Chapter 20

100 Years of Progress on Mountain Meteorology Research

RONALD B. SMITH
Yale University, New Haven, Connecticut

ABSTRACT
Mountains significantly influence weather and climate on Earth, including disturbed surface winds; altered distribution of precipitation; gravity waves reaching the upper atmosphere; and modified global patterns of storms, fronts, jet streams, and climate. All of these impacts arise because Earth’s mountains penetrate deeply into the atmosphere. This penetration can be quantified by comparing mountain heights to several atmospheric reference heights such as density scale height, water vapor scale height, airflow blocking height, and the height of natural atmospheric layers. The geometry of Earth’s terrain can be analyzed quantitatively using statistical, matrix, and spectral methods. In this review, we summarize how our understanding of orographic effects has progressed over 100 years using the equations for atmospheric dynamics and thermodynamics, numerical modeling, and many clever in situ and remote sensing methods. We explore how mountains disturb the surface winds on our planet, including mountaintop winds, severe downslope winds, barrier jets, gap jets, wakes, thermally generated winds, and cold pools. We consider the variety of physical mechanisms by which mountains modify precipitation patterns in different climate zones. We discuss the vertical propagation of mountain waves through the troposphere into the stratosphere, mesosphere, and thermosphere. Finally, we look at how mountains distort the global-scale westerly winds that circle the poles and how varying ice sheets and mountain uplift and erosion over geologic time may have contributed to climate change.

1. Introduction to the field of mountain meteorology

Mountain meteorology is the study of how mountains influence the atmosphere. This subject has drawn curious investigators to it from the earliest days of the physical, geophysical, and geographical sciences. Some investigators are attracted to the subject by their love of mountain adventure: skiing, hiking, and climbing. Some appreciate the beauty of mountain scenery and fast-changing patterns of mountain clouds, winds, and weather. Some are drawn by the excitement of flying over mountains in gliders or aircraft. Many enjoy the challenging physics and mathematics of the unsolved mountain airflow problems. Most important are the practical applications of mountain meteorology, such as predicting damaging winds; forest fires; clear air turbulence; air pollution; wind, solar, and hydropower; water resources; and patterns of regional and global climate.

The influence of mountains on weather has been discussed for at least 2000 years. In about 340 BC, Aristotle speculated about the role of mountain heights in determining the altitude of clouds (Frisinger 1972). In 1647, Blaise Pascal measured the decrease in atmospheric pressure with altitude along a mountain slope to prove that air has weight. Even with this long history, the observational tools of meteorology were poorly developed in the early twentieth century. One hundred years ago, we had only a few weather balloons, analog recording instruments, and permanent mountaintop weather stations (e.g., Mt. Washington, New Hampshire, and Sonnblick, Austria). Now, as we begin the twenty-first century, many powerful observational tools are in use: instrumented aircraft, satellite passive visible and infrared imagery, active lidar and radar remote sensing, and even water isotope analysis. Accordingly, our knowledge has increased manyfold.

In the last 60 years, many small and a few large coordinated field projects have provided a valuable observational database for mountain meteorology. Some examples are given in Table 20-1. These national and
international projects broadened and connected the atmospheric science community.

Equally important to the history of mountain meteorology was the rapid development of numerical simulations of geophysical fluid flows, including microscale, mesoscale, and global weather and climate models. (e.g., Durran 2010). Starting in the 1970s, numerical models with steadily increasing resolution and accuracy have provided a powerful tool for sensitivity and hypothesis testing regarding mountain influences.

Stimulated by these exciting field experiments and advances in numerical modeling, scientists in this field have met annually to discuss new discoveries. These meetings were organized in alternate years by the American Meteorological Society’s Committee on Mountain Meteorology and the European International Conference on Alpine Meteorology (ICAM).

The literature on mountain meteorology is extensive. A guide to the major publications is given in Table 20-2. The books in Table 20-2 provide a deeper account of mountain meteorology than we can give in this short article. As they span nearly 40 years, they also give a sense of how the subject has advanced through time.

The current article is divided into seven main sections. The second section “How high is a mountain?” discusses the physical properties of the atmosphere that control the influence of mountains. The third section summarizes the mathematical description of terrain geometry. The following four sections treat the physics and fluid dynamics of how mountains influence weather and climate. In section 4, the impact of mountains on surface winds is reviewed, including mountaintop winds, downslope winds, barrier jets, gap jets, wakes, thermally driven slope winds, and cold pools. In section 5, orographic precipitation is summarized. In section 6, we describe the generation of mountain waves that propagate deep into the upper atmosphere. In section 7, we consider the global effects of mountains on climate over Earth’s history.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smith (1979)</td>
<td>The influence of mountains on the atmosphere</td>
</tr>
<tr>
<td>Blumen (1990)</td>
<td>Atmospheric Processes over Complex Terrain</td>
</tr>
<tr>
<td>Ruddiman (1997)</td>
<td>Tectonic Uplift and Climate Change</td>
</tr>
<tr>
<td>Barry (2008)</td>
<td>Mountain Weather and Climate</td>
</tr>
<tr>
<td>Chow et al. (2013)</td>
<td>Mountain Weather Research and Forecasting: Recent Progress and Current Challenges</td>
</tr>
</tbody>
</table>
2. Atmospheric reference heights: How high is a mountain?

Mountains are usually ranked by their peak heights. Citizens take pride in their nation’s highest peaks. Climbers rightly brag about the highest peak they have “bagged.” This rank order of mountains, with Mt. Everest (8848 m) at the top, dominates our thinking about mountains.

In this chapter, however, we judge mountain height differently: not by comparison with other mountains, but with reference to various height scales in Earth’s atmosphere (Table 20-3). This simple exercise in “relativism” will carry us a long way toward understanding the way that mountains influence Earth’s atmosphere. Note that these atmospheric reference heights may vary with time and space associated with climate zone, season, and weather. Two mountains with the same geometric height at different latitudes may differ in their height relative to local atmospheric features. The relative height of a mountain may vary season to season or even day to day as the atmospheric conditions change.

a. Earth radius

The ratio of mountain height to Earth radius is, in every case, very small (Table 20-3). For Earth’s highest mountain, Mt. Everest, the “roughness ratio” (i.e., ratio of mountain height to Earth radius) is 8.8 km/6371 km = 0.0014, about a tenth of one percent. A basketball, with its ridges and grooves, has a roughness of about one percent (0.01). A bowling ball or billiard ball roughness is much less, closer to 0.0001. Earth’s roughness falls in between these two examples. From the small roughness ratio, one might expect that mountains are so small that they have very little influence, but this is an incorrect conclusion.

