote an isometric specific heat, *i.e.*, one at constant volume. But as indicated above, isometric processes are rarely relevant to the atmosphere, and are therefore not introduced. Keeping this in mind should help minimise confusion. Table 2.3 provides values of the specific heats for dry air and the different phases of water.

of phase, but is otherwise latent. For water, with one crystalline (ice) phase these phase-change enthalpies are, in specific form, denoted as

$$\ell_v = h_v - h_l \quad (2.12)$$

$$\ell_l = h_l - h_s. \quad (2.13)$$

The first, the specific enthalpy of vaporisation, is the enthalpy required to vaporise a unit mass of liquid, often it is called the latent heat of condensation. The second, the specific enthalpy of fusion, is the enthalpy released by the freezing of a unit mass of liquid, and is often called the latent heat of fusion. Both are positive and their sum is the specific enthalpy of sublimation, ℓ_s . The naming convention is not entirely consistent, but follows historical usage

Usually, it is assumed that the specific enthalpies depend only on T . Actually this is a property of ideal fluids and perfect gases, the latter directly from Joule's classic free expansion experiment, so this assumption is equivalent to approximating the atmosphere as a perfect gas and condensate as an ideal fluid, which for practical purposes is a very good approximation. In this case, for some constituent x ,

$$dh_x = c_{p_x} dT, \quad (2.14)$$

where the constant c_{p_x} is called the isobaric specific heat for the constituent x in its gaseous phase. Thus it follows that

$$d\ell_v = (c_{p_v} - c_l)dT, \quad (2.15)$$

which is known as Kirchoff's relation and also holds for the other phase-change enthalpies. For the various phases of water and for dry air it is also safe to assume that c_{p_x} is constant, at least within the homosphere (below 100 km. In this case,

$$h_x(T) = h_x(T_0) + c_{p_x}(T - T_0) \quad (2.16)$$

and,

$$\ell_v(T) = \ell_v(T_0) + (c_{p_v} - c_l)(T - T_0). \quad (2.17)$$

Because c_{p_v} is less than c_l this implies that the enthalpy of vaporisation decreases with temperature, and would (if the specific heats really were temperature invariant) vanish at a critical temperature

From the above discussion it may have been inferred, that in so far as we speak of the enthalpy of a system, or some other quantity that might depend on the system enthalpy, we must specify what we have adopted for the reference state. In a system that does not admit phase changes, there is little need to take care as to what one assumes for the reference state enthalpies, and this is more of a formal requirement. If phase changes are permitted, differences in the reference enthalpies accompany changes in phase of the matter, and demand more care.

2.1.5 Enthalpies

In the literature one often encounters closed form expressions for the enthalpy, or related variables like static energies. It is often not clear that these expressions depend on an assumed reference state. Historically such expressions have been derived for liquid-gas systems. Below these expressions are derived in a manner that makes the assumed reference state

Table 2.3. *Common thermodynamic constants, where $c_{p_{m,atmrx}}$ denotes the isobaric specific heat of either dry air, vapour, liquid or ice. Isobaric specific heat capacities for water phases given at the triple point. Reference values of entropy and phase-change enthalpies are given at $T = 273.15$ K and at the standard pressure, $p_\theta = 1000$ hPa.*

Quantity	Value	Unit
R_d	0.2870	$\text{kJ kg}^{-1} \text{K}^{-1}$
R_v	0.4615	"
c_{p_d}	1.004	"
c_{p_v}	1.884	"
c_l	4.220	"
c_i	2.097	"
$s_{d,\text{ref}}$	6.783	"
$s_{v,\text{ref}}$	10.321	"
$l_{v,\text{ref}}$	2500.7	kJ kg^{-1}
$l_{f,\text{ref}}$	333.4	"

explicit, and in so doing clarifies the relationships among the different enthalpies. To align the definitions with the historical development of the subject it is also assumed that there is only one condensed phase, that corresponding to liquid. A more general treatment, to account for a solid phase, is presented in §2.1.11.1.

For a closed system $\delta q_t = \delta q_d = 0$. These constraints reduce the degrees of freedom in Eq. (2.11) and are accounted for by rewriting Eq. (2.11) in terms of q_t . To do so requires substituting $1 - q_t$ for q_d and eliminating either q_v , or q_l in favour of q_t . Eliminating q_l and making use of the definition of the phase-change enthalpies allows Eq. (2.11) to be recast as:

$$h = (1 - q_t)h_d + q_t h_l + q_v \ell_v. \quad (2.18)$$

For some (not-necessarily infinitesimal) perturbation about a reference state, the enthalpy change can thus be written as

$$\Delta h = (1 - q_t)\Delta h_d + q_t \Delta h_l + \Delta(q_v \ell_v). \quad (2.19)$$

Assuming a reference state temperature, $T_{0,e}$ such that all the condensate is in the form of liquid (which can only be approximately true), yields the following expression for the enthalpy of a reference state,

$$h_{0,e} = (1 - q_t)h_d|_{0,e} + q_t h_l|_{0,e}. \quad (2.20)$$

By fixing the (arbitrary) values of the constituent specific enthalpies at the reference temperature to be

$$h_d(T_{0,e}) = c_{p_d} T_{0,e} \quad \text{and} \quad h_l(T_{0,e}) = c_l T_{0,e}, \quad (2.21)$$

the enthalpy, $h = \Delta h + h_{0,e}$ can be written as

$$h_e = c_{p_e} T + q_v \ell_v. \quad (2.22)$$

Here the subscript 'e' has been added to the enthalpy as a reminder of the reference state with respect to which it has been defined, and

$$c_{p_e} = c_{p_d} + q_t(c_l - c_{p_d}) \quad (2.23)$$

denotes the specific heat of the system in this reference state.

Alternatively eliminating q_v in Eq. (2.11) and adopting a reference state temperature, $T_{0,\ell}$ wherein all the water is in the vapour phase with the constituent specific enthalpies fixed so that

$$\hat{h}_{0,\ell} = (1 - q_t)\hat{h}_d|_{0,\ell} + q_t\hat{h}_v|_{0,\ell} = c_{p_\ell}T_{0,\ell} \quad (2.24)$$

where

$$c_{p_\ell} = c_{p_d} + q_t(c_{p_v} - c_{p_d}), \quad (2.25)$$

yields another form for the enthalpy. It is called the liquid-water enthalpy and is denoted by subscript ℓ :

$$\hat{h}_\ell = c_{p_\ell}T - q_t\ell_v. \quad (2.26)$$

The subscript “e” or “ ℓ ” serves as a reminder of which reference state has been adopted. The former, which we call the “equivalent” reference state, is somewhat more common, both because it has a somewhat longer history and because it has some advantageous properties – although thermodynamically, particularly if the ice phase is considered, it makes less physical sense. The two enthalpies, \hat{h}_e and \hat{h}_ℓ differ from one another by a constant, reflecting their different reference states.

2.1.6 Entropy

The Second Law postulates the existence of an entropy state function, \mathcal{S} , defined by the property that in equilibrium the state of the system is that which maximises the entropy function. Such a function has the property that $\delta Q \leq Td\mathcal{S}$ with the equality sign holding for reversible transformations. For this case the First Law can be written in the form

$$d\mathcal{H} = Td\mathcal{S} + \mathcal{V}dp, \quad (2.27)$$

thereby identifying the entropy and the pressure as the independent variables in the Enthalpy formulation.

As an extensive state function the entropy, like the enthalpy can be decomposed into its constituent parts:

$$s = q_d s_d + q_v s_v + q_l s_l + q_i s_i. \quad (2.28)$$

Unlike the enthalpy the absolute entropy is not arbitrary to within a constant value. The Third Law specifies that the entropy must go to zero as T goes to zero. Hence the reference entropies cannot be arbitrarily specified. This has consequences for the description of irreversible processes.

For an ideal gas, such as dry air, the specific form of Eq. (2.27) can be integrated to yield an expression of s written in terms of a reference entropy, so that for instance,

$$s_d = s_{d,0} + c_{p_d} \ln(T/T_0) - R_d \ln(p_d/p_0), \quad (2.29)$$

where $s_{d,0}$ is the reference entropy of dry air at the temperature T_0 and pressure p_0 . As in the derivation of an expression for the enthalpy, it is assumed that the specific heats are constant between T and T_0 . An analogous expression can be derived for s_v . For the condensed phases the condensate is assumed to be ideal so that changes in pressure do not contribute to the entropy. Reference values for the entropy of dry air and water vapour at standard pressure and temperature are given in Table 2.3. Reference entropies of the condensed phases can be derived from the reference values

of the phase-change enthalpies and the reference value for water vapour.

A general expression for the composite entropy can thus be derived with respect to the chosen reference state. Here again the basic ideas are developed for a system that does not allow a solid (ice) phase. Relative to a system in an ‘equivalent’ reference state, wherein all the water mass is in the condensed phase and for which the pressure is the standard pressure, $p_\theta=1000$ hPa,

$$s = s_{e,0} + c_{p_e} \ln(T/T_0) - R_e \ln(p_d/p_\theta) + q_v(s_v - s_l), \quad (2.30)$$

with $s_{e,0} = s_{d,0} + q_t(s_{l,0} - s_{d,0})$, and c_{p_e} defined as before. The only gas-phase constituent in the reference state is dry air, hence the gas constant $R_e = (1 - q_t)R_d$. The value of $s_{e,0}$ is thus determined by the amount of water in the system and the reference state temperature and pressure, denoted by T_0 and p_θ respectively.

There are two ways to look at Eq. (2.30). Given a completely specified reference state, it provides an expression for the entropy. This was the sense in which it was derived. Alternatively, one can use this equation to ask what would the reference state temperature need to be, for the system in the reference state configuration (as specified through the pressure, amount and distribution of water mass) to have the same entropy as in the given state. In this case $s_{e,0}$ is set equal to s and Eq. (2.30) becomes an equation for T_0 conditioned on the choice of reference state. For the choice of the equivalent reference state, this alternative application of Eq. (2.30) leads to the interpretation of T_0 as the temperature the system would attain if all of its water is reversibly condensed, and then separated mechanically from the gas but maintained in thermal equilibrium with the dry air as the system is brought reversibly to the reference states pressure – a process that is easier to imagine than to realise.

In the absence of ice-processes, an expression corresponding to Eq. (2.30), but for the liquid-free reference state, follows analogously as

$$s_\ell = s_{\ell,0} + c_{p_e} \ln(T/T_0) - q_d R_d \ln(p_d/p_\theta) - q_v R_v \ln(p_v/p_\theta) - q_l(s_v - s_l), \quad (2.31)$$

with $s_{\ell,0} = s_{d,0} + q_t(s_{v,0} - s_{d,0})$. Physically the reference state temperature is that which the system would attain if reversibly brought to the reference state pressure (by convention p_θ) assuming that the vapour pressure is less than the saturation vapour pressure at this temperature, so that any condensate that may initially be in the system transforms to vapour.

2.1.7 The Clausius-Clapeyron equation

For a closed isobaric and isothermal system it follows from Eq. (2.27) that for a reversible process

$$0 = d(\mathcal{H} - T\mathcal{S}), \quad (2.32)$$

which introduces the Gibbs free-energy, or Gibbs Potential, as $G = \mathcal{H} - T\mathcal{S}$, i.e., the energy available to do work in an isothermal and isobaric system. From this definition it follows that the difference in the Gibbs energy of two constituents is related to the differences in their enthalpies and

entropies, so for example

$$g_v - g_l = h_v - h_l - T(s_v - s_l) \quad (2.33)$$

From the postulates of thermodynamics, whereby in equilibrium \mathcal{H} and T adopt values that maximise \mathcal{S} , it follows that the Gibbs free-energy of a system in equilibrium is a minimum. The condition that, for a closed isothermal and isobaric system, the Gibbs free energy is a minimum determines the partitioning between two phases of matter, say liquid water and water vapour, in equilibrium, whereby for this equilibrium partitioning the vapour state can be said to be saturated, a condition sometimes denoted by subscript ‘s’. The minimisation of the free energy determines the phase partitioning because it requires that the specific Gibbs energy of each phase must be equal, i.e., $g_v = g_l$, otherwise a redistribution of the mass between the phases could lower the total Gibbs energy. This property can be used to derive the temperature dependence of the phase partitioning as follows.

If the temperature of the system changes, this implies a change in the Gibbs energy, such that

$$dg = dh - sdT - Tds = vdp_v - sdT, \quad (2.34)$$

likewise

$$dg_v = v_v dp_v - s_v dT \quad \text{and} \quad dg_l = v_l dp_v - s_l dT. \quad (2.35)$$

But because the maintenance of equilibrium requires that $dg_v = dg_l$ it follows that for such a transformation the vapour pressure changes with temperature as,

$$dp_v = \frac{s_v - s_l}{v_v - v_l} dT. \quad (2.36)$$

This is the Clapeyron equation describing how vapour pressure changes with temperature.

Clapeyron’s equation can be cast in a simpler form by substituting for v_v from the ideal gas law in the denominator of the fraction on its rhs, and additionally noting that $v_l \ll -v_v$ so that $v_v - v_l \approx v_v$, and by rewriting the entropy differences in terms of the vaporisation enthalpy. This latter step is accomplished by realising that as far as the numerator on the rhs of Eq. (2.36) is concerned, for a saturated system $g_v = g_l$, which from Eq. (2.34) implies $s_v - s_l = \ell_v/T$. On the basis of these insights Clausius showed that Clapeyron’s equation can be written in the form

$$d(\ln p_v) \approx \frac{\ell_v}{R_v T} d(\ln T), \quad (2.37)$$

which has come to be known as the Clausius-Clapeyron equation. Fig. 2.1 shows that the Clausius-Clapeyron equation very effectively delimits the distribution of water throughout the atmosphere. The atmosphere sits atop a reservoir of water, which endeavours to bring the air above it into saturation, but if the amount of moisture exceeds the saturation value it condenses, and condensate is effectively removed by precipitation from the system. Hence the saturation specific humidity limits the amount of water in the atmosphere, as seen in Fig. 2.1. Because Eq. (10.1) so strongly controls the distribution of water in the atmosphere, if one had to single out a particular equation as being the most important for the functioning of Earth’s climate, it would be this equation.

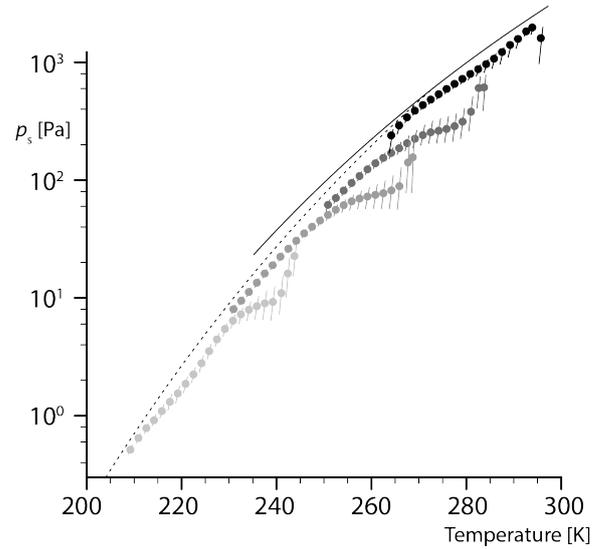


Figure 2.1 Saturation vapour pressure over liquid (solid line) and ice (dashed line). Coloured circles and lines show vapour pressure in the atmosphere, binned according to temperature for different pressure levels (900 hPa, black; 700 hPa, blue, 500 hPa, orange, 300 hPa, red). At $T = 0^\circ\text{C}$ the saturation vapour pressure is 610.15 Pa. At $T = -30^\circ\text{C}$ the saturation vapour pressure over liquid water is 50.8 Pa as compared to 38.0 Pa over ice at the same temperature. Saturation with respect to liquid for $T < 0^\circ\text{C}$ is relevant because super-cooled water is often present in the atmosphere, with homogeneous nucleation of ice particles first occurring at about $T = -38^\circ\text{C}$.

2.1.8 Potential temperatures

In the atmospheric sciences there is a tradition of using temperature variables to measure the system’s entropy. These are usually called potential temperatures as they measure the temperature the system would have to have in a given reference state, for the entropy of this state to be identical to that of the given state. It thus follows that these temperatures are invariant under an isentropic process, but their properties and absolute values depend on the choice of the reference state.

For the equivalent reference state, equating θ_e with the value of T_0 chosen so that $s_{e,0} = s$ in Eq. (2.30) implies that

$$c_{p_e} \ln \theta_e = c_{p_e} \ln T - R_e \ln(p_d/p_\theta) + q_v(s_v - s_l), \quad (2.38)$$

with $p_\theta = 1000$ hPa denoting standard pressure.

Eq. (2.38) can be recast in a more familiar form by expressing the pressure of the dry air in terms of the total pressure and the specific humidity, and by expressing the vapour-liquid entropy difference in terms of the latent heat. For the former,

$$p_d = p \left(\frac{R_e}{R} \right). \quad (2.39)$$

For the latter, by expressing the entropy difference at the end of Eq. (2.38) relative to the vapour entropy in saturation, so that

$$\begin{aligned} s_v - s_l &= s_v - s_s + s_s - s_l \\ &= s_v - s_s + (\ell_v/T). \end{aligned} \quad (2.40)$$

Here the last expression arises because the condition of phase equilibrium is the equality of the specific Gibbs enthalpies of the phases, so that $s_v - s_l = (h_v - h_l)/T = \ell_v/T$. From Eq. (2.29) the difference between the vapour and saturation vapour entropy is measured by the difference in the partial pressures

$$s_v - s_s = -R_v \ln \left(\frac{p_v}{p_s} \right) \quad (2.41)$$

where p_v/p_s defines the relative humidity. So that with Eqs. (2.39) - (2.41), Eq. (2.38) can recast as

$$\theta_e = T \left(\frac{p_\theta}{p} \right)^{\frac{R_e}{c_{pe}}} \Omega_e \exp \left(\frac{q_v \ell_v}{c_{pe} T} \right). \quad (2.42)$$

Here the term

$$\Omega_e = \left(\frac{R}{R_e} \right)^{\frac{R_e}{c_{pe}}} \left(\frac{p_v}{p_s} \right)^{\frac{-q_v R_v}{c_{pe}}} \approx 1 \quad (2.43)$$

has been introduced as a separate factor because by virtue of the smallness of q_t ($\ll 1$) it depends only very weakly on the thermodynamic state. Eq. (2.42) is a complicated expression presented in all its fullness, but it is rarely used in practical applications. For many purposes far simpler expressions captures much of the essential physics. These, and the assumptions they imply, are discussed in § 2.1.11.2.

Choosing instead T_0 so that the liquid-free reference state has the same entropy as the given state introduces the liquid-water potential temperature as

$$\theta_\ell = T \left(\frac{p_\theta}{p} \right)^{\frac{R_\ell}{c_{p\ell}}} \Omega_\ell \exp \left(-\frac{q_l \ell_v}{c_{p\ell} T} \right), \quad (2.44)$$

where

$$\Omega_\ell = \left(\frac{R}{q_d R_d} \right)^{\frac{q_d R_d}{c_{p\ell}}} \left(\frac{R}{q_v R_v} \right)^{\frac{q_v R_v}{c_{p\ell}}} \quad (2.45)$$

with $R_\ell = R_d(1 + \epsilon_2 q_t)$, $c_{p\ell}$ is given by Eq. (2.26) and once more p_θ is the standard pressure. Here too, simpler expressions that provide a reasonably good approximation to θ_ℓ are discussed in § 2.1.11.2.

