

4.1 Background

The parameterization problem involves algorithmically or statistically relating the effects of physical processes that cannot be represented directly in a model to variables that are included. Physical processes are parameterized for a few reasons.

- The small scales involved make it too computationally expensive to represent a process directly.
- The complexity of a process makes it too computationally expensive to represent directly.
- There is insufficient knowledge about how a process works to explicitly represent it mathematically.

The representation of atmospheric processes in models takes place within the dynamical core as well as through the so-called model “physics”. The dynamic processes include the propagation of various types of waves (e.g., advective, Rossby, inertia–gravity). Even though the physics processes are parameterized to a large degree, their correct rendering by a model is nevertheless essential for the prediction of virtually all of the dependent variables. The parameterized processes that are discussed in this chapter include cumulus convection, cloud microphysics, turbulence, and radiation. Land-surface processes are also parameterized because they occur on too small a scale to be represented directly, but they are discussed separately in Chapter 5.

Even though parameterizations are typically developed and discussed independently from each other, and from the dynamical core, this is artificial and should be avoided. This is because parameterizations do interact, and the realism of this interaction determines the accuracy of the model. For example, the parameterized spectral solar radiation represents an energy flux at the land surface, and the land-surface parameterization partitions some of it to the sensible heating of the ground. The resulting land–atmosphere fluxes provide lower-boundary conditions to the surface-layer and boundary-layer parameterizations of turbulence, which distribute heat and moisture throughout the lower atmosphere. And, when the water vapor condenses, parameterizations of convection are relied upon to represent all aspects of the associated subgrid-scale processes. And, microphysical processes that are related to the development of hydrometeors are parameterized for stable precipitation. In turn, convective-cloud and stable-cloud effects on radiation must be parameterized, where this radiation attenuation strongly influences the surface temperature through

the land-surface parameterization. Thus, because of these interdependencies, parameterizations should not continue to be developed in isolation. A more holistic approach is greatly needed if we are to reduce model error.

An issue that will be discussed below is that the performance of some parameterizations can depend on season and the meteorological processes that prevail in specific geographic regions. For example, some convective parameterizations are more appropriate for midlatitudes, while others perform better in the tropics. And, models employed for polar applications will use parameterizations that are different from those that are applicable for midlatitude, coastal-zone simulations, etc. And, the same parameterization is sometimes tuned for specific needs. An obvious related issue is that global models must use the same parameterizations for all geographic areas, thus eliminating the option of choosing ones that best suit a particular region.

Figure 4.1 illustrates how parameterizations fit within the overall framework of a model. The term “resolved” in the upper box refers to grid-scale processes that do not need to be parameterized. An important aspect of this figure is that the primary inputs for parameterizations of any type are the resolved-scale structures of the atmosphere that control the process that is being parameterized. As a very simple example, the resolved static stability near the ground can be used to infer the strength of subgrid-scale turbulence in the boundary layer, which controls the grid-box-average vertical fluxes in the tendency equations for temperature, humidity, and wind. Similarly, layer-average relative humidity from the model can be used to infer the fractional-area coverage of subgrid-scale cumulus clouds, which can be used in the equation that calculates the radiation reaching the surface. Thus, a parameterization relates the resolved-scale input variables to the resolved-scale effects ($\partial\Phi_P/\partial t$ in Fig. 4.1) of the parameterized process. It will be seen that parameterizations can have a wide range of complexities. The middle box in

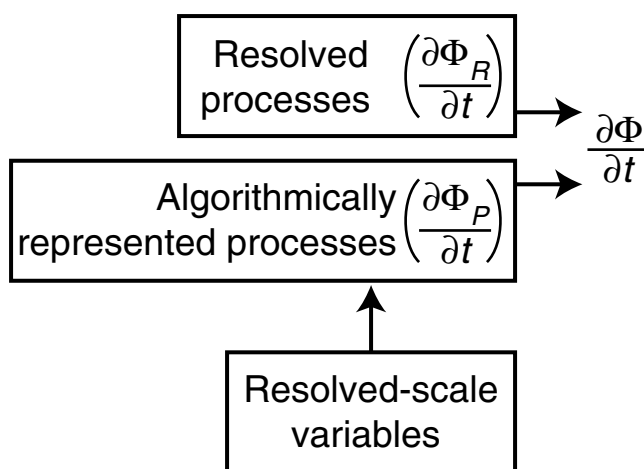


Fig. 4.1

Schematic showing how a predictive equation for a dependent variable Φ has contributions from terms that correspond to resolved processes (subscript R) and parameterized processes (subscript P). The inputs to the parameterizations are the resolved-scale atmospheric variables.

Fig. 4.1 can be a simple look-up table, or it can be so computationally intensive that $\partial\Phi_p/\partial t$ cannot be recalculated at every time step

Refer to Eqs. 2.1–2.6 to see how the effects of parameterized processes are included in the prognostic equations. The momentum equations, Eqs. 2.1–2.3, contain friction terms (Fr), that are separated into viscous and turbulent stresses in Eq. 2.16. The rest of the terms are solved using the methods described in Chapter 3, and their numerical form is part of the dynamical core of the model. But the friction, for example in the boundary layer, results from the existence of the turbulent eddies that cannot be resolved. The boundary-layer parameterization will define these friction terms. Similarly, the diabatic heating–cooling term (H) in the thermodynamic equation, Eq. 2.4, contains contributions from phase changes of water that are defined in the microphysical and convective parameterizations, from heat transport by the parameterized turbulence fluxes near the surface, and from the radiation parameterization.

It is important to be aware that parameterizations are generally developed with certain grid increments in mind. That is, only those aspects of the physical system that are not resolved by the model need to be parameterized. Thus, the modeler should be aware of such assumptions when deciding on parameterizations to employ for a given purpose. There is another general parameterization issue that is related to model resolution: As model grid increments continue to decrease as computing power increases, the models begin to partially (i.e., poorly) resolve some processes that are being parameterized. Thus, there is the risk of “double counting” processes. The consequence of this situation is that there is a range of grid increments for which a process is too poorly resolved to represent explicitly, but there is not a sufficient separation between resolved and parameterized scales for the parameterization assumptions to be valid.

Stensrud (2007) represents the best review available of all the types of parameterizations used in atmospheric models. The discussions in this chapter often follow those in that reference.

4.2 Cloud microphysics parameterizations

Cloud microphysics encompasses all cloud processes that occur on the scales of the cloud droplets and the hydrometeors, rather than on the scale of the cloud itself. The correct modeling of these processes determines the skill with which precipitation type, amount, and spatial distribution are forecast. Similarly, microphysical processes are the cause of the potentially destructive straight-line winds of convective outflow boundaries. And, cloud horizontal and vertical distributions must be modeled well in order for the radiation and surface energy budgets to be predicted with accuracy. Microphysical processes are also critically important in climate modeling. For example, the physical system may respond to increases in greenhouse gases with an alteration in global cloud properties (and albedo), which can have a potential positive or negative feedback relative to the original temperature increase. And, the microphysical impacts of increased natural or anthropogenic aerosols in the atmosphere of a modified climate must be represented in a model because this can change precipitation efficiency.

Historically, stratiform clouds have been explicitly represented in NWP models, with the microphysics parameterized, because their large horizontal extent has allowed them to be resolved by most grids. In contrast, the small horizontal size of most convective clouds relative to typical grid increments has meant that they have been subgrid-scale phenomena, and thus their effects are represented through parameterizations. Thus, the same model parameterizes one type of cloud and explicitly represents another. The two components of the model compete for water vapor, and model output files generally include separate variables for convective precipitation and stable precipitation. This situation still prevails for most operational NWP models, and for climate models.

For research applications, and for some operational LAMs, the horizontal grid increments can be sufficiently small so that the models, called cloud-resolving models, can explicitly represent moist convection on the grid. This allows the same model code, including microphysics parameterizations, to represent all moist processes – a much more appealing situation than the one noted above where different parts of the model apply to convective and stratiform cloud. The horizontal grid increment below which cloud-generating circulations can be explicitly represented by a model is very situation dependent. Weisman *et al.* (1997) suggests that a grid increment of 4 km is sufficient to resolve squall-line convection. However, there will never be a grid increment below which microphysical process will not need to be parameterized. This is because such processes, as we will see, exist on the cloud-droplet and rain-drop scales of micrometers to millimeters; indeed, even the molecular scales are relevant.

4.2.1 Microphysical particles and processes

The following summary is provided for the reader who has not had the benefit of a course in cloud microphysics. The particle types and the microphysical processes that they undergo are important in the context of the generation of precipitation in its various forms, and thus they should be parameterized in some way in atmospheric models. Further information can be obtained from Fletcher (1962), Rogers (1976), Cotton and Anthes (1989), Rogers and Yau (1989), Houze (1993), Pruppacher and Klett (2000), and Straka (2009). The particle types that are involved in microphysical processes are listed below.

- *Cloud droplets* – These are liquid drops, with a typical radius of 10 μm , that form through the condensation of water vapor in the presence of a cloud-condensation nucleus (CCN, small particles that have an affinity for water).
- *Rain drops* – Cloud droplets can grow to rain drops through the accretion mechanism described below, or rain drops can result from the melting of snow crystals. Rain drop radii range from 100 to 1000 μm .
- *Ice crystals* – Water droplets freeze in the presence of an ice nucleus (IN, similar to a CCN) at temperatures below the normal freezing point. Larger droplets freeze at higher temperatures.

- *Aggregates of ice crystals, snow flakes* – These are clusters of ice crystals formed when ice crystals with different terminal velocities collide and coalesce. Snow flakes are formed by this process.
- *Rimed ice particles* – These form when ice crystals collide and coalesce with cloud droplets at temperatures below freezing. If the features of the ice crystal can be distinguished, it is called a rimed ice particle.
- *Graupel particles* – If the crystal features of a rimed ice particle are not recognizable, it is called a graupel particle. Graupel also results from the instantaneous freezing of rain drops, when the sub-freezing drops collide with ice crystals.
- *Hail stones* – As graupel particles fall through the cloud of sub-freezing liquid, they grow by riming. Hail stones result from cases of extreme riming.

Some of the microphysical processes are as follows.

- *Condensation* – Liquid droplets form when water saturation is exceeded at temperatures from -40°C to above freezing. The condensation takes place on CCN that are natural or anthropogenic, typically submicrometer-sized, particles.
- *Accretion* – In the warm-cloud process, droplets with different masses have different terminal velocities, and the resulting collisions between droplets can result in coalescence and droplet growth. As a droplet grows, so does its vertical velocity relative to the smaller cloud droplets, thus increasing the rate of collisions.
- *Evaporation* – Cloud droplets and rain drops evaporate.
- *Ice and snow aggregation* – When ice crystals and snow flakes collide and coalesce, it is called aggregation.
- *Accretion by frozen particles* – Snow, graupel, or hail collect other solid or liquid particles as they fall.
- *Vapor deposition* – The saturation vapor pressure with respect to liquid water is higher than the saturation vapor pressure with respect to ice. Thus, if a cloud that contains both droplets and ice crystals is saturated with respect to water, it is supersaturated with respect to ice. As the ice crystals grow by vapor deposition, the air becomes sub-saturated with respect to the liquid surface, and cloud droplets evaporate. This process is called the Bergeron–Findeisen mechanism.
- *Melting* – As snow flakes fall into the lower troposphere, below the freezing level, they may melt and form rain drops. Similarly, hail and graupel begin to melt as they fall below the freezing level.
- *Freezing* – Water droplets freeze in the presence of IN, riming involves the freezing of water droplets that collide with ice crystals, and rain drops can freeze to form graupel.

Figure 4.2 illustrates the microphysical processes that must be represented in some form in a model in order to predict the types of precipitation shown at the bottom of the figure. A similar diagram can be found in Cotton and Anthes (1989).

4.2.2 Microphysical parameterizations

Microphysical parameterizations aim to represent, as thoroughly as possible, the processes described in the previous section. The parameterizations are divided into two categories, based on how the size distributions of particle types are represented. In *bin models*, the particle size spectrum is divided into intervals, and the particle concentrations are predicted for each interval, or bin. Changes for each bin can result from conversions between particle types, and from the increase or decrease of particle sizes. This requires a predictive equation for each particle type and size bin, which must be solved at each grid point. Thus, the use of bin models is very computationally intensive and is presently limited to research activities, and not operational weather and climate prediction. In contrast, *bulk microphysical parameterizations* assume a prescribed analytic form for the size spectrum of each particle type – e.g., exponential (Kessler 1969) or gamma (Walko *et al.* 1995a) distributions – and the evolutions of the size spectra are obtained by solving predictive equations for the moments. Single-moment, bulk parameterizations only involve prediction of the particle mixing ratio or specific humidity; that is, the ratio of the mass of a particular particle type to the volume or mass, respectively, of the dry air in which the particles are distributed. Double moment schemes predict both the particle mixing ratio and the particle number concentration. Triple-moment schemes add radar reflectivity to the predictive equations, allowing the shape parameter in the gamma distribution to vary independently.

To illustrate an example of how bulk microphysical parameterizations are represented in a single-moment model, the following are predictive equations for the specific humidity of five different forms of water: water vapor (q_v), cloud water (q_{cw}), cloud ice (q_{ci}), snow (q_s), and rain (q_r). Tensor notation is used for notational brevity, such that when a subscript appears twice in the same term it is assumed that the term is summed over all possible subscript values. For example, the first term to the right of the equal sign represents advection in the three space directions, such that u_i equals u , v , and w , and x_i equals x , y , and z for $i = 1, 2, 3$, respectively. The next term to the right is the turbulent mixing term (see Chapter 2) in tensor notation, and is interpreted in the same way. The third terms on the right side in the equations for snow and rain represent the fact that these two types of hydrometeors have significant terminal velocities (V_T). Where there are vertical derivatives in the mass of a species, there will be a contribution to the tendency in proportion to the terminal velocity. The rest of the terms on the right (S) represent various sources and sinks associated with conversions from one type of particle to another.

$$\begin{aligned} \frac{\partial q_v}{\partial t} = & -u_i \frac{\partial q_v}{\partial x_i} - \frac{1}{\rho_0} \frac{\partial}{\partial x_i} \rho_0 \overline{u_i' q_v'} - S_{deps} - S_{depci} \\ & + S_{evapr} - S_{vcondtocw} \end{aligned}$$

$$\frac{\partial q_{cw}}{\partial t} = -u_i \frac{\partial q_{cw}}{\partial x_i} - \frac{1}{\rho_0} \frac{\partial}{\partial x_i} \rho_0 \overline{u'_i q'_{cw}} + S_{vcondtocw} - S_{freezcw}$$

$$-S_{cwtor} - S_{accwbyr} - S_{accwbys}$$

$$\frac{\partial q_{ci}}{\partial t} = -u_i \frac{\partial q_{ci}}{\partial x_i} - \frac{1}{\rho_0} \frac{\partial}{\partial x_i} \rho_0 \overline{u'_i q'_{ci}} + S_{freezcw} + S_{depci} - S_{citos} - S_{accibys}$$

$$\frac{\partial q_s}{\partial t} = -u_i \frac{\partial q_s}{\partial x_i} - \frac{1}{\rho_0} \frac{\partial}{\partial x_i} \rho_0 \overline{u'_i q'_s} - V_{Ts} \frac{\partial q_s}{\partial z} + S_{citos}$$

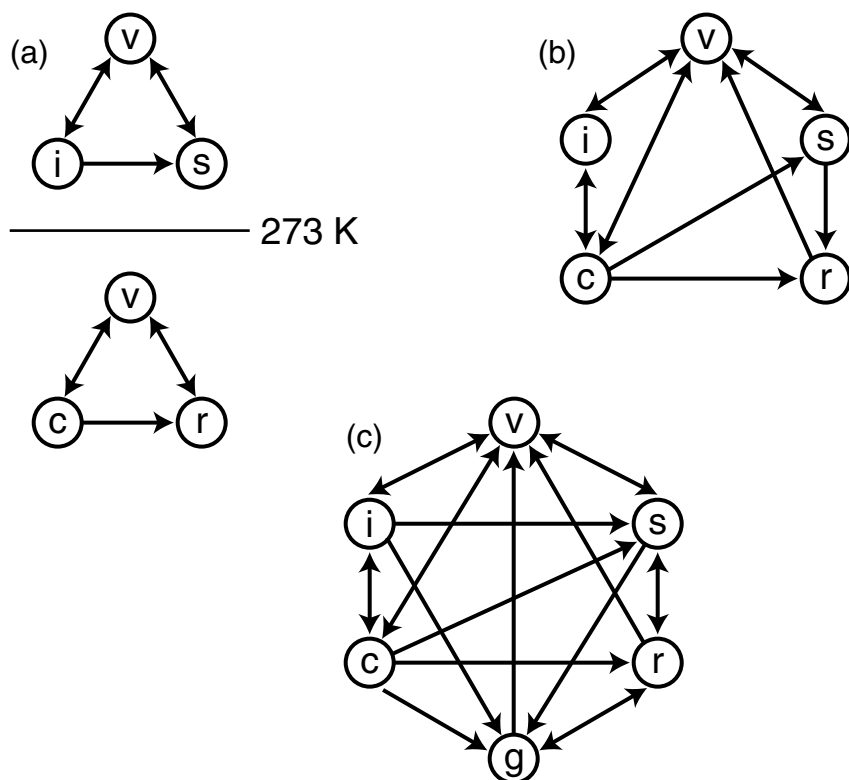
$$+ S_{accibys} + S_{accwbys} + S_{deps} - S_{smelttor}$$

$$\frac{\partial q_r}{\partial t} = -u_i \frac{\partial q_r}{\partial x_i} - \frac{1}{\rho_0} \frac{\partial}{\partial x_i} \rho_0 \overline{u'_i q'_r} - V_{Tr} \frac{\partial q_r}{\partial z} - S_{evapr} + S_{accwbyr} + S_{cwtor} + S_{smelttor}$$

The sources and sinks are defined as follows.

S_{evapr} evaporation of rain drops	$S_{accwbys}$ accretion of cloud water by snow
$S_{accwbyr}$ accretion of cloud-water droplets by rain drops	S_{deps} growth of snow by vapor deposition
S_{cwtor} growth of cloud-water droplets to rain drops by cold-cloud (Bergeron–Findeisen) process	$S_{freezcw}$ freezing of cloud water to produce cloud ice
$S_{smelttor}$ melting of snow to produce rain drops	S_{depci} growth of cloud ice by vapor deposition
S_{citos} growth of cloud ice to snow	$S_{vcondtocw}$ condensation of vapor to form cloud-water droplets
$S_{accibys}$ accretion of cloud ice by snow	

It is within these “S” terms that the parameterizations of microphysical processes are represented. Some schemes are simple, and only a few of the particle types and interactions (conversions) are operative. Others are more complex, with many more interactions. Figure 4.3 illustrates the microphysical processes that are represented in three different example parameterizations, and emphasizes the fact that the number of interactions allowed among particles varies greatly. Shown are the interactions for the Dudhia (1989), Reisner *et al.* (1998), and Lin *et al.* (1983) schemes. The reader should refer to Stensrud (2007) for examples of different approaches for actually representing the different processes.

**Fig. 4.3**

Microphysical processes represented in three different parameterizations: (a) Dudhia (1989), (b) Reisner *et al.* (1998), and (c) Lin *et al.* (1983). Particle types are abbreviated as vapor (v), cloud ice (i), snow (s), cloud water (c), rain water (r), and a combination of graupel and hail (g). The arrows indicate the direction of the particle interactions. The line in panel (a) separates the processes above the freezing level (above the line) from those below the freezing level. From Stensrud (2007).