Earth’s radius varies with latitude due to the distortion from centrifugal forces from its daily rotation. The equatorial and polar Earth radii are 6378 versus 6357 km, respectively. For this reason, we do not measure a hill height from Earth’s center. Such a method would give an unbeatable 21-km advantage to equatorial hills. Instead, we measure hill height relative to an ideal geopotential surface (i.e., geoid) approximating global sea level (see section 3). As the atmosphere also follows the geoid, the sea level reference convention is well suited for the study of mountain influence on the atmosphere. That is, a 5-km high hill at the pole (relative to sea level) will reach to about the same density level in the atmosphere as a 5-km hill at the equator.

b. Density scale height

Under the influence of gravity, with a minor influence from centrifugal forces, the gases in our atmosphere are pulled toward Earth’s surface and compressed into a thin layer. The weight of air caused by gravity is countered by gas pressure from rapid random molecular speeds. The hydrostatic balance between these two forces is reached quickly and naturally in the atmosphere. In the special case of an isothermal atmosphere, this balance gives profiles of air density $\rho(z)$ expressed by

$$\rho(z) = \rho_0 \exp \left( -\frac{z}{H_S} \right),$$  

(20-1)

where $\rho_0$ is the density at the surface and $H_S$ is the density scale height given by $H_S \approx RT/g$. With the gas constant for air $R = 287$ J kg$^{-1}$ K$^{-1}$, typical air temperature of $T = 15^\circ C = 288$ K, and surface gravity $g = 9.81$ m s$^{-1}$, the scale height is $H_S \approx \{(287)(288)/9.81\} = 8400$ m. This is a small value compared to Earth’s radius. According to (20-1), the top of Mt. Everest stands above nearly 65% of the mass of Earth’s atmosphere. This deep penetration of high mountains into the atmospheric mass increases their impact on weather and climate.

The diminished air pressure and density on high mountains is evident in many ways. First is the lack of oxygen for breathing. While oxygen still comprises 21% of the air (by volume), the small overall pressure reduces the partial pressure of oxygen to a value too low for human respiration. Most mountain climbers require supplemental oxygen above 6000 m.

Second is the reduced air drag and air force. A wind gust on Mt. Everest would produce less force than a similar wind at sea level.

Third is the dark sky. At high altitude, with so little air above, the diffuse blue “sky light” that we experience at sea level is reduced. The sky appears black. The sun’s direct beam is more intense at high altitudes and the shadows are darker due to reduced scattering and diffuse radiation.
While less obvious to a mountain climber, the down-going longwave radiation also decreases with altitude. As less carbon dioxide and water vapor lie above a mountaintop, the down-going thermal radiation from these molecules in the wavelength range 3–50 μm decreases quickly with height. Nighttime radiative cooling of the surface grows with altitude. On high peaks, we stand above Earth’s greenhouse effect.

In these four ways, we see that Earth’s mountains have a significant height relative to the atmosphere thickness. This conclusion is nearly independent of location. The atmospheric scale height (20-1) varies by only a few percent from equator to pole and from season to season, so the above discussion would apply equally to any location and time.

c. Flow lifting height

A fundamental question in mountain meteorology is whether the air flows around or over a mountain. In the former case, the air stays nearly in horizontal layers as it flows around the terrain. In the latter case, the air may rise up the windward slopes and over the high terrain. Regarding this question, there is a strong consensus in the literature that the nondimensional mountain height plays the key role

\[ \hat{h} = Nh/U \]  

(20-2)

(e.g., Snyder et al. 1985; Smith 1989a, b ). In (20-2) the stability or Brunt–Väisälä frequency (BVF) is defined as

\[ N^2 = \left( \frac{\partial \theta}{\partial z} \right)\left( \frac{g}{C_p} \right) + \frac{dT}{dz}, \]  

(20-3)

where the potential temperature is \( \theta = T(P/P_0)^\gamma \) and \( \gamma = C_p/C_v \) is the ratio of specific heats. The wind speed is \( U \).

The threshold value for (20-2) is approximately unity. When \( \hat{h} < 1 \), the air can flow over the hill and down the lee slope (Fig. 20-1a). When \( \hat{h} > 1 \), low-level air will split and flow around the hill, gaining and losing little elevation (Fig. 20-1b).

In the existing literature, nondimensional mountain height (20-2) is often referred to as “inverse Froude number.” The author does not favor that terminology as it differs from the original definition of Froude number used in layered flows (see section 4).

The cause of the flow splitting around high hills is the stagnation of the flow on the windward hill slopes caused by relatively high pressure there. Until the flow has stagnated, the center streamline cannot split left and right. The high pressure on the windward slope is caused by an elevated region of lifted dense cold air aloft, directly over the windward slopes. The forward tilt of the lifted air is associated with the generation of a vertically propagating mountain waves (section 6).

Using typical values of \( N = 0.01 \text{ s}^{-1} \) and \( U = 10 \text{ m s}^{-1} \), mountain heights of 100, 1000, and 10 000 m give the nondimensional mountain heights of \( \hat{h} = 0.1, 1, \) and 10, respectively. Thus, air can easily flow over a 100-m hill but never over a 10-km-high hill. For the 1000-m hill, the situation is uncertain and will depend on the wind speed.

A serious limitation to the application of (20-2) is the variability in \( N(z) \) and \( U(z) \) values with altitude, making it difficult to choose representative values (Reinecke and Durran 2008).

d. Water-based reference heights

Water is very important to atmosphere physics, so the vertical distribution of water vapor provides an important reference height. Due to saturation effects, water vapor density decreases much faster with altitude than air density (20-1). For mountain meteorology, it matters whether a hill penetrates up through most of the water vapor in the atmosphere or whether there is substantial water vapor above the mountaintop. For example, the highest mountains on Earth experience little precipitation simply because there is so little water vapor above them.

The quantitative argument begins with the Clausius–Clapeyron equation describing how the saturation vapor pressure \( [e_s(T)] \) depends on temperature (e.g., Wallace and Hobbs 2016; Yau and Rogers 1996; Mason 2010; Lamb and Verlinde 2011):

\[ \left( \frac{1}{e_s} \right) \frac{de_s}{dT} = \frac{L}{RT^2}. \]  

(20-4)
where \( L \) is the latent heat of vaporization and \( R \) is the gas constant for water vapor. With some approximation, this integrates to

\[
e_S(T) = 6.1 \exp(0.0725 \times T) \tag{20-5}\]

with \( e_s(T) \) in hectopascals and \( T \) in degrees Celsius (e.g., Wallace and Hobbs 2016). Now consider an atmosphere with a constant temperature gradient \((\gamma)\), with temperature profile

\[
T(z) = T_0 + \gamma z, \tag{20-6}
\]

where \( T_0 \) is the air temperature at the surface. If the relative humidity \( RH = e_w/e_s \) is constant, the vertical profile of water vapor partial pressure is

\[
e_w(z) = RH \times 6.1\exp[0.0725 \times (T_0 + \gamma z)], \tag{20-7}
\]

so the scale height for water vapor pressure becomes

\[
H_w = \left[ \frac{d \ln(e_w)}{dz} \right]^{-1} = (-0.0725 \times \gamma)^{-1}. \tag{20-8}
\]

Using the standard tropospheric lapse rate of \( \gamma = -6.5^\circ C \text{ km}^{-1} \) and constant RH, we have \( H_w = 2122 \text{ m} \). Thus, a mountain with peak at \( z = 2 \text{ km} \) will have about two-thirds of the atmospheric water vapor below it.