In the absence of saturation a quantity resembling the familiar dry potential temperature arises

$$\theta = T \left(\frac{p_\theta}{p} \right)^{R/c_p}, \quad (2.46)$$

where R and c_p depend on composition of the fluid, and hence q_t . For the particular case of dry air, we denote θ by θ_d as a reminder that $c_p = c_{p_d}$ and $R = R_d$. The non-dimensional pressure describing the proportionality between temperature and potential temperature arises frequently and is known as the Exner function,

$$\Pi \equiv \left(\frac{p}{p_\theta} \right)^{R/c_p}. \quad (2.47)$$

Unlike θ , both θ_e and θ_ℓ are conserved under isentropic transformations of moist air in a way that accounts for the isentropic phase changes between vapour and liquid. For this reason they are often adopted as thermodynamic variables. They differ from one another in that moisture contents, particularly in the lower troposphere cause θ_e to be substantially

larger than θ , while θ_ℓ is typically only slightly less, if at all, than θ by virtue of the typically small amounts of condensate suspended in the air. These differences are most readily evident by recognising that many of the terms in Eqs.(2.42) and (2.44) only contribute small corrections to much simpler forms of the equations, so that

$$\theta < \theta_e \approx \theta \exp \left(\frac{q_v \ell_v}{c_{p_d} T} \right) \quad (2.48)$$

and

$$\theta \geq \theta_\ell \approx \theta \exp \left(\frac{-q_l \ell_v}{c_{p_d} T} \right). \quad (2.49)$$

Further discussion of the advantage of one or the other choice of potential temperature is provided at the end of §2.1.11.3

2.1.9 Static energies

In an atmosphere absent horizontal pressure gradients, and with the vertical distribution of pressure in hydrostatic balance, the change in pressure following parcel displacements follows the geopotential, ϕ , as $v dp = -d\phi$. In this case the adiabatic (literally *no heating*) form of the First Law, Eq. (2.10), becomes

$$0 = d(h + \phi). \quad (2.50)$$

This defines the static energy, $\eta = h + \phi$, as an adiabatic invariant of parcel displacements in such an atmosphere. The name arises because η measures the total energy a parcel has were it static – the kinetic energy is not accounted for. In analogy to the potential temperatures it could just as well be called the potential enthalpy, i.e., the enthalpy a parcel would have were it adiabatically brought to the surface given some specification of the reference enthalpy. The restriction that η is only conserved for an atmosphere in which pressure varies hydrostatically in the vertical, and not at all horizontally might seem restrictive, but the vertical distribution of pressure is (especially on larger scales) well approximated by hydrostatic balance and at least in the tropics horizontal pressure gradients are very small. So that in the more general setting η is very nearly conserved and is often treated as if it were conserved.

Neglecting for a moment corrections implied by the failure of the above assumptions, or the effects of an ice phase, the form of the static energy depends on which reference state the enthalpy is referred to. Adopting the equivalent reference state, in which case $h = h_e$ leads to the following definition of the static energy

$$\eta_e = c_{p_e} T + \ell_v q_v + \phi. \quad (2.51)$$

For the liquid-free reference state, $h = h_\ell$ and

$$\eta_\ell = c_{p_\ell} T - \ell_v q_l + \phi. \quad (2.52)$$

The static energy defined in terms of the equivalent reference state, η_e , is called the moist static energy; η_ℓ is called the liquid-water static energy.

Both η_e and η_ℓ are exactly conserved for ic transforms of the moist system if pressure only varies vertically, and

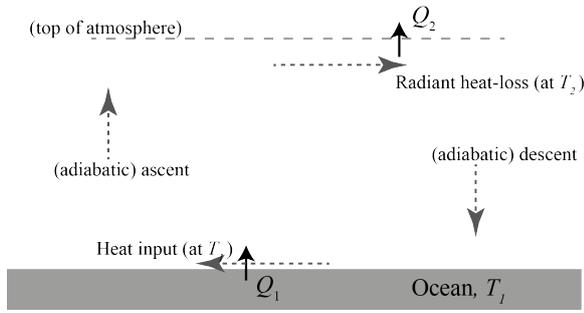


Figure 2.2 The atmosphere as a heat engine, where air is heated by an amount Q_1 over the ocean at some effective temperature T_1 and loses energy (Q_2) radiatively at some cooler temperature, T_2 . It is assumed to ascend and descend adiabatically.

then, hydrostatically. In the absence of water both become identical to the dry static energy,

$$\eta_d = c_{pd}T + \phi. \quad (2.53)$$

The names, moist static energy versus equivalent potential temperature or the liquid water static energy and liquid water potential temperature, emerged historically and not in relation to one another. Because the equivalent reference state contains all the moisture in the condensed form, a more informative terminology would be to refer to the condensation and evaporation potential temperatures (Eqs. (2.42) and (2.44)) and the condensation and evaporation potential enthalpies (Eqs. (2.51) and (2.52)) respectively.

2.1.10 Heat engines and maximum entropy production

2.1.10.1 Heat engines

Many thermodynamic processes can be interpreted in terms of a heat engine, which converts heat into work. The atmosphere is one of them. The limits the amount of work that can be done by a heat engine to that which can be done by a reversible heat engine, i.e., the Carnot Cycle. Hence the idea of heat engines arises in many places as a way to bound the work that can be done by a system.

A Carnot cycle, shown schematically in a fashion applicable to the atmospheric circulation (Fig. 2.2), is an ideal process composed of four stages. In the first stage the system extracts an amount of energy Q_1 from a reservoir at some temperature T_1 increasing the entropy by the amount $dS = Q_1/T_1$. In the second stage the system performs an amount of work, $W_{12} > 0$ on its environment without a change in entropy, until it reaches the temperature of a second reservoir at some temperature T_2 , where $T_2 < T_1$. In the third stage the system loses an amount Q_2 of energy to its environment at a temperature T_2 , decreasing its entropy by the amount Q_2/T_2 . Finally, in the fourth stage an amount W_{21} of work is done on the system to isentropically return it to its initial state at temperature T_1 . For a reversible process the entropy gained in the first stage equals what is lost in the third stage, so that $Q_2/T_2 = Q_1/T_1$. More generally, if

entropy is produced by irreversible processes, then

$$\frac{Q_1}{T_1} = \frac{Q_2}{T_2} - \Delta S_{\text{irr}}, \quad (2.54)$$

where $\Delta S_{\text{irr}} > 0$ accounts for the entropy by irreversible processes.

The net work done over the course of one cycle is given by $W = W_{1,2} - W_{2,1}$. By conservation of energy

$$W = Q_1 - Q_2 = Q_1 \left(1 - \frac{T_2}{T_1}\right) - T_2 \Delta S_{\text{irr}}. \quad (2.55)$$

The term in the parentheses defines the efficiency of the Carnot Cycle, it defines how much of the heat extracted from the warmer reservoir can be converted to work in an irreversible process. From the Second Law, the entropy production by irreversible processes is positive, so that irreversible processes further limit how much of the energy extracted from the warm reservoir can be used to do work. The atmosphere does not do work on the Earth or space, but one can imagine the system doing work by converting heating into the kinetic energy of the circulation (as schematically shown in Fig. 2.2), which is then removed from the system (either by dissipation or through conversion to another form of energy, for instance wind-power) to maintain a steady state.

More generally one can apply the question of how much work can be done to the question of the magnitude of the irreversible entropy production by a variety of processes. If no work is done, or any work that is done simply balances dissipation, and the system is stationary, $Q_2 = Q_1 = \Delta Q$, and from Eq. (2.54)

$$\Delta S_{\text{irr}} = \Delta Q \left(\frac{1}{T_2} - \frac{1}{T_1}\right). \quad (2.56)$$

An equation of this form can thus be used constrain how much entropy production the system can support. For instance, when applied to the tropical circulation one can think of energy being added to the system through surface fluxes, at $T_1 = T_{\text{sfc}}$ and energy being lost to space through radiation, and an effective radiative temperature of T_{rad} . Taking ΔQ to be about 240 W m^{-2} , $T_{\text{rad}} = 255 \text{ K}$ and $T_{\text{sfc}} = 300 \text{ K}$ determines the irreversible entropy production to be $0.14 \text{ W m}^{-2} \text{ K}^{-1}$. The challenge then becomes to associate ΔS_{irr} with specific processes in the hope that Eq. (2.56) can be used to bound the strength of these processes.

As an example, assume that the dissipation of kinetic energy, to create power for wind-turbines, was the main source of irreversible entropy production. Then the Carnot cycle could be used to bound the strength of the circulation driven by adding energy to the fluid at some temperature, T_1 , and extracting it at a temperature T_2 . All that would be needed would be a relationship between the rate of dissipation, and the strength of the circulation.

The above discussion has taken for granted that Q is given. In principle the amount of heating depends also on the circulation, which extracts energy more or less efficiently from the heat-reservoir, i.e., stronger winds increase the flux of moisture and internal energy from the surface. Such system's can arrive at stationary solutions which differ in terms of the amount of power they produce. Often it is found that nature

tends to prefer the solutions that maximise power, an idea closely related to the idea of maximum entropy production.

2.1.10.2 Entropy production by irreversible processes

The relationship between the potential temperatures and entropy becomes evident upon substitution of θ_e back into Eq. (2.30), whereby

$$\theta_e = T_0 \exp\left(\frac{s - s_{e,0}}{c_{pe}}\right), \quad (2.57)$$

wherein T_0 now adopts some fixed value which also fixes $s_{e,0}$ independently of the value of s . The differential of Eq. (2.57) is

$$c_{pe} \frac{d\theta_e}{\theta_e} = c_{pe} \frac{dT_0}{T_0} + ds - ds_{e,0} - \left(\frac{s - s_{e,0}}{c_{pe}}\right) dc_{pe}, \quad (2.58)$$

which could have just as easily been formed by taking the differential of Eq. (2.30).

For a closed system q_t is constant, hence composition dependent reference properties ($s_{e,0}, c_{pe}$), are (along with T_0) also constant, so that

$$ds = c_{pe} d(\ln \theta_e) = c_{pe} d(\ln \theta_\ell). \quad (2.59)$$

However if the relative composition of the system changes, through a change in q_t , then those reference properties, which depend on the composition of the system, also change and the above equalities no longer hold. Instead,

$$ds = c_{pe} d(\ln \theta_e) + [(c_l - c_{pd}) \ln(\theta_e/T_0) + (s_{l,0} - s_{d,0})] dq_t. \quad (2.60)$$

From this equation the importance of mixing of water substance arises both directly, through the term dq_t , and indirectly through its contributions to changes in $\ln \theta_e$.

Even in a closed system, in which no work is done at the boundaries, the approach to equilibrium, for instance by mixing air masses at different temperatures, or with different amounts of constituent matter, because it happens in a state of disequilibrium, irreversibly increases the entropy. In a system with condensate, or experiencing phase changes, evaporation in subsaturated air, or condensation in supersaturated air is also a non-equilibrium processes, and associated with an irreversible production of entropy. Also dissipative processes, for instance the dissipation of kinetic energy also is a source of entropy, and enthalpy. As it turns out, apart from radiative processes, most of the entropy production in the atmosphere is associated with the mixing of moisture, and that fraction which is associated with the dissipation of kinetic energy comes from a surprising source. Hydrometeors falling through the atmosphere reach a terminal velocity when the drag they experience balances their gravitational acceleration. The drag represents work being done on the fluid by the hydrometeors, work which is dissipated in the wake of the droplet and which is substantially larger than the dissipation of kinetic energy from large-scale fluid motions.

2.1.11 Further thoughts on thermodynamic variables

2.1.11.1 Incorporating the ice phase

In the above discussion the ice-phase was neglected because the static energies and the potential temperatures are usually defined, for historical reasons, to only incorporate the liquid phase. Neglecting the ice phase also makes it easier to understand the basic concepts underlying the construction of conserved moist thermodynamic variables such as θ_e or η_ℓ .

Once the basic idea of the moist entropies and enthalpies is clear however, it is possible to generalise their definition to incorporate the ice-phase. Doing so results in the following generalisation of Eqs (2.22) and (2.26) so that the static energies become

$$\eta_e = c_{pe}T + \ell_v q_v - \ell_f q_i + \phi, \quad (2.61)$$

$$\eta_\ell = c_{pe}T - \ell_v q_l - \ell_s q_i + \phi. \quad (2.62)$$

The presence of ice also encourages the definition of an ice-only reference state, wherein $h_{d,0} = h_{i,0} = 0$ so that for a system in thermal equilibrium

$$\eta_i = c_i T + \ell_s q_v + \ell_f q_l + \phi. \quad (2.63)$$

with $c_i = c_{pd} + q_t(c_i - c_{pd})$. This form of the moist static energy is thermodynamically more sensible than (2.61) as a reference state in which all the water exists as solid (ice) condensate is asymptotically more accessible than the ‘equivalent’ state wherein all water is in the liquid form. The reference state corresponding to (2.62) is even easier to work with as it is accessible absolutely, not just asymptotically, i.e., there are (atmosphere-like) temperatures at which all the condensate will exist as vapour, where as all of the vapour is extinguished in favour of ice only as the temperature goes to zero.

Introducing expressions for the potential temperatures that account for the ice phase is somewhat more involved. Analogous to the generalisation of the enthalpy implicit in Eq. (2.61), the expression for the entropy becomes

$$s = s_{e,0} + c_{pe} \ln(T/T_0) - R_e \ln(p_d/p_\theta) + q_v(s_v - s_l) - q_i(s_l - s_i). \quad (2.64)$$

The entropy difference between the condensed phases is calculated by assuming each is an ideal condensate, so that its entropy is independent of pressure. Integrating the entropy for each phase from its value at the triple-point temperature, T_* , and noting that at the triple point $s_l - s_i = \ell_f/T_*$, it follows that

$$s_l - s_i = \frac{\ell_f}{T_*} + (c_l - c_i) \ln\left(\frac{T}{T_*}\right). \quad (2.65)$$

Given this expression for the entropy difference between the condensed phases the expression for the generalisation of the equivalent potential temperature to account for the ice phase thus introduces an additional multiplicative factor

$$\left(\frac{T}{T_*}\right)^{-q_i \frac{(c_l - c_i)}{c_{pe}}} \exp\left(-\frac{q_i \ell_f}{c_{pe} T_*}\right). \quad (2.66)$$

in the definition of θ_e . With this additional term the ice-equivalent potential temperature becomes

$$\theta_e = T \left(\frac{p\theta}{p} \right)^{\frac{R_e}{c_{pe}}} \left(\frac{T}{T_*} \right)^{-q_i \frac{(c_1 - c_i)}{c_{pe}}} \Omega_e \exp \left(\frac{q_v \ell_v}{c_{pe} T} - \frac{q_i \ell_f}{c_{pe} T_*} \right). \quad (2.67)$$

which reduces to unity, as expected, in the absence of ice.

2.1.11.2 More approximate descriptions

The application of the Laws of thermodynamics to any real system necessarily involves approximations. The atmosphere is no exception. Even as developed up until this point a great number of approximations have been made. Not only have ideal mixtures of ideal constituents been assumed, but the condensate phase been assumed to occupy zero volume. Some assumptions have not even been mentioned, for instance the neglect of surface effects or of the effects of impurities in the condensate phase; electric dipole and magnetic moments of the matter under consideration have also been neglected.

Of the assumptions that have been made, many are tenable only in a certain range of temperatures. For example, if the specific heats were actually constant irrespective of temperature then Kirchoff's relation and the requirement that the vaporisation enthalpy be zero at absolute zero, implies that $\ell_{v,0} = (c_1 - c_{p_v})T_0$, which from Table 2.3 is clearly not the case. Variations in the specific heats with temperature are crucial to explain the value the enthalpy of vaporisation attains at ambient temperatures. For a perfect gas the specific heats are related to the degrees of freedom of a molecule over which thermal energy is equally distributed. However not all the degrees of freedom of molecules in the atmosphere are accessible at all temperatures, and even for simple molecules degrees of freedom are frozen out at lower temperatures, which explains why the specific heat for the diatomic constituents of dry air, such as N_2 and O_2 are not influenced by vibrational degrees of freedom, which only become important at much higher temperatures.

Notwithstanding the departures from exactness entailed by all of the above approximation, for many applications the book-keeping entailed by a fully two component description of the atmosphere as a mixture of water and dry air may unnecessarily complicate the physical description and obscure physical relationships. For the study of clouds and cloud processes latent heating is essential, but for many questions the influence of moisture on the specific heat (and even the gas constant) may not be important. With the assumption that $c_1 = c_{p_v} = c_{p_d}$

$$\eta_e \approx \eta_d + \ell_{v,0} q_v \quad (2.68)$$

$$\eta_\ell \approx \eta_d - \ell_{v,0} q_\ell. \quad (2.69)$$

Here the vaporisation enthalpy has been replaced by its value at the triple point consistent with the fact that in assuming $c_{p_v} = c_1$ Kirchoff's relation implies that ℓ_v is independent of temperature. For many purposes, as shall be evident through the remainder of this book (and even the later sections of this chapter). Equations (2.68) and (2.69) are preferred as they simplify the analysis and more clearly illustrate physical

ideas that don't depend on differences in the gas constants and specific heats.

Similarly simple expressions for the potential temperatures require two additional assumptions: (i) difference between the gas constants of water vapour and dry air can be neglected; (ii) moisture variations are only important in so far as they multiply the phase-change entropy. Adopting these additional assumptions leads to the the following approximations to the potential temperatures

$$\theta \approx \theta_d \quad (2.70)$$

$$\theta_e \approx \theta_d \exp \left(\frac{q_v \ell_v}{c_{p_d} T} \right) \quad (2.71)$$

$$\theta_\ell \approx \theta_d \exp \left(\frac{-q_i \ell_f}{c_{p_d} T} \right), \quad (2.72)$$

and in so doing helps highlight the close relationship between the potential temperatures and the static energies.

In addition, exact expressions for the saturation vapor pressure, over liquid or solid (ice) surfaces to not exist. This has given rise to a rich literature of approximations (as reviewed by Murphy and Koop, 2005). A formula that for many purposes strikes a good balance between accuracy and simplicity, was suggested by O Tetens in 1930, and reformulated later by Bolton (1980) as

$$p_s = 6.112 \exp \left[\frac{17.67(T - 273.15)}{T - 29.65} \right]. \quad (2.73)$$

For $238.15 \text{ K} < T < 308.15 \text{ K}$, Eq. (2.73) differs by less than 0.5% from the more accurate, albeit much more complex, formulations, reviewed by Murphy and Koop. It is similarly accurate to what one would derive by assuming that ℓ_v varied from its triple point value linearly in T following Kirchoff's relation. Similar, expressions for ice, and their relation to more complex expressions are explored further in the exercises at the end of this chapter.

Although approximations such as those outlined above are often adopted, care is warranted. For instance neglecting differences in the gas constants implies that moisture fluctuations no-longer contribute to density fluctuations, i.e., that $T_\rho = T$. In the tropical boundary layer, moisture fluctuations are much larger than temperature fluctuations and contribute to roughly half the variability in T_ρ . Deriving a consistent thermodynamic framework requires an asymptotic approach, for instance by expanding the equations about a small parameter equal to the difference between the specific heats and gas constants, so as to better appreciate at which order they contribute to one or the other expression. Such an analysis would then identify what level of thermodynamic description is necessary to describe a system for a particular application, but has yet to be developed.

2.1.11.3 Choice of thermodynamic variables

The question often arises as to which thermodynamic variables are most appropriate to adopt for the study of a given process. There is a tradition of calling these thermodynamic coordinates as they define the space within which thermodynamic processes are studied, something that is especially

evident in the diagrammatic methods reviewed in the next section. Enthalpy based variables are well suited to the treatment of mixing processes, because they are so naturally linked to the extensive variables, and except for the dissipation of kinetic energy there are no sources of enthalpy through irreversible processes in the isobaric system.

It would appear that solving for the evolution of the potential temperature, θ_e , rather than the entropy, s_e would make no difference. However mixing processes are an irreversible source of entropy production which must be explicitly accounted for if the entropy equation is being solved. These are implicitly included when the potential temperature equation is solved for. Consider the mixing between two air-masses, so that the mixed state can be denoted by an over-bar, identifying intensive properties as the (appropriate) average of the intensive properties of the two systems. The difference between averaging entropies, as compared to potential temperatures, becomes readily apparent when it is considered that $\ln \bar{\theta} \geq \bar{\ln \theta}$. The homogenisation of θ by a system that solves for θ increases the entropy (which is proportional to $\ln \theta$ of the system as one would expect for an isobaric process which conserves enthalpy. In this sense θ behaves like an enthalpy variable.

Deciding whether the condensed (equivalent), η_e or η_i or the evaporated (condensate-free) η_ℓ , reference state representation is more or less favourable, depends on what is being described. Traditionally systems in which liquid water is explicitly accounted for favour η_ℓ because it reduces to the dry static energy in the absence of condensate. Systems in which water vapour is the other thermodynamic coordinate tend to favour η_e representations, in part because η_e is, given the approximations in Eq. (2.68), roughly constant even for open systems in which precipitation converges or diverges from a parcel of air. In the more exact representations of η_e precipitation affects the value of the total water, and hence c_{pe} , and thus acts as an enthalpy source. But these effects are usually small and as a result θ_e or η_e are widely used to describe large-scale systems, particularly in so far as they involve precipitation, and θ_ℓ or η_ℓ find favour in studies of non-precipitating boundary layer clouds, as discussed in Chapter 5.