4.2.3 Initialization of microphysical variables

Ideally it would be possible to initialize the microphysical variables just as we do the other dependent variables. However, there are a couple of impediments to success. First, the existence of specific types of microphysical particles in the atmosphere can only be roughly inferred, at best, based on satellite cloud imagery (cloud ice and water) and various sources of precipitation observations (rain and snow). The vertical distributions of the particles, and their horizontal spatial detail at the cloud scale, are unknown, thus making their initialization very problematic. In addition, the microphysical variables respond quickly to forcing by atmospheric circulations on the cloud scale and larger, so initializing the variables without also including corresponding compatible circulations would be futile. For example, if cloud and precipitation observed along a front are used to initialize microphysical variables, the variables will only be retained by the model if it has the front and associated vertical circulation in the correct location. Without the frontal lifting, the cloud and precipitation will dissipate. Because of the above issues,

microphysical variables are sometimes assumed to have zero concentrations at the initial time of a forecast, where the expectation is that they will spin up to realistic values within the initial 3–6 h. When sequential data-assimilation methods are employed (see Chapter 6), the model simulation that provides the first-guess for the analysis can be used to define the microphysical variables. Or, when continuous data-assimilation, such as Newtonian relaxation, is used, model-generated microphysical variables are automatically part of forecast initial conditions.

4.2.4 Modeling the effects of anthropogenic and natural aerosols on microphysical processes

The atmospheric CCN represent the subset of the general population of aerosols that can nucleate a cloud droplet at a particular water saturation. The ability of a particle to act as a CCN depends primarily on chemical composition and size. Thus, an important issue related to the inclusion of microphysical processes in NWP and climate models is the correct representation of the details of the CCN from natural and anthropogenic sources (Rosenfeld *et al.* 2008). This is challenging from a number of respects. One is that atmospheric aerosol properties are not systematically observed at all. Another is that the complex interactions among aerosol particles, hydrometeors, and cloud dynamics, including the dynamic competition for water vapor among nuclei of different sizes and composition, mean that predicting the specific response of the system to the types and amounts of available aerosols (even if we had that information) can be challenging. An example of the importance of knowing simply the approximate amount of CCN follows. If there are many CCN available, cloud-droplet concentrations can be large. For a given liquid-water content, this means that droplets are smaller, the cloud optical thickness and albedo are higher, and precipitation efficiency is reduced. The lower precipitation efficiency leads to higher cloud liquid-water content, cloud lifetime, and cloud thickness (Albrecht 1989). We thus can have the situation where an increase in CCN availability can lead to a reduction in rain-drop concentration. Complicating the situation are the facts that some aerosols can decrease cloud albedo (Kaufman and Nakajima 1993) and their chemical compositions influence their activation as cloud droplets (Raymond and Pandis 2002). Thus, the lack of our ability to operationally predict aerosol properties has implications for the predictability of microphysical processes and clouds, and weather in general. For example, Taylor and Ackerman (1999) found that the elevations of cloud tops and the microphysical structure of stratus clouds were significantly affected by aerosols emitted by ships into an otherwise clean maritime environment.

Because of long-term trends in anthropogenic aerosols from pollution, and mineral aerosols from desertification, and the fact that aerosols, clouds, and precipitation are critical components of the climate system, it is especially important in climate modeling for aerosols to be simulated in terms of their sources, sinks, and transport. See Levin and Cotton (2009) for a complete summary of aerosol effects on microphysics, and Heintzenberg and Charlson (2007) for their role in climate.

4.3 Convective parameterizations

It is important to be able to accurately simulate moist convection¹ with models for a variety of practical reasons. Intense moist convection can lead to flash flooding, gust fronts, and tornadoes. And, the aggregate effect of individual convective elements is an important component of monsoon circulations, the Hadley and Walker circulations, and the ENSO. These large-scale processes need to be simulated properly in climate predictions and long-range weather predictions, and thus models need to be able to accurately represent the effects of the convection on their resolved scales. And, shallow cumulus clouds dominate the tropics, and are common in other latitudes, greatly impacting the global albedo. The effects of these clouds thus need to be represented in weather and climate models, which must reasonably render the radiation budget.

In general, convective parameterizations activate moist convection at relative humidities lower than water-vapor saturation at a grid point. This is because the convective columns are subgrid scale, so the grid-box-average relative humidity will be sub-saturated even though there are saturated regions within the grid box. In addition to generating grid-box-average values of convective precipitation, the schemes also define the effects of the subgrid-scale convection on other grid-scale variables. The overall objective is for these parameterizations to define convection in the right place and at the right time (with the correct diurnal cycle if applicable), and with the correct evolution and intensity. And, the parameterization should define the appropriate modification by the convection to the large-scale environment, so that subsequent convection can be accurately predicted. Figure 4.4 shows the overall concept, even though Mapes (1997) points out how it is a simplification. In general, the large-scale processes (e.g., low-level convergence, destabilization of a deep layer) control the moist convection, and, in turn, the convection will

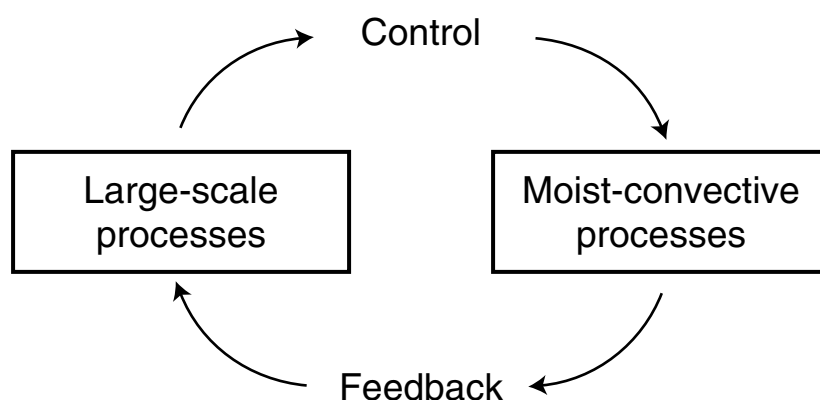


Fig. 4.4

Schematic of the interaction between large-scale processes and moist convection. Adapted from Arakawa (1993).

¹ The term moist convection refers to convection that leads to cloud, and possibly precipitation, formation.

modify the large scale, possibly with the aggregate latent heating from the convective cells maintaining the larger-scale circulation. The purpose of the parameterization is to emulate this process.

Moist convection in the atmosphere can be classified into two types. Deep convection extends vertically over a large fraction of the depth of the troposphere, and is associated with (1) low-level convergence that exists on a scale larger than the individual updrafts and (2) deep conditional instability. In contrast, shallow convection spans only a small fraction of that depth, with cloud tops perhaps a few kilometers above the surface. Precipitating deep convection dries the environment by removing water vapor, and warms it as a result of the compensating subsidence. However, nonprecipitating shallow moist convection has no direct net influence on the environment. Its existence does indirectly impact the environment, however, because the clouds reflect solar energy and the resulting shading of the ground means a cooler boundary layer.

Cloud-resolving models, that are capable of explicitly resolving convective-scale circulations, employ grid increments of perhaps 1 km and are commonly used in research (see Wu and Li (2008) for a review and a comprehensive list of references). For example, Weisman *et al.* (1997) use a model with a grid increment of 4 km that they state explicitly resolves squall-line convection. However, because moist convection consists of a mix of updrafts and downdrafts that often have scales of a few hundred meters to a few kilometers, it will be years before operational global and limited-area weather-forecast models are capable of resolving them. And, it will be decades before global climate models have sufficient horizontal resolution to resolve moist convection. Thus, convective parameterizations will be needed well into the foreseeable future.

4.3.1 Types of convective parameterizations

A common feature of most convective parameterizations is that they calculate the Convective Available Potential Energy (CAPE) and the Convective INhibition (CIN) of the environment in order to estimate the characteristics of convection. Figure 4.5 graphically illustrates these two variables in terms of a typical warm-season sounding on a thermodynamic chart. In this sounding (thin line), the lapse rate is dry adiabatic below 800 hPa, it is isothermal between 800 and 700 hPa, and nearly dry adiabatic above that to about 500 hPa. The heavy solid line shows the temperature of a parcel that is lifted from the surface, through the Lifting Condensation Level (LCL) to the Level of Free Convection (LFC). Between the LCL and LFC the parcel is colder and more dense than its environment, and is thus negatively buoyant. Energy is required to lift the parcel against this downward force. This amount of energy is, by definition, the CIN, and is proportional to the striped area in the figure. Mathematically, CIN is defined as follows:

$$CIN = -g \int_{SL}^{LFC} \frac{\theta(z) - \bar{\theta}(z)}{\bar{\theta}(z)} dz,$$

where θ is the potential temperature of a parcel rising dry or moist adiabatically from its Starting Level (SL) to the LFC, $\bar{\theta}$ is the potential temperature of the environment, and the

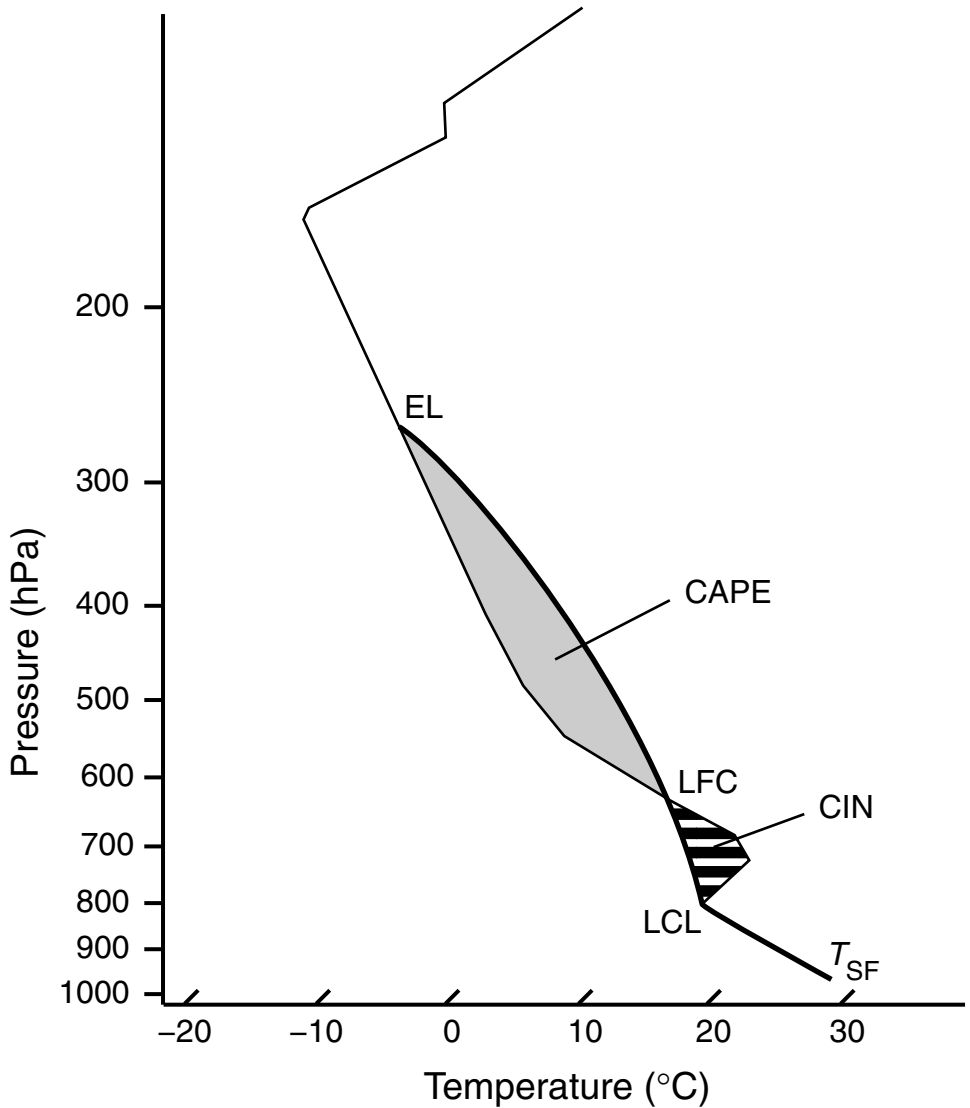


Fig. 4.5

Example of a warm-season temperature sounding (fine line) with a dry adiabatic layer below 800 hPa, an isothermal layer from 800 to 700 hPa, and another near-dry-adiabatic layer up to about 500 hPa. The heavy solid line defines the temperature of a parcel that is lifted from the surface through the Lifting Condensation Level (LCL) and the Level of Free Convection (LFC), to the Equilibrium Level (EL). The areas are shown that define the Convective Available Potential Energy (CAPE) and the Convective INhibition (CIN). The surface temperature is T_{SF} .

negative sign exists so that CIN is positive when energy is required to lift the parcel to its LFC. If there is sufficient energy to lift the parcel above the LFC, it will rise buoyantly along the heavy line until it reaches the Equilibrium Level (EL) where the parcel is neutrally buoyant. It is assumed with this simple parcel theory that there is no mixing between the parcel and its environment. The CAPE is the buoyant energy available to an ascending

parcel as it rises from the LFC to its EL where it loses its buoyancy, and is proportional to the shaded area in the figure. It is defined mathematically as

$$CAPE = g \int_{LFC}^{EL} \frac{\theta(z) - \bar{\theta}(z)}{\bar{\theta}(z)} dz.$$

Thus, for convection to exist there must be available CAPE to provide the buoyant energy to accelerate parcels upward, and there must be a method by which parcels overcome the prevailing CIN.

There are a large number of ways of categorizing convective-parameterization schemes, including the following.

- There are general approaches to the convective-parameterization problem, and schemes are identified in terms of whether they follow a particular one. For example, schemes that adjust the environmental vertical temperature profile when the relative humidity exceeds a threshold and the temperature profile is unstable are called moist-convective adjustment schemes. Sometimes schemes are identified in terms of the first author to publish an approach. For example, methods that produce convection based on resolved-scale moisture convergence are called Kuo-type schemes.
- A classification suggested by Mapes (1997) is based on whether the development of convection is controlled by the creation of CAPE or the removal of CIN. So-called deep-layer-control schemes, also termed equilibrium-control schemes, tie the development of convection to the creation of CAPE by large-scale processes. In these methods, the convection is assumed to maintain the instability in the large-scale environment in a state of equilibrium that is near neutrality. Alternatively, low-level-control schemes, also called activation-control schemes, relate convection to the removal of CIN. In fact, many approaches include elements of both low-level and deep-layer controls.
- Some convective parameterizations represent the effects of only deep moist convection (most schemes discussed here), while others apply to only shallow convection (e.g., Albrecht *et al.* 1979, Deng *et al.* 2003, Bretherton *et al.* 2004). Some apply to both types (Tiedtke 1989, Gregory and Rowntree 1990, Betts and Miller 1993, and Kain 2004).
- The schemes can be classified in terms of the environmental, grid-scale, variables that are affected by the convection. Most schemes only define the impact on the environmental temperature and humidity, but some also treat effects on the momentum (e.g., Fritsch and Chappell 1980, Han and Pan 2006).
- Some methods directly define the final state of the environment after the convection has effected the change, while others attempt to simulate the process by which the change takes place. The former, generally more simple, approaches are called static schemes, while the latter are referred to as dynamic schemes.
- A distinction among methods is the nature of the so-called trigger function. This is the set of criteria in the parameterization that prescribes where and when the parameterized convection will be activated. The importance of this component of a convective parameterization was demonstrated by Kain and Fritsch (1992), who tested five different trigger functions in the same model and in the same parameterization, the

Kain–Fritsch scheme (Kain and Fritsch 1993), for the same meteorological case. There were substantial differences in the simulated parameterized convection for the different trigger functions. A similar dependence was found by Stensrud and Fritsch (1994).

- The scales resolved by the models are a way of classifying these schemes, such that there are mesoscale-model parameterizations and coarse-grid-model parameterizations. The particular distinction is that the mesoscale models (grid increments of 5–50 km) have sufficient horizontal resolution to explicitly resolve mesoscale circulations associated with the convection, where examples include thunderstorm outflow boundaries, mesohighs and mesolows, rear-inflow jets, and midlevel vortices associated with mesoscale convective systems. Thus, the mesoscale models need to only parameterize the convective-scale processes (e.g., Stensrud and Fritsch 1994, Zheng *et al.* 1995), whereas coarser-grid models must parameterize both the mesoscale and the convective-scale processes, and their many interactions (Frank 1983).

A given convective parameterization can be classified in terms of a number of the methods in the above incomplete list.

4.3.2 Scale considerations

In the above discussion is mentioned the relationship between model resolution and the nature of the convective parameterization. But, the parameterization problem also depends on other scale issues. In particular, Frank (1983) elaborates on a discussion in Ooyama (1982) regarding the relationships between convective parameterizations and the scale of the convective process. Figure 4.6 illustrates the scale regions that have relevance to the parameterization of convection. The abscissa is the physical length scale (L) of the process, and the ordinate is the dynamic length scale expressed in terms of the Rossby radius of deformation (R). The latter length scale is defined as

$$R = \frac{NH}{(\zeta + f)^{1/2}(2Vr^{-1} + f)^{1/2}},$$

where N is the Brunt–Väisälä frequency, H is the scale height of the circulation, ζ is the relative vorticity, f is the Coriolis parameter, V is the rotational component of the wind, and r is the radius of curvature of streamlines. Within region I, with length scales of less than 10 km, are individual convective cells and clouds. Regions I and II pertain to phenomena, referred to as dynamically small, where $L < R$, and region III pertains to dynamically large phenomena where $L > R$. Frank (1983) states that the parameterization problem is somewhat simpler for region-III processes because there is a stronger relationship between the large-scale flow and convection. For example, if a cold-front's strength and position are simulated correctly, the associated convective rainfall will be straightforward to represent with a parameterization. A parallel argument is that, when latent heating from convection affects the mass field in dynamically large systems, the system will adjust through changes

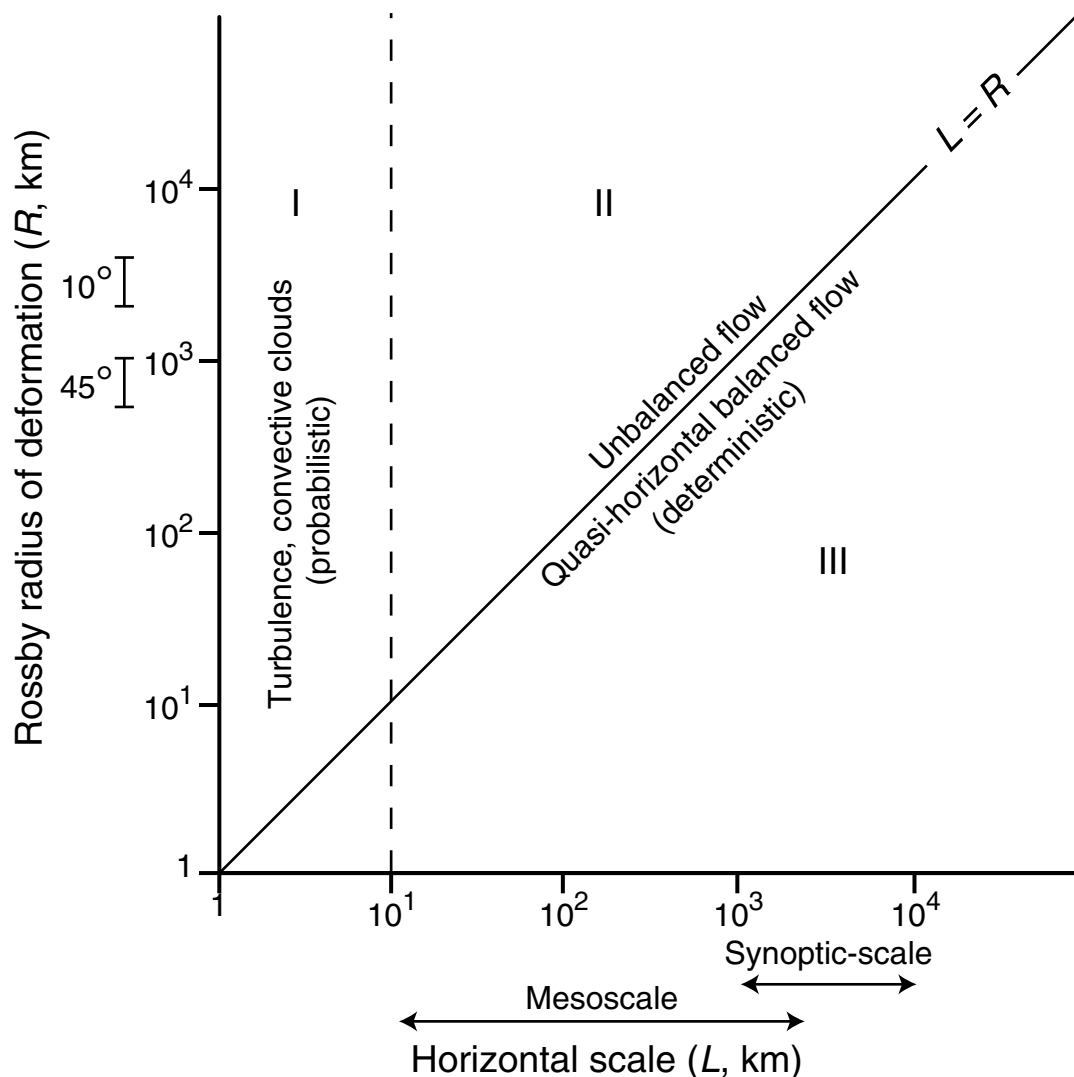


Fig. 4.6

Schematic showing three regimes (I, II, III) of atmospheric circulation, defined in terms of the prevailing physical length scale (abscissa) and the dynamic length scale (ordinate), where the latter is represented by the Rossby radius of deformation. The intervals defined on the ordinate indicate the values of R that are typical at the noted latitudes. Adapted from Frank (1983), and originally from Ooyama (1982).

in the rotational component of the wind.² Because the rotational wind component is not strongly related to convection, there are no links, or only weak links, to secondary convective processes. In contrast, for dynamically small systems, adjustment of the mass field to

² On dynamically large scales, the mass field will “dominate” the geostrophic-adjustment process, and the imbalance caused by the latent heating will produce an adjustment toward a geostrophic value in the rotational component of the wind. See Section 6.10.1 for a discussion of the geostrophic-adjustment process.

the latent heating will cause divergent circulations that will influence the future evolution of convection, making the process more difficult to parameterize.