A second important water-based height scale is the lifting condensation level (LCL). As a subsaturated air parcel rises adiabatically from sea level, it cools at the rate \( \Gamma = -(g/Cp) = -0.0098^\circ C \text{ m}^{-1} \). At the LCL, the relative humidity reaches unity and vapor condensation begins. This altitude marks the base of cumulus clouds. The LCL altitude depends mostly on the relative humidity at Earth’s surface. An approximate expression is

\[
Z_{\text{LCL}} = (25)(100 - RH) \tag{20-9}
\]

(Lawrence 2005). As an example, a surface relative humidity of \( RH = 50\% \) gives \( Z_{\text{LCL}} = 1250 \text{ m} \).

The LCL enters mountain meteorology in two ways, depending on the cause of the vertical air motion and the cloud type. First, if cumulus clouds are being generated by thermal convection from the surrounding plains, the \( Z_{\text{LCL}} \) marks the cloud base of the cumulus. If \( h > Z_{\text{LCL}} \), a mountain peak will reach preexisting cumulus clouds. Second, if the wind is strong and air is forced upward by the terrain, \( h > Z_{\text{LCL}} \) is the condition for the mountains to generate clouds.

The third water-based reference height \((Z_F)\) is the freezing/melting level where \( T = 0^\circ C \). If the lapse rate is constant aloft \((20-6)\), the \( 0^\circ C \) “freezing/melting” level will be found at the altitude

\[
Z_F = -T_S/\gamma. \tag{20-10}
\]

If \( \gamma = dT/dz = -6.5^\circ C \text{ km}^{-1} \) and surface temperature \( T_S = 20^\circ C \), the melting level will be at \( Z_F = (20/6.5) = 3.1 \text{ km} \). Above this level, precipitation will fall as snow. Below \( Z_F \), precipitation will fall as rain. This snow/rain boundary is often evident in mountain scenery (Fig. 20-2). Because it depends so sensitively on surface temperature, \( Z_F \) varies greatly with latitude, season, and weather type. In cold Arctic climates (e.g., Denali, Alaska) in winter, the freezing level would be “below sea level” so all precipitation falls as snow. In tropical lands (e.g., Mt. Kilimanjaro, Kenya), only the highest elevations receive snow.

The in-cloud temperature influences precipitation in two ways. First, at subzero temperatures, the Bergeron ice-phase mechanism may accelerate the formation of hydrometeors (e.g., Colle and Zeng 2004a,b; see also Wallace and Hobbs 2016; Yau and Rogers 1996; Mason 2010; Lamb and Verlinde 2011). Second, falling snowflakes will melt to form rain (Lundquist et al. 2008; Minder et al. 2011). The latent heat absorbed in the melting process may alter the air motion (Unterstrasser and Zängl 2006).

The difference between rain and snow is important in mountain meteorology because of their different fall speeds (e.g., \( 5 \text{ m} \text{s}^{-1} \) versus \( 1 \text{ m} \text{s}^{-1} \)). In cases with strong winds, snow could be carried several kilometers before reaching Earth, because of its slow fall speed. This can be associated with “spillover” where snow falls on the lee side of a mountain range. Rain, on the other hand, cannot be carried downwind more than a kilometer or so from where it is created (see section 6).

The storage of water on Earth’s surface is influenced by the phase of the precipitation. Falling rain will...
immediately start moving through the soil toward a nearby stream. It may be flowing down a river within a few hours of “landing” on the mountain surface. In contrast, precipitation falling as snow will be stored on the ground surface for several days, and possibly months or years. It may accumulate into glaciers. Fallen snow will not enter the soil and rivers until warm air arrives and the snow melts.

e. Natural layers in the atmosphere

The real atmosphere generally has a complex temperature profile $T(z)$ with height, unlike (20-6). Of particular importance is the presence of warm layers of air riding over colder layers: a so-called inversion. As warm air is less dense than cold air at the same pressure, it is difficult to push warm air down into heavy cold air and difficult to push heavy cold air up into warm air. Such a temperature interface is said to be statically stable, as it resists vertical motion. Occasionally, surface-derived pollutants are trapped below these interfaces, making them visible to the human eye. Inversions can also cause rising cloudy air to detrain at those levels. These stable layers can intersect a mountain slope or, if the wind is strong enough, the layer may be lifted to pass over a hill crest. Five common stable layers in Earth’s atmosphere are sloping midlatitude fronts, Arctic surface inversions, the tropopause, the trade wind inversion, and the continental fair-weather boundary layer inversion. All of these layers impact mountain airflows. An example of a trade wind inversion near Hawaii is shown in Fig. 20-3. One should not miss the experience of standing atop Mauna Loa and looking out at the trade wind inversion lapping up to the mountain peak, like the ocean waves surging up against your feet on a gently sloping beach.

f. Climatological reference heights

Two other mountain reference heights depend not on the current atmospheric conditions but on the regional annual climatology. These heights are the glacier “equilibrium line” and the “tree line.”

An important reference height in glaciology is the equilibrium line (EL; Hooker 2005; Cuffey and Paterson 2010). At elevations above EL, snow accumulation in winter exceeds mass loss to melting and sublimation in summer. With no glacial motion, snow depth would grow year after year without limit. In steady state, this annual accumulation is balanced by glacial sliding motion carrying ice to lower elevations. Below the EL, there is a net loss of snow by melting and sublimation.

The tree line is a concept from forest ecology that describes the maximum altitude that alpine tree species can grow (e.g., Körner 2012). An example is given in Fig. 20-4. Due to the slow multiyear growth of trees, the tree line is not seasonal nor does it vary week by week due to weather systems. Instead, it integrates the effect of cold winters and warm summers, variable winds, soil moisture and soil thickness. Both the EL and the tree line influence the atmosphere by modifying the surface albedo, roughness, and evaporation potential.
The terrestrial planets also have high terrain that influence their atmospheres. Some of the criteria we used to judge the effective height of hills on Earth (Table 20-4) will apply to hills on these other planets. We skip Mercury here, as it has only a trace atmosphere. We also skip the large Jovian planets, as they have no solid surfaces and therefore no terrain. We include Pluto, in spite of its questionable “planet” status, as it has newly discovered terrain. Note that mountain elevations on Earth are relative to sea level while other planets have no ocean reference. On those other planets, a computed geoid must be used as a reference.

The highest mountain in the solar system is Olympus Mons on Mars with $h = 22$ km. Relative to the planet radius ($h/R$) and scale height ($h/H_s$) it is still the highest mountain in the solar system. Standing at the top of Olympus Mons, one would look down at about 86% of the atmospheric mass. On Venus, the penetration of Maxwell Montes is roughly comparable to Mt. Everest on Earth in regard to its penetration into the atmosphere.

A distant example is Pluto, visited in 2015 by the New Horizons deep space probe. It found mountains reaching to 3500 m, now given names honoring Tenzing Norgay and Edmund Hillary, who first reached the summit of Mt. Everest in 1953. Using its ratio of height to planet radius ($h/R$), it reaches higher than Mt. Everest. On the other hand, due to Pluto’s small surface gravity, the atmospheric scale height $H_s$ is large and its hills have a small ratio of $h/H_s$. These mountains surmount only a few percent of the atmospheric mass. Thus, in spite of the large roughness ratio, mountains probably play little role in Pluto’s atmosphere.

### 3. Mountain geometry

A prerequisite for the study of mountain meteorology is a mathematical framework for describing the distribution and geometry of natural terrain. As the existing literature on this methodology is scattered and small (e.g., Petkovsek 1978, 1980; Wood and Snell 1960; Young et al. 1984; Stahr and Langenscheidt 2015), we provide a few basic ideas below.