2.2 Thermodynamic diagrams

Thermodynamic diagrams are used to represent the state of a system, as well as thermodynamic processes. An examples of such a process might be how temperature changes as air rises in the absence of heating, i.e., *adiabatically* with or without condensational processes. Such diagrams can say much about the state of the atmosphere and how it might have come into being as they can quickly convey a wide range of quantitative information to the trained eye. This information can be helpful in determining the subsequent evolution of the atmosphere and is routinely used by weather forecasters to evaluate the likelihood of different events, ranging from fog formation to the energy available for deep convective overturning.

Many of these diagrams predate the widespread use of electronic computers. Even though today it is possible to compute many quantities directly, thermodynamic diagrams are still very useful for encapsulating the state of the atmosphere and remain widely used. The most common thermodynamic diagram is the Skew- T diagram, and its cousin the Tephigram. Other diagrams include the Clapeyron, the Emagram or Neuhoff, and the Stüve diagram. Less familiar, but also useful is the Paluch diagram, and moist-static energy diagrams.

One difficulty that all diagrams share is that they are two dimensional, and the most compact description of the state of the atmosphere encompasses three dimensions, for instance, $\{T, p, q_t\}$. Two ways have been devised for getting around this problem. One is to recast moisture as a temperature variable, as is done in constructing the dew-point temperature. The other is to link moisture to temperature and pressure by a saturation assumption. Examples of both approaches are evident through this book.

2.2.1 Skew- T and related diagrams

One of the most widely used thermodynamic diagrams is the Skew- T . It adopts temperature, T , and $\ln p$ as its thermodynamic coordinates. The logarithm of pressure is chosen for the vertical coordinate rather than the pressure itself because in an isothermal atmosphere height varies with $\ln p$, and hence for a realistic temperature profile the ordinate is roughly proportional to height. Isotherms are skewed at an angle of about 45° from the vertical, so T varies along vertical lines with $\ln p$. The exact angle of skewness is chosen so that the isentropes of a dry atmosphere and isotherms are orthogonal at 1000 hPa and 0°C , as illustrated in Fig. 2.3. Skewing the isotherms in this fashion thus better differentiates isotherms from isentropes.

Closely related to the Skew- T , and somewhat predating it, is the Tephigram, literally the T φ gram, where φ was originally used to denote potential temperature. Hence in the Tephigram the coordinate system is orthogonal, with T measured by the abscissa and $\ln \theta_d$ by the ordinate. Usually Tephigrams are right-rotated so that the ordinate becomes roughly proportional to $\ln p$ and hence height. This follows because along a dry adiabat the First Law dictates that

$$\ln \theta - \ln T = -\frac{R_d}{c_p} \ln p + \text{const.} \quad (2.74)$$

Hence associating the y -axis with $\ln \theta$ and the x -axis with T , isobars can be shown to satisfy an equation of the form $y = \ln x + C$. At temperatures of practical interest (which, measured in Kelvin, are large) the curvature in the isobars (which vary with $\ln T$) in such a space is small, so that if the diagram is appropriately rotated isobars become approximately horizontal. Hence Tephigrams and Skew- T diagrams are very similar, the chief difference being that isobars are slightly curved on the former and adiabats are slightly curved on the latter.

Both the Skew- T and the Tephigram are derived from the Emagram, which uses an orthogonal T - $\ln p$ coordinate system. The Emagram is thought to be the first thermodynamic

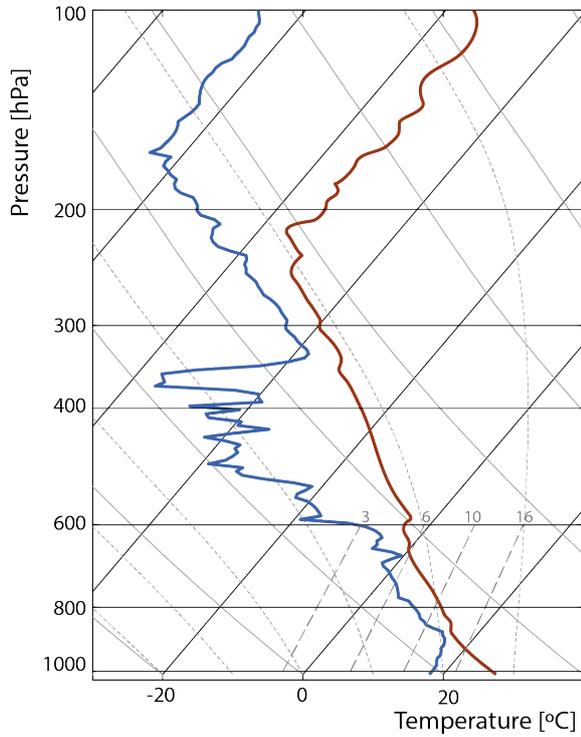


Figure 2.3 Skew- $T \ln p$ Diagram: saturated pseudo-adiabats (grey, short dashed) are almost perpendicular to the abscissa near the surface and curve to become parallel to dry adiabats (grey solid) as p decreases. Equisaturation lines are shown by long dashes, in grey. Plotted is the July 3, 2009 sounding from Cabauw in the Netherlands, prior to a period of severe weather. The blue line denotes the locus of dew-point temperatures measured at different pressures as the sounding rose through the atmosphere, the blue line the temperature.

diagram to be routinely used to visualise and quantify the state of the atmosphere. It was proposed in the late 19th century by Heinrich Hertz who is mostly known for his contributions to understanding electromagnetic radiation.

Isobars on a Tephigram and the isentropes of a dry atmosphere on a Skew- T diagram are examples of fundamental lines. Fundamental lines are isolines whose shape is decided by the thermodynamic coordinates. Other examples of fundamental lines include equisaturation curves and isotherms. In a dry atmosphere, adiabats and isentropes are used interchangeably to describe a line of constant θ_d . Through a coordinate transform the fundamental lines of one thermodynamic diagram can serve as the thermodynamic coordinates of another. The relative orientation of the fundamental lines on the Skew- T diagram help visualise how different processes are related to one another, and are evident upon a closer inspection of Fig. 2.3. Isobars are horizontal lines, which decrease logarithmically upwards so that the relation between height and distance along the ordinate is nearly linear. The isotherms are straight lines, slanted at roughly 45° toward the upper right, and are plotted in 20 K bands. Dry adiabats (lines of constant θ_d) are thin grey lines with negative slope and slight concavity, which intersect the isotherms nearly perpendicularly. Pseudo-adiabats, which we define below,

Table 2.4. *Frequently Used Temperatures*

	Name	Equation
T	Temperature	
θ	Potential T	(2.46)
T_ρ	Density T	(2.7)
θ_ρ	Density θ	(2.46) with T_ρ
θ_e	Equivalent θ	(2.42); (2.71) as approx.
θ_s	Saturation θ_e	θ_e with $q_v = q_s$, $q_l = 0$
θ_ℓ	Liquid-Water θ	(2.44); (2.72) as approx.
$\theta_{\ell,\rho}$	Density θ_ℓ	(2.44) with T_ρ

are thin dotted curves that are approximately vertical at high pressure and temperatures, but curve to the left and become nearly parallel to dry adiabats at cold temperatures. Equisaturation lines, describing how the dew-point temperature (defined below) varies with $\ln p$ and T given constant specific humidity, are marked by short dashed-lines, and inclined to the left of the isotherms.

To visualise how moisture is distributed in the atmosphere, on a Skew- T diagram one plots the dew-point of the air, as is shown by the blue line in Fig. 2.3. The dew-point temperature is the temperature air would have if it were cooled to the point of saturation, and as such it depends only on the moisture content of the air and the ambient pressure. If the dew point temperature equals the temperature, this implies that the air is saturated, $p_v = p_s$. Thus the dew-point depression, measured as the difference between the temperature and the dew-point temperature, measures the relative humidity, so that in Fig. 2.3 the air is relatively dry between 500 hPa and 350 hPa, and nearly saturated at around 900 hPa and again near 680 hPa. For an adiabatic process the dew point temperature follows an equisaturation curve. Hence the close alignment between the dew-point temperature and lines of equisaturation along with the alignment of temperature with lines of constant potential temperature is a signature of a layer that is well mixed by turbulence, as for instance is the case near the surface (lower 150 hPa) in Fig. 2.3.

Another useful way to visualise the state of the atmosphere is in terms of θ_e and its value if the atmosphere were saturated at the same temperature, plotted versus height, or temperature, in the atmosphere. Using temperature as a vertical coordinate is particularly helpful for climate change studies because the top of the troposphere more closely maintains a constant temperature than it does a constant height. The value that θ_e would adopt were its specific humidity set equal to the saturation specific humidity, without changing the temperature or pressure, is called the saturation equivalent potential temperature. It is denoted by θ_s and defined as

$$\theta_s = T \left(\frac{p\theta}{p} \right)^{\frac{R_s}{c_s}} \Omega_s \exp \left(\frac{q_s \ell_v}{c_s T} \right). \quad (2.75)$$

with

$$\Omega_s = \left[1 + \frac{q_s R_v}{R_s} \right]^{\frac{R_s}{c_s}}. \quad (2.76)$$

where $c_s = c_{p_d} + q_s(c_1 - c_{p_d})$ and $R_s = (1 - q_s)R_d$. The saturation equivalent potential temperature adds yet another temperature to the bewildering array of temperatures adopted by meteorologists (Table 2.4). Because q_s depends only on T and p , lines of constant θ_s are fundamental lines on a Skew- T , as well as the other thermodynamic diagrams discussed above, and define the pseudo-adiabat. The word pseudo arises because formally the process corresponding to constant θ_s is similar to a reversible adiabat, but the removal of condensate upon condensation, as implied by the use of q_s instead of q_t implies a loss of condensate enthalpy by the system, hence it is not truly adiabatic. The difference between θ_s and θ_e measures the subsaturation.

2.2.2 Isentropes, adiabats, and lapse rates

For the moist system, irrespective of whether the moisture condenses, the state vector is three dimensional. In addition to the thermodynamic variables required to describe the dry system, an additional variable is required to specify the composition of the system, q_t for instance. An assumption that constrain the third degree of freedom is thus required to render the state of the system as a point in a two dimensional space. This third degree of freedom also leads to a distinction between isentropes and adiabats. In a dry system, where the composition of the system is fixed, an adiabatic process is an isentropic process. In a moist system, where the composition of the system weights the different contribution of water versus dry-air to the system entropy, isentropes and adiabats are not the same. In a dry atmosphere the loss of mass does not change its specific entropy, the loss of condensate from a moist system does.

In the dry system, isentropes (or adiabats) are described by lines of constant potential temperature, θ_d . In the moist system isentropes are described by lines of constant θ_e . The rate at which the temperature falls off (or lapses), as a parcel is lifted to lower pressure, or greater altitude, is called the lapse rate. In a dry atmosphere, the decrease of temperature with height along a line of constant θ_d is called dry adiabatic lapse rate, Γ_d . From the enthalpy form of the First Law, for an adiabatic process,

$$0 = d\mathcal{h}_d - vdp. \quad (2.77)$$

For hydrostatic changes in pressure, and given the definition of the dry enthalpy, this implies that $c_{p_d}dT + gdz = 0$ which yields the dry adiabatic lapse rate as

$$\Gamma_d \equiv -\frac{dT}{dz} = \frac{g}{c_{p_d}} \quad (2.78)$$

In a moist, but *unsaturated*, atmosphere an analogous process yields a slightly modified lapse rate of g/c_{p_e} . But because c_{p_e} depends on q_t the lapse rate depends on the composition of the system, and its derivation assumed that c_{p_e} is constant, i.e., the system is closed. It thus is best described as the unsaturated moist isentropic lapse rate.

The *saturated* moist isentropic lapse rate is derived similarly, but starting from the form of the first law valid for a saturated system,

$$0 = d\mathcal{h}_e - vdp \quad \text{where} \quad \mathcal{h}_e = c_{p_e}T + \ell_v q_s \quad (2.79)$$

where the expression for \mathcal{h}_e is taken from Eq. (2.22). One could have equivalently started from Eq. (2.38), and assumed that θ_e is constant, but doing so makes the derivation of the lapse-rate more cumbersome. For an isentropic process c_{p_e} is constant, and the main complication arises from the term involving the saturation specific humidity, q_s , as

$$q_s = \frac{p_s}{p} \left(\frac{\epsilon_1}{1 - (1 - \epsilon_1)\frac{p_s}{p}} \right). \quad (2.80)$$

Irrespective of the term in parentheses, which is only a minor correction, q_s depends on both p and (through p_s) T . To simplify the notation let's define the partial derivatives

$$\beta_p \equiv -\frac{\partial \ln q_s}{\partial \ln p} = (1 + \epsilon_2 q_s) \approx 1, \quad (2.81)$$

$$\beta_T \equiv \frac{\partial \ln q_s}{\partial \ln T} = \frac{\ell_v}{R_v T} \beta_p \approx \frac{5400 \text{ K}}{T}, \quad (2.82)$$

Adapting this notation to Eq. (2.79) allows us to write the isentropic form of the first law as

$$\left[c_p + \ell_v \left(q_s \frac{\beta_T}{T} \right) \right] dT - \left(1 + q_s \frac{\beta_p}{RT} \right) vdp = 0. \quad (2.83)$$

where $c_p \equiv c_{p_d} q_d + c_{p_v} q_s + c_1 q_l$ is the specific heat for the composite system, which we have assumed to be saturated. Substituting with the hydrostatic relation for vdp above, the lapse rate (or decrease of temperature with height) for a saturated isentropic process follows as

$$\Gamma_s \equiv -\frac{dT}{dz} \Big|_{\theta_e} = \gamma \Gamma_d \quad (2.84)$$

with

$$\gamma \equiv \frac{c_{p_d}}{c_p} \left[\frac{1 + q_s \beta_T \left(\frac{R_v}{R} \right)}{1 + q_s \beta_T \left(\frac{\ell_v}{c_p T} \right)} \right]. \quad (2.85)$$

In deriving Eq. (2.85), β_T is expressed in terms of β_p using Eq. (2.82), and fusion enthalpy has been neglected. The saturated moist isentropic lapse rate (or saturated isentrope) thus depends on q_t as well as T and p .

The dimensionless lapse rate γ measures the relative role of internal energy as compared to the vaporisation enthalpy in doing the work required to lift a parcel. It decreases with temperature from a value very near unity at the colder temperatures of the upper troposphere, where little water is available to condense, to a value near 0.4 at 300 K (Fig. 2.4), with an inflection point near 280 K.

In discussing lapse rates, terminology can be confusing. Γ_d is called the dry adiabatic lapse rate, i.e., the lapse rate of dry air following an adiabatic process. In contrast Γ_s is the saturated isentropic lapse rate: 'isentropic' because its derivation additionally assumes that mass was conserved, so that q_t is constant, which is more stringent than simply assuming that no heat is added; and 'saturated' because it was assumed that $q_t \geq q_s$ to that $q_v = q_s$. To contrast with the dry adiabat it is tempting and common to call Γ_s the moist adiabat, or reversible moist adiabat. The former is common,² but can be confusing because the moist system need not be

² There is a certain tradition of imprecision in the usage of the phrase "moist adiabat" as it is mostly intended to draw distinction

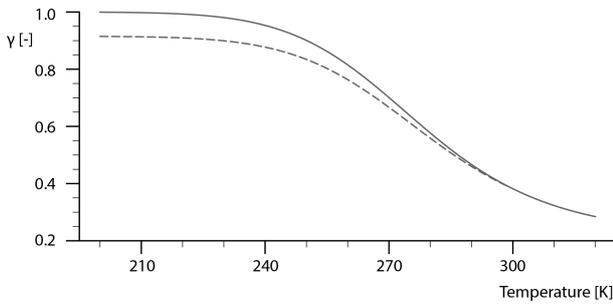


Figure 2.4 Non-dimensional lapse-rate, γ , following Eq. 2.85 for a fixed atmospheric pressure of 1000 hPa. The dashed curve shows the pseudo-adiabat (for which it is assumed that condensed water is removed) and the solid curve the saturated isentropic (or reversible moist adiabat). The pseudo-adiabat asymptotes to a value less than unity because of the specific heat of condensed water, whose contribution depends on the temperature of the air at initial saturation, here taken as 300 K.

saturated, and an adiabatic process need not be isentropic. The latter is more precise, but uses two words ('reversible' and 'adiabatic') when one ('isentropic') would work just fine. Given that systems are in general moist, it would be both simpler and more sensible to talk about saturated and unsaturated isentropes.

For large parcel displacements an increasing fraction of the initial vapour will be converted to condensate, and the assumption that this condensate remains in the parcel becomes increasingly untenable. A better approximation would be to adjust Γ_s , at every temperature and pressure, to that which it would have if the system were just saturated, i.e., $q_t = q_s$. This is equivalent to assuming that condensate is immediately lost as precipitation, hence its enthalpy content can no longer contribute to the expansional work done by the rising parcel, which leads to a greater change in temperature with decreasing pressure, i.e. a larger lapse rate. Such a process is irreversible and non-adiabatic and so the resultant lapse-rate, denoted $\Gamma_{\bar{s}}$, is often called the pseudo-adiabat, see Fig. 2.4. A straightforward calculation of the difference between pseudo-adiabats and saturated isentropes shows that they differ most appreciably in the upper troposphere. This is to be expected because T decreases with height, hence q_s decreases and q_1 (whose effects are neglected in the pseudo-adiabat) increases correspondingly. Overall the differences in T tend to be less than 0.5 K below 400 hPa increasing to as much as 3 K to 5 K between 100 hPa and 200 hPa.

One advantage of the pseudo-adiabat is that it depends only on T and p and is representable as a fundamental line on a standard atmospheric thermodynamic diagram. Additionally, it more naturally accommodates an ice phase. Whereas accounting for freezing (which was not done above) would lead to a discontinuity in the saturated isentrope at the triple point temperature, the pseudo-adiabat is only discontinuous in its derivative at this point. Nonetheless, the saturated

to a dry adiabat. In this sense its usage (also elsewhere in this book) is agnostic to fine distinctions arising from whether the process is strictly reversible, the fineness of the thermodynamic approximation, or how precisely, if at all, ice-phase processes are accounted for.

isentropes (accounting for ice formation, and thus computed numerically rather than with Eq. (2.85)) defines a thermodynamic limit – following any adiabatic process an air-parcel rising to a given altitude must be colder than its saturated isentropic value. In the lower troposphere deviations from the saturated isentrope are most likely to be associated with mixing processes, simply by virtue that the absolute humidity differences between ascending parcels and their environment are largest there. In the upper troposphere precipitation and the retardation of ice formation contribute the most to differences between rising parcels and their saturated isentropic value. An interesting footnote is that the temperature dependence of γ , which is important for climate change, is greatest at temperatures around 280 K, which are characteristic of the present day lower tropical troposphere.

2.2.3 Soundings

Soundings in the atmosphere are measurements of its state as a function of altitude. Generally they are made with sensors lofted by a balloon, or dropped with a parachute. In either case the sensor package drifts with the mean wind as it rises and uses global positioning systems (GPS) to measure the wind vector, and *in situ* sensors to measure temperature and relative humidity along its trajectory. In some special cases also other quantities, e.g., ozone. An example of a standard meteorological sounding is shown in Fig. 2.3, taken from a summer day in the Netherlands, foreshadowing severe weather. Plotted are two lines, the red, right-most, line demarcates the temperature, and the blue, left-most, line demarcates the dew-point temperature. Sometimes wind-vectors are plotted alongside the sounding, but not in the present case.

A variety of thermodynamic processes can be inferred from a sounding. For instance, well-mixed layers (or dry-adiabatic processes) will have the dew-point follow the equisaturation line, and temperature will, in the absence of condensation, follow the dry adiabat. This is more or less the case near the surface (pressures greater than 900 hPa) in the sounding plotted in Fig. 2.3. It suggests that adiabatic lifting of surface air would lead to condensation (an equal temperature and dew-point) at a pressure near 900 hPa. This level, where lifted air condenses, is called the lifting condensation level, or LCL. Further ascent would follow the saturation isentrope, here approximated by the pseudo-adiabat which falls off somewhat less rapidly with height than the temperature in the sounding. This implies that air parcels condensing at the LCL and rising along the saturated adiabat will be warmer than their environment and, modulo the contribution of condensate to density, more buoyant. Only at about 250 hPa does the environmental temperature begin to increase again, in association with the tropopause, limiting the buoyancy of saturated ascent from the LCL. Above this level the atmosphere is more or less isothermal, and one is in the stratosphere. Some of the other interesting features in this sounding will be discussed in subsequent sections.