4.3.3 Relationship between the subgrid-scale (convective) precipitation parameterization and the resolved-scale precipitation

For all but very-high-horizontal-resolution models that can explicitly represent convective cells, models generally employ both convective and microphysical (Section 4.5.1) parameterizations. This means that precipitation can be produced by the model both when the convective parameterization is triggered and when the explicit processes represented in the microphysical parameterization produce resolved-scale precipitation that reaches the ground. In the former case, subgrid-scale precipitation is represented by the parameterization at the model-resolved grid scale, for sub-saturated grid-box conditions. In the latter case, grid-box saturation is required at some point in the column. There are two precipitation variables defined on the model grid at the surface; one is the parameterized convective precipitation and one is the resolved-scale precipitation. Even though these are summed to produce a total-rainfall field, model developers often look at both output fields separately to help them better understand internal model processes. This dual treatment of precipitation processes, by two generally distinct components of the model, leads to some conceptual and real difficulties. For example, the convective parameterization often does not produce cloud water and ice on the grid scale, even though precipitation has been generated, and thus no radiative effects of the clouds are rendered in the model. And, precipitation generated in one geographic region of a meteorological event will be produced by the parameterization, and by the explicit microphysics code in another geographic area. For example, in a mesoscale convective system, the microphysics parameterization may produce precipitation in the trailing stratiform-precipitation region, while the convective parameterization represents the precipitation elsewhere. Or, in an extra-tropical cyclone the microphysics and convective parameterizations could predominate in the precipitating regions of the warm and cold fronts, respectively. Figure 4.7 illustrates how the partitioning of precipitation from these two sources can depend on the meteorological event and the convective parameterization. Shown in both panels is the ratio (percentage) of the convective precipitation to the total precipitation, for 36-h simulations that employed four different convective parameterization schemes in a model that was otherwise the same. The left panel applies to simulations of a mesoscale convective system that occurred in the spring season, and the right one pertains to an Arctic front in the winter, with some convection in the warm air mass. There were clearly great differences among the simulations in terms of the partitioning of the precipitation between the resolved and subgrid components. For example, for the mesoscale convective system (a), the use of the Anthes–Kuo parameterization caused virtually all of the precipitation to be produced by the subgrid-scale mechanism, whereas when the Grell parameterization was used a large percentage of the precipitation was produced by the resolved-scale mechanism. Relationships between the two parts of the model that simulate precipitation processes are clearly not simple, nor is it always easy to anticipate which one will dominate for a particular case. And, these results make it clear that it is not reasonable to equate

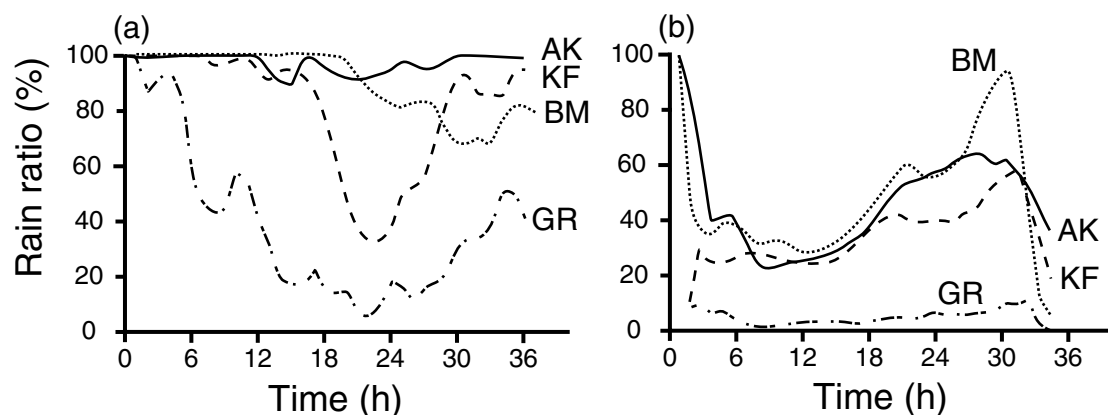


Fig. 4.7

Ratio (percentage) of subgrid-scale precipitation to total precipitation (sum of subgrid and resolved precipitation) for simulations of two meteorological cases, where four different convective parameterizations were used for each case. The horizontal grid increment was 36 km, and the ratios are based on totals for the computational area. Panel (a) pertains to simulations of a mesoscale convective system that occurred in May, and panel (b) pertains to simulations of an Arctic front in February, with some convection in the warm air mass. The four convective parameterizations were the Grell (GR; Grell 1993, Grell *et al.* 1994), Kain–Fritsch (KF; Kain and Fritsch 1993), Betts–Miller (BM; Betts and Miller 1986), and Anthes–Kuo (AK; Anthes 1977, Grell *et al.* 1994) schemes. Adapted from Wang and Seaman (1997).

resolved and subgrid precipitation produced in the model to stratiform and convective precipitation in the atmosphere.

In spite of the common lack of a direct link between the convective and resolved-scale precipitation code in a model, there are exceptions. For example, hybrid methods partition a fraction of the parameterized precipitation from the convective scheme to the grid-scale precipitation defined in the microphysics scheme (e.g., Frank and Cohen 1987).

4.3.4 Summary of example convective-precipitation parameterizations

Many different convective parameterizations have been developed and used in models of various scales (e.g., Arakawa and Schubert 1974; Kuo 1974; Kreitzberg and Perkey 1976; Anthes 1977; Brown 1979; Fritsch and Chappell 1980; Molinari and Corsetti 1985; Betts and Miller 1986; Frank and Cohen 1987; Tremback 1990; Grell 1993; Kain and Fritsch 1993; Janjić 1994, 2000; Grell and Dévényi 2002; Kain 2004).

The following are commonly used schemes. The very brief descriptions are meant only to illustrate some of the high-level properties of the methods. Stensrud (2007) and Wang and Seaman (1997) should be consulted for more-lengthy summaries.

Grell scheme

This is a variant of the Arakawa–Schubert parameterization (Arakawa and Schubert 1974), and is a deep-layer-control scheme. In the Arakawa–Schubert method, shallow and deep cumulus clouds, with a spectrum of sizes in each grid box, are idealized as plumes. In

contrast, the Grell scheme only uses one cloud size, which is justified given that the applications are on the mesoscale. The subgrid precipitation is calculated by

$$P = Im(1 - \beta),$$

where I is the condensate in the updraft, m is the updraft mass flux at cloud base, and $(1 - \beta)$ is the precipitation efficiency. The latter is assumed to be a function of the resolved environmental wind shear in the lower troposphere. Convective downdrafts are parameterized.

Anthes–Kuo scheme

This parameterization (Anthes 1977, Grell *et al.* 1994) is a variation of one of the earliest convective schemes (Kuo 1965, 1974), and uses a column-integrated moisture convergence (M) to determine the location and intensity of convection. When conditional instability exists and the moisture convergence exceeds a threshold, convection is initiated. Because the parameterized convection is based on the source of buoyant energy, it is classified as a deep-layer-control scheme. The moisture that is converging in the column is partitioned into convective precipitation and moistening of the column. The precipitation rate (P) is calculated by

$$P = (1 - b)M,$$

where

$$b = 2(1 - \overline{RH})$$

and \overline{RH} is the column-mean relative humidity. This is a computationally undemanding scheme, which is one reason for its somewhat enduring popularity. It is, however, not especially well founded because moisture convergence does not necessarily result in convective activity. There are better schemes now available.

Betts–Miller scheme

The Betts–Miller scheme (Betts and Miller 1993, Janjić 1994) is another deep-layer-control scheme that, upon initiation of convection, adjusts the model profiles of temperature and moisture in each grid column toward specified reference profiles that correspond to a quasi-equilibrium condition that is associated with deep convection (Betts 1986). The parameterized precipitation is calculated by

$$P = \int_{P_B}^{P_T} \frac{q_R - q}{\tau g} dp,$$

where q is the grid-point specific humidity, q_R is based on the deep-convection reference profile for specific humidity, τ is the time scale over which the adjustment occurs, and the P_T and P_B are the pressures at the top and bottom of the cloud.

Kain–Fritsch scheme

The Kain–Fritsch scheme (Kain and Fritsch 1993) is an updated version of the Fritsch–Chappell scheme (Fritsch and Chappell 1980). Here, the activation of convection is defined by low-level forcing, and is also a function of the CAPE at a grid point. So it is both a low-level- and deep-layer-control scheme. The convective precipitation is calculated as

$$P = ES,$$

where E is the precipitation efficiency and S is the sum of the vertical fluxes of vapor and liquid at about 150 hPa above the LCL.

4.3.5 The choice of convective parameterization, and its impact on the simulation

Convective parameterizations employ a wide variety of approaches to the problem, and assumptions, and they inevitably perform best for those situations where the assumptions are better satisfied. This can make them dependent on the geographic area and the prevailing meteorological process. For example, some parameterizations seem to work best in the tropics, or in midlatitudes, or in high latitudes. Unfortunately, the parameterizations in global models must perform adequately for all climates and weather scenarios.

The point was made earlier that parameterization methods used for coarse-resolution models must represent both the convective-scale processes as well as the mesoscale processes that are related to the convection. In contrast, schemes used in mesoscale models only need to parameterize the convective scales. In a nested system of grids in a LAM, which may span resolutions from the synoptic scale to the mesogamma scale, it is not unreasonable to use different parameterizations on the different grids. On a very-high-resolution grid in a nest, which can explicitly resolve convection, it would be appropriate to not use any parameterization. Unfortunately, most convective parameterizations were designed for models with grid increments of 20–30 km or larger. Even though there is evidence that some can still be used with grid increments as small as 10 km, there currently is no good solution to the problem of how to represent convection between that resolution and those that are needed to explicitly resolve convection.

To illustrate the potential sensitivity of the accuracy of the precipitation forecast to the choice of the convective parameterization, see Fig. 4.8. The average total (convective and resolved) rain rate is plotted for a spring-season convective event (panel a), based on observations, and for five simulations that used different treatments for the convection – four different parameterizations, and no parameterization. For all simulations, the model grid increment was 12 km. At specific times in the simulations, the rain rate varied by as much as a factor of three or four among the different parameterizations. Also depicted is the bias score averaged for three warm-season convective events (panel b), again for each of the four parameterizations and for the use of no parameterization. The horizontal grid increment was 36 km, representing an appropriate resolution for the use of any of the parameterizations tested. Both the simulation-average scores on the right, as well as the

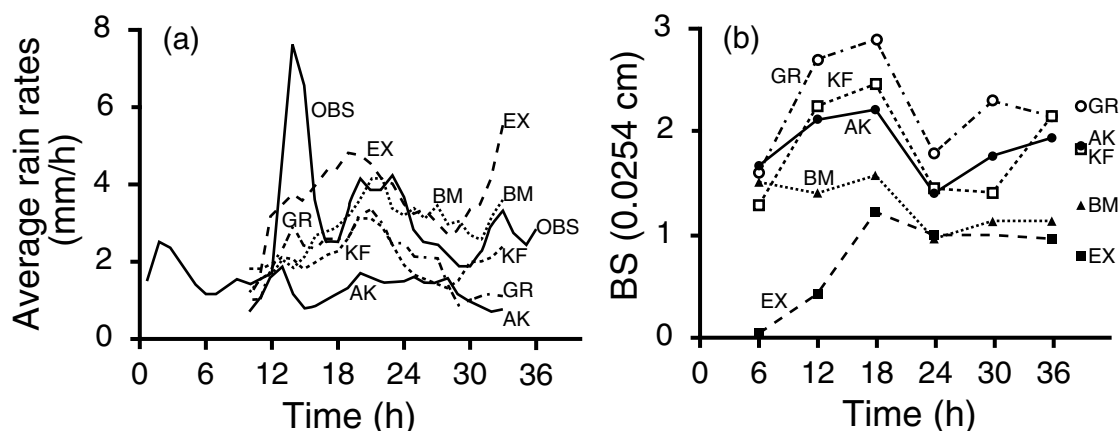
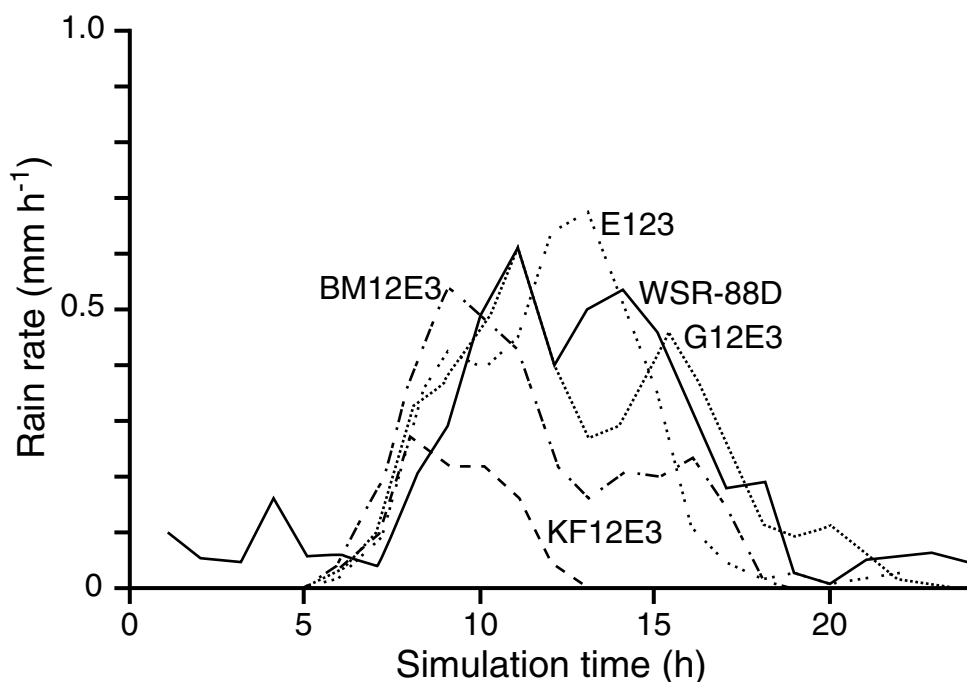


Fig. 4.8

Average rain rate for a spring-season convective event (a), based on observations (OBS), and for five simulations that used different treatments for the convection – four different parameterizations, and no parameterization (EX). For all simulations the model grid increment was 12 km. Also depicted is the bias score averaged for three warm-season convective events (b), again for each of the four parameterizations and for the use of no parameterization. The horizontal grid increment was 36 km. The four convective parameterizations were the Grell (GR), Kain–Fritsch (KF), Betts–Miller (BM), and Anthes–Kuo (AK) schemes. Adapted from Wang and Seaman (1997).

time-dependent curves, show a substantial dependence of the precipitation amount on the parameterization that was employed. The Betts–Miller scheme produced a simulation-average bias of close to unity, whereas the Grell scheme had a bias that exceeded two. The simulation that used no convective parameterization (explicit – EX) severely under-predicted the early precipitation amounts because of the time required to develop grid-scale saturation. This is one of hundreds of examples in the literature of the dependence on many factors of convective-parameterization performance.

Not only is model-simulated precipitation sensitive to the parameterization scheme employed on that grid, it has been shown that precipitation from cloud-resolving simulations on a fine grid in a nest is sensitive to the convective parameterization that is employed on the surrounding coarser grids. For example, Warner and Hsu (2000) describe tests with an operational LAM having three grids in a nest. The two coarser grids had grid increments of 10 km and 30 km, and thus required the use of a convective parameterization. In contrast, the innermost grid, with a grid increment of 3.3 km, explicitly represented the convection (i.e., no parameterization was used). For a model simulation of summer convection in the southwestern USA, Fig. 4.9 shows the grid-average hourly rain rate produced on the convection-resolving grid of this model when three different convective parameterizations were used on the surrounding grids. Also shown is the rainfall on that grid when no parameterization was used on the outer grids, as well as the rainfall estimated by the reflectivity from the WSR-88D radar. Even though the model used on the inner grid was identical in all four simulations, there clearly was a large impact of the choice of the convective parameterization used on the other grids. Through LBC effects, the parameterizations on the coarser grids caused stabilization and drying to various degrees on the convection-resolving grid, resulting in substantial differences in the simulated precipitation.

**Fig. 4.9**

Grid-average hourly rain rates on the inner, convection-resolving grid (grid 3) of a nested LAM, when three different convective parameterizations were used on the surrounding grids (grids 1 and 2). Also shown is the rainfall on that grid when no parameterization was used on the outer grids (E123), as well as the rainfall estimated by the reflectivity from the WSR-88D radar. The lines are labelled in terms of the convective parameterization used: BM is Betts–Miller, G is Grell, KF is Kain–Fritsch, and E is explicit. The model configuration used on the inner grid was identical in all four simulations. Adapted from Warner and Hsu (2000).