#### a. The geoid

The description of Earth’s terrain requires a careful specification of a reference surface or “datum” (Meyer 2018). This surface is usually chosen by geodesists as an oblate ellipsoid, formed by rotating an ellipse around an axis coincident with Earth’s rotation axis. Lines of constant latitude on the ellipsoid are circular while meridians are ellipses with primary radii of 6357 km at the pole and 6378 km at the equator. The datum is chosen to approximate a “geoid,” a geopotential surface, everywhere perpendicular the gravity force vector. Of all the possible concentric geopotential surfaces, the reference geoid is chosen to coincide with the average water surface of a quiescent ocean. All terrain heights are measured relative to the geoid to give a function of longitude $\lambda$ and latitude $\phi$:

$$h(\lambda, \phi)$$ (20-11)

(see Fig. 20-5). The advantage of using a geopotential surface to define terrain, for mountain meteorology, is that the atmosphere is a massive fluid which, when left alone, will seek to flatten itself along a geopotential surface. Thus (20-11), with its geoid reference, is a nearly

---

**Table 20-4. Mountains on terrestrial planets.**

<table>
<thead>
<tr>
<th>Planet</th>
<th>Venus</th>
<th>Earth</th>
<th>Mars</th>
<th>Pluto</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest mountain</td>
<td>Maxwell Montes</td>
<td>Mt. Everest</td>
<td>Olympus Mons</td>
<td>Tenzing, Hillary Montes</td>
</tr>
<tr>
<td>Mountain height (km) $h$</td>
<td>11</td>
<td>8.8</td>
<td>22 km</td>
<td>3.5</td>
</tr>
<tr>
<td>Planet radius (km) $R$</td>
<td>6052</td>
<td>6371</td>
<td>3391</td>
<td>1187</td>
</tr>
<tr>
<td>Surface pressure</td>
<td>90 bars</td>
<td>1013 hPa</td>
<td>4–9 hPa</td>
<td>10 microbar</td>
</tr>
<tr>
<td>Atmosphere gas</td>
<td>CO₂</td>
<td>N₂, O₂</td>
<td>CO₂</td>
<td>CH₄, N₂</td>
</tr>
<tr>
<td>Scale height (km) $H_s$</td>
<td>15.9</td>
<td>8.5</td>
<td>11.1</td>
<td>60</td>
</tr>
<tr>
<td>$h/R$</td>
<td>0.0018</td>
<td>0.0014</td>
<td>0.0065</td>
<td>0.0029</td>
</tr>
<tr>
<td>$h/H_s$</td>
<td>0.69</td>
<td>1.04</td>
<td>1.98</td>
<td>0.058</td>
</tr>
</tbody>
</table>

---

**Fig. 20-5. Terrain map of the world in geographic (i.e., latitude–longitude) coordinates.** (Source: GMTED 2010, U.S. Geological Survey; Danielson and Gesch 2011.)
true measure of how terrain penetrates into a quiescent atmosphere.

Over the last hundred years, our knowledge of Earth terrain has advanced considerably. Using manual geodetic surveys, aerial and satellite stereo imagery, radar and lidar altimetry, and GPS, the elevation errors have been reduced to a few tens of centimeters and the spatial resolution enhanced to a few tens of meters. One-kilometer gridded global datasets such as GTOPO30 (Danielson and Gesch 2011) can be supplemented by local finescale digital elevation models (DEMs). The remaining minor definitional and measurement problems regarding Earth’s terrain are negligible in the study of mountain meteorology.

b. Coordinate systems

For small regions of Earth’s surface, it is convenient to define an approximate local Cartesian coordinate system \((x, y)\). The usual procedure for converting a spherical coordinate system to Cartesian is to use a map projection where a virtual light source is imagined to project each point on Earth’s surface to a developable surface. The Universal Transverse Mercator (UTM) projects points onto a cylinder touching Earth on a meridian. A Lambert conformal project points onto a cone with axis on the rotation axis, cutting Earth on two standard parallels. These projections are used for regions as large 5000 km on a side. They use “eastings” and “northings” as a Cartesian coordinate system \((x, y)\).

For still smaller regions, a simple planar projection can be used, touching Earth at a reference location \(\lambda_0, \phi_0\). The local Cartesian coordinates are

\[
\begin{align*}
  x &= C(\lambda - \lambda_0) \cos(\phi_0), \quad (20-12a) \\
  y &= C(\phi - \phi_0), \quad (20-12b)
\end{align*}
\]

where the coefficient is \(C \approx 111.1 \text{ km per degree}\). With any of these projections, the global description \((20-11)\) can be replaced by the local description

\[
h(x, y) \quad (20-13)
\]

as shown in Fig. 20-6.

c. Peaks, basins, and cols

If the terrain height above the datum is represented by a smooth differentiable mathematical function \((20-13)\), the methods of analytic geometry allow us to define peaks, basins, and cols. We first define the slope vector, using \((20-13)\),

\[
S(x, y) = \nabla h = \frac{\partial h}{\partial x} \hat{i} + \frac{\partial h}{\partial y} \hat{j}. \quad (20-14)
\]

A mountain peak (i.e., local maximum) will have a negative definite \(H\) matrix with negative eigenvalues. A depression (i.e., local minimum) will have a positive definite \(H\) with positive eigenvalues. In both cases, the determinant \(\text{det}(H) > 0\). At a col (i.e., saddle point), \(H\) will have one positive and one negative eigenvalue. The determinant \(\text{det}(H) < 0\). The eigenvectors of \((20-16)\) provide information about the orientation of long ridges, valleys, or saddle points.

An important issue in complex terrain is the distinction between mountain peak height versus mountain “prominence.” The peak height is easily defined as \(h_M = \max[h(x, y)] = h(x_p, y_p)\). In contrast, the prominence of the peak (Helman 2005) is defined as “height of a mountain above the saddle on the highest ridge connecting it to a peak higher still.” A hiker, descending
from the highest peak “A” to the highest col connecting to secondary peak “B,” must still ascend an amount $\Delta h$ to reach “B,” so $\Delta h$ is the prominence of “B.” By this definition, the highest point on any island or continent has equal height and prominence, whereas lower peaks have a prominence smaller than their height above sea level. The concept of terrain prominence is often used by “peak baggers” who rank the value of a mountain ascent by its prominence. In meteorology, prominence can be important as it quantifies the extra ascent an airstream must take in climbing over a high peak compared to passing over a lower nearby col.