2.3 Convective instability

Archimedes is credited with having developed the concept of buoyancy. According to the Archimedean Principle an object will rise if its mass is less than that of the fluid it displaces. If it is denser, and hence has a greater mass than the fluid it displaces, it will sink. The rising or sinking is a result of an imbalance from between the pressure and the gravitational force which either allow the denser object to do work on its environment as it sinks, or the environment to do work on the object as it rises. This imbalance between the hydrostatic pressure and the gravitational force is measured by the buoyancy, as introduced in Eq. (2.9).

In this section the buoyancy is used to explore the convective stability of different configurations of a static fluid. First the stability of a layer to infinitesimal displacements within that layer is considered. Next, the stability of two layers whose properties change discontinuously at an interface, for instance a cloud boundary is explored. Finally the response of the stability to finite displacements is considered. One of the things that makes moist fluids fascinating is the variety of convective instabilities that they support.

2.3.1 Infinitesimal displacements within a layer

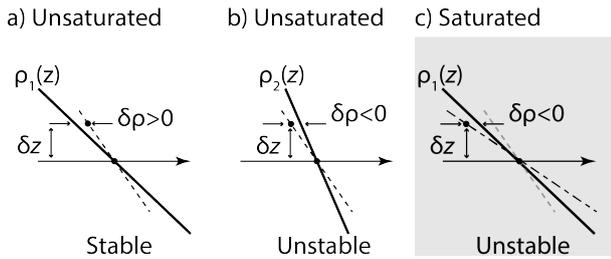


Figure 2.5 Schematic showing how buoyancy changes for an infinitesimal isentropic displacement, δ , of a fluid parcel relative to the environmental density profile, $\rho(z)$, within the layer and denoted by the solid line. Panel (a) and (c) share the same density stratification, $\rho_1(z)$, but the fluid in (c) is saturated. Panel (a) and (b) both show an unsaturated case, but $\rho_2(z)$ in panel (b) decreases less strongly height than $\rho_1(z)$ in panel (a). Lines also denote unsaturated (dashed) and saturated (dash-dot) isentropic density change with height.

For a homogeneous fluid, an infinitesimal displacement, δ of a fluid parcel will be unstable if the buoyancy arising from a displacement is in the same sense of the displacement. For instance if an upward displacement is accompanied by a relative increase in buoyancy the displacement will be amplified. The buoyancy due to an infinitesimal displacement is given as the difference between the density change that would arise from an isentropic displacement of a fluid parcel (denoted by subscript s , to denote the idea of a reversible adiabatic, or isentropic, displacement) and the environmental density gradient, such that

$$b = -g \frac{\rho(z + \delta z)|_s - \rho(z + \delta z)}{\rho(z)} = -g (\partial_z \ln \rho|_s - \partial_z \ln \rho) \delta z. \quad (2.86)$$

If the environmental density decreases with height less than the isentropic change in the density of a displaced fluid parcel, then the fluid parcel will, after a small upward displacement, find itself less dense than the environmental density at that same level and begin to accelerate upward. This situation is shown in panels (b) and (c) in Fig. 2.5 If the environmental density decreases with height more than it does for an isentropic change than the fluid parcel will find itself more dense than the environment and accelerate downward, as depicted in panel (a). In this situation the environmental density profile is said to be stable and in the absence of dissipation the displaced fluid parcel will oscillate about its equilibrium level with the frequency N , where

$$N^2 = g (\partial_z \ln \rho|_s - \partial_z \ln \rho). \quad (2.87)$$

N is called the Brunt-Väisälä Frequency. In the case that $N^2 < 0$, the fluid is said to be convectively unstable. Because the density change following an isentropic displacement of a fluid parcel is a thermodynamic property of the fluid, N is given by the environmental stratification and the state of the fluid. The density change following an isentropic displacement of a fluid parcel also depends on the state of the fluid, in particular whether or not it is saturated. Hence, and as shown in Fig. 2.5, a given density profile which is stable if the fluid is unsaturated may be unstable if the fluid is saturated. A density gradient that is stable in the case the fluid is saturated, but unstable in the case the fluid is unsaturated is called conditionally unstable. To make clear whether or not the density gradients in a fluid are being evaluated in comparison to saturated, versus unsaturated, isentropic parcel displacements N_s is used to define the former, and N_d the latter.

If it is assumed that the pressure felt by the disturbed parcel adjusts instantaneously to the pressure at the new height, $z + \delta z$, then pressure differences between the displaced parcel and its environment vanish. In this case it is sufficient to consider

$$\frac{\rho'}{\rho} = -\frac{T'_\rho}{T_\rho} = -\frac{T'}{T} - \frac{R_v}{R} q'_v + \frac{R_d}{R} q'_t. \quad (2.88)$$

In the absence of saturation, $q'_v = q'_t$, and q_t is conserved for the parcel displacement. Environmental gradients of q_t do, however, contribute to the density difference between the displaced parcel and the environment, such that

$$N^2 = g \left[\frac{1}{T} (\partial_z T - \partial_z T|_s) + \left(\frac{R_v - R_d}{R} \right) \partial_z q_t \right]. \quad (2.89)$$

The difference between the isentropic lapse rate and the environmental lapse rate can be expressed in terms of the entropy, or potential temperature, gradient, such that the Brunt-Väisälä Frequency for the unsaturated fluid becomes

$$N^2 = c_p \Gamma \frac{d \ln \theta}{dz} + g \left(\frac{R_v - R_d}{R} \right) \frac{dq_t}{dz}. \quad (2.90)$$

In deriving Eq. (2.90) $c_p \Gamma$, where Γ denotes the adiabatic lapse rate, substitutes for g in the first term on the rhs. Writing N^2 in terms of Γ anticipates the derivation of its expression for a saturated layer. In the completely dry case the expression simplifies further because the second term vanishes, Γ reduces to Γ_d and c_p reduces to c_d .

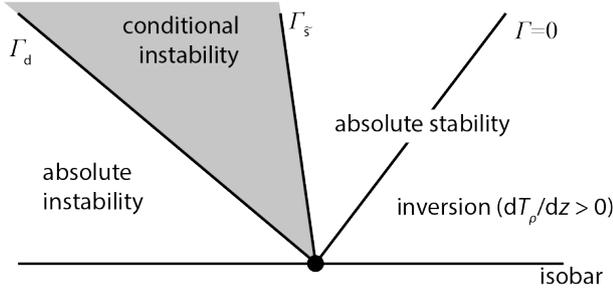


Figure 2.6 Different regions of atmospheric stability, as delineated by the fundamental lines demarcating different lapse rates on a Skew- T diagram.

For a saturated layer, changes in q_v vary as q_s . Changes in q_s can be related to temperature fluctuations through the Clausius-Clapeyron equation, as described by Eq. (2.82). Pressure fluctuations make no contribution because the displaced parcel is assumed to adjust to the pressure in the environment, so that terms depending on the difference between its pressure and that of the environment vanish. Accounting for the temperature related q_s variations as well as gradients in q_t thus results in the following expression for the Brunt-Väisälä frequency in a saturated layer:

$$N_s^2 = g \left[\frac{1}{T} \left(1 + q_s \frac{R_v}{R} \beta_T \right) (\partial_z T - \partial_z T|_s) - \frac{R_d}{R} \partial_z q_t \right]. \quad (2.91)$$

As in the dry case, the difference between the actual lapse rate and the saturated isentropic lapse rate is, following Eq. (2.38), proportional to the vertical gradient of the value of θ_e in the saturated layer. Hence the Brunt-Väisälä Frequency in a saturated layer, can be expressed analogously to the expression for unsaturated layers, such that

$$N_s^2 = c_p \Gamma_s \frac{d \ln \theta_e}{dz} - g \frac{R_d}{R} \frac{dq_t}{dz}. \quad (2.92)$$

As is expected N_s^2 reduces to N_d if q_t vanishes. Because $d\theta_e/dz < d\theta/dz$ the saturated Brunt-Väisälä Frequency is less than that of the dry fluid, and perturbations that would be stable for an unsaturated fluid ($N^2 > 0$) may be unstable ($N_s^2 < 0$) for a saturated fluid.

The difference between N^2 and N_s^2 define three stability regimes of the atmosphere. A region of absolute stability and instability, irrespective of the saturation state of the atmosphere, and between them a region where the atmosphere is said to be conditionally unstable. A conditionally unstable layer is one which would be unstable if it were saturated, but stable if it is unsaturated. These different stability regimes, demarcated roughly by the sign of Γ_d and Γ_s , are illustrated schematically in Fig. 5.3.1.4, as they would appear in relationship to one another in a Skew- T diagram. This demarcation is only rough because N^2 and N_s^2 , also depend on moisture, and thus are not fundamental lines on a Skew- T . Because Γ_s decreases in magnitude with increasing temperature, as roughly illustrated by the tendency of pseudo adiabats in the Skew- T diagram (Fig. 2.4) to become more perpendicular to dry adiabats at warmer temperatures, the

domain of conditional instability increases with temperature. It is easier to destabilise a warm atmosphere.

2.3.2 Stability across interfaces

In the atmosphere very sharp gradients are frequent, and arise from different air-masses coming together to form a contact discontinuity. Perhaps the best example of this is a cloud. In such situations the stability of the interface can be explored by comparing the change of density across the interface, or the effect of mixing on the fluid properties on one or the other side of the interface. Both situations are depicted schematically in Fig. 2.7 for the most interesting case of when one fluid is saturated and the other is not.

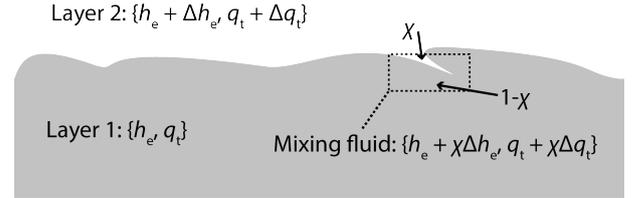


Figure 2.7 The interface between an unsaturated fluid overlying a saturated fluid, and their thermodynamic state. Because the boundary between the fluids is assumed to be infinitesimally thin there is no difference in the pressure between the two layers.

For the case of an overlying unsaturated fluid separated from an underlying saturated fluid by an interface, as depicted in Fig. 2.7, the stability of the interface depends on whether or not the upper fluid (layer 2) is more or less dense than the lower layer. If the upper fluid is denser, i.e., $\Delta \rho > 0$ the interface will be unstable. This condition can be expressed in terms of the thermodynamic variables, $\{h, q_t\}$. Assume that the differences between the fluid states is small so that,

$$\frac{\Delta \rho}{\rho} \approx - \left(\frac{\Delta T}{T} + \frac{\Delta R}{R} \right) \quad (2.93)$$

where $R = R_d + q_v R_v - q_t R_d$, hence its change reflects changes in fluid composition. For the saturated fluid $q_v = q_s$ and for the unsaturated fluid $q_v = q_t$. The change in the temperature and gas constants across the layers can be expressed in terms of changes in q_t and h_e given

$$\Delta h_e \approx (\Delta c_{p_e}) T + c_{p_e} \Delta T + (\Delta \ell_v) q_v + \ell_v \Delta q_t + \ell_v q_c \quad (2.94)$$

and

$$\Delta R \approx R_v (\Delta q_t + q_c) - R_d \Delta q_t. \quad (2.95)$$

In deriving these expressions it is assumed that the state of the upper layer fluid can be expressed in terms of a Taylor series expansion about the state of the lower layer fluid. Such an approximation is only roughly correct as for many situations Δq_t is of the same order as q_t , nonetheless it gives some insight into the stability of the layer.

Rearranging Eq. (2.94-2.95) results in an expression for the density difference between the two layers as

$$\frac{\Delta \rho}{\rho} = \left[\frac{\ell_v}{c_{p_e} T} - \frac{R_v}{R} \right] (\Delta q_t + q_c) + \left[\frac{c_l - c_{p_d}}{c_{p_e}} + \frac{R_d}{R} \right] \Delta q_t - \frac{\Delta h_e}{c_{p_e} T}. \quad (2.96)$$

For $\delta\rho > 0$ the stratification of the two layers is unstable, hence the instability criterion can, assuming $\Delta q_t < 0$ and $q_c \ll |\Delta q_t|$, be written as

$$\frac{\Delta h_e}{\ell_v \Delta q_t} > \kappa_1. \quad (2.97)$$

where

$$\kappa_1 \approx 1 + \frac{T}{\ell_v} \left[(c_1 - c_{pd}) - \left(\frac{R_v - R_d}{R} \right) c_{pe} \right]. \quad (2.98)$$

For conditions typical of the subtropics, $\kappa_1 \approx 1.3$. If, one adopts the common approximation of neglecting differences between the specific heats $\kappa_1 \approx 1$. In either case it means that for the situation depicted in Fig. 2.7 the moist enthalpy, or equivalently the moist static energy, must decrease across the interface; more strongly the greater the change in q_t . Eq. (2.98) can be thought of as a form of conditional instability, whereby an increase in the moist-static energy across the layer is sufficient for the interface to be stable in the absence of mixing.

The case in which the two layers mix introduces additional possibilities. Isobaric mixing of fluid elements from the two layers as depicted in Fig. 2.7 can, through non-linearities in the equation of state, lead to a mixture whose density is not bounded by the temperatures of the constituent air masses. Consider how small, isobaric, changes in the density, defined in Eq. (2.93), depend on changes in the thermodynamic coordinates $\{h_e, q_t\}$. We first consider the case where the mixing fraction, χ , of the unsaturated fluid is small, so that the mixture remains saturated. This case differs from the static situation in that now $q_v = q_s$ which changes with T . Denoting the small change by δ , so for instance $\chi \Delta h_e = \delta h_e$, this implies that

$$\delta R = R_v \delta q_s - R_d \delta q_t = R_v \frac{q_s}{T} \beta_T \delta T - R_d \delta q_t. \quad (2.99)$$

Similarly an expression involving q_s arises in expanding the expression for δh_e such that

$$\frac{\delta \rho}{\rho} = \left[\frac{c_1 - c_{pd}}{c_p + \frac{\beta_T \ell_v q_s}{T}} \left(1 + \frac{R_v}{R} \beta_T q_s \right) + \frac{R_d}{R} \right] \delta q_t - \frac{\delta h_e}{c_{pd} T} \gamma \quad (2.100)$$

For the usual case in which saturated air mixes with drier, air so that $\delta q_t < 0$, then this implies that density fluctuations will be negative if

$$\frac{\Delta h_e}{\ell_v \Delta q_t} > \kappa_2 \quad (2.101)$$

where

$$\kappa_2 = \frac{c_{pd} T}{\gamma \ell_v} \left[\frac{c_1 - c_{pd}}{c_p + \frac{\beta_T \ell_v q_s}{T}} \left(1 + \frac{R_v}{R} \beta_T q_s \right) + \frac{R_d}{R} \right] \quad (2.102)$$

is the buoyancy reversal parameter. Because we have assumed that Δq_t is negative this implies that Δh must be sufficiently negative for mixtures to be more dense than their constituent components. For conditions typical of subtropical stratocumulus $\kappa_2 \approx 0.56$. If, as is common, the thermodynamics is developed without considering differences in specific heats, then in such a system $\kappa_2 = 0.23$. Because $\kappa_2 < \kappa_1$ this implies the existence of stable interfaces, which

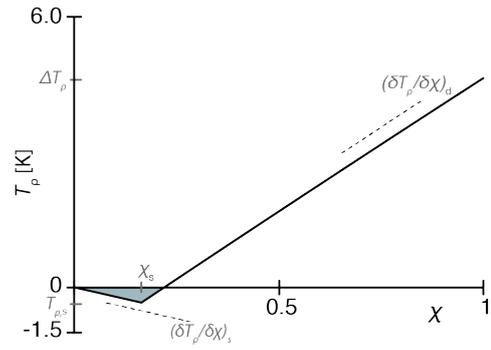


Figure 2.8 Mixing diagram showing how the buoyancy of a saturated layer changes upon mixing with a fraction χ of unsaturated air, for two air masses whose buoyancy differs by $\Delta b = 0.154$. The saturation mixing fraction χ_s , is the mixture with the minimum buoyancy, $b_s = -0.011$. The different rate of change of buoyancy with mixing fraction is illustrated by the dashed lines for saturated and unsaturated processes respectively.

as a result of mixing processes, for instance by diffusion, become unstable.

The situation whereby

$$\kappa_1 > \frac{\Delta h_e}{\ell_v \Delta q_t} > \kappa_2 \quad (2.103)$$

defines the case of buoyancy reversal, expressed here in the thermodynamic coordinates $\{h_e, q_t\}$, but because for an isobaric process $\Delta \eta = \Delta h$ the buoyancy reversal criterion is equivalently formulated in terms of moist static energy. It says that for η'_e sufficiently negative, the accompanying density fluctuation will be less than zero assuming the perturbed fluid parcel remains saturated.

Physically this condition just expresses the fact that the evaporation of condensate from the saturated layer, when it mixes with the drier layer, causes cooling in the mixture which can more than offset the warming from the increase in dry enthalpy that accompanies the mixing. This leads to the curious phenomenon of a mixing induced instability. This type of instability is thought to be important for stratocumulus decks in the very dry subtropics, and likely inhibits them from becoming too deep (e.g., as discussed in Chapter 5).

The situation for buoyancy reversal based on the state of a typical subtropical stratocumulus cloud is illustrated with the help of Fig. 2.8. Here one sees that for small mixing fractions the density temperature decreases, reaching its most negative value when the mixing fraction takes on a critical value associated with that mixture which just evaporates all the condensate contributed to the mixture by the saturated component of the mixture. This is called the saturation mixing fraction, χ_s . For $\chi > \chi_s$ the mixed parcel warms within increasing χ . The figure shows that for typical stratocumulus layers there is not sufficient liquid water to support strongly negatively buoyant mixtures, i.e., χ_s is small.

Thus in the atmosphere all conditionally unstable interfaces are also subject to buoyancy reversal. As saturated air rises through the environment the cloud-clear-air interface

is necessarily destabilised by mixing processes at the cloud front. As the cloudy air rises, clear air mixes at the cloud edge, in ways that are very effective of consuming the available potential energy of the rising plume. This makes clouds very effective mixing entities, something that is hidden from view by their visual appearance.

2.3.3 Subcritical instabilities, and CAPE

A sub-critical instability is one in which the basic state is only unstable to perturbations greater than a certain amplitude. There is a tradition of considering the stability of finite displacements of fluid parcels as a way of measuring the stability of the atmosphere. Here the thinking is that disturbances, for instance a spreading gust front, can bring layers of the atmosphere to a state, and into an environment, where they become unstable.

The susceptibility of the atmosphere to finite amplitude parcel displacements is often measured by the Convective Available Potential Energy, or CAPE. CAPE measures the amount of work the atmosphere is capable of doing on a parcel lifted to its level of free convection. Denoting the CAPE by the symbol \mathcal{A} :

$$\mathcal{A} = \int_{z_f}^{z_n} b \, dz, \quad (2.104)$$

where b is the buoyancy as defined in Eq. (2.9), and the limits of integration are z_f , the level of free convection (LFC) and z_n , the level of neutral buoyancy (LNB). This is the level at which the parcel lifted following some specified process, for instance along a saturated isentrope, or pseudo-adiabat, ceases to be buoyant relative to the environment. Most of the time z_n is near the tropopause.

Substituting (2.9) into (2.104) and using the hydrostatic equation to replace the integration in height by an integration in pressure yields

$$\mathcal{A} = \int_{p_n}^{p_f} R_d T'_\rho \, d(\ln p). \quad (2.105)$$

Thus on the Skew- T diagram (e.g., Fig. 2.3) \mathcal{A} is roughly equal to the area between the environmental temperature and the dashed line. ‘Roughly’ because the potential acceleration depends also on the available moisture, so as to account for differences between T and T_ρ .

CAPE as defined by (2.104) depends sensitively on the properties of the parcel being lifted and the manner in which it is lifted. Small differences in the initial state of a parcel can lead to large differences in \mathcal{A} . For instance, a slight drying of the boundary layer in Fig. 2.3 will raise the LCL and decrease the temperature of the saturated adiabat, and raise the level of free convection, thereby reducing \mathcal{A} .