4.4 Turbulence, or boundary-layer, parameterizations

4.4.1 Boundary-layer structure

At the lower boundary of the troposphere is the turbulent layer through which the influence of the surface is directly transmitted to the free atmosphere above. Through this boundary layer, or mixed layer, turbulent eddies transport water vapor and heat upward from their source at the surface. Also, the frictional stress exerted by the surface on the atmospheric fluid is transmitted by the turbulence. There are two causes of turbulence, or sources of turbulent energy. One is the buoyancy that creates rising parcels of air, or convection, and the compensating subsidence, when the land surface is heated during the day. The other source is related to the rate of change of the horizontal wind speed with height – i.e., the vertical shear of the horizontal wind. When this shear is small, and there is no buoyancy, the flow is nonturbulent, or laminar. When the shear exceeds a threshold, the flow becomes turbulent, with the turbulent energy derived from the mean wind. The

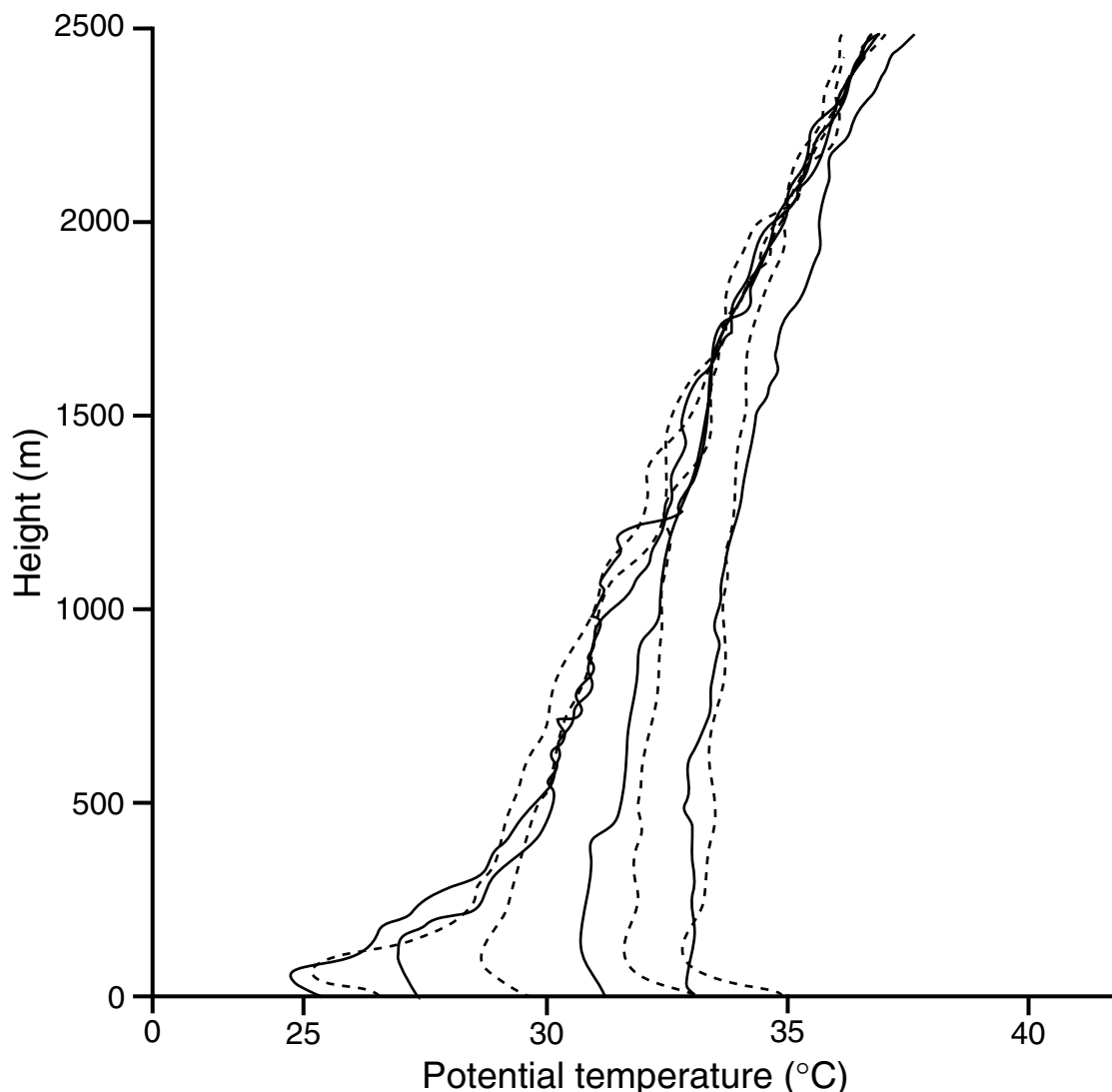
buoyancy-driven convective turbulence is dominant during the day, while the shear-driven turbulence is more common at night. Because the wind speed perpendicular to any surface, including the ground, must be zero, turbulence cannot exist at the surface and cannot transport heat or moisture there. Thus, in a very shallow layer of a few molecules to a few millimeters above the surface, called the laminar sublayer (also, microlayer), transfers of heat, moisture, and surface-frictional effects are through molecular processes. Thus, the laminar sublayer is the nonturbulent interface between the ground and the turbulent mixed layer, and the mixed layer is the turbulent interface between the laminar sublayer and the free atmosphere above the mixed layer. The lower 50–100 m of the mixed layer, where the turbulent transport of heat, moisture, and momentum vary relatively little (compared to the situation in the mixed layer above), is called the surface layer.

The vertical extent of the turbulent mixing defines the daytime (convective) boundary-layer depth. At night, the atmosphere near the surface cools, and the source of the buoyant energy is eliminated. Any new turbulent energy in this stable layer near the ground must now be derived from the vertical shear in the horizontal wind. Unless the horizontal wind is exceptionally strong, this nocturnal boundary layer is much more shallow than the daytime one. The turbulence and the well-mixed profiles of the different meteorological variables penetrate progressively upward during the daytime heating cycle. Figure 4.10 shows measured profiles of the vertical structure of potential temperature at a few times during the daytime heating cycle in the Great Basin Desert in the USA. As the heating continued during the day, the depth of the approximately constant potential-temperature layer progressively increased. At night, as the cooling land surface in turn cools the lowest layer of the atmosphere, a temperature inversion forms.

Figure 4.11 illustrates distinctions between nocturnal turbulence that is generated from wind shear alone, and daytime turbulence that results from both buoyant motion as well as wind shear. Both curves show the variation with time of the vertical inclination of the wind at 29 m Above Ground Level (AGL), based on bivane measurements.³ The lower curve, with relatively small-amplitude and high-frequency excursions from the horizontal direction, shows the effect of nocturnal turbulence that results only from the vertical shear of the horizontal wind. In contrast, the upper daytime curve shows similar high-frequency variability, but it is superimposed on a lower-frequency change with a period of perhaps 15–60 seconds. The longer-period changes during the day are associated with larger horizontally moving turbulent eddies.

In purely laminar flow, the layers of air slide over each other without much mixing between them. The only mixing that occurs is through the exchange of molecules between layers. Molecules from a slower-moving layer nearer the ground enter a faster-moving layer above, and the slower speed of the molecules represents a drag that slows down the upper layer. Analogously, faster-moving molecules move downward, with the drag effect causing the lower layer to speed up. Molecules of water vapor move between layers, causing a net transfer from moist layers to dry layers, and heat is transferred by virtue of the

³ A bivane is a wind vane with two axes of rotation, one horizontal and one vertical.

**Fig. 4.10**

Four simultaneous pairs of radiosonde-based potential-temperature profiles for a salt flat (solid line) and a vegetated, sandy site (dashed line) in the Great Basin Desert in September 1997 on a relatively clear calm day. There were four pairs of radiosonde ascents during the day at 0830 LT (0850 for the salt-flat sounding), 1000 LT, 1200 LT, and 1400 LT (left to right in the figure). The two locations are about 20 km apart. Provided by Elford Astling, US Army Dugway Proving Ground.

different kinetic energies of the molecules exchanged between layers. In nonturbulent flow, this is how the layers of air “feel” each other. With turbulent flow, there are eddies that mix the air between the layers, with this type of mixing being much more efficient than the molecular mixing of laminar flow. One can imagine the vertical exchange of properties with turbulent mixing in the same way as with molecular mixing. For example,

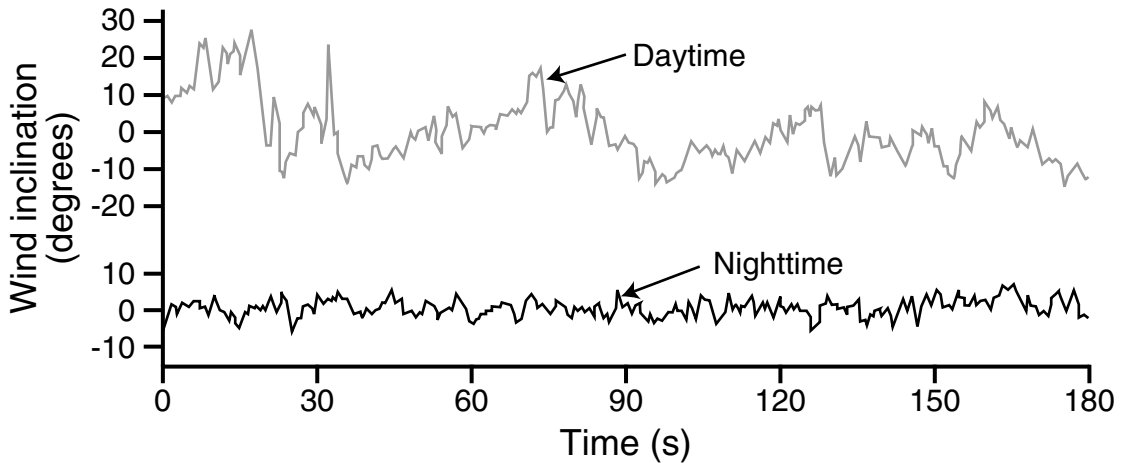


Fig. 4.11 The variation with time of the vertical inclination of the wind at 29 m AGL, based on bivane measurements, for daytime and nighttime conditions. The mean horizontal wind speed was $3\text{--}4\text{ m s}^{-1}$. From Priestley (1959).

because water-vapor content near the surface generally decreases with height, upward-moving air in the turbulent eddies will contain more water vapor and downward-moving air will contain less water vapor.

Figure 4.12a is a schematic of the geometry of the daytime (convective) and nocturnal (stable) mixed-layer structure, and of the transitions between the two regimes. During the daylight hours, the convective mixed layer will increase in depth as the surface heating generates buoyancy-driven turbulence that erodes upward into the troposphere. The depth will typically reach about 1 km, but may span the entire troposphere in strongly heated deserts. After sunset, the ground and the lower atmosphere cool, and the buoyant source of turbulent energy diminishes. The nocturnal, or stable, mixed layer derives most of its turbulent energy from the wind shear, with the depth of the layer being considerably less than in the daytime. In contrast to the daytime, at night the mixing can be intermittent. The shear will develop to a critical value; mixing will abruptly ensue and decrease the shear to a subcritical value, shutting off the mixing; the shear will then increase again; etc. At the ground, this process is manifested as periods of calm that are occasionally interrupted when moderate or strong winds are briefly mixed downward from above. Above the stable, nocturnal boundary layer there exists residual turbulence from the daytime mixed layer, with the intensity decaying with time as a result of internal friction within the fluid.

Figure 4.12b shows typical vertical daytime profiles of wind speed (u), water-vapor density (ρ_v), and potential temperature (θ). Parcels of unsaturated air that are mixed upward or downward by turbulence, cool and warm, respectively, at the dry adiabatic lapse rate. Thus, the well-mixed vertical profile of temperature is the dry adiabatic lapse rate, which is known as a “neutral” temperature profile. In these conditions, the potential temperature is uniform with height. The temperature itself decreases at about $10\text{ }^{\circ}\text{C km}^{-1}$ in this region. Within the surface layer, closer to the ground, the temperature decreases even more rapidly

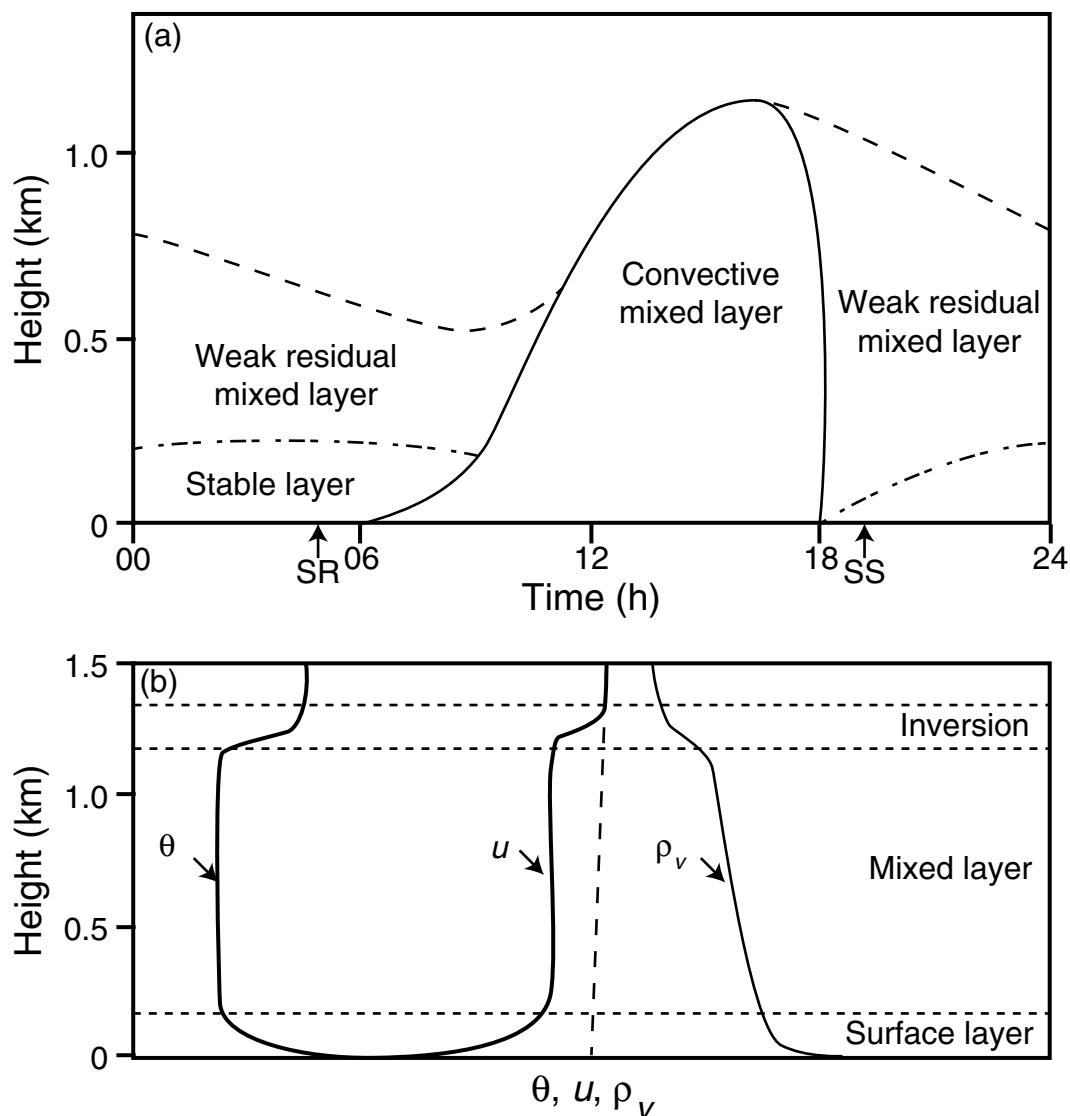


Fig. 4.12

Schematics of (a) the typical diurnal variation of the boundary-layer structure and (b) typical vertical profiles within the daytime boundary layer of potential temperature (θ), horizontal wind speed (u), and water-vapor density (ρ_v). The times of sunrise (SR) and sunset (SS) are shown in (a). The dashed line in (b) represents the wind speed that would exist without friction between the atmosphere and ground. Adapted from Oke (1987).

with height at a superadiabatic rate. At the top of the mixed layer is often a potential temperature inversion, within which potential temperature increases with height. The transition from uniform potential temperature to the inversion above is often used to define the depth of the mixed layer based on radiosonde soundings. Wind speed increases rapidly with height within the surface layer, from zero at the ground, and remains relatively uniform

within the mixed layer. Throughout the midlatitude troposphere above the mixed layer, the climatological north–south temperature contrast causes the wind speed to increase with height up to the tropopause. The dashed line represents the value that the wind speed would attain without the retarding effect of friction, which is transmitted through the boundary layer by turbulence. Above the mixed layer in the free atmosphere, where turbulence does not transmit the frictional stress of Earth's surface, the wind speed is greater. The water-vapor content, here defined in terms of the density of the water vapor, is fairly uniform within the mixed layer, but it does decrease somewhat with height because the source is at the surface and entrainment mixes in drier air from above the boundary layer.

Internal structures within boundary layers

Daytime convective boundary layers and the nighttime residual mixed layers are represented in Fig. 4.12 as simple structures with smooth temperature lapse rates. However, various factors can cause a considerable amount of internal structure to exist. First, when the boundary layer contains layers of dust, the radiative heating and cooling effects of the dust impact the vertical temperature profile. Even if the dust appears to be uniformly distributed throughout the boundary layer, the vertical sorting of different particle sizes and mineral types (having different optical properties) can produce vertical differences in heating and cooling rates. Another factor is the development of internal boundary layers that result from air flowing over surfaces with contrasts in properties such as the heat flux or roughness. For example, if the horizontal wind transports boundary-layer air from a smooth, hot surface to a cooler, rougher one, an internal boundary layer that is forced by the rougher surface develops within the original boundary layer. That is, the different surface roughness and heat flux would produce an internal boundary, within the mixed layer, across which the wind and temperature profiles would differ. This boundary would intersect the surface at the edge of the temperature and roughness contrast, and rise with increasing distance downstream. There are always subtle to major contrasts in surface properties, so whenever the boundary-layer air moves horizontally these internal boundary layers will complicate the structure. For example, the potential-temperature profiles in Fig. 4.10 show considerable variability within the boundary layer.

Aerodynamic roughness and the vertical wind profile

The near-surface turbulent fluxes of heat, moisture, and momentum are influenced by the structure and spacing of surface-roughness elements such as rocks, vegetation, and soil grains. In general, rougher surfaces cause more-intense turbulence. Expressions representing the effect of the turbulence on the vertical wind profile in the surface layer employ a parameter called the roughness length (z_0) to describe the roughness characteristics of the surface. In particular, it can be shown that, for conditions of neutral stability (strong convective mixing),

$$u(z) = \frac{u_*}{k} \ln \frac{z}{z_0}, \quad (4.1)$$

where u is the speed of the mean wind at height z , u_* is the friction velocity, k is the von Karman constant with a value that is thought to be 0.35–0.40, and z is the height above the

ground. The friction velocity represents the drag of the atmosphere against Earth's surface, or the frictional stress. Recall that the surface layer, where this equation applies, is the lower 50–100 m of the mixed layer where the turbulent fluxes of heat, moisture, and momentum vary relatively little in the vertical.

If z in Eq. 4.1 is set to z_0 , u is equal to zero. Thus, z_0 is the height above the ground at which the mean wind speed goes to zero in neutral conditions, and is proportional to the roughness of the surface. Because u_* is not a function of height in the surface layer and k is a constant, u increases logarithmically with increasing z . Figure 4.13 shows a typical vertical profile of horizontal wind speed for neutral conditions (i.e., a solution to Eq. 4.1). If the equation is solved for u_* , it is clear that u_* is linearly related to the slope of the line ($u/\ln(z/z_0)$). The y -intercept of the line is z_0 . Also shown are profiles of u with height for unstable and stable conditions; that is, with temperature lapse rates greater than and less than, respectively, the neutral (dry adiabatic) value. For stable conditions the profile is concave downward, and for unstable conditions it is concave upward.

Obviously it is important to be able to estimate the roughness length in order to apply this equation. Bagnold (1954) determined that the roughness length over a bare flat sand surface is approximately equal to the mean diameter of the sand grains, divided by 30 (roughly 10^{-5} m for typical sand). When vegetation is present, the roughness lengths are larger and more difficult to obtain (Driese and Reiners 1997). Even though studies that estimate roughness lengths over agricultural fields with regularly spaced plantings are abundant, there are fewer studies of less-homogeneous natural environments.