A useful, but less universal, terrain description is the relative mountain height using a smoothed terrain as a reference surface. Defining the smooth reference surface to be $h_{\text{REF}}(x, y)$, the local “relative” terrain is

$$h'(x, y) = h(x, y) - h_{\text{REF}}(x, y; L).$$  \hspace{1cm} (20-17)

The choice of smoothing algorithm and smoothing scale ($L$) is up to the user. As the reference surface for relative terrain is not a geopotential surface, $h'(x, y)$ is not a measure of how high terrain will penetrate into a quiescent atmosphere. In its favor, relative terrain may provide an estimate of vertical motion over terrain as the slope of the perturbation terrain would usually dominate over the smoothed reference terrain. An accurate analysis of both absolute and relative terrain is important for mountain meteorology.

d. Terrain statistics

In considering a broad area of Earth’s surface, it is convenient to have integral or statistical descriptions of the terrain characteristics. The volume of the terrain above sea level is

$$\text{volume} = \int \int h(x, y) \, dx \, dy. \hspace{1cm} (20-18)$$

The variance (above sea level) is

$$\text{variance} = \int \int h^2(x, y) \, dx \, dy. \hspace{1cm} (20-19)$$

In (20-18) and (20-19), it makes a great difference whether one uses absolute terrain (20-11) or relative terrain (20-17). If relative terrain is used, the volume (20-18) may be nearly zero and the variance (20-19) is more sensitive to roughness than to mean height above sea level.

A widely used statistic is the terrain histogram, showing how much land area lies between each interval of elevation ($\Delta z = z_2 - z_1$). An example of a histogram for the South Island, New Zealand, is shown in Fig. 20-7.

The appearance of the histogram depends on the grid size and the elevation bin width ($\Delta z$) chosen. Most actual terrain histograms are concave, as seen in Fig. 20-7, whereas a simple cone would have a linear histogram.

The histogram contains the information needed to compute terrain volume (20-18) and variance (20-19), but it provides no information about the geographic distribution of the terrain. In most cases, we need to know the mean location of a group of peaks, how widely are they spread over the landscape, and how anisotropic is the distribution. To obtain this information, we compute the first and second spatial moments.

The planform centroid location $(\bar{x}, \bar{y})$ (relative to coordinate system origin) is given by the first moments

$$\bar{x} = \int \int x \, h(x, y) \, dx \, dy / \text{volume}, \hspace{1cm} (20-20a)$$

$$\bar{y} = \int \int y \, h(x, y) \, dx \, dy / \text{volume}. \hspace{1cm} (20-20b)$$

Note that one should use absolute terrain (20-11) in (20-18) rather than relative terrain. The second planform moments (independent of origin) are

$$M_{xx} = \int \int (x - \bar{x})^2 \, h(x, y) \, dx \, dy, \hspace{1cm} (20-21a)$$

$$M_{xy} = \int \int (x - \bar{x})(y - \bar{y}) \, h(x, y) \, dx \, dy, \hspace{1cm} (20-21b)$$

$$M_{yy} = \int \int (y - \bar{y})^2 \, h(x, y) \, dx \, dy. \hspace{1cm} (20-21c)$$
These moments describe the horizontal extent of the terrain. The matrix of second moments is

\[
M = \begin{bmatrix}
M_{xx} & M_{xy} \\
M_{xy} & M_{yy}
\end{bmatrix}.
\]  

(20-22)

Because this matrix is symmetric, it has real eigenvalues \((\lambda_1, \lambda_2)\) and orthogonal eigenvectors \((v_1, v_2)\). These values are found from

\[
Mv = \lambda v
\]

(20-23)

(Strang 2006). The anisotropy and orientation of a mountain range can be determined from the eigenvalues and eigenvectors of matrix \(M\) (Smith and Kruse 2018). These properties are important for mountain meteorology as airflow parallel to a narrow ridge will be less disturbed than airflow across the ridge (Smith 1989b). Also, terrain orientation relative to the sun can influence solar heating of the slopes.

An alternative measure of terrain anisotropy is the slope variance matrix proposed by Baines and Palmer (1990), Lott and Miller (1997), and Smith and Kruse (2018):

\[
S = \begin{bmatrix}
S_{XX} & S_{XY} \\
S_{XY} & S_{YY}
\end{bmatrix}
\]

(20-24)

with elements

\[
S_{ij} = \frac{\partial h}{\partial x_i} \frac{\partial h}{\partial x_j} dx dy.
\]

(20-25)

This matrix \(S\) describes the variance and covariance of the east–west and north–south terrain slopes. It is more sensitive to the smaller terrain scales than \(M\) in (20-22).

As \(S\) is symmetric, it will have real eigenvalues \((\lambda_1, \lambda_2)\) and orthogonal eigenvectors \((v_1, v_2)\), describing the anisotropy and orientation of the terrain. The ratio of eigenvalues \((\lambda_1/\lambda_2)\) for \(S\) is a measure of anisotropy. The orthogonal eigenvectors give the orientation of the long and short axes of the terrain. These statistical measures of terrain may be sensitive to smoothing.

e. Transects and projected profiles

A common graphical representation of complex 3D terrain is a transect across the terrain or a projected profile of the terrain. Examples of three terrain projected profiles are given in Fig. 20-8 for New Zealand. In these examples, the viewing angle is chosen to be from the northwest, a common direction from which the wind blows. A terrain data file can be rotated into a different orientation using a rotation matrix. The profile for the rotated terrain shows what an air parcel approaching the islands would see as it approaches the terrain. Defining rotated coordinates \(x'\) and \(y'\), three useful profiles are defined by

\[
F_1(y') = \max_x [h(x', y')],
\]

(20-26a)

\[
F_2(y') = \int h(x', y') dx',
\]

(20-26b)

\[
F_3(y') = \int \max_x \left( \frac{dh}{dx} \right) dx'.
\]

(20-26c)

Profile \(F_1\) in meters (Fig. 20-8a) is useful for comparing mountain heights with various atmospheric reference heights (section 2) such as the LCL, where clouds will form or the blocking height controlling flow splitting. The sharp high peaks are evident in this profile, but the airflow may divert around them. The volume profile \(F_2\) in meters to kilometers (Fig. 20-8b) is smoother and gives a sense of how long an air parcel may remain elevated as it passes over the terrain. This measure may be important for orographic precipitation or long mountain wave generation. The \(F_3\) profile in meters in Fig. 20-8c is the sum of all the upslope terrain. It indicates the amount of repetitive ascent and descent as air passes over the terrain. For New Zealand, the net rise exceeds 10 000 m. The ratio of \(F_3\) to \(F_1\) is a nondimensional measure of terrain roughness.

f. Terrain spectra

A useful representation of terrain is the power spectra. It provides estimates of how different horizontal
scales contribute to the shape and roughness of terrain. This information is important because the atmosphere may respond differently to each terrain scale.

For a one-dimensional terrain, we use the single Fourier transform

\[
\hat{h}(k) = \int_{-\infty}^{\infty} h(x) \exp(ikx) \, dx \tag{20-27}
\]

and the inverse transform

\[
h(x) = \left(\frac{1}{2\pi}\right) \int \hat{h}(k) \exp(-ikx) \, dk. \tag{20-28}
\]

Note that the position of the normalizing factor \(2\pi\) varies from author to author. A simple example of the Fourier method is the popular Witch-of-Agnesi ridge profile

\[
h(x) = \frac{a^2h_M}{(x^2 + a^2)} \tag{20-29}
\]

with maximum hill height \(h_M\), half-width \(a\), and cross-sectional area \(\pi ah_M\). Its Fourier transform (20-27) is

\[
\hat{h}(k) = \pi ah_M \exp(-a|k|) \tag{20-30}
\]

(Fig. 20-9). The wider the hill (20-29), the narrower the Fourier transform (20-30).