Because \mathcal{A} describes the work the atmosphere can do on a parcel, or alternatively the potential energy available to a convecting parcel, it can be related to the maximum kinetic energy. That is it bounds the amount of kinetic energy a parcel could have, thus defining a velocity scale:

$$w_{\max} = \sqrt{2\mathcal{A}} \quad (2.106)$$

which is another measure of the intensity of convection.

2.3.3.1 Other CAPE-like measures

CAPE is the most common measure of the atmosphere to support overturning through the instability of saturated ascent. Whether or not CAPE is present in the atmosphere depends on the properties of the parcel being lifted, and the process whereby it is lifted. Even then many other factors come into play in deciding on the capacity of the atmosphere to overturn. The amount of work that must be invested to access the CAPE of a sounding is highly variable. In some cases one does not have to do a large amount of work on a parcel before the atmosphere starts returning the favour. To measure the work that must be invested to realise CAPE, another parameter, the Convective Inhibition (or CIN), is introduced. It is the analog to CAPE, but measures the amount of work that must be done to lift a parcel from some reference level, p_* , to its level of free convection:

$$\mathcal{I} = - \int_{p_*}^{p_f} R_d T'_\rho \, d(\ln p). \quad (2.107)$$

Thus the ability of the atmosphere to do work on an air parcel depends not on CAPE alone, but also on other factors such as CIN. In Fig. 2.9, CIN corresponds to the area between the environmental temperature and the saturated isentrope rising from the LCL to the LFC.

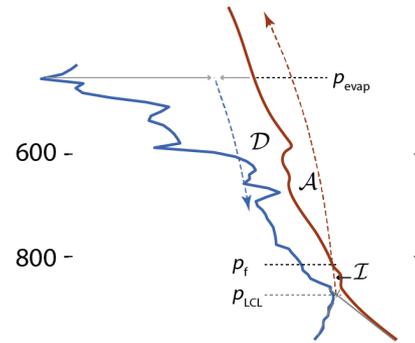


Figure 2.9 Illustration of different types of CAPE using the sounding of Fig. 2.3. The CAPE is illustrated by the area between the red dashed line showing the saturated adiabatic lapse rate for a parcel rising from the LCL. The Convective Inhibition, or CIN, is defined by the negative area between p_f and the LCL. The blue dashed line shows the temperature of air formed by saturating very dry air near 500 hPa, and maintaining it in saturation as it descends, and the area between the blue dashed line and the environmental temperature demarcates downdraft CAPE.

Another CAPE like measure of the atmosphere is called down-draft CAPE, or \mathcal{D} . It and the other forms of CAPE are illustrated in Fig. 2.9. \mathcal{D} measures the stability of saturated downward displacements of air first brought to its wet-bulb temperature by evaporation of falling rain. The wet-bulb temperature (sometimes denoted T_w) is bounded by the dew-point temperature and the actual temperature. It is the temperature that one gets by isobarically bringing air to saturation through evaporation. Because water is evaporated into the air, the air both cools and moistens increasing its mixing ratio while decreasing its temperature.

Hence,

$$\mathcal{D}(p_i) = \int_{p_{\text{evap}}}^{p_{\text{sf}}} R_d T'_\rho d(\ln p), \quad (2.108)$$

where here $T_\rho(p_{\text{evap}}) = T_w$ and T'_ρ is the difference between the virtual temperature of a parcel at some level p_{evap} brought to saturation by evaporating water into it and the environmental value of T_ρ . Physically \mathcal{D} measures the stability of air to evaporation of rain. In environments with large values of \mathcal{D} vigorous down-drafts can be formed by evaporating water (from precipitation) into a dry ambient environment. The analogy to buoyancy reversal, whereby negatively buoyant parcels can be created through isobaric mixing, should be apparent.

2.3.3.2 Caveats on CAPE

As alluded to above the actual CAPE of a given atmospheric sounding is not a number without ambiguity. This ambiguity stems mostly from the varied definitions associated with it. Above CAPE has been defined to be the positive area on the thermodynamic diagram. Others define it to be the *net* positive area associated with a parcel lifted dry adiabatically from the surface to the level of neutral buoyancy. Others compute the CAPE associated with a parcel characterised by the *mean* properties of the lower 10 hPa to 50 hPa of the atmosphere. Yet others adjust the surface properties of the sounding to reflect what they anticipate conditions will be like at some particular point in time. Because it measures a conditional process, different definitions arise naturally, and these can be both quantitatively and qualitatively different, as how much energy is available to a parcel raised to its level of free condition depends very much on the parcel being lifted.

The thermodynamic processes which govern the evolution of the state of the parcel above the lifting condensation level also play a large role in determining CAPE. For instance parcels in which ice forms will have different values of CAPE than for parcels in which ice is not allowed to form. Likewise, rising parcels usually mix with their environment, in fact the more unstable they are the more they are likely to mix. Thus CAPE can be seen as a function of the state of a parcel, the environment, and the specified type of thermodynamic process for the rising parcel. These ambiguities do not diminish the value of a measure like CAPE, but they do indicate that if it is to be used quantitatively the particular use of the concept must be made precise.

2.3.4 Slice method

The ability of CAPE to characterise the potential energy available to convection is based on a number of idealisations. These include: (i) that the parcels being lifted follow the specified thermodynamic process; (ii) that the parcels being lifted are characteristic of the air from which the convection actually develops; (iii) that the response of the environment can be neglected. More refined measures of convective instability attempt to address one or more of these limitations. Most notable among these is the slice method introduced

by Jacob Bjerknes, one of the pioneers in the development of meteorology as a branch of physics, in the 1930s. Unlike in parcel theory, where infinitesimal parcels are assumed to move through a quiescent environment, Bjerknes' slice method considers the finite size of convective motions and hence the compensating downward motions they induce. By doing so it identifies an effective stability, which depends on the area fraction occupied by the ascending motion.

The starting point for this method is to assume a convecting atmosphere at some reference height z_0 above cloud base. It is envisioned that in some sufficiently large area, the fraction of the convecting area is a robust quantity which can be denoted by a . Integrating the vertical velocity over the convecting region allows one to define a mean convective velocity w_c . Mass continuity across the reference level allows one to express the mean subsiding velocity of the environment w in terms of (a, w_c) :

$$aw_c + (1 - a)w = 0. \quad (2.109)$$

In convective modelling the quantity aw_c is related to the convective mass flux, \mathcal{M}_c , by the density, i.e.,

$$\mathcal{M}_c = \rho aw_c. \quad (2.110)$$

In what follows the density is assumed to be constant, and hence dropped from the discussion. This is consistent with the treatment of shallow convection, wherein density differences are small, and Eq. (2.109) is a statement of mass conservation.

Assuming that the air rising in the convecting region is rising along a saturated adiabat, while the subsiding air follows a dry adiabat, the temperature difference between the convecting region and the environment at the reference level can be expressed as follows:

$$T_c(z_0) - T(z_0) \approx T_{01} - \Gamma_s \Delta z_\uparrow - (T_{02} + \Gamma_d \Delta z_\downarrow), \quad (2.111)$$

where both Δz_\uparrow and Δz_\downarrow are defined as positive

$$T_{01} = T(z_0) + \Delta z_\uparrow \Gamma \quad (2.112)$$

$$T_{02} = T(z_0) - \Delta z_\downarrow \Gamma, \quad (2.113)$$

and $\Gamma = -dT/dz$ the environmental lapse rate. Note that the above implies that the model is most appropriate if Δz_\uparrow is associated with the distance above cloud base.

The above relations (illustrated schematically in Fig. 2.10) describe the temperature in the convecting region as the temperature that the environmental air at a distance Δz_\uparrow below z_0 would have if it were lifted to z_0 along a saturated isentrope (denoted Γ_s) while the temperature of the environment at z_0 is that which the environmental air a distance Δz_\downarrow above the reference level would have were it brought dry adiabatically to z_0 . The distances Δz_\uparrow and Δz_\downarrow are given by $w_c \Delta t$ and $-w \Delta t$ respectively, with the sign convention chosen to ensure that a positively measured downward displacement leads to a negative, or downward, velocity.

Substituting from above yields an expression for the temperature difference in terms of a, w_c and Γ ,

$$\frac{T_c - T}{\Delta t} = w_c(\Gamma - \Gamma_s) - w(\Gamma - \Gamma_d) \quad (2.114)$$

$$= w_c \left[\Gamma - \Gamma_s + \frac{a}{1 - a} (\Gamma - \Gamma_d) \right]. \quad (2.115)$$

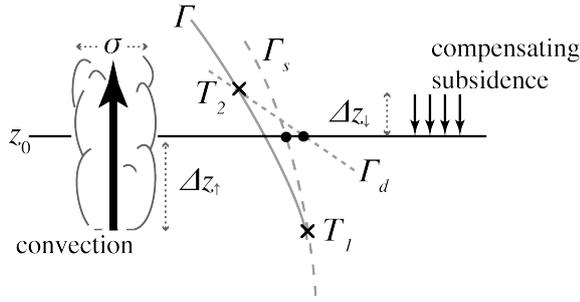


Figure 2.10 Situation modelled by the slice method. The convecting air temperature cools following the saturated adiabat, Γ_s as it rises a distance Δz_\uparrow . The compensating subsidence covers a greater area and descends a distance Δz_\downarrow and warms following the dry adiabat, Γ_d . The convection is energetically inhibited because the response of the environment stabilises the atmosphere.

Thus the criterion for convective instability, $T_c - T > 0$ is equivalent to requiring that

$$\Gamma > \Gamma_s + a(\Gamma_d - \Gamma_s) > \Gamma_s. \quad (2.116)$$

This requirement is more severe than that given by the parcel method, or for the case of infinitesimal displacements in the absence of a humidity gradient. The physical difference being that the compensating downward motion stabilises the environment. However, in the limit when a goes to zero the original criterion, $\Gamma > \Gamma_s$, that one derives from parcel theory is recovered.

A further implication of the above result is that moist convection is most unstable if its fractional area is smallest. Physically this is not surprising, as for a given w_c vanishing a implies that the compensating environmental motion is minimised and hence the environment is stabilised the least. This type of analysis might help explain the spacing of convective systems, or why nature prefers to concentrate convection over rather small areas. Mathematically this is apparent by considering the neutral limit of (2.115):

$$\Gamma - \Gamma_s(1 - a) - a\Gamma_d = 0. \quad (2.117)$$

Hence for a given unstable lapse rate, only for

$$a \leq \frac{\Gamma - \Gamma_s}{\Gamma_d - \Gamma_s}. \quad (2.118)$$

will the atmosphere be convectively unstable.

2.4 Fluid dynamics

For studying conservation laws and force balances in flows, the concept of the substantial derivative, denoted

$$D_t := \partial_t + \nabla \cdot \mathbf{v} \quad (2.119)$$

needs to be mastered. It describes how a quantity changes following its motion, here denoted by \mathbf{v} , with ∇ the gradient operator. In an inertial coordinate system in a gravitational field, $\mathbf{g} = (0, 0, -g)$, the laws governing the conservation of mass, momentum and energy of the two component fluid can

(with some assumptions to be discussed later) be expressed as

$$D_t \rho = -\rho \nabla \cdot \mathbf{v}, \quad (2.120)$$

$$D_t \mathbf{v} = -\frac{1}{\rho} [\nabla p + \mu \nabla^2 \mathbf{v}] + \mathbf{g}, \quad (2.121)$$

$$D_t s_\ell = \frac{1}{\rho T} [\mu (\nabla \mathbf{v}) : (\nabla \mathbf{v}) - \nabla \cdot \mathbf{F}_h], \quad (2.122)$$

$$D_t \rho_t = -\rho_t \nabla \cdot \mathbf{v}. \quad (2.123)$$

Equation (2.120) describes conservation of total mass, where the density, $\rho = \rho_d + \rho_t$, depends on the constituent densities for dry air and total water. The second is the Cauchy equation, describing Newton's law for a fluid parcel, with velocity, $\mathbf{v} = (u, v, w)$. Surface forces acting on the parcel are written in terms of the pressure gradient (for isotropic forces). Deviatoric surface forces are assumed to behave in a diffusive fashion described by the dynamic viscosity, μ . The only body force is that arising from \mathbf{g} . The Second Law of Thermodynamics includes the divergence of an enthalpy flux, \mathbf{F}_h , associated with radiative and conductive enthalpy transfer, and heating through the dissipation of kinetic energy, which is positive definite and proportional to the dynamic viscosity. Water mass conservation is described in terms of the total water density, ρ_t , by Eq. (2.123).

Eqs. (2.120)-(2.123) can be derived by assuming that the two components of the fluid constitute an ideal mixture, and the velocity is the mass averaged velocity of the two components of the fluid, and the ideal gas law provides an additional constraint to close the system. The above system holds in the case of a multi-component fluid in the limit where the condensed phase must be assumed to be comprised of sufficiently small particles so that it can be described as a continuous field, which diffuses and flows, and has the same temperature, as the other constituents of the fluid. These assumptions become difficult to maintain for the development of a precipitate phase, wherein large particles are, by definition, falling through the flow, and their thermal inertia can lead to large temperature differences to that of the mean flow. An approximate way of treating such situations will be discussed at the end of the next section. An introduction to a literature providing a more exact description, by individually tracking the motion of the different fluid components and exchanges of momentum and energy between them, is provided in the section on further reading.

2.4.1 Soundproof equations

The above set of equations are rarely used as a basis for the investigation of fluid processes in meteorology or oceanography. The main reason is that despite the many simplifications in their derivation, they still include a phenomenon not thought to be essential for many processes to which the equations may be applied – sound waves. Compared to most processes of interest, sound waves (acoustic modes) are very fast, and mostly serve to adjust the pressure field so as to maintain some form of incompressibility. For most applications of meteorological interest, assuming the flow satisfies an incompressibility condition following its evolution (rather

Table 2.5. *Soundproof Equation Sets*

Assumption	Name
$\rho_a = \rho_0$	Boussinesq
$\rho_a = \rho_0(z)$	Anelastic
$\rho_a = \frac{\rho_0(z)\theta_0(z)}{\theta(x,y,z,t)}$	Pseudo-incompressible

than trying to simulate the adjustment process) makes the equations easier to handle. Such equations sets are usually arrived at by neglecting the contribution of pressure fluctuations on density in the continuity equation, such that it can be written in terms of a sound-proof density ρ_a ,

$$D_t \rho_a = -\rho_a \nabla \cdot \mathbf{v}. \quad (2.124)$$

with different forms of incompressibility arising based on how ρ_a is approximated, e.g., Table 2.5.

The most restrictive of these arises from the Boussinesq approximation, but it is also the most widely used and the most straightforward for outlining the main form of the scaling arguments to arrive at a sound-proof, or soundproof, equation set. Hence the derivation for the Boussinesq case is sketched out below.

The Boussinesq approximation involves solving for the velocity \mathbf{v} , density ρ , potential temperature θ_ℓ , pressure p , and total-water specific humidity q_t which are assumed to vary about a dry and resting reference state $\{\rho_0, \theta_0, p_0(z)\}$. Only the reference state pressure varies, and then only vertically.

In the limit of small perturbations, which for the moment we denote by a prime,

$$\rho'(p, T_\rho) \approx \left(\frac{\partial \rho}{\partial p} \right)_0 p' + \left(\frac{\partial \rho}{\partial T_\rho} \right)_0 T_\rho' \quad (2.125)$$

so that

$$\frac{\rho'}{\rho_0} = \frac{p'}{p_0} - \frac{T_\rho'}{T_{\rho,0}}. \quad (2.126)$$

The principle assumptions of the Boussinesq system are that density perturbations, ρ' are, relative to the mean state, small, such that

$$\varepsilon \equiv \frac{\rho'}{\rho_0} \ll 1, \quad (2.127)$$

and that relative pressure perturbations are in most cases of interest much smaller than relative temperature perturbations. With these two assumptions,

$$\frac{\rho'}{\rho_0} \approx -\frac{T_\rho'}{T_{\rho,0}}. \quad (2.128)$$

Eq. (2.128) is a good approximation because the pressure is an integral quantity, in that through the hydrostatic balance it depends on the integral of the temperature field through the atmosphere. So local temperature perturbations are expected to be much larger than pressure perturbations. Alternatively, this approximation follows because $p'/p_0 \sim M^2$ where the Mach Number, M is much less than unity, where as temperature fluctuations are influenced by entropy fluctuations, which are independent (but generally larger) than the Mach Number.

Consider the case when the flow is governed by a single velocity-scale U and length-scale H (which imply a time-scale H/U) determined by the inertial motions. In this case the continuity equation (2.120) can be non-dimensionalised by the time, velocity and distance, scales (with non-dimensional quantities denoted by tilde),

$$(\tilde{t}, \tilde{\mathbf{v}}, \tilde{\mathbf{x}}) \quad \text{where} \quad t = H\tilde{t}/U, \quad \mathbf{v} = \tilde{\mathbf{v}}U, \quad \mathbf{x} = H\tilde{\mathbf{x}},$$

such that

$$\tilde{D}_t(\varepsilon) + (1 + \varepsilon)\tilde{\nabla} \cdot \tilde{\mathbf{v}} = 0, \quad (2.129)$$

where \tilde{D}_t and $\tilde{\nabla}$ denote the non-dimensional form of the convective (substantial) derivative and gradient operator, e.g., Eq. (2.119). With (2.127) the leading order balance is simply non-divergence, which in terms of dimensional variables becomes:

$$\nabla \cdot \mathbf{v} = 0. \quad (2.130)$$

To analyse the balances in the momentum equation Eq. (2.121) can be multiplied by the density, here expanded as $\rho_0 + \rho'$, and a similar procedure followed, such that:

$$(1 + \varepsilon)\tilde{D}_t \tilde{\mathbf{v}} = -\tilde{\nabla}(\tilde{p}_0 + \tilde{p}') + (1 + \varepsilon)\tilde{\mathbf{g}} + \frac{1}{\mathcal{R}_\ell} \tilde{\nabla}^2 \tilde{\mathbf{v}}, \quad (2.131)$$

where $\tilde{p}' = (H/\rho_0 U^2)p'$, $\tilde{\mathbf{g}} = (0, 0, -gH/U^2)$. $\mathcal{R}_\ell \equiv UH/\nu$ is the Reynolds Number, with the kinematic viscosity, $\nu = \mu/\rho$. \mathcal{R}_ℓ is typically very large, 10^8 , for geophysical flows. The leading order balance in this equation is $dp_0/dz = -\rho_0 g$, i.e., hydrostatic balance, which is here expressed in dimensional form, so that

$$D_t \mathbf{v} = -\frac{1}{\rho_0} \nabla p' - \left(\frac{T_\rho'}{T_{\rho,0}} \right) \mathbf{g} + \nu \nabla^2 \mathbf{v}. \quad (2.132)$$

A simple application of the scale analysis outlined above would suggest that the last term, which involves the viscosity, should be negligible. This is not the case, as turbulence leads to the production of velocity fluctuations across very small scales, much smaller than used to scale the pressure gradients, and these ensure that the viscous term contributes at leading order. The form of Eq. (2.132) makes clear the manner in which fluctuations in density temperature, T_ρ , drive density fluctuations. The term $(-T_\rho'/T_{\rho,0})\mathbf{g}$ has units of acceleration, it is a reduced gravity and is called the buoyancy term and denoted as $\mathbf{b} = (0, 0, b)$, whereby $b = gT_\rho'/T_{\rho,0}$. The last term in Eq. (2.132) is the diffusion, which acts on small scale velocity fluctuations.

The thermodynamic equation, (2.122), when written in terms of θ_ℓ becomes, at leading order,

$$D_t \theta_\ell = \frac{\theta_0}{c_{p_\ell} T_0(z)} [\nu(\nabla \mathbf{v}) : (\nabla \mathbf{v}) - \nabla \cdot \mathbf{F}_\ell]. \quad (2.133)$$

Given the continuity equation the expression for water mass conservation can be written in terms of the specific humidity, as

$$D_t q_t = 0. \quad (2.134)$$

Taken together, Eqs.(2.130), (2.132), (2.133) and (2.134) define the moist Boussinesq equations. They form a closed set given an equation of state. The equation of state is given in the form of the buoyancy function, $b(T, p, q_t)$. Both T and p

are diagnostic functions. T follows from the definition of θ_l , given $p' + p_0$, and p' is given by taking the divergence of the momentum equation, in which case the incompressibility of the flow, Eq. (2.130), results in an equation for $\nabla^2 p$ which can be inverted for p' .