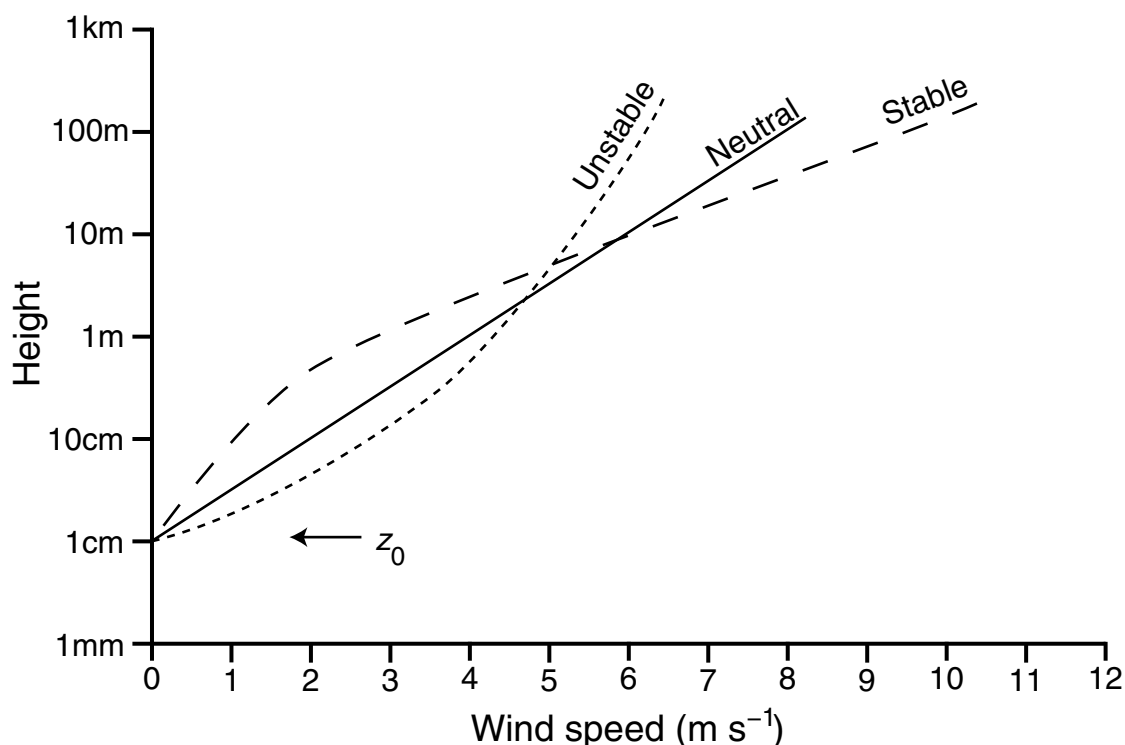


Fig. 4.13 Wind speed as a function of height within the surface layer, for neutral, stable, and unstable conditions. From Stull (1988).

Boundary layers that are detached from the surface

All boundary layers originate because of the influence of a surface with which they are in contact. However, boundary layers do not necessarily remain in contact with the surface over which they form, with important consequences for the weather. An example of one such situation is shown in Fig. 4.14, where a boundary layer first develops over the high desert plateau of northern Mexico. When southwesterly lower-tropospheric winds cause this heated layer of air to move toward the lower terrain elevations to the northeast, the boundary layer becomes detached from the surface. When this happens, a temperature inversion, i.e., a stable layer of air, forms at the base of the Elevated Mixed Layer (EML). This stable layer can inhibit the

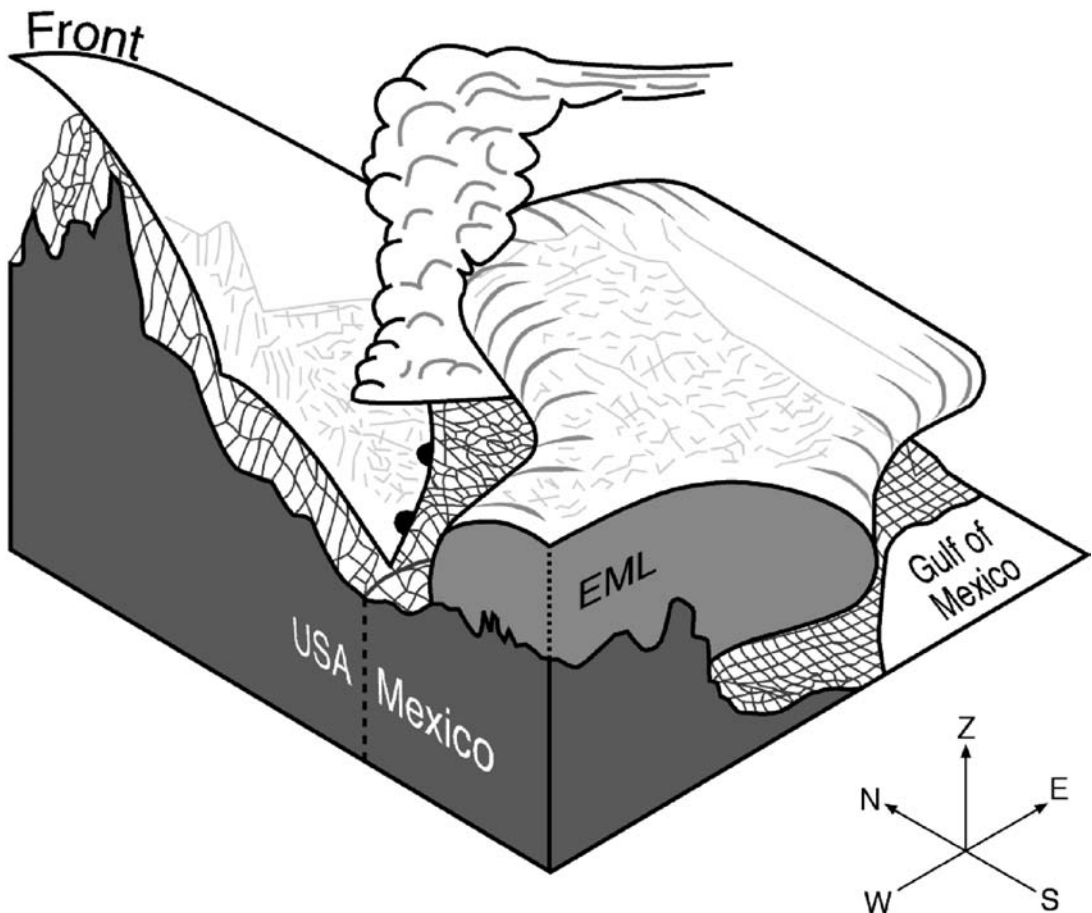


Fig. 4.14

Schematic of an EML-related severe-weather environment over the southern Great Plains of North America. In the foreground is the high plateau of Mexico, from which a mixed layer is flowing to the northeast and becoming elevated over the Great Plains of the USA. The northwest quadrant of the area is the southern Rocky Mountains. Convective clouds are seen on the northwest edge of the EML, associated with low-level southeasterly flow causing moist, unstable air to run out from under the convection-inhibiting inversion at the base of the EML. The meteorological case discussed in the text had a front positioned to the east of the Rocky Mountains, as illustrated. Adapted from Lakhtakia and Warner (1987).

development of convective rainfall because parcels of air that are buoyant in the neutral lapse rates near the surface are no longer buoyant when they encounter this warm elevated mixed layer. The large CIN associated with this thermal “lid”, represented by the hot boundary-layer air, allows moisture and heat to build up in the layer between it and the ground. If the convective motions erode through the inversion, or leak out around it, the consequence is a sudden release of energy that can lead to severe convective storms. This process produces severe convective weather in the semi-arid Great Plains of the USA in the spring. There are many other areas in the world where mixed layers become elevated, with important potential consequences for downstream precipitation. These situations must be represented in models in order for precipitation forecasts to be accurate.

4.4.2 Boundary-layer parameterization closures

The momentum equations in tensor notation are

$$\frac{\partial u_i}{\partial t} = -u_j \frac{\partial u_i}{\partial x_j} - \delta_{i3} g + f \epsilon_{ij3} u_j - \frac{1}{\rho} \frac{\partial p}{\partial x_i}, \quad (4.2)$$

where the viscous-stress terms have not been included. As noted earlier, we sum over repeated indices that appear with variables that are multiplied by each other. The Kronecker delta, δ_{mn} , equals 1 when $m = n$ and is zero otherwise. The alternating unit tensor is defined as

$$\begin{aligned} &+1, \quad \text{if } i, j, k \text{ are in ascending order} \\ \epsilon_{ijk} &= -1, \quad \text{if } i, j, k \text{ are in descending order} \\ &0, \text{ otherwise,} \end{aligned}$$

where ascending order means that i, j, k are 1, 2, 3 or 2, 3, 1 or 3, 1, 2. Descending order is 3, 2, 1 or 2, 1, 3 or 1, 3, 2. The value is zero if any of the i, j, k are the same.

In the calculation of the Reynolds' equation for the u component of momentum, shown in Eq. 2.15, the first step was to separate each dependent variable in Eq. 2.1 into a perturbation and a mean component, where it was assumed that the perturbation is associated with turbulence. Making this same substitution into Eq. 4.2, produces

$$\frac{\partial}{\partial t}(\bar{u}_i + u'_i) = -(\bar{u}_j + u'_j) \frac{\partial}{\partial x_j}(\bar{u}_i + u'_i) - \delta_{i3} g + f \epsilon_{ij3}(\bar{u}_j + u'_j) - \frac{1}{(\bar{\rho} + \rho')} \frac{\partial}{\partial x_i}(\bar{p} + p').$$

Expanding this, and making the assumption that $\rho' \ll \bar{\rho}$, produces

$$\frac{\partial \bar{u}_i}{\partial t} + \frac{\partial u'_i}{\partial t} = -\bar{u}_j \frac{\partial \bar{u}_i}{\partial x_j} - u'_j \frac{\partial \bar{u}_i}{\partial x_j} - \bar{u}_j \frac{\partial u'_i}{\partial x_j} - u'_j \frac{\partial u'_i}{\partial x_j} - \delta_{i3} g + f \epsilon_{ij3} \bar{u}_j + f \epsilon_{ij3} u'_j - \frac{1}{\bar{\rho}} \frac{\partial \bar{p}}{\partial x_i} - \frac{1}{\bar{\rho}} \frac{\partial p'}{\partial x_i}. \quad (4.3)$$

Now, applying the Reynolds' postulates as well as the continuity equation, as in Section 2.2, produces

$$\frac{\partial \bar{u}_i}{\partial t} = -\bar{u}_j \frac{\partial \bar{u}_i}{\partial x_j} - \delta_{i3} g + f \epsilon_{ij3} \bar{u}_j - \frac{1}{\bar{\rho}} \frac{\partial \bar{p}}{\partial x_i} - \frac{\partial}{\partial x_j} \overline{u'_i u'_j}. \quad (4.4)$$

The Reynolds' stress term on the right represents the effects of turbulence on the mean motion, where the quantity $u'_i u'_j$ is called a double correlation or a second statistical moment. Given that the model equations only predict the mean quantities, these covariances need to be defined in some way. There are two approaches for accomplishing this. One involves linking the magnitude of this (second-moment) term to the resolved-scale, or mean (first-moment), variables. The other requires the development of a predictive equation for the covariances.

To obtain a predictive equation for the covariances, begin by subtracting Eq. 4.4 from Eq. 4.3, producing a predictive equation for the turbulent velocity components:

$$\frac{\partial u'_i}{\partial t} = -u'_j \frac{\partial u'_i}{\partial x_j} - \bar{u}_j \frac{\partial u'_i}{\partial x_j} - u'_j \frac{\partial \bar{u}_i}{\partial x_j} + f \epsilon_{ij3} u'_j - \frac{1}{\bar{\rho}} \frac{\partial p'}{\partial x_i} + \frac{\partial}{\partial x_j} \overline{u'_i u'_j}. \quad (4.5)$$

Multiply this by u'_k , apply an averaging operator and Reynolds' postulates, to produce an equation for the second term on the right in the following predictive equation for the covariance:

$$\frac{\partial}{\partial t} \overline{u'_i u'_k} = \overline{u'_i \frac{\partial}{\partial t} (u'_k)} + \overline{u'_k \frac{\partial}{\partial t} (u'_i)}. \quad (4.6)$$

Then change all i indices to k in Eq. 4.5, multiply every term by u'_i , and again apply an averaging operator and Reynolds' postulates, producing an equation for the first term on the right in Eq. 4.6. Add the two equations to produce the following predictive equation for the covariance:

$$\begin{aligned} \frac{\partial}{\partial t} \overline{u'_i u'_k} = & -\bar{u}_j \frac{\partial}{\partial x_j} \overline{u'_i u'_k} - \overline{u'_i u'_j} \frac{\partial \bar{u}_k}{\partial x_j} - \overline{u'_k u'_j} \frac{\partial \bar{u}_i}{\partial x_j} - \frac{\partial}{\partial x_j} \overline{u'_i u'_k u'_j} \\ & + f(\epsilon_{kj3} \overline{u'_i u'_j} + \epsilon_{ij3} \overline{u'_k u'_j}) - \frac{1}{\bar{\rho}} \left[\frac{\partial}{\partial x_i} \overline{p' u'_k} + \frac{\partial}{\partial x_k} \overline{p' u'_i} - p' \left(\frac{\partial u'_i}{\partial x_k} + \frac{\partial u'_k}{\partial x_i} \right) \right]. \end{aligned}$$

This represents six different predictive equations, for $\overline{u'u'}$, $\overline{u'v'}$, $\overline{u'w'}$, $\overline{v'v'}$, $\overline{v'w'}$, and $\overline{w'w'}$. Unfortunately, a triple-correlation or third-moment term now exists on the right side of the equation. If predictive equations are derived for these moments, quadruple correlations will appear on the right side. This situation where there are always more unknowns than equations, requiring that the unknown terms be represented as some function of the known variables, defines the turbulence-parameterization closure problem. Table 4.1 summarizes the equations for the different statistical moments.

Table 4.1 Example of the prognostic equations for the first three statistical moments, indicating the number of equations and the number of unknowns.

Example prognostic variable	Statistical moment	Equation	Variable parameterized	Number of equations	Number of unknowns
\bar{u}_i	First	$\frac{\partial \bar{u}_i}{\partial t} = \dots - \frac{\partial}{\partial x_j} \overline{u'_i u'_j}$	$\overline{u'_i u'_j}$	3	6
$\overline{u'_i u'_j}$	Second	$\frac{\partial}{\partial t} \overline{u'_i u'_j} = \dots - \frac{\partial}{\partial x_k} \overline{u'_i u'_j u'_k}$	$\overline{u'_i u'_j u'_k}$	6	10
$\overline{u'_i u'_j u'_k}$	Third	$\frac{\partial}{\partial t} \overline{u'_i u'_j u'_k} = \dots - \frac{\partial}{\partial x_m} \overline{u'_i u'_j u'_k u'_m}$	$\overline{u'_i u'_j u'_k u'_m}$	10	15

When there are predictive equations for the first moments of the state variables (u , v , w , T , etc.), and the covariance terms (e.g., $\overline{u'v'}$) are parameterized in terms of the first moments, it is called a first-order closure. With second-order closure methods, there are predictive equations for both the state variables and the covariances, and the triple correlations are parameterized in terms of the first and second moments. Thus, the order of the closure is defined in terms of the highest-order prognostic equations that are retained. Stull (1988) illustrates the use of correlation triangles to summarize the unknowns for different orders of closure (Table 4.2). Note that these two tables apply only to the momentum

Table 4.2 Correlation triangles illustrating the unknowns associated with the different levels of the turbulence closure, for the momentum equations only

Order of closure	Correlation triangle of unknowns
Zero	$\begin{array}{c} \bar{u} \\ \bar{v} \quad \bar{w} \end{array}$
First	$\begin{array}{c} \overline{u'u'} \\ \overline{u'v'} \quad \overline{u'w'} \\ \overline{v'v'} \quad \overline{v'w'} \quad \overline{w'w'} \end{array}$
Second	$\begin{array}{c} \overline{u'u'u'} \\ \overline{u'u'v'} \quad \overline{u'u'w'} \\ \overline{u'v'v'} \quad \overline{u'v'w'} \quad \overline{u'w'w'} \\ \overline{v'v'v'} \quad \overline{v'v'w'} \quad \overline{v'w'w'} \quad \overline{w'w'w'} \end{array}$

Source: From Stull (1988).

equations, and thus there are more unknowns when considering the full set of equations. A motivation for higher-order closures is the assumption that the more moments that are predicted, the more accurate will be the solution for the state variables. That is, the higher the moment that is defined by a parameterization, the less the approximation will affect the principal forecast variables, the first moments.

There also are closure methods wherein some of the terms in a particular moment category are parameterized and some are explicitly predicted. For example, in the prognostic equations for the first moments, some second moments on the right side may be parameterized while others are predicted. If all the second moments are predicted, the closure would be second order. If they are all parameterized, it would be first order. Thus, in this case it would be referred to as a 1.5 order method. There are other noninteger closures.

Regardless of the order of the closure, there are two different approaches that can be used in the parameterization. In one, the unknown quantity at a grid point is defined in terms of known quantities, or their vertical derivatives, at the same grid point. Of course the derivative would have to be calculated using adjacent grid points in the vertical. This is called a *local closure*, where such methods have been used through the third order. Alternatively, the unknown variable at a grid point can be estimated using known quantities that are defined at locations that are a significant distance away in the vertical. This is a *nonlocal closure*. The nonlocal closures and the higher-order local closures generally yield more-accurate solutions than do the lower-order local closures, but the former involve greater computational expense and model-code complexity (Stull 1988). An advantage of the second-order or higher closure methods is that the second moments of the wind components can be used to quantify the total Turbulent Kinetic Energy (TKE), such that

$$\frac{TKE}{m} = \frac{1}{2} \overline{u_i u_i},$$

where m is mass. The TKE variable is useful for any application that requires knowledge of the turbulence intensity (such as dispersion of air pollution and turbulence loading on structures).

4.4.3 Local closures

There are many, many different local and nonlocal closure approximations in use in models, and this section and the next one will only provide a couple of illustrations. A common local closure approximation is referred to as K-theory or gradient-transport theory. In a first-order closure, we must parameterize the second moments. Assume the following generic predictive equation for a variable ξ :

$$\frac{\partial \bar{\xi}}{\partial t} = \dots - \frac{\partial}{\partial x_j} (\overline{\xi' u_j'}).$$

A closure approximation for the flux $\overline{\xi' u_j'}$ is

$$\overline{\xi' u_j'} = -K \frac{\partial \bar{\xi}}{\partial x_j}, \quad (4.7)$$

where the parameter K is a scalar with units of m^2s^{-1} . For positive K , the above equation states that the flux $\bar{\xi}'u'_j$ is down the local gradient of $\bar{\xi}$. Combining the above equations leads to

$$\frac{\partial \bar{\xi}}{\partial t} = \dots + \frac{\partial}{\partial x_j} K \frac{\partial \bar{\xi}}{\partial x_j},$$

which now has no undefined variables on the right. The entire system of equations would be closed because there is a prognostic or diagnostic equation for each variable. The coefficient K is referred to by a variety of names, including the eddy viscosity, the eddy diffusivity, the eddy-transfer coefficient, the turbulent-transfer coefficient, and the gradient-transfer coefficient. Different K values are sometimes associated with the transfer of different variables, so that we sometimes see the K s written as K_m , K_E , and K_H , for momentum, moisture, and heat, respectively. Obviously the values of K control the turbulent flux, and intuition tells us that a parameterization of K in terms of the first moments would link it to the wind shear (Richardson number) and the static stability, quantities that are related to the mechanical and buoyant production of turbulence, respectively. Stull (1988) reviews many methods for the parameterization of K in the context of this local closure. That reference also reviews many other local closures.

Because the turbulent flux is defined by local conditions, this simple closure works best for situations when the turbulent eddies are small and locally generated. For large eddies, the nonlocal closures described in the next section provide better results.