A second example is the double bump

\[
h(x) = \begin{cases} \frac{h_M}{2} \left[1 - \cos \left(\frac{2\pi x}{d}\right)\right] & \text{for } -d < x < d, \\ 0 & \text{for } |x| > d \end{cases} \tag{20-31a}
\]

\[
h(x) = 0 \quad \text{for } |x| > d, \tag{20-31b}
\]

with area \(h_Md\). Its Fourier transform is

\[
\hat{h}(k) = \left(\frac{h_M}{2}\right) \left[\frac{2}{k} - \frac{1}{(k + k_M)} - \frac{1}{(k_M - k)}\right] \sin(kd), \tag{20-32}
\]

where \(k_M = 2\pi/d\). As both terrains (20-29) and (20-31) are even functions, both transforms (20-30) and (20-32) are real.

The terrain power spectrum displays the length scales of the terrain. In one dimension, this is the product of the Fourier transform and its complex conjugate

\[
P(k) = \hat{h}(k)\hat{h}(k)^* \tag{20-33}
\]

shown for both ridge shapes (20-29) and (20-31) in Fig. 20-9. The “Witch” power spectrum decreases monotonically while the double-bump terrain shows a spectral peak at \(k = k_M\).

Using Parseval’s theorem, the variance (20-19) can be related to the power spectrum (20-33) by

\[
\text{Var}(h) = \int_{-\infty}^{\infty} \hat{h}^2(x) \, dx = \left(\frac{1}{2\pi}\right) \int_{-\infty}^{\infty} P(k) \, dk. \tag{20-34}
\]

For the Witch of Agnesi, (20-29), (20-30), and (20-34) give

\[
\text{Var}(h) = \frac{\pi h_M^2 a}{2}. \tag{20-35}
\]

For two-dimensional terrain we use the double Fourier transform

\[
\hat{h}(k, l) = \iint h(x, y) \exp(ikx + ily) \, dx \, dy \tag{20-36}
\]

with the inverse Fourier transform

\[
h(x, y) = \left(\frac{1}{2\pi}\right)^2 \iint \hat{h}(k, l) \exp(-ikx - ily) \, dk \, dl. \tag{20-37}
\]

The terrain power spectrum is given by

\[
P(k, l) = \hat{h}(k, l)\hat{h}(k, l)^*. \tag{20-38}
\]

Using Parseval’s theorem, the variance (20-19) computed from (20-38) is

\[
\text{Var}(h) = \iint h^2(x, y) \, dx \, dy = \left(\frac{1}{2\pi}\right)^2 \iint P(k, l) \, dk \, dl. \tag{20-39}
\]

These Fourier representations are useful in treating the scale dependence of orographic precipitation (section 5)
and mountain wave momentum and energy fluxes (section 6).

g. Terrain in numerical models

As numerical models currently play such a large role in mountain meteorology research, it is important that Earth’s terrain be properly represented in these models. Most models use terrain flowing coordinates (Gal-Chen and Somerville 1975; Schar et al. 2002) but the distortions in this grid system can introduce errors in fluid dynamical computations. An alternate approach is a Cartesian grid with level surfaces that intersect the terrain (Mesinger et al. 1988; Gallus and Klemp 2000; Purser and Thomas 2004; Lundquist et al. 2012). The advantages of each coordinate system are discussed by Shaw and Weller (2016).

h. Mathematical models of terrain evolution

A deeper physical understanding of terrain geometry is found by attempting to mathematically model the temporal evolution of terrain $h(x, y)$ over geologic time scales. The governing equation is

\[ \frac{dh}{dt} = U - E, \]

where $dh/dt$ is the rate of local elevation change and $U$ and $E$ are the local uplift and erosion rates (Goren et al. 2014). While tectonic uplift $U(x, y, t)$ is often taken as a smooth or even regionally constant factor in space and time, the erosion rate $E(x, y, t)$ is highly variable and physically represents rock weathering, soil sliding, stream incision, and glacial scraping and transport. As illustrated in Fig. 20-10, erosion can generate a wide variety of terrain scales. Fast erosion on steep slopes with river incision can isolate and preserve high ground, until the rivers slowly cut back into the remaining high ground.

Erosion rates are influenced by atmospheric conditions such as precipitation amount and rates, temperature, sunlight, vegetation, and glaciers (Anders et al. 2006; Han et al. 2015). Thus, there is a slow interaction

![Fig. 20-10. Modeled terrain evolution for New Zealand over 26 million years: (left) elevation (m) and (right) erosion rate (mm yr$^{-1}$). [From Goren et al. (2014); copyright 2014 John Wiley & Sons, Ltd.]]
between terrain evolution and the impact of terrain on the atmosphere.

4. Disturbed surface winds

An important branch of mountain meteorology is the study of how mountains influence the patterns of wind near the surface of Earth. In this section, we review seven aspects of this subject: mountaintop winds, severe downslope winds, barrier jets, gap jets, wakes, thermally driven winds, and cold-air pools. These disturbed winds are common near mountainous terrain. Important applications of this section are to wind power, forest fires, and air pollution. Already we see that successful wind farms are being installed on mountaintops, in downslope wind areas, and in gaps. Recent forest fires in California were fanned and pushed by downslope winds. Air pollution is problematic in stagnant cold-air pools but seldom on windy mountaintops.

a. Mountaintop winds

It is a common experience to find strong winds at the top of hills. These strong winds can be understood using three simple ideas. First, in the absence of terrain, wind speed increases with height in the atmospheric boundary layer due to surface friction. Anemometers on the top of a tall thin tower on flat terrain almost always record faster winds than at the base. Any hill able to penetrate up through the lower boundary layer will find stronger winds aloft.

If the mountain is tall and steep enough, it may penetrate out of the frictional layer and into other wind layers. In a baroclinic atmosphere, with fronts and inversions, wind speed and direction often vary with height. A good example is in China, where the hilltops reached the low-level jet (Yuan and Zhao 2017). Tall mountains in the subtropics (e.g., Mauna Loa) reach above the easterly summer trade winds into the westerlies.

Hills also disturb the airflow. Usually, elevated terrain will accelerate the flow at hilltop. In potential (i.e., irrotational) flow theory, a smooth circular cylinder will double the ambient flow speed at its crest. A sphere will accelerate the airspeed by a factor of $3/2$. In both cases, as the airflow streamtubes narrow to squeeze by the barrier, the flow speed must increase to conserve the volume flow rate in the tube. According to Bernoulli’s equation, the air pressure at hilltop drops below the hydrostatic prediction.

One of the first theoretical treatments of turbulent airflow over terrain was by Jackson and Hunt (1975) and Carruthers and Hunt (1990). Their methods drew on decades of aerodynamic and turbulent flow research. An example of a hill in a turbulent boundary layer is shown in Fig. 20-11. As air approaches the hill, it feels an adverse pressure gradient that decelerates the flow. As the air approaches the crest, the pressure gradient becomes favorable and the flow accelerates. On the lee slope, the adverse pressure gradient reappears even more strongly. If the lee slope is steep enough, flow separation can occur. In this disturbed flow, gravity and stratification play no role.