Allowing for non-equilibrium condensate phase introduces additional complexity and usually involves a greater degree of approximation. For instance these phases will introduce entropy source terms, and in general will not follow the flow. Increasingly large hydrometeors, like hail, will have temperatures that increasingly depart from that of the ambient flow. The hydrometeors also exert a force on the flow, but this is accounted for through the condensate loading term that appears in the equation for the density temperature, e.g., Eq. (2.7). In the subsequent discussion we also drop the primes, taking it for granted that the thermodynamic quantities denote departures from the reference state. So doing facilitates an exploration of how these departures in turn depart from an expected value as is customarily done in the analysis of turbulence

2.4.2 Quasi-static (primitive) equations

For large-scale atmospheric flows the quasi-static equations prove to describe the dynamics to a good degree of approximation. In some ways these are the most original dynamic equations for the study of atmospheric motions, having been derived by Vilhelm Bjerknes, Jakob Bjerknes's father and in some ways the father of modern meteorology, as the basis for numerical weather prediction at the turn of the last century. When written with pressure as a vertical coordinate these equations take on a particularly convenient form, and formally resemble the equations used to study small-scale flows. The starting point for deriving the equations is again the Eqs (2.120)-(2.123). Considerations of large-scale flows in a non-inertial reference frame (such as the rotating Earth), cannot, almost by definition, neglect the Coriolis acceleration, $\mathbf{f} \times \mathbf{v}$, and are frequently written in a coordinate system appropriate to the sphere (such as spherical coordinates) or some segment of a sphere (for instance by mapping the cartesian coordinate system conformally onto a Mercator projection). For completeness in the discussion that follows the apparent acceleration from the Coriolis force is represented through the inclusion of the vertical component of the inertial frequency $\mathbf{f} = \{0, 0, f\}$, but a Cartesian geometry is retained for simplicity. Consistent with the scale of the analysis, viscous transport terms are neglected.

The crucial assumption of the quasi-static equations is that the circulations are shallow, equivalently that the vertical length scale is comparable to the depth of the troposphere and is much smaller than the horizontal length scale. This implies that the vertical velocities first appear at higher order. To follow the implications of this assumption; let us represent the velocity vector by a two dimensional (horizontal) vector on a geopotential surface \mathbf{v}_ϕ and a component, w perpendicular to this surface,

$$\mathbf{v} = (\mathbf{v}_\phi(x, y, z, t), w(x, y, z, t)),$$

where $w \ll v_\phi$. For this scaling the leading order momentum equations (2.121) become

$$D_t \mathbf{v}_\phi = -\mathbf{f} \times \mathbf{v} - \frac{1}{\rho} \nabla_\phi p, \quad (2.135)$$

$$0 = -\frac{1}{\rho} \frac{\partial p}{\partial z} - g, \quad (2.136)$$

where ∇_ϕ denotes the two dimensional (x, y) , gradient operator along surfaces of constant geopotential. These equations can be combined with the thermodynamic equation (2.122) and continuity equations (2.120) and (2.123) to form a closed set; although because w has been eliminated (2.136) further manipulation is required to derive an equation for w .

Rather than following this path it proves convenient to rewrite the equations with pressure as a vertical coordinate, as this greatly simplifies the equations, at the expense of the boundary conditions which become more challenging – the pressure velocity, ω , unlike the kinematic velocity, w , does not generally vanish at the surface. The hydrostatic constraint, (2.136) forms the basis for such a transformation as in a hydrostatic atmosphere pressure effectively measures the mass of the atmosphere above a given point in space, so it can be thought of as a mass coordinate. Upon integration, and by virtue of the fact that ρ is positive definite

$$p(x, y, z, t) = p(x, y, 0, t) - g \int_0^z \rho(x, y, \zeta, t) d\zeta$$

is a strictly decreasing function of z , with ζ substituting for z in the integral to avoid confusion.

In carrying out the transform it proves useful to note that for some generic field ϑ

$$v = A(x, y, z, t) = A(x, y, z(x, y, p, t), t) \quad (2.137)$$

$$= B(x, y, p, t), \quad (2.138)$$

where A and B will generally have different forms. Writing terms in this manner helps illustrate that

$$\left(\frac{\partial \vartheta}{\partial x} \right)_z = \frac{\partial A}{\partial x}$$

will not in general equal

$$\left(\frac{\partial \vartheta}{\partial x} \right)_p = \frac{\partial B}{\partial x}$$

as the terms on the lhs imply a change in the variable along a constant height surface, and the terms on the rhs imply a change along an isobaric surface. Rather

$$\left(\frac{\partial \vartheta}{\partial x} \right)_z = \left(\frac{\partial \vartheta}{\partial x} \right)_p + \left(\frac{\partial z}{\partial x} \right)_p \frac{\partial \vartheta}{\partial z} \quad (2.139)$$

$$= \left(\frac{\partial \vartheta}{\partial x} \right)_p + \frac{1}{g} \left(\frac{\partial \phi}{\partial x} \right)_p \frac{\partial \vartheta}{\partial z}, \quad (2.140)$$

where the final expression is written in terms of the geopotential, ϕ . Similar forms follow for the partial derivatives with respect to the other coordinates, y and t . Above the subscript z and p are used to denote in which coordinate system the differentiation is taking place. To discriminate between the vertical velocity in the z -coordinate system, $w \equiv D_t z$ the pressure velocity is defined by the symbol ω such that $\omega \equiv D_t p$.

Applying this coordinate transformation greatly simplifies the continuity equation, making it divergence free,

$$\nabla \cdot \mathbf{v} = \nabla_p \cdot \mathbf{v}_p + \partial_p \omega = 0 \quad (2.141)$$

This makes physical sense, because working in terms of pressure, a mass coordinate, mass conservation implies that the mass field must be divergence free. Applying the transformation to the pressure gradients terms in the horizontal momentum equations transforms these to linear functions of the geopotential, such that

$$D_t \mathbf{v}_p = -\mathbf{f} \times \mathbf{v} - \nabla_p \phi, \quad (2.142)$$

where in pressure coordinates, the differential operator D_t takes the form

$$D_t = \partial_t + \mathbf{v}_p \cdot \nabla_p.$$

Eq. (9.1) shows that adopting pressure as the vertical coordinate eliminates density from the momentum equations.

For large-scale systems it is common to work with temperature as the thermodynamic coordinate, this is convenient in pressure coordinates where the pressure term in the enthalpy form of the thermodynamic equation (2.10) can simply be replaced by the vertical velocity, such that

$$D_t T = -\frac{1}{c_{p_e}} [\omega \partial_p \phi - \nabla \cdot \mathbf{F}_h - Q_{\text{cnd}}], \quad (2.143)$$

where a term arising from changes in ℓ_v following the flow has been neglected by virtue of the fact that $q_v \ll 1$. The net condensation heating, Q_{cnd} , on the rhs of (2.143) thus is the only term that remains to represent the effect of phase changes on the enthalpy, and it is proportional to changes in the water vapour specific humidity,

$$D_t q_v = -\frac{Q_{\text{cnd}}}{\ell_v}. \quad (2.144)$$

Equations (2.141)-(2.144) are the quasi-static equations, although the modern literature often refers to them as the primitive equations in pressure coordinates, one could just as well call them the Bjerknes equations. Both the Boussinesq and the quasi-static equations in pressure coordinates share a common structure. In both cases the approximations or coordinate transforms render the momentum sources linear in thermodynamic quantities, and in both cases the flow is purely solenoidal, i.e., non-divergent. As a practical matter this non-divergence is, however, handled very differently in the different systems. In the quasi-static equations ω becomes a diagnostic quantity, except at the surface. For both systems the equations are non-linear as a consequence of the quadratic form introduced through the advective derivative. This non-linearity has the effect that upon averaging the equations, for instance to eliminate smaller scales, additional terms, involving correlations between fluctuating quantities, are introduced into the governing equations. This is discussed further below.

For moist flows the primitive equations may be supplemented by explicit equations for the condensate mass, of the form $D_t q_c = C$ similar to what was proposed for the moist Boussinesq equations. Likewise, by casting (2.143) and (2.144) in enthalpy form, neglecting variations in c_{p_e} and ℓ_v ,

and summing, a moist form of the conservation law emerges,

$$D_t \eta_e = \nabla \cdot \mathbf{F}_h + (\partial_t + \mathbf{v}_p \cdot \nabla_p) \phi. \quad (2.145)$$

The equation illustrates that in addition to the divergence of the enthalpy flux, \mathbf{F}_h , the time rate of change of the geopotential following the isobaric component of the flow contributes to a change of the moist static energy, as anticipated in the derivation of the static energy. Often the latter term will be neglected, in which case the equivalent moist static energy, η_e becomes an adiabatic invariant of the flow, much like θ_ℓ for the moist Boussinesq system.

2.5 Reynolds averaged equations

Reynolds averaging is a particular type of averaging that facilitates the isolation of large-scale degrees of freedom. For a variety of reasons it makes sense to restrict oneself to some subset of, usually larger, scales when studying fluid motion. Computationally, this is motivated by a desire to limit the degrees of freedom that must be solved for. Theoretically, some assumptions are more readily justified (for instance that pressure is everywhere hydrostatic) when small scales are neglected. And on simple practical grounds, the smallest scales of the flow, on the order of the Kolmogorov scale, are simply not relevant for many questions. Even in turbulent flows the fluctuations associated with a particular realisation of a flow are often not of interest, rather the mean properties.

Let the Reynolds average of a quantity be denoted by an over-bar. The over-bar is defined through a filtering, or averaging, procedure with the following properties:

$$\overline{\overline{\varphi}} = \overline{\varphi} \quad \text{and} \quad \overline{\overline{\varphi\vartheta}} = \overline{\varphi\vartheta}, \quad (2.146)$$

where φ is some field, and ϑ is a second field, which could also be identical to φ . The filter is further assumed to commute with other linear operators, such as differentiation in the time or space domain. Defining the deviation from the Reynolds averaged by the prime, so that,

$$\varphi = \overline{\varphi} + \varphi' \quad (2.147)$$

the above implies that

$$\overline{\varphi'} = 0 \quad \text{and} \quad \overline{\overline{\varphi\varphi'}} = 0. \quad (2.148)$$

Even with these assumptions, an application of the Reynolds average to a non-linear quantity, such as the convective derivative, introduces additional terms. For the case of the substantial, or advective, derivative, which given the assumption of incompressibility, Eq. (2.130) can be written in flux form, such that

$$\overline{D_t u} = \partial_t \overline{u} + \nabla \cdot (\overline{\mathbf{v}u}) + \nabla \cdot (\overline{\mathbf{v}'u'}). \quad (2.149)$$

The last term on the rhs had no counter part in the original equations, and is the divergence of a quantity called the Reynolds stress, or in the case that u is a scalar, the Reynolds flux. For small scale flows the Reynolds fluxes act in all directions; in large scale flows it is often assumed that the averaging is over much larger areas in the horizontal than in the vertical. In this case it can be assumed that the flow

is sufficiently homogeneous that the horizontal fluxes can be neglected relative to the resolved quantities, i.e.,

$$\nabla_p \cdot (\overline{\mathbf{v}_p \bar{u}}) \gg \nabla_p \cdot (\overline{\mathbf{v}'_p u'}),$$

and only the vertical transport by fluctuating quantities is retained.

2.5.1 Types of filtering

Reynolds averaging is a conceptual procedure, where the filtering of the equations is thought of in terms of an average over an ensemble of realisations of a flow. For most applications of interest a particular flow realisation is solved, for instance by simulation, and filtering is employed to separate small from large scales. In the case that such filtering is applied in spectral space, with a wave-number cutoff filter that truncates all scales smaller than a specified scale, the Reynolds averaging rules can also be shown to hold. However, many flows are solved for in the physical domain, and in conditions (for instance over limited areas) where spectral filtering is not possible. Here the filter that is used to separate large from small scales is usually defined implicitly, through the numerical methods employed to solve the equation. These methods distort scales near the grid-scale and truncate all the scales smaller than the grid-scale. But filters based on a local stencil in physical space (like a running average) are not spectrally sharp, and thus influence a range of scales at, and larger than, the grid-scale. For such filters the Reynolds averaging rules usually cannot hold, but the additional terms that arise from these imperfections of the filters are normally neglected or assumed small compared to the errors in the parameterisation of the terms that are retained.

2.5.2 Turbulence kinetic energy equation

Clouds are usually turbulent, and an application of Reynolds averaging can be used to illustrate the basic mechanisms which control the production and dissipation of turbulence. The resultant equation, called the turbulence kinetic energy (TKE) equation plays an important role in the parameterisation of clouds and boundary layer processes. Because turbulence kinetic energy is conceptualised as being concentrated in eddies whose scale is much smaller than a scale height, a starting point for its derivation is the Boussinesq equations. An equation for the Reynolds averaged velocity can be derived by writing the velocity equation, (2.132), in terms of fluctuating and averaged quantities, averaging, and then applying the Reynolds averaging rules. So doing results in an equation of the form

$$\overline{D}_t \bar{\mathbf{v}} + \overline{(\mathbf{v}' \cdot \nabla) \mathbf{v}'} = -\frac{1}{\rho_0} \nabla \bar{p} + \bar{\mathbf{b}} + \nabla \cdot \bar{\boldsymbol{\tau}}. \quad (2.150)$$

where here the viscous term in Eq. (2.121) has been written in terms of the stress tensor, which is defined as twice the dynamic viscosity times the rate-of-strain tensor,

$$\boldsymbol{\tau} = \nu \left[(\nabla \mathbf{v}) + (\nabla \mathbf{v})^T \right]. \quad (2.151)$$

The additional term on the lhs of Eq. 2.150 arises because the substantial derivative for the Reynolds averaged flow is only defined with respect to the averaged flow, that is

$$\overline{D}_t \equiv (\partial_t + \bar{\mathbf{v}} \cdot \nabla). \quad (2.152)$$

Subtracting Eq. (2.150) from Eq. (2.132) results in an equation for the velocity fluctuations of the form

$$\overline{D}_t \mathbf{v}' - \overline{(\mathbf{v}' \cdot \nabla) \mathbf{v}'} = -\frac{1}{\rho_0} \nabla p' + \mathbf{b}' + \nabla \cdot \boldsymbol{\tau}' - \mathbf{r}', \quad (2.153)$$

where

$$\mathbf{r}' = (\mathbf{v}' \cdot \nabla) \mathbf{v}' + (\mathbf{v}' \cdot \nabla) \bar{\mathbf{v}}. \quad (2.154)$$

Taking the inner product between (2.153) and \mathbf{v}' and averaging, results in a scalar equation for the small-scale (or turbulence) kinetic energy,

$$\bar{e} = \frac{1}{2} \overline{\mathbf{v}' \cdot \mathbf{v}'} \quad (2.155)$$

such that

$$\overline{D}_t \bar{e} = -\frac{1}{\rho_0} \nabla \cdot (\overline{\mathbf{v}' p'}) + \overline{\mathbf{v}' \cdot (\nabla \cdot \boldsymbol{\tau}')} + \overline{w' b'} - \overline{\mathbf{v}' \cdot \mathbf{r}'}. \quad (2.156)$$

From basic relations in vector calculus and the rules of Reynolds averaging the second and fourth terms on the rhs can be readily simplified, yielding the common form of the equation for turbulence kinetic energy, wherein the evolution of \bar{e} following the mean flow is given as the balance between the turbulent transport, production and dissipation of turbulence energy,

$$\overline{D}_t \bar{e} = \nabla \cdot \bar{\boldsymbol{\mathcal{T}}} + \bar{\mathcal{P}} - \varepsilon. \quad (2.157)$$

The individual terms in this balance are given respectively as

$$\boldsymbol{\mathcal{T}} = \left[\overline{\mathbf{v}' e'} + \frac{1}{\rho_0} \overline{\mathbf{v}' p'} + \mathbf{v}' \cdot \boldsymbol{\tau}' \right] \quad (2.158)$$

$$\mathcal{P} = \overline{w' b'} - \overline{\mathbf{v}' u' \cdot \partial_x \bar{\mathbf{v}}} - \overline{\mathbf{v}' v' \cdot \partial_y \bar{\mathbf{v}}} - \overline{\mathbf{v}' w' \cdot \partial_z \bar{\mathbf{v}}} \quad (2.159)$$

$$\varepsilon = \nu \overline{(\nabla \mathbf{v}') \cdot (\nabla \mathbf{v}')}. \quad (2.160)$$

An important limit of the above equations is when the averaging is anisotropic, such that the horizontal scale is much larger than the vertical scale of the averaging, as might be the case when considering boundary layers which are homogeneous along the boundary. This is also the case in large-scale models where grid-cells may be thousands of times larger in the horizontal direction than it is in the vertical direction. If one further assumes that the energy is contained in scales much smaller than the largest (horizontal) averaging scale then it is often justified to assume homogeneity in these directions, so that terms involving horizontal derivatives may be neglected. In this case the transport and production terms simplify to the following:

$$\nabla \cdot \boldsymbol{\mathcal{T}} = \partial_z \left[\overline{w' e'} + \frac{1}{\rho_0} \overline{w' p'} \right] \quad (2.161)$$

$$\mathcal{P} = \overline{w' b'} - \overline{u' w' \partial_z \bar{u}} - \overline{v' w' \partial_z \bar{v}} \quad (2.162)$$

$$(2.163)$$

where viscous transport has also been neglected.

2.5.3 Buoyancy production in multi-phase flows

The buoyancy term can be related to perturbations in thermodynamic variables, as buoyancy fluctuations are carried by both moisture and temperature fluctuations. For instance, given θ_ℓ and q_t as thermodynamic coordinates, the buoyancy flux can be written in terms of contributions from each, such that

$$\overline{w'b'} = \frac{1}{T_{\rho,0}} \left(a_{\theta_\ell}|_0 \overline{w'\theta'_\ell} + a_{q_t}|_0 \overline{w'q'_t} \right), \quad (2.164)$$

where for a thermodynamic coordinate ϑ ,

$$a_{\vartheta} \equiv \frac{\partial T_\rho(\vartheta, \dots)}{\partial \vartheta}. \quad (2.165)$$

This partial derivative provides a more formal basis for defining the Brunt-Väisälä frequency, as $N^2 = (g/z)a_\phi$.

A definition of a first order phase transition is that the partial derivatives describing the material properties of the system are discontinuous at the phase boundary. The familiar example being the compressibility of liquid being manifestly different than that of vapour. Hence, the partial derivatives relating perturbations in the density temperature to those in the state variables will take on different values according to whether the flow is saturated or not. For the case of an unsaturated flow

$$a_{\theta_\ell} \approx \frac{\theta}{T} \quad \text{and} \quad a_{q_t} \approx T\epsilon_2. \quad (2.166)$$

These equations imply that for unsaturated flow, for instance for the surface fluxes in the convective boundary layer, sensible heat fluxes are about fifteen times more efficient, per unit of surface cooling, at accelerating the flow, as compared to moisture fluxes. However over great expanses of the tropical oceans latent heat fluxes are more than fifteen times larger, meaning that latent heat fluxes contribute as much, or more, to the surface buoyancy flux as do sensible heat fluxes.

In saturated flow, $q_v = q_s(T, p)$ and $q_c = q_t - q_s$. Hence in saturated air fluctuations in the specific humidity terms that appear in the expression for T_ρ will result from fluctuations in θ_ℓ , likewise fluctuations of q_t will result in fluctuations in T , so that

$$a_{\theta_\ell} \approx \left(\frac{T}{\theta} \right) \gamma \quad \text{and} \quad a_{q_t} = T \left[\left(\frac{\ell_v}{c_p T} \right) \gamma - 1 \right] \quad (2.167)$$

where γ , the ratio of the saturated to dry adiabatic lapse rate, was defined in Eq. (2.85). Given typical values this implies that, unlike for unsaturated flows, moisture fluctuations are very efficient in generating buoyancy fluctuations in saturated flows, but entropy fluctuations are not.