4.4.4 Nonlocal closures

Nonlocal-closure methods are motivated by the recognition that much of the mixing in the boundary layer can be associated with large eddies whose vertical dimension is approximately that of the boundary-layer depth, and that these eddies are not related to the local static stability or wind shear at some point in the middle of the boundary layer. Rather, these eddies are driven by the deeper mean stability that spans the boundary layer, which in turn responds to the surface heat flux. Figure 4.15 is a schematic that illustrates the distinction between local closures and a couple of different types of nonlocal closures. In Fig. 4.15a, which applies to local closures, the unknown higher-moment terms at the middle grid point in the column are parameterized through the use of known variables at that point, or through derivatives of the known variables whose calculation requires values at adjacent grid points.

One framework for treating the nonlocal-closure problem is called transilient turbulence theory (Stull 1988, 1993). Imagine a column of model grid boxes, where a small subset of the boxes is shown in Fig. 4.15, and identify a single reference grid box within the column. And, assume that we can define the grid boxes above and below from which turbulent eddies mix air into the reference grid box, as well as the grid boxes that receive air from the reference grid box. Figure 4.15b shows mixing between a reference grid box at the midpoint of the column, and all the surrounding boxes. The same process can be used to define the vertical turbulent mixing between all the other grid points and their

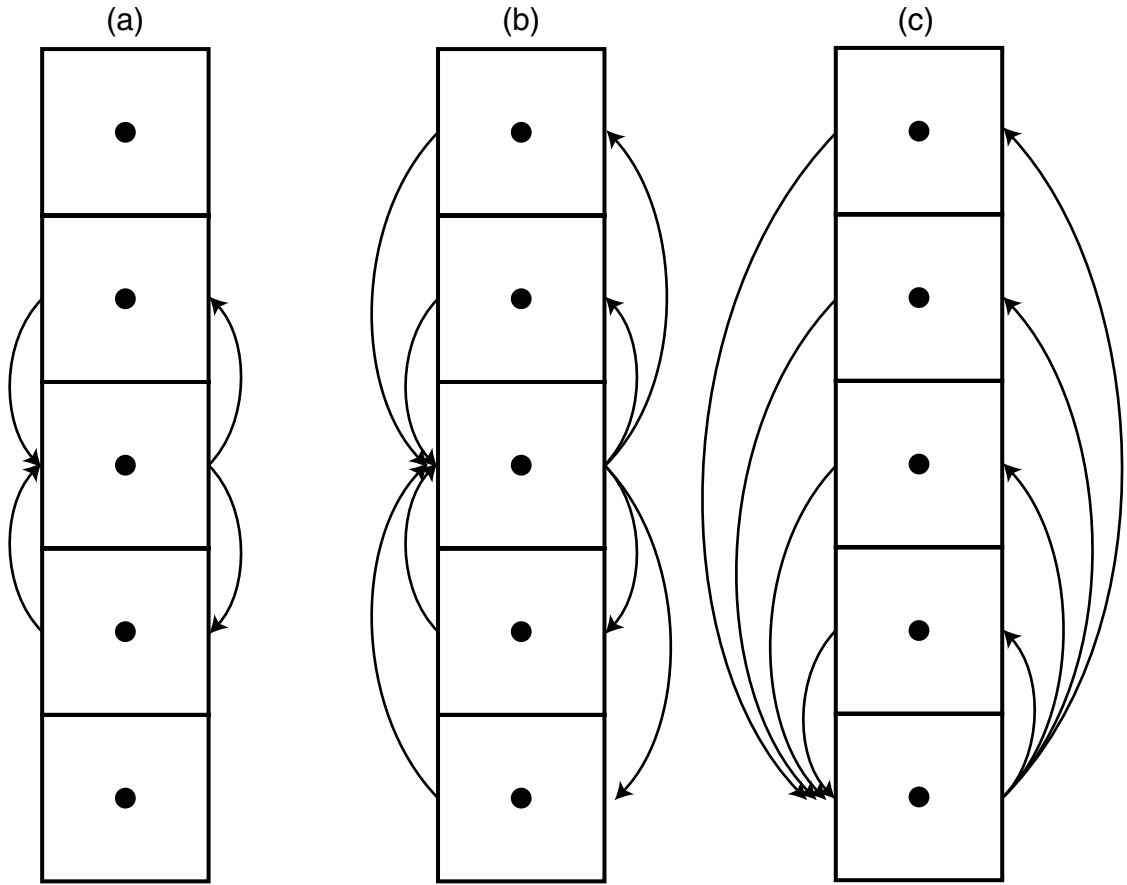


Fig. 4.15

Schematic of the distinction between local closures and two different types of nonlocal closures. Panel (a) applies to local closures, and (b) and (c) pertain to nonlocal closures. See the text for details.

surroundings. Given this conceptual view of the situation, the process by which heat, water vapor, or a passive tracer are transferred among boxes can be quantified using the transilient-turbulence concept and the notation of Stull (1988). Let $\bar{\xi}_i$ represent the concentration of a passive tracer in a reference grid box i , let c_{ij} be the fraction of air in box i that is transported by turbulence from the donor box j during a time period Δt , and let $\bar{\xi}_j$ be the concentration of the tracer in donor box j . To find the new concentration in box i after the elapsed time, we simply sum the contributions from all N grid boxes in the column, such that

$$\bar{\xi}_i(t + \Delta t) = \sum_{j=1}^N c_{ij}(\Delta t) \bar{\xi}_j(t).$$

This equation defines the exchange between each box i and all the other boxes j . It is a matrix equation, where c_{ij} is an $N \times N$ matrix of mixing coefficients called a transilient

matrix, and $\bar{\xi}_i$ and $\bar{\xi}_j$ are $N \times 1$ matrices (vectors). Because the transilient matrix is a function of the turbulence, it is the same for all variables ξ . See Stull (1993) for discussions of many nonlocal parameterizations in the context of transilient-turbulence theory.

An example of a simple nonlocal closure is described in Blackadar (1978) and Zhang and Anthes (1982). Here, the intensity of the vertical convective transfer of heat, moisture, and momentum during the day is determined from the surface heat flux and the thermal structure of the entire mixed layer (not the local thermal structure). Figure 4.15c, which applies to this method, shows that the vertical exchange is visualized as taking place between each model layer in the boundary level and the lowest model layer. That is, both the small and the large eddies have their roots in the surface layer. An example of a turbulent-transfer term is as follows:

$$\frac{\partial \theta}{\partial t} = \dots m(\theta_a - \theta),$$

where θ is the local potential temperature at some model level in the boundary layer, θ_a is the potential temperature at the top of the surface layer (~ 10 m AGL), and m is a function of the surface heat flux and represents the fraction of the mass in a grid box in the column that is exchanged with the surface grid box during a specific time interval.

Figure 4.16 shows another way of viewing the distinction between nonlocal and local closures. On the left (a) is a vertical profile of the mean potential temperature within and above a forest canopy during the daylight hours (a convective mixed layer). There is a shallow inversion within the canopy, with an unstable layer above that, which transitions to a near-neutral layer. Above this, near the top of the boundary layer, the profile becomes stable. The vertical dashed lines show the deep movement of air parcels (open circles) within the boundary layer. The demarcations along the three vertical lines to the right of the sounding, also in panel (a), show how the layer is divided into turbulent and nonturbulent (laminar) flow regimes, and into stability regimes according to whether local or nonlocal methods are used in the definition. In panel (b) are shown heavy vertical arrows (heat flux) that indicate the directions and magnitudes of the vertical fluxes of heat that are observed under these conditions. Also shown are the heat fluxes estimated using local and nonlocal closures. A local-closure approximation for the vertical heat flux, such as Eq. 4.7 with $\bar{\theta} = \bar{\xi}$ and $j = 3$, would be

$$\overline{\theta'w'} = -K_H \frac{\partial \bar{\theta}}{\partial z},$$

such that the direction and magnitude of the flux is defined by the vertical gradient of the potential temperature. Use of this approach results in the “Local static stability” layering (a) and the “Local interpretation” of the heat flux directions (b). These heat fluxes are clearly inconsistent with the observed fluxes shown with the large arrows (b). In contrast, when the static stability is defined across the depth of the boundary layer by comparing the potential temperature of the near-surface parcel with those of the parcels above, to reflect the deep-layer stability, the more-realistic nonlocal static stability and heat-flux profiles

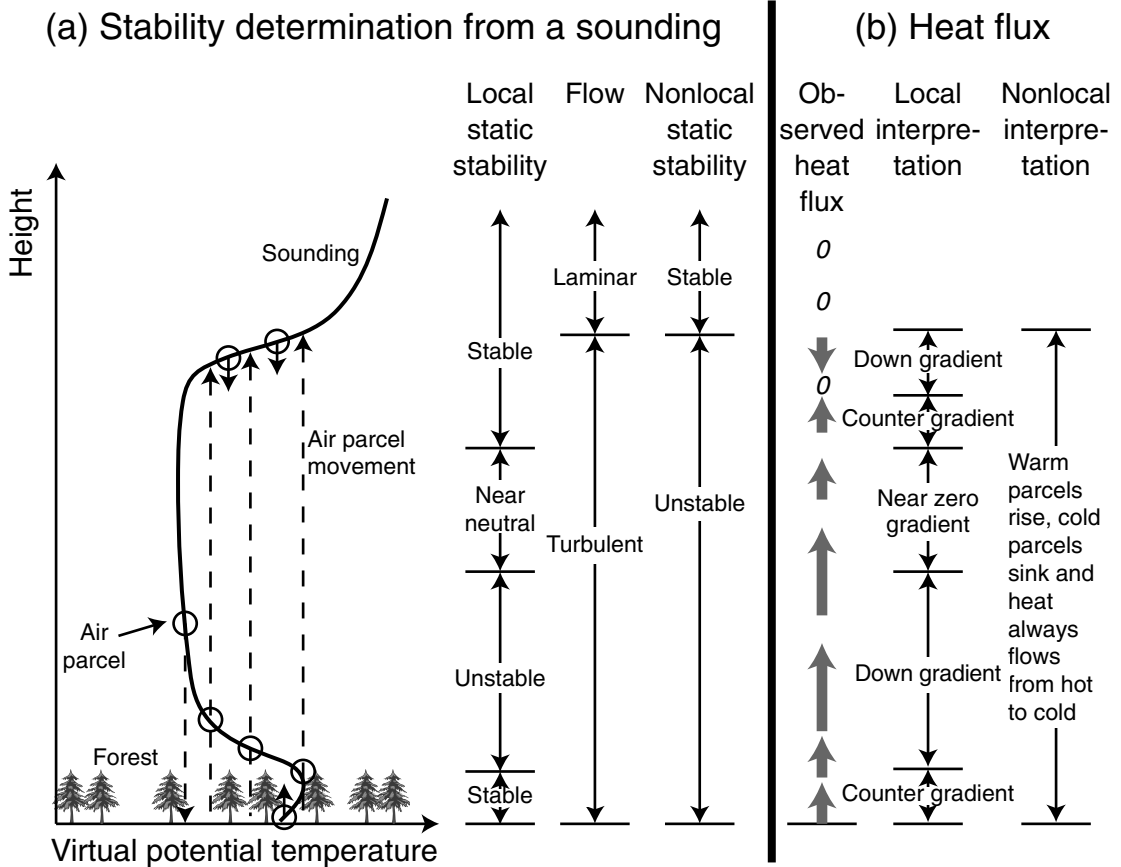


Fig. 4.16

Schematic showing stability defined based on local and nonlocal methods, and associated heat fluxes for each type of method. See the text for details. Adapted from Stull (1991).

result. Equations such as the above one for the local-closure approximation to the vertical heat flux can incorporate a correction term to the local gradient that incorporates the effects of the large eddies, such that

$$\overline{\theta'w'} = -K_H \left(\frac{\partial}{\partial z} \bar{\theta} - \gamma \right).$$

This correction defines an effective vertical lapse rate that is more unstable than the actual one, maintaining an upward heat flux through a deeper layer.

4.5 Radiation parameterizations

Electromagnetic radiation from the Sun is responsible for the existence of all processes in the atmosphere, including the midlatitude westerlies and cyclones, monsoons, tropical cyclones, and the Hadley circulation on the global scale, to convection and coastal

circulations on the mesoscale. Less obvious, but nevertheless important, consequences of radiative processes include radiation fogs, the strong surface-based temperature inversions that impact air quality, evaporation of water at Earth's surface, the extreme katabatic winds in polar latitudes, the development of buoyant instabilities that lead to convective weather, etc. In order to simulate these processes correctly, models must represent the interaction of the radiation with the land, oceans, vegetation, clouds, air molecules, and mineral aerosols of natural and human origin. A critical component of the calculations is to provide the radiative flux at Earth's surface, because it is the spatial distribution of this surface heating that is responsible for the differential heating of the atmosphere. In addition, the radiative flux divergence of energy within the atmosphere must be calculated in order to define the radiative heating and cooling in a column. Because radiation sometimes interacts with the atmosphere and the surface at a molecular level (e.g., molecular scattering, interaction with cloud droplets) and because this interaction is a complex function of the prevailing spectrum of wavelengths, the processes are too small in scale and too complex to simulate directly. Thus they need to be parameterized. The first section below briefly reviews the relevant radiative processes that must be simulated by models, and the remaining ones describe the parameterization of shortwave and longwave radiative fluxes.

4.5.1 Processes that must be represented

This section reviews some basic concepts regarding the transmission of radiation within the atmosphere, and illustrates the sorts of physical processes that must be represented by a radiation parameterization. Figure 4.17 shows the disposition of the solar energy that enters the Earth–atmosphere system, in terms of global-average values. Of 100 units of radiation entering the atmosphere annually from the Sun, 31 units are both reflected and scattered back to space. This includes 6 units that are reflected from Earth's surface (land and ocean), 17 units that are reflected and scattered from clouds, and 8 units that are reflected and scattered from molecules and dust in the atmosphere. Twenty units are absorbed by the atmosphere and clouds. The remaining 49 of the 100 units are direct and diffuse solar radiation that are absorbed by Earth's surface. The partitioning of radiant energy at Earth's surface between the absorbed (49 units) and reflected (6 units) components is typically calculated within the land-surface model (Chapter 5) that is part of the atmospheric model. The other processes – absorption by the atmosphere, clouds, and dust; scattering by air molecules and dust; and reflection by clouds and dust – are represented in the radiation parameterization. Figure 4.18, again based on a global annual average, summarizes what happens to the 69 units of energy that are shown in Fig. 4.17 to be absorbed by both Earth's surface and atmosphere. Energy gains at the surface, totaling 144 units, are 49 units from direct and diffuse solar radiation and 95 units of infrared from the clouds and gas of the atmosphere. Surface energy losses, which also must total 144 units, include 114 units of infrared emitted to space and the atmosphere, 23 units lost by evaporation, and 7 units lost to the atmosphere through sensible heat fluxes. The net radiation is the radiative gain of 144 units minus the radiative loss of 114 units. Also, the 155 units of energy gains by the atmosphere from the various sources on the left must equal the losses from

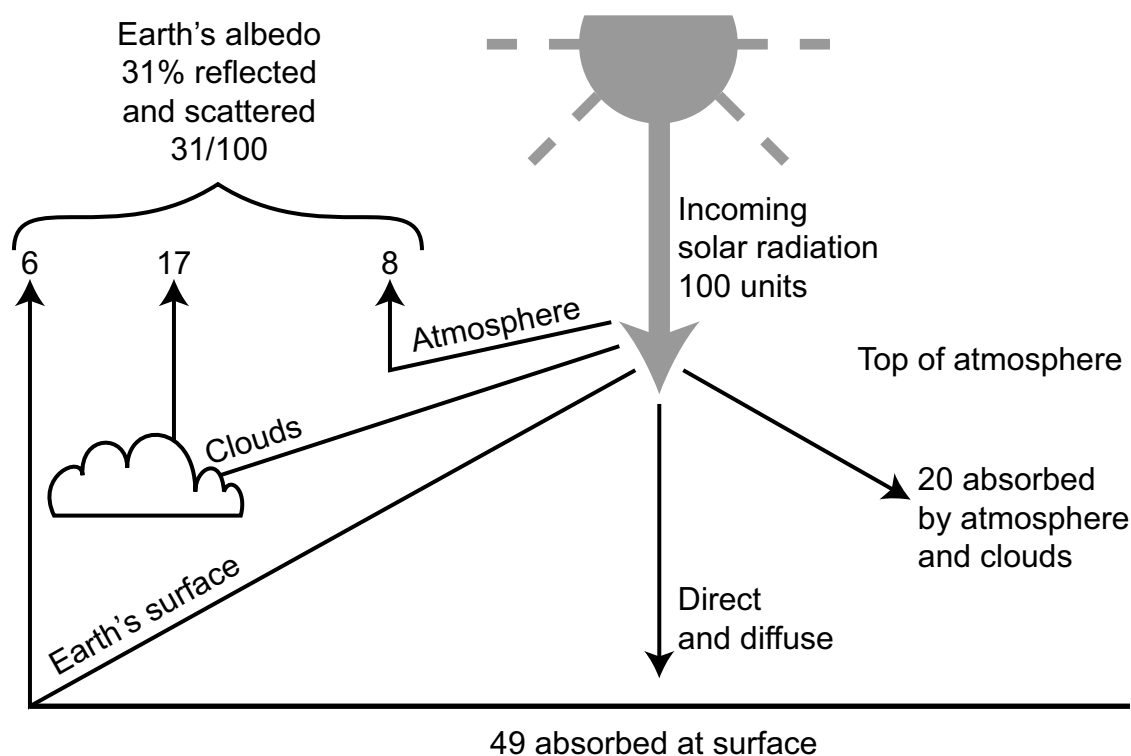


Fig. 4.17

The estimated atmospheric energy budget in terms of the global annual-average components. The estimated partitioning of energy is based on various sources.

the processes on the right. The land-surface component of the modeling system is responsible for calculating the evaporation rate, and the emission and absorption of infrared energy; the boundary-layer parameterization calculates the turbulent fluxes of heat by convection; the precipitation processes (both parameterized and resolved-scale) define the latent heating resulting from water condensation; and the radiation parameterization defines the remaining processes. These estimates of the various components of the budgets are being routinely revised, and there are some important differences among the estimates (e.g., Mitchell 1989, Kiehl and Trenberth 1997). It is important to remember that the magnitudes of these components are global averages, and thus they represent a large contribution from conditions over the oceans. Some energy-budget diagrams of this sort show an actual energy flux at the top of the atmosphere, rather than an arbitrary base value of 100 units. The global-average, top-of-the-atmosphere energy flux on a horizontal surface is about 342 W m^{-2} . The individual components of the budget thus represent partitions of this base value.