A third reason for strong hilltop winds applies to forested landscapes such as the mountains of New England or Scotland. Here, some of the most popular hill hikes reach “bald tops” above the tree line (section 2). The local lack of trees reduces friction and allows for strong winds. A well-known example of hilltop winds is Mt. Washington in the White Mountains of New Hampshire. For 62 years, Mt. Washington held the surface wind speed record of 231 mph ($103 \text{ m s}^{-1}$), recorded 12 April 1934.

![Fig. 20-11. Theoretical perturbation wind speed profiles over a smooth ridge in the atmospheric boundary layer with no influence of stratification. Maximum wind is at the ridge top. [From Jackson and Hunt (1975); copyright Royal Meteorological Society.]]
All three of the hilltop wind speed-up mechanisms (i.e., existing shear, hill-induced disturbance, and bald top) probably contributed to that record.

As we consider wider hills, gravity and stratification become more important. The key parameter controlling this transition is nondimensional mountain width

\[ \tilde{a} = Na/U, \quad (20-40) \]

where \( N \) is the stability frequency \((20-3)\), \( a \) is the ridge half-width, and \( U \) is the ambient wind speed. As \( \tilde{a} \) increases toward unity, the disturbed air has time to feel the buoyancy restoring force associated with stable stratification, and mountain waves may be generated (section 6). As discussed in section 6, this wave generation introduces an asymmetry into the airflow distortion. When \( \tilde{a} \gg 1 \) the flow develops a hydrostatic force balance in the vertical.

A nice example of the transition toward more influence of stratification is the airflow observation of Vosper et al. (2002), shown in Fig. 20-12. In the faster wind case (Fig. 20-12a) the maximum wind speed is seen at the high points of the terrain. In the slower, with larger \( \tilde{a} \), the fastest winds are seen on the lee slopes (Fig. 20-12b). This downwind shift of the maximum wind is a precursor of the severe downslope wind phenomena discussed below.

b. Severe downslope flows

An important phenomenon in mountainous terrain is the “severe downslope wind” or “plunging flow.” In this case, the strongest winds blow down the lee slope of a ridge. While less common than hilltop winds, it is frequent in certain mountainous regions. Researchers agree that it requires special conditions to occur. First, the ridge must be wide enough to give \( \tilde{a} > 1 \) so that gravity and fluid buoyancy play a role. Second, it usually requires a ridge higher than 500 m, but the exact required height depends on the wind and stability environment. Aloft, it helps to have a stable inversion or wind reversal in midtroposphere.

Interest in downslope winds was stimulated in the atmospheric science community by the aircraft observations of the Boulder windstorm in 1972 (Fig. 20-13) by Lilly and Zipser (1972). This discovery stimulated an exciting period of research attempting to identify a simple theory for downslope winds (Klemp and Lilly 1975; Clark and Peltier 1977, 1984; Peltier and Clark 1979; Smith 1985; Durran 1986; Durran and Klemp 1987; Bacmeister and Pierrehumbert 1988). These ideas were reviewed by Smith (1989a) and Durran (1990, 2015a). The theories of severe downslope winds fall into two broad categories: layered and continuous. Both seem to have some validity.

1) LAYER THEORY OF DOWNSLOPE WINDS

The layer theories of downslope winds idealize the atmospheric temperature profile into a cold layer below a warm layer with a sharp interface (e.g., Armi 1986). The temperature jump \( (\Delta \theta) \) at the interface is described by its “reduced gravity”

\[ g' = \frac{\Delta \theta}{\bar{\theta}} \quad (20-41) \]

With the hydrostatic assumption, the problem simplifies to the classical “shallow water theory” already well established in fluid mechanics. In this body of theory, the upstream Froude number plays an important role.
Fr
  \frac{\text{flow speed}}{\text{wave speed}} = \frac{U}{\sqrt{g'H}}
\end{align}

with layer depth $H_\infty$ and speed $U_\infty$. The Froude number is the ratio of flow speed to the speed of long gravity waves. The nondimensional hill height is

$$H = \frac{h_M}{H_\infty}.$$  

The most important case is when the upstream Froude number (20-42) is less than unity but exceeds the "critical value"

$$1 > Fr = f(H).$$  

The flow will accelerate up the terrain slope while the layer depth will decrease. At the hill crest, critical condition is $Fr_{\text{crest}} = 1$. On the lee slope, the speed will continue to increase and the layer thickness will further decrease. In most cases, the flow will undergo a "hydraulic jump" over the lee slope or over the flat terrain downwind (Fig. 20-14).

As an example, consider a 1-km-deep layer of cold air underlying a deep atmosphere 10°C warmer. This contrast gives a "reduced gravity" $g' = 0.327 \text{ m s}^{-2}$. If the incoming flow speed is $U = 10 \text{ m s}^{-1}$, the upstream Froude number is

$$Fr = \frac{10}{\sqrt{(0.981)(1000)}} = \frac{10}{31.1} = 0.55.$$  

When the ridge height is greater than about 300 m, the critical condition will be met and air will plunge down the lee side of the ridge.

The underlying physics behind this remarkable windward–leeward asymmetry has to do with the ability of a stratified fluid to send wave signals upstream. In "super-critical" flow (i.e., $Fr > 1$) the layer cannot send wave impulses upwind against the flow and adjustments of speed and depth occur differently than in subcritical flow (i.e., $Fr < 1$).

The advantage of the shallow layer (or "hydraulic") theory is the elegance of the physics and the ease of demonstration. Water spilling over a dam or pouring from a pitcher is a nearly perfect analogy. The drawback is that the atmosphere never acts exactly like a two-layer fluid. Thus, the quantitative application of hydraulic theory is seldom if ever justified in the real atmosphere (Durran 1990). Three-dimensional aspects of shallow water theory were investigated by Armi and Williams (1993) and Schär and Smith (1993a,b).
2) CONTINUUM THEORIES OF DOWNSLOPE WINDS

The tendency for strong downslope flow is also seen in the theory of steady mountain waves (section 6). Mountain wave theory is a sturdier foundation upon which to build a theory of plunging flow as it accounts for the continuous nature of the atmosphere. In the simple case of constant $N$ and $U$, the flow is characterized by the non-dimensional height \( \hat{h} \) and width \( \hat{a} \) of the ridge:

\[
\hat{h} = \frac{Nh}{U} \quad \text{and} \quad \hat{a} = \frac{Na}{U}.
\]

(20-46)

When $\hat{a} = (Na/U) \gg 1$, the flow is nearly hydrostatic and only $\hat{h} = Nh/U$ matters. When $\hat{h} \ll 1$, linear theory is valid, such as the Queney solution in section 6. For higher ridges, with $\hat{h} \approx 1$, linear theory is invalid. Some insight comes from Long’s model solutions (Long 1955; Huppert and Miles 1969) for finite $\hat{h}$, but those elegant solutions break down when wave breaking begins. It is a challenging nonlinear problem (e.g., Durran 1986, 1990; Wang and Lin 1999).

A big step forward was the 2D numerical simulation of severe downslope winds by Clark and Peltier (1977, 1984) and Peltier and Clark (1979). They found that, for a sufficiently high ridge, when wave breaking began aloft the flow would readjust itself into a severe wind configuration (Fig. 20-15). Clark and Peltier suggested that the mountain wave might be amplified by reflecting from the “self-induced” wind reversal (i.e., critical level) at the wave breaking altitude. To support this idea of downward wave reflection at a critical level, Clark and Peltier (1984) carried out further numerical experiments with reversed wind with above some altitude $Z_c$; that is, an “environmental critical level.” They found that for certain values of $Z_c$, plunging flow could be obtained with smaller mountain heights than in the no-shear cases.