At warm temperatures the difference between a_{q_t} in saturated versus unsaturated flows is about a factor of five, increasing to twice that value at cold temperatures, where $\Gamma_s \approx \Gamma_d$. The relative effectiveness of moisture fluctuations, in saturated versus unsaturated situations, makes physical sense, because an increase in q_t is, all things else being the same, associated with condensation, which warms the parcel, thereby increasing its buoyancy. Similar reasons explain why a_{θ_ℓ} can be as much as a factor of two smaller in saturated, relative to unsaturated conditions. Overall moisture fluxes are an important source of boundary layer turbulence

in tropical regions, and an essential source in saturated flows, i.e., clouds.

2.5.4 Effective heat and moisture sources, and the gross moist stability

In this section we present some concepts that have been developed to understand the gross, or bulk, properties of large-scale flows. The starting point for doing so are the quasi-static equations, written in a form that neglects the isobaric transport of geopotential. This is generally a good approximation in the tropics, where horizontal pressure gradients are weak. In this case the quasi-static thermodynamic equations for a large-scale moist flow can be written for the dry static energy, η_d , and the specific humidity, q_v as

$$\overline{D_t \eta_d} = -\partial_p \overline{\omega' \eta'_d} + Q_{\text{cnd}} + Q_{\text{rad}} = Q_1, \quad (2.168)$$

$$\overline{D_t q_v} = -\partial_p \overline{\omega' q'_v} - \frac{Q_{\text{cnd}}}{\ell_v} = -\frac{Q_2}{\ell_v}. \quad (2.169)$$

Here the over-bar denotes a Reynolds average, and horizontal homogeneity is assumed. The source terms in the dry static energy budget are associated with condensational, Q_{cnd} , and radiative heating, Q_{rad} respectively. This formulation of the thermodynamic equations separates the large-scale adiabatic processes from the small scale and diabatic processes, which are then grouped together in single terms, the apparent enthalpy source, Q_1 , and the apparent moisture sink, Q_2 . As could be anticipated from the equation for the moist static energy (neglecting changes in the geopotential and vaporisation enthalpy following along the isobaric flow), the sum of (2.168) and the enthalpy form of (2.169) shows that

$$\overline{D_t \eta_e} = Q_{\text{rad}} - \partial_p \overline{\omega' \eta'_e} = Q_1 - Q_2. \quad (2.170)$$

This equation provides a powerful constraint for many diagnostic studies, but also for modelling (see for instance Chapter 5). For instance, by integrating over the troposphere, from some p_{top} to p_{sfc} it demonstrates that on large-scales, where the net import or export of moist static energy by the advective flow can be neglected, changes in the moist static energy are driven by surface moist static energy fluxes (which are dominated by evaporation, equivalently precipitation), and the net transfer of radiant energy out of the column, i.e.,

$$\langle D_t \eta_e \rangle = \langle Q_{\text{rad}} \rangle + \overline{\omega' \eta'_e} \Big|_{\text{sfc}} \quad (2.171)$$

where the angle brackets denote a vertical integral, so that in pressure coordinates,

$$\langle \vartheta \rangle = \frac{1}{\Delta p} \int_{p_{\text{top}}}^{p_{\text{sfc}}} \overline{\vartheta}(\dots, p) dp, \quad (2.172)$$

where $\Delta p = p_{\text{sfc}} - p_{\text{top}}$ denotes the pressure depth of the integral and Reynolds averaging is implicit in the definition of the vertical average.

Expanding the first term on the lhs of Eq. (2.171) shows that the

$$\partial_t \langle \eta_e \rangle + \langle \mathbf{v}_p \cdot \nabla_p \eta_e \rangle + \langle \omega \partial_p \eta_e \rangle = \langle Q_{\text{rad}} \rangle + (\Delta p)^{-1} \overline{\omega' \eta'_e} \Big|_{\text{sfc}}. \quad (2.173)$$

The vertical pressure velocity can be decomposed using Galerkin methods, a typical example of which is the projection of the continuous vertical structure of a variable onto orthogonal basis functions. Such a transformation, in terms of ω is equivalent to writing

$$\omega = \sum_{i=1}^{\infty} \hat{\omega}_i(x, y, t) \Omega_i(p) \quad (2.174)$$

where the basis functions are given by the Ω_i . For the simple case in which the variance of the vertical motion is well represented by a single vertical basis function, Ω_1 ,

$$\omega = \hat{\omega}_1(x, y, t) \Omega_1(p) + \omega^*. \quad (2.175)$$

In this case the vertically averaged moist static energy equation becomes

$$\begin{aligned} \partial_t \langle \eta_e \rangle + \langle \mathbf{v}_p \cdot \nabla_p \eta_e \rangle + \langle \omega^* \partial_p \eta_e \rangle + \langle \omega \rangle \mathcal{G}(\eta_e) \\ = (\Delta p)^{-1} \overline{\omega' \eta_e'} \Big|_{\text{sfc}} + \langle Q_{\text{rad}} \rangle, \end{aligned} \quad (2.176)$$

which introduces the gross vertical averaging operator, \mathcal{G} . It is defined such that for an arbitrary variable, ϑ ,

$$\mathcal{G}(\vartheta) \equiv \langle \Omega_1 \partial_p \vartheta \rangle. \quad (2.177)$$

In Eq. (2.173) the term $\mathcal{G}(\eta_e)$ has come to be called the gross moist stability. If, in Eq. (2.173), to the extent that the first three terms on the lhs are small and can be neglected relative to the other terms, an equation for the amplitude of the gravest mode of vertical motion can be derived in terms of the energy input into the column and its gross moist stability, such that

$$\hat{\omega}_1 \approx \frac{1}{\mathcal{G}(\eta_e)} \left[(\Delta p)^{-1} \overline{\omega' \eta_e'} \Big|_{\text{sfc}} + \langle Q_{\text{rad}} \rangle \right]. \quad (2.178)$$

This equation helps illustrate the physical meaning of $\mathcal{G}(\eta_e)$ as it defines the proportionality between the moist enthalpy fluxes that drive the flow and the strength of the mean vertical circulation. It is the moist analog to the balance in clear skies between the dry static stability, the local radiative heating, and the vertical velocity at a point.

A similar analysis may be applied to the moisture equation. In the case advective fluxes are small compared to evaporative fluxes (as might be expected on sufficiently large scales), a simple balance arises such that

$$P - E_{\text{sfc}} - \langle \omega \rangle \mathcal{G}(q_v) \approx 0 \quad (2.179)$$

where E_{sfc} denotes the surface evaporation, $P = \langle Q_{\text{cnd}} \rangle$ is the net precipitation, and $\mathcal{G}(q_v)$ defines a gross moisture lapse rate. Some authors prefer to directly relate precipitation to energy fluxes, and in so doing they define a different form of the gross moist stability as the ratio

$$\frac{Q_1}{Q_2} - 1. \quad (2.180)$$

This is sometimes called the Normalised Gross Moist Stability, and its apparent similarity to the efficiency of heat engines is purely coincidental. Defining the 'Normalised Gross Moist Stability' in this fashion avoids having to make assumptions about the structure of the circulation when discussing the relationship between fluxes of energy and precipitation. These and related ideas are discussed further in

Chapter 5, they also prove to be important in Chapter 13 where they are applied to understanding patterns of precipitation change in the tropics.

2.6 Turbulence

2.6.1 Phenomenology of turbulence

A hallmark of turbulent flows is that they distribute energy over a broad range of scales. This can be thought of in terms of a Fourier representation of their energy spectrum, as energy is densely distributed across scales through the nonlinearities of inertial interactions. In a flow where only one scale is forced, for instance through an instability, the resulting fluid motions excites other scales of motions through non-linear interactions and secondary instabilities. This results in a rich range, or spectrum, of length scales becoming evident, for instance as revealed through a Fourier decomposition of the motion which quantifies how much variability is carried by a particular scale of motion. These interactions have a purpose, as they transport inviscid invariants of the flow (e.g., energy) from scales where energy is excited to scales where molecular processes operate efficiently and can thus dissipate the energy. This type of transport across scales is called a cascade, and it is illustrated schematically in Fig. 2.11 According to this view, energy is injected into the system at a macroscopic scale l_0 that drives the largest eddies, inviscidly transferred to smaller eddies in the inertial subrange, and finally dissipated into heat at the Kolmogorov scale l_K



Figure 2.11 Cartoon of energy cascade whereby energy “cascades” inviscidly from large scales to small scales where it is then dissipated. Here the scale of the motion is denoted by its wavenumber $k = 2\pi/l$ where l is the wavelength.

The description of fully developed turbulence as a cascade of interacting eddies dates back to the work of Lewis Fry Richardson in the early part of the 20th Century, and the self-similar character of the eddies have been confirmed in numerous experiments. Yet as of today, there is still no deductive theory that derives the observed self-similar scaling laws from the Navier-Stokes equations. Fortunately, much of the basic behaviour of fully developed turbulence characterised by a high Reynolds Number can be understood on a simple phenomenological basis

For a turbulent flow in a stationary state, the TKE equation can be approximated as a balance between the rate of

energy injection E_I at the macroscale and energy dissipation rate ε at the microscale

$$\partial_t e \approx E_I - \varepsilon \approx 0 \quad (2.181)$$

Consider now an eddy of size l with a relative velocity v_l , be defined as a velocity difference over a distance l and that is characterised by an eddy turnover time through

$$t_l \sim \frac{l}{v_l}. \quad (2.182)$$

where the symbol \sim means “equal up to an order of magnitude”. The main hypothesis, stated in the celebrated 1941 paper by Andrej N. Kolmogorov is that the rate of energy transfer cascading from the largest eddies down to the smallest eddies is constant and approximated by the mean dissipation rate ε . The transfer of energy in the inertial range $l_K < l < l_0$ can be intuitively interpreted as the flux of turbulent kinetic energy $e \sim v_l^2$ from an eddy of size l cascading down to the smaller eddies.

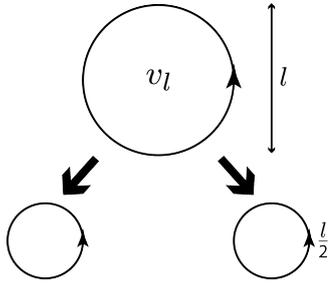


Figure 2.12 Schematics of the breakdown of a turbulent eddy of size l into smaller eddies.

Taking the eddy turnover time t_l as the typical timescale during which the eddy breaks up into smaller eddies (Fig. 2.12), the transfer of energy can be written as

$$\frac{e_l}{t_l} \sim \frac{v_l^3}{l} \sim \varepsilon \sim \text{cst} \quad (2.183)$$

so that

$$v_l \sim \varepsilon^{1/3} l^{1/3} \propto l^{1/3} \quad (2.184)$$

which is a shorthand for one of the main results of Kolmogorov’s 1941 theory, K41. A number of key properties directly follow from K41. First K41 shows that the velocity field in turbulence is self-similar and described with an exponent $h = 1/3$. This implies that the gradient of the velocity field would be singular everywhere for small l if there would not be a Kolmogorov dissipation scale where diffusion becomes dominant. Diffusive effects can be neglected only if the eddy turnover time t_l is much smaller than the diffusion time scale $t_{l,\text{diff}} \sim l^2/\nu$, but because the diffusion time scale goes to zero faster than the eddy turnover time, there will always be a scale at which diffusion becomes dominant. Equating the diffusion and the eddy turnover time gives as an estimate for this Kolmogorov dissipation scale

$$l_K \sim \left(\frac{\nu^3}{\varepsilon} \right)^{1/4}. \quad (2.185)$$

The range of scales between the outer scale at which the energy is injected and the Kolmogorov scale is dictated by the

integral-scale Reynolds Number $\mathcal{R}_\mathcal{E} = UL/\nu$ where U denotes the relative velocity scale related to the largest eddies. By estimating the energy dissipation rate as the energy transfer rate at the outer scale, i.e. $\varepsilon \sim U^3/L$ and using Eq. (2.185) the ratio between L and l_K scales with the Reynolds Number as

$$\frac{L}{l_K} \sim \mathcal{R}_\mathcal{E}^{3/4}. \quad (2.186)$$

A typical convective atmospheric boundary layer of $L \sim 1$ km depth and a typical relative velocity scale at this outer scale of 1 m s^{-1} is characterised by a $\mathcal{R}_\mathcal{E} \sim 10^8$, which justifies the value presented in the scale analysis of §2.4.1. This value of $\mathcal{R}_\mathcal{E}$ implies a Kolmogorov scale of the order of 1 mm. This simple example shows the enormous range of spatial scales, six orders of magnitude, over which the K41 scaling behaviour of the velocity field extends. The scaling behaviour K41 can be easily reformulated into the famous 5/3 energy spectrum by taking the Fourier transform of the turbulent kinetic energy $e_l \sim v_l^2$

$$e(k) \sim \varepsilon^{2/3} k^{-5/3} = \alpha_K \varepsilon^{2/3} k^{-5/3} \quad (2.187)$$

where the wavenumber k can be associated to the inverse eddy size, i.e. $k \sim 1/l$. The energy spectrum demonstrates that the largest eddies are the most energetic. Universality requires that the proportionality constant α_K to be independent of the particular flow. This prediction is well supported experimentally with $\alpha_K = 1.6$ being called the Kolmogorov constant. At present there exists no accepted theory which explains the value of Kolmogorov’s constant. The lack of such a theory is the quintessential problem of turbulence.

As will be demonstrated in Chapter 5, the self similarity of atmospheric turbulence allows the possibility of realistically numerical modelling at relative coarse resolutions of around 100 m. Such Large Eddy Simulations resolve only the largest, most energetic eddies, whereas the smaller unresolved eddies can be realistically parameterised in terms of the resolved large eddies through the use of K41

2.7 Models of clouds and circulation

For the most part, the equations required to describe clouds are known. This means that a great deal of insight into the role of clouds in the climate system can be attained by exploring the behaviour of these equations under different circumstances. This approach, whereby equation systems describing cloud systems are used to construct virtual laboratories from which insights are derived, is the dominant research methodology at the present time.

Two challenges however limit the insights that can be derived from these approaches, and continue to necessitate a strong connection with observations. One is that not all the equations are known. In particular on microscopic scales many of the interactions among particles, and between the gas and particle phases especially with respect to ice-processes, are still poorly understood. This issue is even more manifest when it comes to describing the interaction

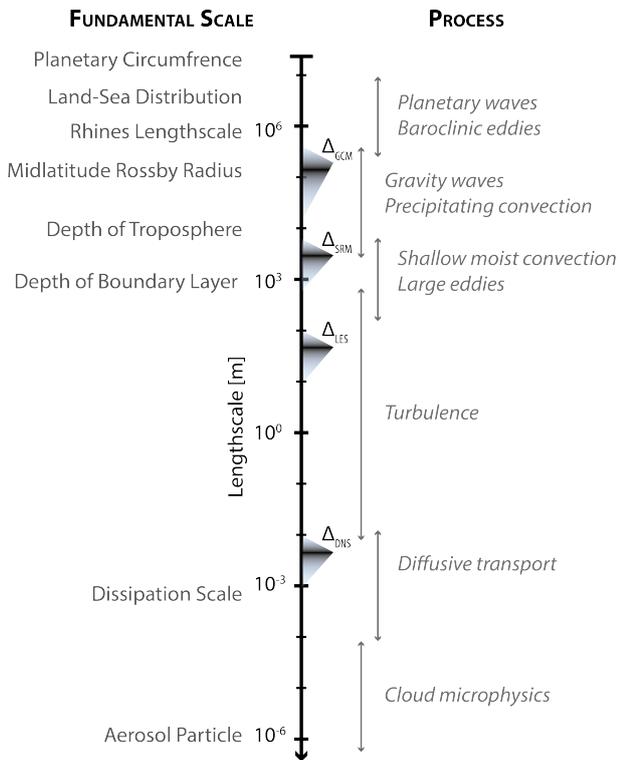


Figure 2.13 Physical length-scales in the atmosphere, and processes associated with them. Different classes of models are here characterised by the size, Δ , of their grid mesh, which is related to the smallest scales that they can represent.

of the atmosphere with other components of the climate system, for instance the biosphere. The other challenge is that even if the equations were all known they would encompass far too many degrees of freedom to ever contemplate solving, even given computers with computing capacities millions of times more powerful than what is presently possible. These challenges mean that the simulation systems used as virtual fluid dynamical laboratories of cloudy atmospheres are necessarily approximate.

The approximations made to render a fluid-dynamical description of a cloudy atmosphere computationally tractable are manifold: (i) some truncation of scale; (ii) an aggregated description (parameterisation) of the behaviour of particulate matter and condensate; (iii) a simplification in the treatment of radiative transfer. In addition, to close the equations inevitably involves developing approximate models describing interactions with other components of the climate system, for instance the surface, as well as the behaviour of degrees of freedom that, because of the scale truncation, are not explicitly represented. The different types of cloud dynamical models are distinguished mostly by the first approximation, namely at what scale they truncate the fluid-dynamical equations. Fig. 2.13 illustrates the range of scales within the atmosphere, processes associated with different scales, and the names associated with models that truncate their equations at a given scale. Because the range of processes included by a particular truncation differs depending

on the scale of the truncation these different types of models differ, as discussed below, qualitatively from one another.

2.7.1 Direct numerical simulation

Simulations based on fluid-dynamical descriptions that resolve eddies down to the Kolmogorov scale are called Direct Numerical Simulation, or DNS. DNS is the most fundamental description of fluid dynamical transport, but because computational restrictions limit the largest space and time scales that can be represented, it is also the most special purpose. Advances in computing power has, however, begun to make it possible to use DNS over large-enough domains that it has become applicable to problems of interest to atmospheric scientists, up to and including boundary layer scales. Because it is applied to problems whose largest scales are still relatively small, DNS is often based on the Boussinesq approximation, which on the scales usually represented can be asymptotically justified. But for some DNS the full compressible equations, or even two-fluid equations, may be applied. In DNS a main hypothesis is that of Reynolds Number similarity, which states that once a sufficiently large Reynolds Number is achieved then important statistical properties – like vertical profiles of mean and variances of temperature and specific humidities – cease to depend on the Reynolds Number. Even in the case that Reynolds Number similarity is a good assumption, DNS remains limited by many of the same limitations of models that more approximately represent the fluid-dynamical equations, namely simplifications in the coupling to realistic boundaries, or necessarily approximate descriptions of microphysical processes and radiative transfer. Despite these limitations which DNS shares with other types of cloud circulation models DNS is emerging as not just an exciting method for studying classic boundary layer flows, but also for studying cloud mixing processes and the interactions among turbulence, cloud-microphysical processes and radiant energy transfer.

2.7.2 General circulation models

General Circulation Models, or GCMs, are usually based on the primitive equations, and applied to the global atmosphere in a way designed to explicitly represent the main mechanism, baroclinic instability, of meridional heat transport in the mid-latitude troposphere (see Chapter 10). Because baroclinic eddies have a scale of about 1000 km, they can be well resolved by models using a grid whose mesh-size is on the order of a few hundreds of kilometres or finer (Fig. 2.13). General Circulation Models are a cornerstone of numerical weather prediction, and also serve as the dynamical core of the atmosphere in Global Climate Models. The identical initialism (GCM) for Global Climate Models and General Circulation Models – upon which they are often based – is a source of ambiguity, here it is restricted to mean the latter. In GCMs the vertical heat transport, associated with a cascade of processes ranging from near surface turbulent eddies, to boundary layer circulations, to deep convective clouds, and even cloud systems, are not at all

resolved, and must be approximated through parameterisations. The approximate, and generally inadequate, representation of vertical heat and moisture transport and of cloud processes in GCMs also influences their ability to properly represent radiant energy transfer through the atmosphere, something which is vital for an adequate representation of the climate system. Over the years a great deal of effort, as reviewed in Chapter 6, has been expended in better understanding how the process of vertical heat and momentum transport by eddies (and waves) is related to the mean state of the larger-scale flow that GCMs explicitly represent. Their chief limitation relative to smaller-scale models is the uncertainty of these processes, and their importance for the large-scale flow. This is especially true in the tropics, where moist cumulus convection and radiant energy transfer interact strongly with the circulation. The main advantage of the GCMs is that they provide a closed fluid-dynamics description of the transient dynamics of the large-scale flow, whose statistical properties define Earth's climate.

2.7.3 Storm-resolving models

Storm-resolving models, SRMs, typically forgo a global description so as to allow for sufficiently fine scales to resolve the vertical overturning of the troposphere. This then permits them to crudely, but explicitly, represent convective storms that reach through the depth of the troposphere – even when such storms adopt commensurately fine horizontal scales, as for the case of precipitating deep convection in the tropics. This implies that an SRM must have a spatial resolution commensurate with the depth of the tropical troposphere, and hence grid-mesh spacings on the order of a few kilometres. The requirement of such a fine grid-mesh generally limits the domain of SRMs to be on the order of hundreds to a few thousands of kilometres. SRMs are often based on the anelastic equations, but sometimes solve the fully compressible equations.