All substances with a temperature greater than absolute zero emit radiation. The rate at which energy is radiated at a given temperature, summed over all wavelengths, is proportional to the fourth power of the absolute temperature, which is the basis for the Stefan–Boltzmann law. With an added dimensionless coefficient, ϵ , the law states that

$$\text{Energy emitted} = \epsilon \sigma T^4, \quad (4.8)$$

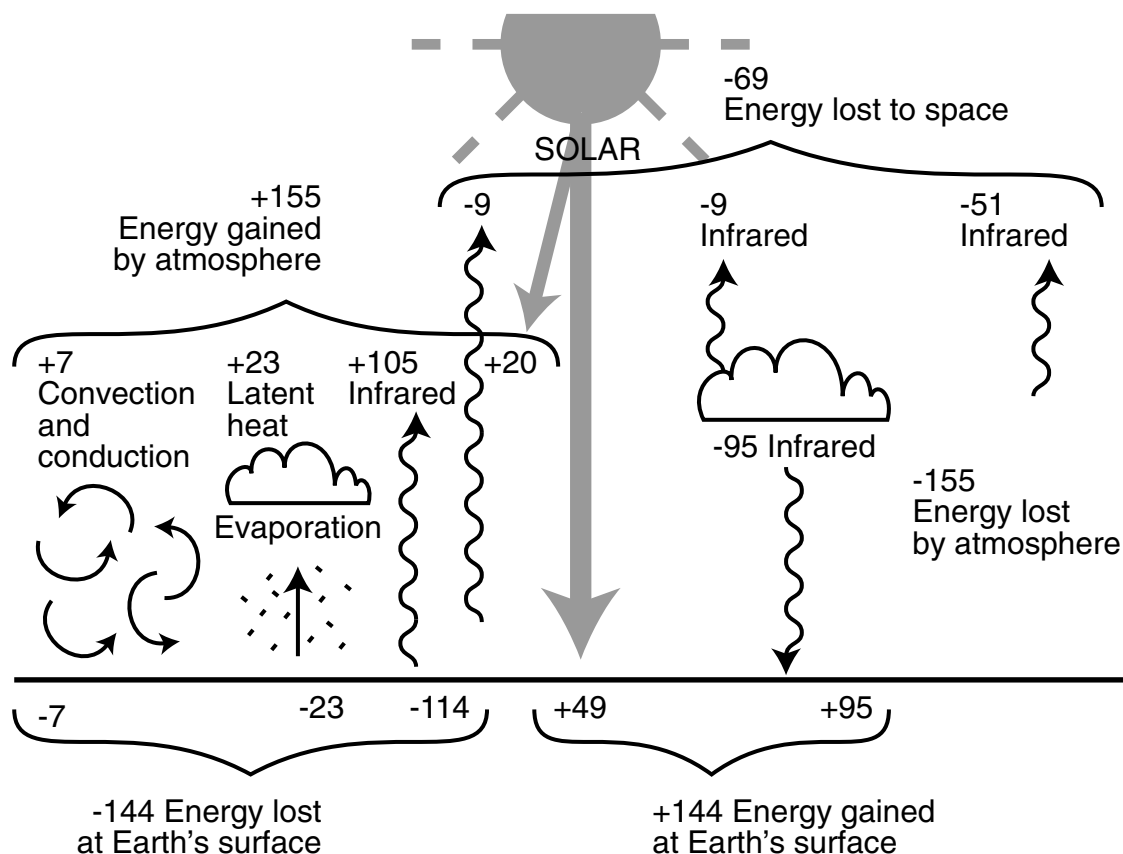


Fig. 4.18

The components of the energy exchange between the surface and the atmosphere, based on a global annual average. The estimated partitioning of energy is based on various sources.

where σ is the Stefan–Boltzmann constant, equal to $5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$. If the radiating body emits the maximum possible radiation per unit area per unit time, the emissivity, ϵ , is equal to unity, and the emitter is called a black body. Less efficient radiators have emissivities between zero and unity. The intensity of the radiation emitted by a black body at different wavelengths is a function of temperature, and is prescribed by Planck's law. This intensity distribution with wavelength has a very similar shape for emitters of any temperature (see the examples of Fig. 4.19), and has a single maximum. The particular wavelength composition of the emitted energy depends on the temperature of the emitter, such that a temperature increase not only increases the total amount of energy emitted (Eq. 4.8), but it also increases the fraction from the shorter wavelengths. That is, the curve in Fig. 4.19 shifts to the left as temperature increases, and the wavelength of the peak emission, λ_{max} , moves accordingly, such that

$$\lambda_{\text{max}} = 2.88 \times 10^{-3}/T,$$

where λ_{max} is in meters and T is in kelvin.

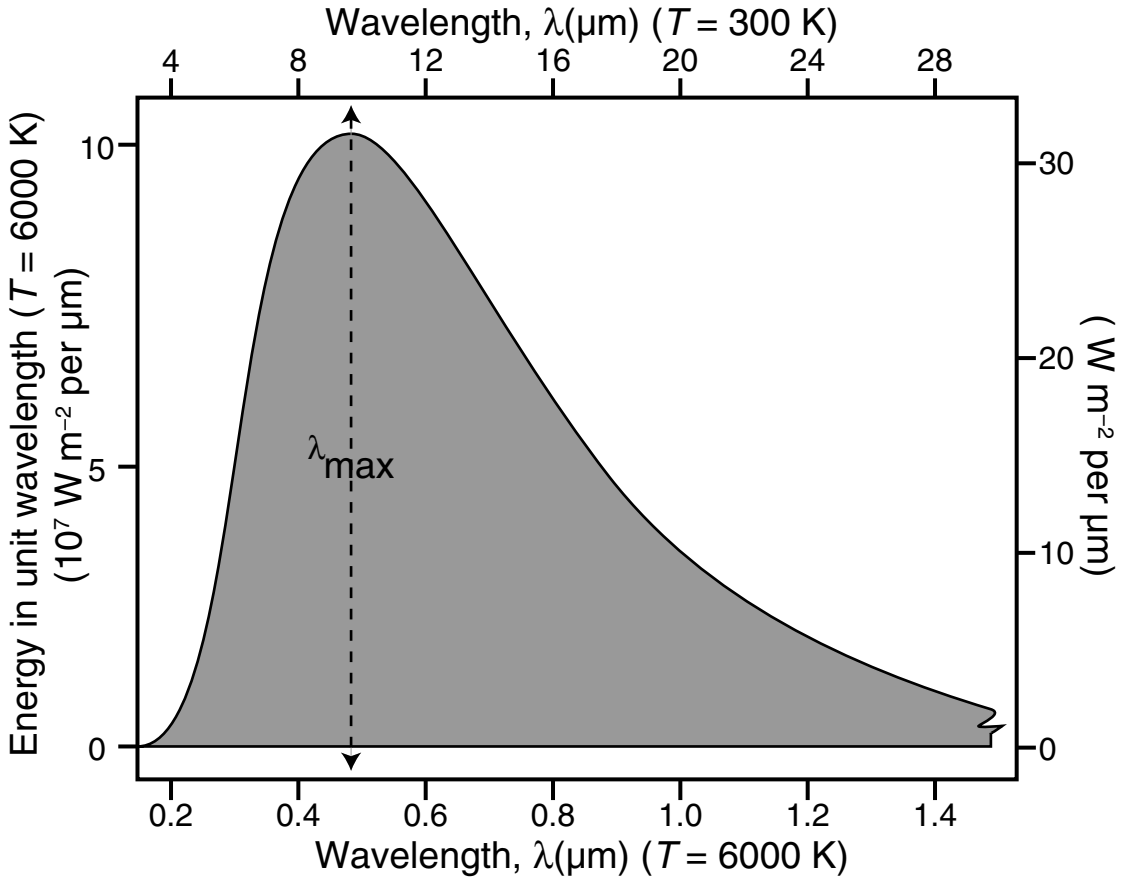


Fig. 4.19

The spectral distribution of radiant energy emitted from a black body at (a) 6000 K, left vertical and lower horizontal axes and (b) 300 K, right vertical and upper horizontal axes. The λ_{\max} is the wavelength at which the energy output per unit wavelength is a maximum. Approximately 10% of the energy is emitted at wavelengths longer than those shown. Adapted from Monteith and Unsworth (1990).

In most atmospheric applications, we are only concerned with wavelengths in the ultra-violet, visible, and infrared portions of the full electromagnetic spectrum. Much of the Sun's energy is in the visible part of the spectrum, from 0.36 μm (violet) to 0.75 μm (red), as shown in Fig. 4.19, which depicts the energy intensity for the different wavelengths emitted by the Sun which has a surface temperature of about 6000 K. There is also significant solar energy emitted in the wavelength bands that are shorter than the violet (the ultraviolet) and longer than the red (the infrared), with the total solar ultraviolet–visible–infrared band extending from about 0.15 μm to about 3.0 μm . The infrared energy in this solar spectrum is referred to as the solar infrared, to contrast it with the longer wavelength infrared that is emitted by the cooler Earth and its atmosphere. Figure 4.19 also shows the spectrum of the energy emitted by the Earth and its atmosphere, which have a temperature of roughly 300 K. The wavelength band is all within the infrared, and extends from about 3 μm to 100 μm . Thus, the solar spectrum from 0.15 to 3.0 μm is referred to as shortwave

radiation, while Earth's spectrum in the range 3.0–100 μm is longwave radiation. Note that the maximum intensity of the Sun's radiation is about a factor of 10^7 greater than that for Earth. The peak intensity of the solar spectrum is at 0.48 μm (λ_{max} is green, in the middle of the visible spectrum), whereas for the Earth–atmosphere system it is at about 10 μm .

Figure 4.20 shows the seasonal and latitudinal variation of the possible daily total (without atmospheric attenuation) solar radiation receivable on a horizontal surface. The

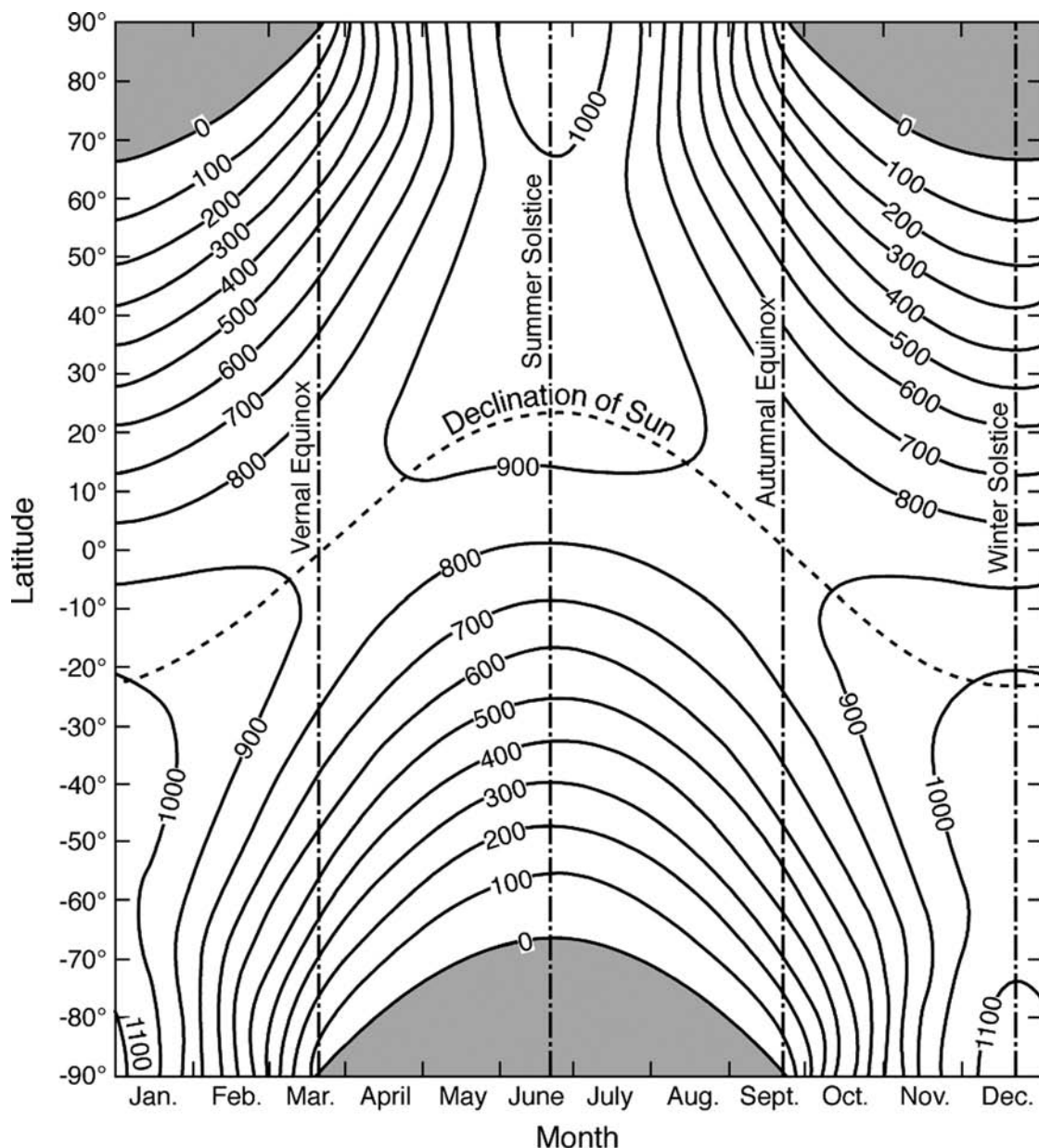


Fig. 4.20

Daily-total radiant energy received on a horizontal surface at the top of the atmosphere, by time of year and latitude. The isopleths are labelled in cal cm^{-2} . From List (1966).

values represent the time integral of the unattenuated direct solar energy flux during daylight hours, which is based on the intensity of the Sun's radiation, and the Earth–Sun geometry that controls the daily evolution of the Sun angle. No meteorological effects are accounted for (reflection by clouds, scattering, etc.), so this is the maximum possible energy receivable. It is the effect of the atmosphere on this radiation that must be parameterized.

Because of the existence of clouds, dust, and optically active (absorbing and emitting) gases, Earth's atmosphere is far from transparent to the passage of the radiant energy emitted by Earth and the Sun. Each of the constituents has its own unique effect on each individual wavelength band of the radiation. In terms of the bulk effect of the atmosphere, part of the radiation is reflected or scattered, part is absorbed, and the remainder is transmitted. The fraction of the incident radiation that is absorbed (absorptivity) individually by the atmosphere's major gaseous components, and by the total atmospheric mixture of gases, is shown in Fig. 4.21. It is clear that the gaseous medium totally absorbs some wavelength

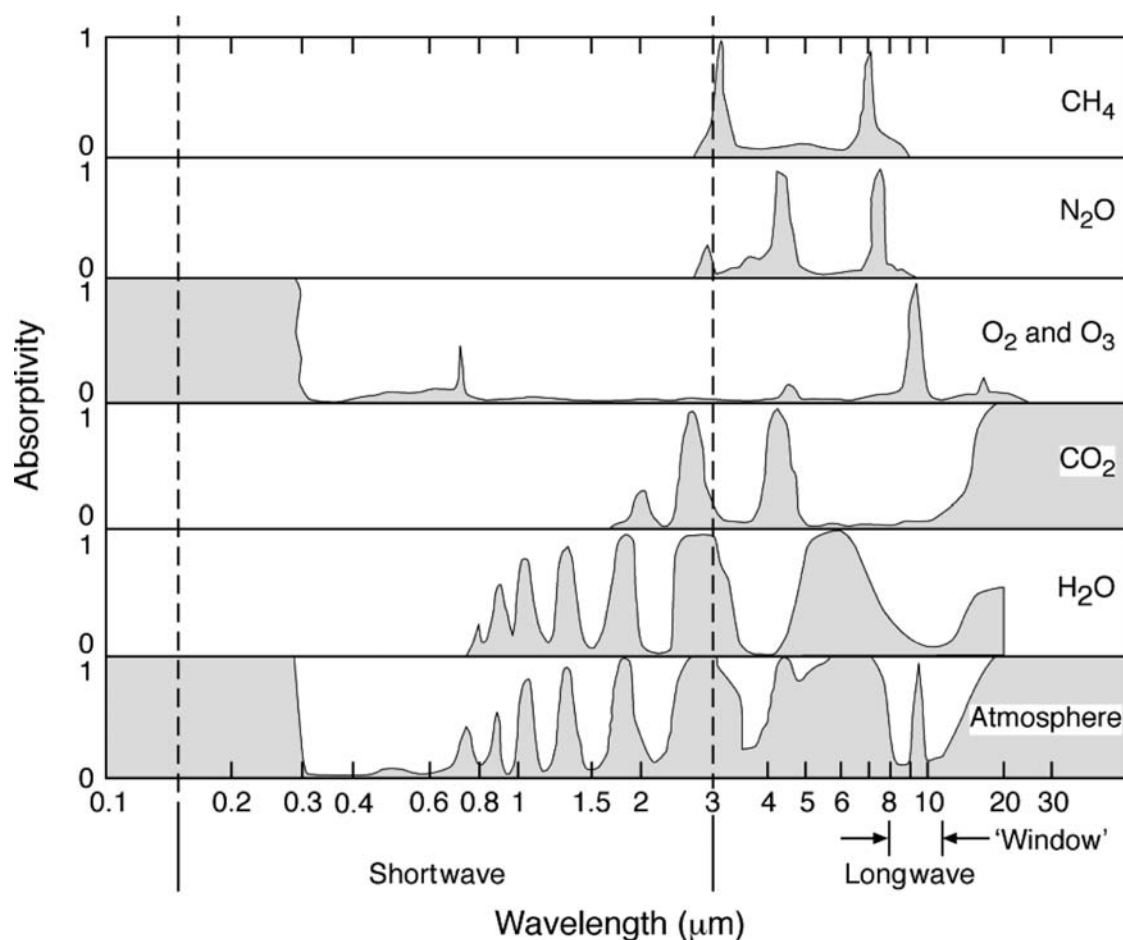


Fig. 4.21

The absorptivity (fraction absorbed) of the atmosphere's major gaseous components, and of the total atmospheric mixture of gases. Adapted from Fleagle and Businger (1963).

bands of radiant energy, while it is relatively transparent to others. Within the visible wavelengths of the solar spectrum, from 0.36 to 0.75 μm , where the greatest energy is represented, not much absorption occurs from gases. At ultraviolet wavelengths shorter than 0.30 μm , ozone effectively absorbs virtually all of the energy, and water vapor becomes an important absorber at wavelengths longer than 0.80 μm . There is a “window” in the longwave absorption from about 8 to 11 μm , which encompasses the wavelength of the peak emission at 10 μm from the Earth–atmosphere system that emits at about 300 K (Fig. 4.19). Clouds absorb and emit radiation in this window, and thus their correct parameterization is important to the modeled radiation balance. Figure 4.22 shows the effects of both scattering and absorption on the attenuation of the solar beam as it penetrates the cloud-free atmosphere. The outer curve, A, is the solar spectrum at the top of the atmosphere, which differs from the smooth blackbody spectrum of a 6000 K radiator shown in Fig. 4.19. Here, the degree to which the emissivity departs from unity depends on wavelength. The remainder of the curves show the progressive modification of the top-of-the-atmosphere spectrum by ozone absorption, molecular scattering, aerosol scattering, and water vapor and oxygen absorption. The lower curve (E) shows the energy that survives the downward transit of the solar beam through the atmospheric medium. It is the purpose of radiation parameterizations to estimate the three-dimensional time-of-day-, seasonal-, weather-, and climate-dependent details of these processes whose annual-average effects are shown in these figures.

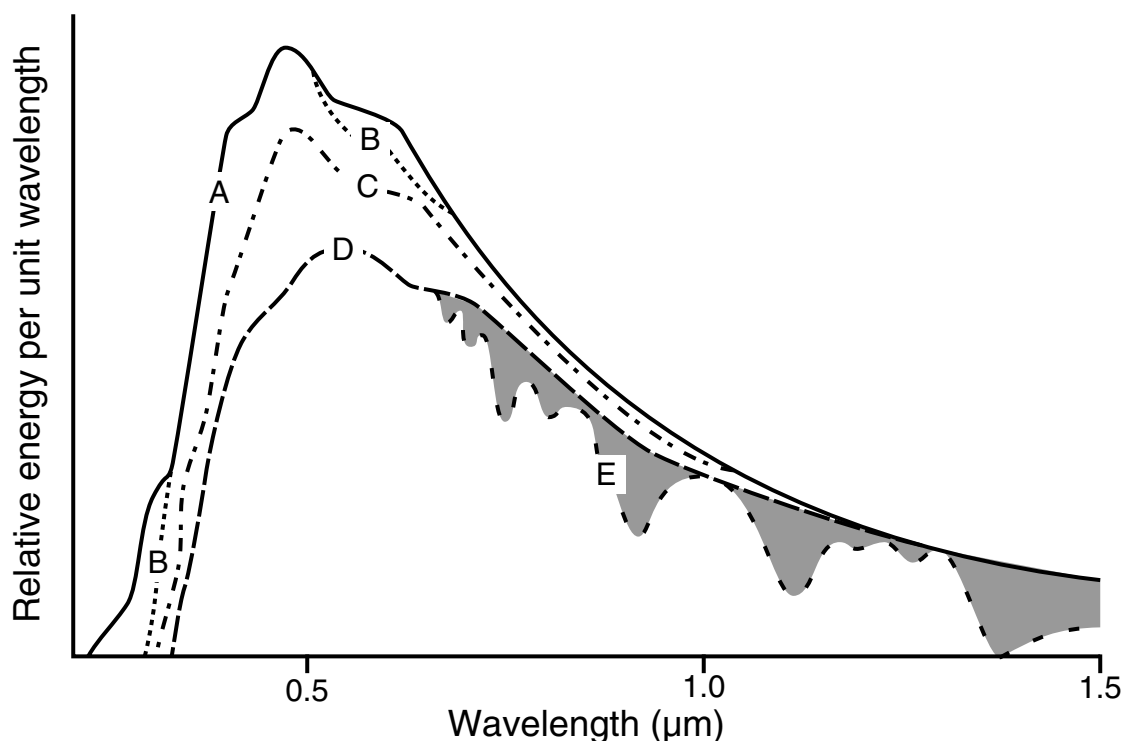


Fig. 4.22

Attenuation of the extraterrestrial solar spectrum (A) by ozone absorption (B), molecular scattering (C), aerosol scattering (D), and water vapor and oxygen absorption (E). Curve E is terrestrial sunlight. From Henderson (1977).

4.5.2 The general framework for representing radiation in models

The above physical effects of absorption, emission, scattering, and reflection are accounted for in calculations of the vertical fluxes of longwave and shortwave radiation at every grid point in a model. The vertical convergence of these energy fluxes is then employed in the thermodynamic energy equation (Eq. 2.4), such that

$$\frac{\partial T}{\partial t} \propto \frac{1}{\rho c_p} \frac{\partial}{\partial z} (F_D - F_U),$$

where F_D is the downward flux and F_U is the upward flux of radiant energy of all wavelengths. This equation is converted to finite-difference form, and solved at every grid point with the other terms in the equation. In addition, the downward flux at Earth's surface is also calculated and used as input to the model's land-surface parameterization (Chapter 5). The challenge is to calculate the longwave and shortwave fluxes with sufficient accuracy, especially for climate models where a small percentage error could be damaging, yet with sufficient efficiency so that the models can execute in the required period of time. This latter issue can be critical because radiation parameterizations can be very time consuming, and thus a variety of alternative approaches and approximations have been developed for solving the radiative-transfer equations. This is a sufficiently important issue that these parameterizations are often not called every time step, in order to reduce the computational demand.

Stephens (1984) lists a few factors that make the accurate modeling of radiation processes challenging, including the following.

- Radiation can simultaneously influence the atmospheric dynamics in multiple ways, and the required accuracy of the approximations in the parameterization can be dependent on the prevailing meteorological process. This makes it difficult to choose a set of approximations that is satisfactory in all situations.
- The atmospheric dynamics respond to diabatic heating from radiative flux convergence, phase changes, and sensible heating. However, these processes are sometimes coupled, for example when radiative cooling leads to condensation, leading to complex nonlinear interactions that may be difficult to approximate.

Another challenge is the following one. Required for calculations of shortwave and longwave radiative transfer are vertical profiles of temperature, pressure, and water-vapor mixing ratio from the model. In addition, concentrations of many other optically active gases, as well as natural and anthropogenic aerosols, are needed. In some forecasting situations, climatological distributions of the gases and aerosols are assumed. But, it is well known that there is considerable temporal and spatial (horizontal and vertical) variability in these quantities, and this is motivating efforts to include estimates of their concentrations in model initial conditions. For simulation of future climate, experimental approaches involve the specification of future increases in carbon dioxide and other gases according to different scenarios (see Chapter 16).

Note that there will be no attempt in this text to go into detail about radiative-transfer calculation methods, and associated approximations used in various parameterizations. Readers interested in the details should consult Liou (1980), Stephens (1984), and Stenrud (2007).

4.5.3 Parameterization of longwave fluxes

The simplest, least computationally demanding, and generally least accurate approach relates bulk properties of the resolved-scale atmosphere to the radiative flux. For example, there are simple, empirical approaches to estimating the downwelling longwave radiation (F_{LD}) at the ground using near-surface values of temperature. For example, Unsworth and Monteith (1975) suggest that

$$F_{LD} = c + d\sigma T_a^4,$$

where T_a is the 2-m air temperature, $c = -119 \pm 16 \text{ W m}^{-2}$, and $d = 1.06 \pm 0.04$ for a location in the UK. A similar relationship was proposed by Anthes *et al.* (1987) to calculate the net longwave flux at the surface:

$$F_{Lnet} = \epsilon_g \epsilon_a \sigma T_a^4 - \epsilon_g \sigma T_g^4,$$

where ϵ_g is the substrate, or ground, emissivity, T_g is the ground temperature, and T_a and ϵ_a are the temperature and the emissivity, respectively, at 40 hPa above the surface.

The general solution of the radiative transfer problem involves estimation of the following integral equations:

$$F_U(z) = \int_0^\infty \pi B_\nu(z=0) \tau_\nu^f(z, z=0) d\nu + \int_0^\infty \int_0^z \pi B_\nu(z') \frac{d\tau_\nu^f}{dz'}(z, z') dz' d\nu \quad (4.9)$$

and

$$F_D(z) = \int_0^\infty \int_z^\infty \pi B_\nu(z') \frac{d\tau_\nu^f}{dz'}(z, z') dz' d\nu, \quad (4.10)$$

where $F_U(z)$ and $F_D(z)$ are the upward and downward longwave fluxes through level z , respectively, ν is frequency, B_ν is Planck's function, τ_ν^f is the diffuse transmission function defined by the hemispheric integral

$$\tau_\nu^f(z, z') = 2 \int_0^1 \tau_\nu(z, z', \mu) \mu d\mu,$$

where μ is the cosine of the zenith angle and

$$\tau_\nu(z, z', \mu) = \exp \left[-\frac{1}{\mu} \int_{u(z)}^{u(z')} k_\nu(p, T) du \right].$$

The quantity $k_\nu(p, T)$ is the absorption coefficient and u is the concentration of the attenuating gas defined along the path from z to z' . The first term on the right in Eq. 4.9 represents the attenuation of the radiation emitted by Earth's surface. The second term in Eq. 4.9 and the single term in Eq. 4.10 correspond to the analogous emittance of longwave radiation by the atmosphere. The various parameterizations used for calculation of the longwave fluxes employ a variety of approximations for calculating the integrals in the above four equations. See Liou (1980), Stephens (1984), and Stensrud (2007) for discussions of specific techniques, and of how the effects of clouds on longwave fluxes are parameterized.

4.5.4 Parameterization of shortwave fluxes

The transfer of solar radiation in the atmosphere is less complex than that for longwave radiation because there is not the complexity of the simultaneous absorption and emission from layer to layer in the vertical. That is, the atmosphere does not emit in these wavelengths. However, shortwave radiative transfer has additional complexity in the sense that molecular scattering is important, unlike the situation with longwave transfer.

As with longwave parameterizations, there are relatively simple empirical methods for estimating the shortwave fluxes at Earth's surface. Examples are described in Anthes *et al.* (1987), Savijärvi (1990), and Carlson and Boland (1978). For example, Anthes *et al.* (1987) employ the following expression for the shortwave flux absorbed by the ground surface:

$$H_S = S_0(1 - \alpha)\tau \cos \zeta,$$

where S_0 is the solar constant, α is the albedo, ζ is the solar zenith angle, and τ is the shortwave transmissivity. The transmissivity calculation is based upon Benjamin (1983), and accounts for absorption and scattering by direct and diffuse radiation, and the effects of multiple cloud layers.

For more-complex calculations, the direct irradiance at a level z is defined by Beer's law as

$$F_D(z, \mu_0) = \mu_0 \int_0^\infty S_\nu(\infty) \tau_\nu(z, \infty, \mu_0) d\nu,$$

where $F_D(z, \mu_0)$ is the downward radiant flux through level z for a beam of radiation with a zenith angle θ_0 ($\mu_0 = \cos \theta_0$), the integral is over frequency (ν), $S_\nu(\infty)$ is the solar irradiance at the top of the atmosphere, and the monochromatic transmittance function is

$$\tau_\nu(z, \infty, \mu_0) = \exp\left(-\frac{1}{\mu_0} \int_{u(z)}^{u(\infty)} k_\nu du\right).$$

See Stensrud (2007) and Stephens (1984) for a discussion of approximations to these integrals used in parameterizations, as well as to the expressions that represent shortwave absorption and scattering. These references also discuss the parameterization of cloud effects on the shortwave fluxes.

4.6 Stochastic parameterizations

The influence of unresolved scales on the resolved scales is typically parameterized through the use of deterministic formulae. In contrast, stochastic parameterizations recognize that there are multiple possible states of the unresolved processes corresponding to a given state of the resolved variables, and that these can feed back differently to the resolved scales. These stochastic approaches take various forms. One example is described in Lin and Neelin (2000), wherein a random component is added to the CAPE that is calculated from the temperature and moisture profiles in the deep-layer-control scheme of the Betts and Miller (1986) convective parameterization. Addition of this random component improved the tropical intraseasonal variability of modeled convection. Similarly, Grell and Dévényi (2002) developed a parameterization that can use a large ensemble of closure assumptions and parameter values, and statistical techniques are then used to define the proper feedback to the resolved model variables. This method has been used operationally in the NCEP Rapid Update Cycle (RUC) model. Other examples of the many applications of stochastic parameterizations are described in Palmer (2001), Jung *et al.* (2005), and Plant and Craig (2008).

4.7 Cloud-cover, or cloudiness, parameterizations

With high-resolution cloud-resolving models, e.g., with grid increments of 1 km, it is possible to reasonably assume that an entire grid box is either cloudy or cloud-free. This is clearly not a reasonable assumption for global weather and climate models having horizontal grid increments of 10–100 km. Thus, for such coarser-resolution models there is a need to parameterize the cloud geometry in order to properly allow for the effects of the clouds on the radiation and surface-energy budgets. See Tompkins (2002, 2005) and Tompkins and Janisková (2004) for background and additional references.

Geometric cloud properties that might be estimated include:

- the horizontal fractional coverage of the grid box by cloud,
- the vertical fractional coverage of the grid box by cloud, and
- the overlap of the clouds in each vertical column.

Because the microphysics and convection parameterizations do not directly yield the above information, it must be independently estimated. Two approaches are used: relative-humidity-based methods and statistical methods.

Imagine a single grid volume whose horizontal area is partially filled with clouds, where such clouds extend in the vertical through the model layer. In the subvolume filled with cloud, the air is saturated. Elsewhere, the region is cloud-free and unsaturated. In this case, the grid-volume-mean relative humidity is obviously less than 100%. Thus, if a model predicts the grid-resolved relative humidity to be 100%, it can be assumed that

the grid-box area is filled with cloud. If the model grid-resolved relative humidity is less than 100%, it remains to be inferred by the parameterization whether there is any cloud in the grid box whose effects on the radiation and surface-energy budget should be accounted for.

All parameterization approaches assume the existence of fluctuations of humidity and/or temperature on the subgrid scale. Figure 4.23 shows an example distribution of mixing ratio and saturation mixing ratio over a grid increment, where the mean grid-box condition is unsaturated. Without the fluctuations, there would be no regions where $q_s < q$, leading to saturation. There are a number of different approaches for relating subgrid cloud cover to resolved-scale variables. One set of methods is based on diagnostic relationships between subgrid-scale fractional cloud cover and relative humidity. They generally apply to only the first property in the above list – the horizontal fraction of the grid box that is covered by cloud. A common mathematical relationship between relative humidity and fractional cloud cover, proposed by Sundqvist *et al.* (1989), is

$$C = 1 - \sqrt{\frac{1 - RH}{1 - RH_{crit}}},$$

where C is cloud fraction and RH_{crit} is the critical Relative Humidity (RH) above which cloud is assumed to form. Here $0.0 \leq C \leq 1.0$ and $RH_{crit} \leq RH \leq 1.0$, so that C increases monotonically from 0.0 to 1.0 as RH increases from RH_{crit} to 1.0. Because such a simple relationship is not likely to be equally suitable over a wide range of cloud types and climates and weather regimes, many alternative approaches have been suggested. For example, Slingo (1980, 1987) proposes separate RH-based relationships for convective clouds and high-, medium-, and low-level stratiform clouds. And Xu and Randall (1996) include both the total cloud-water and cloud-ice mixing ratio, and RH as predictors.

Another general approach to the problem involves specification of the subgrid-scale probability distribution function for the humidity, as well as possibly that of the temperature. Various symmetrical and asymmetrical distributions have been assumed, where Fig. 4.24 shows a

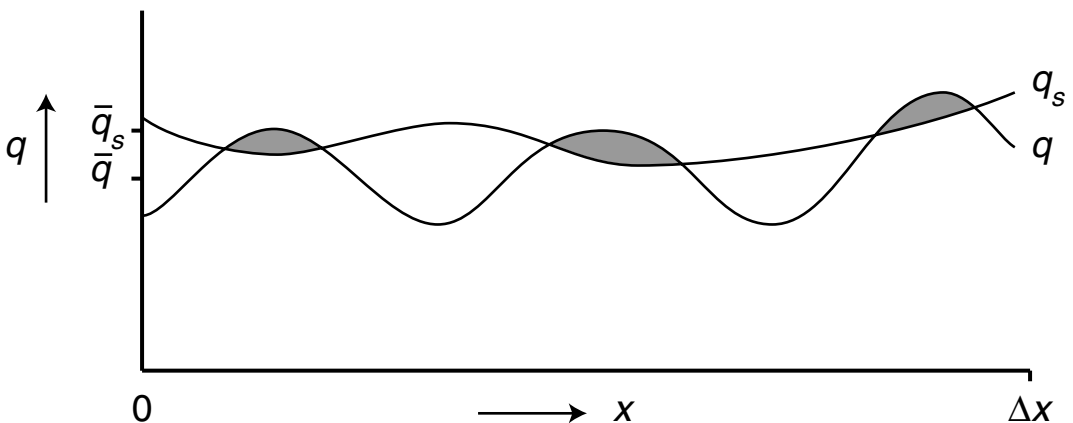


Fig. 4.23

Example of the variation of mixing ratio (q) and saturation mixing ratio (q_s) over a grid increment, where regions of saturation exist in spite of the fact that the grid-average condition is unsaturated.

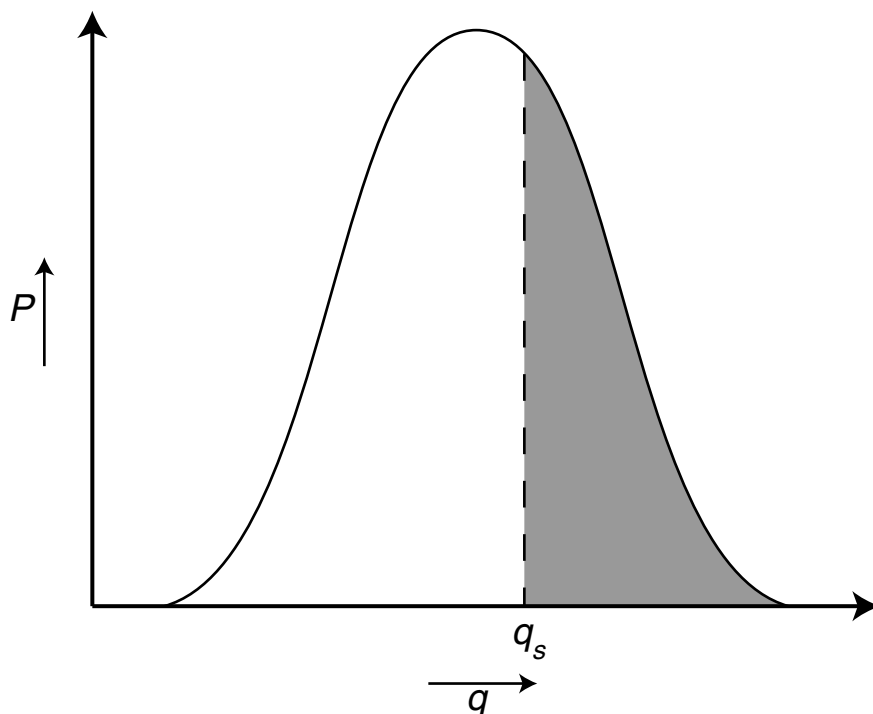


Fig. 4.24

Schematic example of the probability distribution function of the mixing ratio on subgrid scales. Shown is the saturation mixing ratio (q_s), which is based on the assumption of a uniform temperature. The region is shaded for which saturation implies the existence of cloud cover.

schematic example of how this can be used to estimate the cloud fraction. In the figure, it is assumed, as a simplification, that the temperature does not fluctuate on these scales, so that there is a constant saturation mixing ratio. The cloud cover is defined by the integral over the part of the distribution for which q exceeds q_s .

SUGGESTED GENERAL REFERENCES FOR FURTHER READING

General

Stensrud, D. J. (2007). *Parameterization Schemes: Keys to Understanding Numerical Weather Prediction Models*. Cambridge, UK: Cambridge University Press.

Convective parameterizations

Emanuel, K. A., and D. J. Raymond (eds.) (1993). *The Representation of Cumulus Convection in Numerical Models of the Atmosphere*. Meteorological Monographs, No. 46, Boston, USA: American Meteorological Society.

Frank, W. M. (1983). Review: The cumulus parameterization problem. *Mon. Wea. Rev.*, **111**, 1859–1871.

Smith, R. K. (ed.) (1997). *The Physics and Parameterization of Moist Atmospheric Convection*. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Turbulence parameterizations

Stull, R. B. (1988). *An Introduction to Boundary Layer Meteorology*. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Radiation parameterizations

Liou, K.-N. (1980). *An Introduction to Atmospheric Radiation*. London, UK: Academic Press.

Stephens, G. L. (1984). The parameterization of radiation for numerical weather prediction and climate models. *Mon. Wea. Rev.*, **112**, 826–867.

Cloud microphysics parameterizations

Cotton, W. R., and R. A. Anthes (1989). *Storm and Cloud Dynamics*. London, UK: Academic Press.

Fletcher, N. H. (1962). *The Physics of Rain Clouds*. Cambridge, UK: Cambridge University Press.

Houze, R. A., Jr (1993). *Cloud Dynamics*. London, UK: Academic Press.

Pruppacher, H. R., and J. D. Klett (2000). *Microphysics of Clouds and Precipitation*. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Rogers, R. R. (1976). *A Short Course in Cloud Physics*. Oxford, UK: Pergamon Press.

Rogers, R. R., and M. K. Yau (1989). *A Short Course in Cloud Physics*. 3rd edn. Oxford, UK: Butterworth-Heinemann.

Straka, J. M. (2009). *Cloud and Precipitation Microphysics: Principles and Parameterization*. Cambridge, UK: Cambridge University Press.

Cloud-cover parameterizations

Tompkins, A. M. (2005). *The Parameterization of Cloud Cover*. ECMWF Technical Memorandum.

PROBLEMS AND EXERCISES

1. Why might it be problematic for global models to use different parameterizations for different geographic regions, even though this could mean that the parameterizations perform better?
2. At about what horizontal grid increment do you imagine that parameterizations for sub-grid-scale cloud cover will not be needed? How might the answer to this be weather-regime dependent?
3. Suggest types of sensing systems that might be useful for initializing microphysical variables.

4. Describe how the existence of mineral aerosols of natural and human origin can influence microphysical and radiative processes in the atmosphere.
5. Given that mineral aerosols have important impacts on microphysical and radiative processes, how can their effect be practically represented in model forecasts?
6. What physical mechanisms might be responsible for convective parameterizations on the outer grids of an interacting nest influencing the resolved convective precipitation on the inner grid (e.g., as in Fig. 4.9)?