The great advantage of storm-resolving models is that their explicit representation of at least the gravest modes of convective heat transport in the troposphere allows for a more consistent representation of circulation features that often accompany such storms, from cold pools and gust fronts, to stratiform cloud shields. It also more naturally allows these circulation systems to be coupled with parameterisations of cloud microphysical processes and radiant energy transfer. Unlike a GCM, whereby most parameterisations must be coupled with one another, in SRMs, at least for the deepest clouds, parameterisations of cloud microphysical processes and radiation can interact with circulations that the model explicitly, albeit crudely, resolves. In the literature other SRMs are often referred to using different names: for instance Cloud Resolving Models (CRMs); Cloud System Resolving Models (CSRMs); or Convection Permitting Models (CPMs). Some authors distinguish between ‘Cloud Resolving’ and ‘Convection Permitting’ Models by virtue of how they parameterise yet finer scale processes. Because the distinguishing factor of SRMs is their ability to resolve storms, meaning primarily the circulation (wind) systems and geometric features of tropical convection, but not necessar-

ily the finest scales of associated clouds or convection, the SRM nomenclature is adopted in this text. By definition, an SRM does not include a parameterisation of deep convection. However, vertical heat transport arises on many other scales, for instance from the large eddies within the atmospheric boundary layer, or from shallow convective clouds which might reach a depth of a few kilometres, and must be parameterised in an SRM, a process that shares many of the complexities and pitfalls of the parameterisation of deep convection, albeit in a way that is hopefully less strongly coupled to the large-scale flow. Most SRM simulations also need to parameterise (or specify) the large-scale flow, for instance by relaxing it back to some presumed state, or by nudging it to lateral boundary conditions that are taken from another source. SRMs are being increasingly applied to study large-scale problems in geophysical fluid dynamics, and in so doing provide a better foundation for understanding the ways in which cloud processes affect these dynamics.

2.7.4 Large-eddy simulation

Large-eddy simulation, or LES, differs from SRM simulations in that it endeavours to resolve also the large eddies responsible for the heat and momentum transport within the atmospheric boundary layer, which for convective situations may be between 500 m and a few kilometres. This implies a spatial resolution of about 500 m, and hence a grid mesh size of about 100 m or less, which is still much larger than the Kolmogorov scale. As long as the large-eddies are well resolved, LES has the advantage that the smaller scale motions which are not resolved can often be assumed to be representable in a universal way based on an understanding of the phenomenology of homogeneous and isotropic turbulence, as discussed in Section 2.6.1. Because LES resolves the turbulent circulations on the cloud scale, it is thought to provide an adequate description of mixing processes associated with cumulus clouds, and to provide a strong foundation for linking radiative transfer to explicitly resolved clouds, and microphysical processes to explicitly resolved circulations. However, even when run at resolutions of a few meters, some mixing processes remain poorly represented within LES, especially those in regions of strong stratification, as one finds atop stratocumulus topped boundary layers (Chapter 5) and in nocturnal boundary layers. Such a fine spatial resolution also precludes a consideration of many meso- and large-scale processes, as historically LES domains have been on the order of kilometres. This means that LES usually neglects, or at least distorts, the way in which processes on the small scale influence the mesoscale in ways that may feed back and influence the small-scales. Like SRMs, LESs are also using equation sets that allow them to simulate flows over a deeper atmosphere and are being applied to ever larger domains, thereby providing insight into how cloud and boundary layer processes induce and influence circulations on meso (10 km to 100 km) scales.

2.8 Outlook

Although it might seem that the foundations of a fluid dynamical description of clouds is more or less worked out, there are several interesting and active areas of open research. In terms of the thermodynamics, basic questions relating to a consistent but simplified representation of non-equilibrium components of multi-phase flows remain to be clarified. This involves not only thermal effects, for instance associated with multi-phase particles (melting snow and graupel), but also dynamic effects when particles decouple from the flow. Only recently has research begun to attempt a full accounting of the entropy production in association with non-equilibrium phases and much work remains to be done, for instance in association with the entropy production by cloud microphysical processes.

Better understanding how to measure the energy available to convection, even in idealised frameworks, is an area where more research would be beneficial. Is there an optimal definition of convective available potential energy, the state to which a convecting atmosphere attempts to relax to, or the time-scale on which they relax? In this context many basic problems related to the stability of convecting atmospheres, and the implied asymmetries between upward (saturated) and downward (unsaturated) motion remains open. Here the advancing power of DNS to explore instabilities in moist atmospheric flows, and the nature of turbulence in multiphase atmospheric flow is very promising. At the same time more research is needed on conceptual models capable of better elucidating constraints on moist systems arising from the Second Law.

An exciting open area of research, called forth by the increasing computing power, is how to create soundproof equation systems that are energetically consistent on very large domains, and which consistently represent the multi-phase thermodynamics. Most asymptotic approaches to these problems remain focused on dry atmospheres and fail to consistently represent moist and microphysical processes. On the larger scales the moist static energy framework is promising and a better understanding of how the energetics of the flow couple to the circulation using this framework is an exciting and popular area of research, but here the role of shallow circulations, which such a framework endeavours to neglect, emerges as a key question.

Advances in computing power are beginning to blur the lines between traditional modelling frameworks. Recent years have witnessed the advent of hybrid models that embed LES or SRMs in GCMs, through an approach referred to as super-parameterisation (Chpt 6). Even more recently, very highly parallelised computations run on tens and even hundreds of thousands of processors have begun to make it possible to perform LES on grid-scales of a hundred meters, across domains of more than a thousand kilometres, for periods of weeks, or to run SRMs over the entire globe, thereby superseding GCMs, for periods of days to weeks. Along with DNS, the ability of these very computationally intensive approaches to encompass the interaction of a wide range of

Table 2.6. *Constants for Tetens' formula, as given in Eq. 2.190, for saturation vapor pressure over liquid water and ice.*

state	a	b [K]
water	17.2693882	35.86
ice	21.8745584	7.66

scales make them very exciting tools that are beginning to shed new light on how clouds and circulations interact.

Exercises

1. Show that choosing a reference state enthalpy for both h_e and h_ℓ determines the reference state vaporisation enthalpy. What is $\ell_{v,0}$ if one assumes that $h_1(T_{0e}) = c_1 T_{0e}$ and $h_v(T_{0,\ell}) = c_{p,v} T_{0,\ell}$?
2. Show that if in equilibrium h and T adopt values that maximise the entropy function s , then this implies that the Gibbs potential, g is minimised.
3. Assuming that the specific heats are constant, use the values in Table 2.3 to show that the reference entropy for liquid water at 0 °C is 3.518 kJ kg⁻¹ K⁻¹ and that for solid ice at the same temperature it is 2.298 kJ kg⁻¹ K⁻¹
4. If the energy is removed by internal, turbulent, dissipation (\mathcal{E}) then this represents an additional source of heating that must be accounted for, i.e., the first law of thermodynamics becomes

$$W = \mathcal{E} + Q_1 - Q_2. \quad (2.188)$$

and the second law

$$\frac{Q_1}{T_1} = \frac{Q_2}{T_2} - \frac{\mathcal{E}}{T_{\mathcal{E}}} - \Delta s_{\text{irr}}. \quad (2.189)$$

Show how in steady state, if the dissipation happens near the lower boundary, at a temperature $T_{\mathcal{E}} = T_1$ this implies a system that does more work than a Carnot engine. Why doesn't this contradict the second law?

5. In deriving the Clausius-Clapeyron equation it is assumed that $\ell_v/T = s_v - s_l$. Demonstrate that this is true in the case of a saturated system.
6. Tetens' originally formulated the saturation vapor pressure as

$$p_s = p_s(T_*) \exp \left[\frac{a(T - T_*)}{T - b} \right] \quad (2.190)$$

where $T_* = 273.16$ K is the triple point temperature, so that $p_s(T_*) = 610.78$ Pa. The fitting parameters, a and b , are given in Table 2.6 for saturation over liquid water and ice respectively. Compare this form to that given in Eq. (2.73).

7. Derive an expression for the saturation pressure over liquid by assuming: (i) ℓ_v is constant, and (ii) by allowing ℓ_v to varies with T following Kirchoff's relation. Compare these to Tetens's formula as give in Eq. (2.73) and the reference fit by Goff and Gratch as summarized by Murphy and Koop (2005).

8. Derive γ_ℓ and Ω_ℓ , which appear in Eq. (2.46) which defines the liquid-water potential temperature.
9. Fill in the steps of the derivation of the ice-liquid potential temperature as given by Eq. (2.66).
10. Starting from Eq. (2.38), derive Eq. (2.83), stating all assumptions.
11. By deriving an equation for the equisaturation line on a Skew-T diagram calculate the change in the dew-point depression with height for adiabatic ascent. Use this to derive an expression for the LCL of a parcel with a given dew-point depression.
12. Assume that surface fluxes are well represented by the bulk aerodynamic formulae,

$$\overline{w'\vartheta'} = C\|\mathbf{v}_{10m}\|(\vartheta_{10m} - \vartheta_{\text{sfc}})$$

where ϑ is a generic scalar quantity, like temperature or moisture. Estimate the height of the lifting condensation level given the surface latent and sensible heat fluxes, assuming a drag coefficient, C , of 0.001. State any additional assumptions that you may have to make to arrive at a solution. Does the expression depend on wind speed, or sea-surface temperature?

13. Assuming that in the tropics the temperature profile follows the saturated isentrope corresponding to saturation over the warmest surface waters. If $\delta\eta_e$ measures the difference between the surface moist static energy and that in the free troposphere derive an expression for $\delta\eta_e$ as a function of the deviation from the warmest sea-surface temperatures. Does this expression depend on the temperature of the warmest waters, or only on the temperature difference?
14. Derive an expression for the rate at which q_1 increases with height for adiabatic ascent.
15. Assuming an atmospheric thermal structure wherein $\eta_{e,s}) = 335 \text{ kJ kg}^{-1}$ assuming $c_e = c_d$, at what altitude is q_s less than q_{CO_2} assuming a pre-industrial concentration of CO_2 of 280 ppm.
16. Derive an energy principle for the Boussinesq system.
17. For a more general fluid the deviatoric forces can be expressed more generally in terms of a stress tensor τ so that the viscosity term in Eq. 2.121 is instead written in terms of the divergence of this stress tensor, cf. Eq. (2.150). Using the definition of the incompressible stress tensor demonstrate that the dissipation term (Eq.2.160), is non-negative. For these manipulations it proves useful to use tensor notation, such that

$$\tau = \mu \left[\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right].$$

18. In Eq (2.145) show that the term involving the gradient of the geopotential can be related to the change in the kinetic energy of the flow. How can one interpret the energetic effect of the time derivative of the geopotential?
19. By adopting typical pressure gradients in the tropics, or the extra tropics, calculate to what extent it is reasonable to neglect geopotential gradients along pressure surfaces when calculating the evolution of η_e in Eq. (2.145).
20. Using the reanalysis calculate the EOFs of the monthly mean vertical velocity field in the tropical atmosphere

over the ocean. How much variance does the leading mode, ω_1 explain. What is the wind profile associated with this vertical mode? How do the EOFs vary if one compares EOF calculated over climatological subsidence versus convective regions?

Further Reading

Section 2.1 Thermodynamics:

The subject is treated in many introductory meteorological texts, but in most cases poorly. The best introductory guide to classical thermodynamics is given by the short series of lectures by Enrico Fermi (1956), which is published by Dover. This book provides a readable and elegant presentation of basic thermodynamic concepts, from a classical perspective. Although the treatment is general, many of the particular concepts of atmospheric thermodynamics, for instance the derivation of the potential temperature or the Clausius-Clapeyron equation, are readily extracted from this book. The definitive modern treatment of classical thermodynamics, with extensions to statistical mechanics, is provided in Herbert B. Callen's beautiful book entitled *Thermodynamics and an Introduction to Thermostatistics* (Callen, 1985) in particular for its more elegant and modern presentation of the laws and their relationship to the existence of quantities like entropy or enthalpy. For a discussion of particular concepts related to atmospheric thermodynamics, and in particular the special role played by moisture, students and practitioners will be well served by the classic (but out of print) book by Iribarne and Godson (1981) or the introductory chapters of *Atmospheric Convection* by Kerry Emanuel (1994). The latter has particularly clear discussions of the moist potential temperatures.

Papers by Pauluis and Held (2002a,b) on the entropy budget of radiative convective equilibrium are an excellent and definitive introduction to an entropy view on the atmospheric circulation. The paper by Kleidon and Renner (2013) provides a very clear overview of the use of entropy and maximum power concepts to understand the hydrological cycle.

Section 2.2 Thermodynamic Diagrams

Visualising the state of the atmosphere, and the development of thermodynamic processes on thermodynamic diagrams can be a useful way to develop understanding. In addition to Iribarne and Godson (mentioned above) a recent paper by Böing et al. (2014) provides insight into the use of thermodynamic diagrams for the study of mixing processes.

Section 2.3 Moist convective instability:

This section is presented in a rather untraditional way, starting from the small scale mixing process associated with what has come to be known a buoyancy reversal. This basic concept has a long history in the literature, but a beautiful and influential presentation of the ideas can be found in a 1980

paper by David A. Randall (1980) The title alludes to the fact that buoyancy reversal is related to traditional ideas of convective instability, the treatment and different measures of which are well presented in the aforementioned book by Emanuel. Buoyancy reversal has always been thought to be an important concept for stratocumulus clouds, and recent work on the topic by Juan Pedro Mellado (2010; 2017) provides the most insightful contributions on the topic from the point of view of fundamental fluid dynamics, as this work has helped clarify long-standing uncertainty related to the role of buoyancy reversal in cloud mixing.

Section 2.4 Fluid Dynamics:

Fluid dynamics is a vast topic. A very good basic text on basic fluid mechanics is that by Kundu and Cohen (2002) for which the early editions are perfectly adequate. For a complete treatment of multi-phase flows, including discussion of the many intricacies associated with the treatment of particles, the article by Peter Bannon (2002) is a good reference. In deriving the Boussinesq or other sound-proof equation systems asymptotic techniques are employed. For a guide to asymptotic approaches in fluid mechanics, through which many of the equation sets used to describe flows in the atmosphere and ocean can be derived, the reader is referred to a recent review article by Rupert Klein (2010). Although work on these topics might seem academic, the question of sound-proof equation sets that relax the requirements on the basic state are coming increasingly into focus as numerical simulation is capable of solving flows with an increasing number of degrees of freedom, which allows for a treatment of small scales over very large domains wherein the background state might be expected to vary greatly.

For an introduction to Reynolds averaging (or filtering approaches in general) and turbulence most books on turbulence will provide a good introduction. Perhaps the best book on the topic with an atmospheric perspective in mind (for a consideration of Reynolds Averaging, and higher order equations) is the book by Wyngaard (2010). Stephen Pope's book (Turbulent Flows, Cambridge University Press, 2000) has become a standard text on turbulence, but the beauty of the subject is perhaps better captured by the Uriel Frisch's Book (Turbulence, the Legacy of A. N. Kolmogorov, Cambridge University Press, 1996) which influenced the presentation of the turbulence cascade in §2.6.1. G. I. Barenblatt's book (Scaling, self similarity, and intermediate asymptotics, Cambridge, 1996) is an excellent introduction to similarity theory. None of these texts adequately cover moist, or multi-phase, turbulent flows as appropriate for clouds.

The discussion of effective heat and moisture sources, originates in the pioneering analysis of Michio Yanai, beginning with his 1973 paper (Yanai et al., 1973). The idea of gross moist stability was introduced by Neelin and Held (1987). A useful and more general review is provided by David Raymond et al. (2009).

Table 2.7. Symbols and key variables used in this chapter

Symbol	Description	Units (SI)
v, \mathcal{V}	specific volume, volume	$\text{m}^3 \text{kg}^{-1}, \text{m}^3$
s, \mathcal{S}	specific entropy, entropy Eq. (2.28)	$\text{J kg}^{-1} \text{K}^{-1}, \text{J K}^{-1}$
h, \mathcal{H}	specific enthalpy, Eq. (2.11)	$\text{J kg}^{-1}, \text{J}$
g, \mathcal{G}	specific Gibbs free energy, Gibbs free energy $G = \mathcal{H} - TS$	$\text{J kg}^{-1}, \text{J}$
δQ	Heating	J
a_θ	$\partial_\theta T_\rho$, Eq. (2.165)	
b	buoyancy	m/s^2
c_e	Eq. (2.22), equivalent-state c_p	J kg^{-1}
c_ℓ	Eq. (2.26), liquid-state c_p	J kg^{-1}
\bar{e}	Turbulence kinetic energy	$\text{m}^2 \text{s}^{-2}$
η_d	Dry Static Energy	J kg^{-1}
η_e	Moist Static Energy Eq. (2.51)	J kg^{-1}
η_ℓ	Liquid-Water Static Energy Eq. (2.52)	J kg^{-1}
η_s	Saturation η_e	J kg^{-1}
κ_1	Interfacial stability criterion	–
κ_2	Buoyancy Reversal parameter, Eq. (2.102)	–
k	Wavenumber, $2\pi/\text{wavelength}$	–, or rad m^{-1}
l	Size, of an eddy.	m
E	Turbulence energy spectral density	m s^{-2}
N^2	Brunt-Väisälä frequency, Eq. (2.89)	s^{-1}
N_s^2	N^2 in saturated air Eq. (2.91)	s^{-1}
Q_1	Apparent enthalpy source	$\text{J kg}^{-1} \text{s}^{-1}$
Q_2	Apparent moisture sink	$\text{kg kg}^{-1} \text{s}^{-1}$
Q_{cnd}	Net condensational heating	$\text{J kg}^{-1} \text{s}^{-1}$
Q_{rad}	Radiative heating	$\text{J kg}^{-1} \text{s}^{-1}$
R_e	Equivalent-state gas constant	$\text{J kg}^{-1} \text{K}^{-1}$
R_ℓ	Liquid-state gas constant	$\text{J kg}^{-1} \text{K}^{-1}$
Γ_s	Saturated adiabatic lapse rate	K m^{-1}
Γ_d	Dry adiabatic lapse rate	K m^{-1}
α_K	Kolmogorov constant	–
β_θ	$(\partial/q_s) \partial_\theta q_s$ e.g., Eq.(2.81) and (2.82)	–
ϵ_1	$R_d/R_v \approx 0.622$	–
ϵ_2	$R_v/R_d - 1 \approx 0.608$	–
ϵ	Dissipation rate of turbulence kinetic energy	$\text{m}^2 \text{s}^{-3}$
γ	Γ_s/Γ_d , Eq. (2.85)	–
χ	Mixing Fraction	–
μ	Dynamic viscosity	N s m^{-2}
p_θ	Reference pressure, $10 \times 10^5 \text{ Pa}$	Pa
ϕ	Geopotential, gz	J kg^{-1}
l_K	Kolmogorov microscale	m
ν	Kinematic viscosity	$\text{m}^2 \text{s}^{-1}$
ϑ	Generic Variable	$[\vartheta]$
a	Convective area fraction Fig. 2.10	–
\mathcal{M}	Convective Mass Flux Eq. (2.110)	$\text{kg m}^{-2} \text{s}^{-1}$
\mathcal{A}	CAPE Eq. (2.105)	J kg^{-1}
\mathcal{I}	Convective Inhibition, Eq. (2.107)	J kg^{-1}
\mathcal{D}	Eq. (2.108), Downdraft CAPE	J kg^{-1}
$\mathcal{G}(\vartheta)$	Gross lapse rate operator	$\text{Pa}^{-1} [\vartheta]$
\mathcal{P}	Mean production of turbulence kinetic energy	$\text{J kg}^{-1} \text{s}^{-1}$
\mathcal{T}	Mean flux of turbulence kinetic energy	$\text{J m kg}^{-1} \text{s}^{-1}$
F_h	Diabatic enthalpy flux	$\text{J m s}^{-1} \text{kg}^{-1}$
τ	Stress tensor	$\text{m}^2 \text{s}^{-2}$
v	Velocity vector (u, v, w)	m s^{-1}
\mathcal{R}_e	Reynolds Number UH/ν	–