Smith (1985) used Long’s equation to explain the Clark and Peltier discoveries. He postulated a region of stagnant, well-mixed fluid associated with wave breaking. One incoming streamline divides vertically to encompass this region (Fig. 20-16). Solving Long’s equations gave a prediction for the height of the dividing streamline $Z_d$, the shape of the stagnant region, and the wind speed near the ground. Severe downslope winds could occur for dividing streamline heights of

\[
Z_d = \frac{U}{N} \left( \frac{3}{2} + 2\pi n \right) \quad \text{with} \quad n = 1, 2, 3, \text{etc.}
\]

(20-47)

The fact that these special altitudes are spaced at $2\pi n$ rather than $\pi n$ intervals suggests that the wave resonance is not the same as linear resonance caused by wave reflection from a rigid lid. We could call it a “nonlinear resonant reflection.” In the case of reversed wind aloft, the theory equates $Z_c$ and $Z_d$.

Currently, our understanding of severe downslope winds is a combination of the layered and continuous theories. Numerical models of stratified airflow are able to simulate downslope winds (Doyle et al. 2000; Nance and Colman 2000; Reinecke and Durran 2009), but not always to predict them.

3) EXAMPLES OF SEVERE DOWNSLOPE WINDS

As mentioned above, the Boulder downslope wind, often called the “Chinook,” is considered to a prototype case because of the pioneering aircraft survey by Lilly and Zipser (1972, their Fig. 30). We mention a few others here as examples.

The foehn wind in the Alps has received much attention for bringing warm, gusty wind down into Alpine valleys (Hoinka 1985; Zängl 2002; Gohm and Mayr 2004). A south foehn occurs as a cold front approaches the Alps from the northwest, dropping the pressure on the north side of the Alps. This dropping pressure helps to draw deep southerly air across the Alps, and warm leeside descent follows. Northerly foehn can also occur in the Alps (Jiang et al. 2005).

In California, the Santa Ana wind blows from the northeast when an anticyclone develops over Oregon and Washington State. Turning winds aloft promote decoupling and severe wind formation (Fig. 20-17).
Santa Ana wind hugs the surface but skips upward on occasion, causing intermittency (Hughes and Hall 2010; Abatzoglou et al. 2013; Cao and Fovell 2016).

The Bora wind, a plunging northeasterly wind on the Dalmatian coast of the Adriatic, has a well-defined shallow cold layer spilling over a ridge with a slight saddle, down to the sea. It has a brutal region of turbulence in a hydraulic jump over the sea. It was first probed by aircraft during the ALPEX project (Table 20-1) in 1982 (Smith 1987) and later in the MAP project in 1999 (Grubišić 2004).

In the Southern Hemisphere, one of the known downslope winds is the Zonda in Argentina (Norte 2015). This is a deep westerly wind that climbs over the southern Andes and then plunges onto the steppes of Argentina. In seasonality, latitude, and dynamics it is similar to the Chinook.

Our final example is a surprise discovery of plunging flow on the western slopes of Dominica in the eastern Caribbean (Minder et al. 2013). The easterly trade winds exhibit a strongly stratified layer from 500 to 3000 m and a wind maximum at \( z = 700 \) m near cloud base. Over the mountains of Dominica, the trade wind inversion plunges and the whole layer accelerates down the lee side (Fig. 20-18). There does not appear to be a hydraulic jump, but there is intense turbulence above the leeside jet.

Let us compare these six examples of downslope winds: Chinook, foehn, Bora, Santa Ana, Zonda, and Dominica. The Chinook and Zonda are deep, with westerlies throughout the troposphere. The other four have strong directional shear. In fact, the last three, Bora, Santa Ana, and Dominica, are easterly winds with westerlies aloft. They satisfy the theoretical prediction that even small hills can produce strong downslope winds if there is an “environmental critical level.” Many, if not most, severe winds have strong wave breaking in the troposphere and thus have reduced generation of deeply propagating mountain waves.

Other examples of downslope winds are from the Falkland Islands (Mobbs et al. 2005) and Antarctica (Grosvenor et al. 2014).
c. Gap jets

Gap jets are strong winds passing through gaps between hills or through elevated saddles (section 3). Like most other strong local winds, they accelerate down a pressure gradient. This tendency is approximately captured by the simplest form of the Bernoulli equation. Along a level surface in loss-less incompressible flow, the Bernoulli function \( B \) is conserved:

\[
B = P + \left( \frac{1}{2} \right) \rho U^2 \quad (20-48)
\]

(Gaberscek and Durran 2004). With \( B \) conserved along a streamline, the strongest wind will occur at the lowest pressure. The pressure gradient is usually caused by one of three mechanisms: mountain waves aloft, low-level cold air advection, or differential heating from the sun (e.g., a sea breeze situation).

An example of a sea-breeze-driven gap wind is shown in Fig. 20-19 near Monterey Bay on the California coast (Banta 1995). With daytime heating of air over the continent, the higher pressure over the sea pushed air through the gap reaching speeds of 5 m s\(^{-1}\). Much of the wind power generation in California is derived from sea-breeze-driven gap jets.

A second example of gap flow is near the Gulf of Tehuantepec along the Pacific Coast of Mexico (Fig. 20-20). With the arrival of a cold barrier jet from the north, high pressure pushes the air through the gap to the Pacific Ocean (Steenburgh et al. 1998).

Other gap wind studies are Juan de Fuca Strait (Overland and Walter 1981; Colle and Mass 2000), Alaska (Colman and Dierking 1992; Lackmann and Overland 1989), British Columbia (Jackson and Steyn 1994a,b), Columbia Gorge (Sharp 2002), Brenner Pass in the Alps (Gohm and Mayr 2004; Mayr et al. 2007; Marić and Durran 2009), the Aleutians (Pan and Smith 1999), the Mistral (Jiang et al. 2003), Prince William Sound (Liu et al. 2008), Hawaii (Smith and Grubišić 1993), and Gibraltar (Dorman et al. 1995).

An interesting dynamical question concerns the gap wind after it leaves the gap. When the jet leaves the terrain and passes out over a flat plain or ocean, how far does it persist? Relatedly, does it end abruptly in a hydraulic jump or does it slowly spread and mix into the environment? This is probably the most difficult aspect of gap jets to predict.

A second interesting question relates to the width of the gap. Overland (1984) argues persuasively that when the gap width exceeds the Rossby radius (RR),
the flow resembles a barrier jet (see the next section) more than a gap jet. In (20-49) the Coriolis parameter is \( f = 2\Omega \sin(\varphi) \). This idea is similar to the argument by Pierrehumbert and Wyman (1985) regarding the distance of upstream blocking. In both cases, the Coriolis force seems to limit the width of blocking and barrier jets.

d. Barrier jets

Barrier jets are defined here as low-troposphere mountain-parallel jets with the Coriolis force pushing air into the barrier. By this definition, barrier jets must have their barrier to the right of the flow in the Northern Hemisphere and to the left in the Southern Hemisphere. Some barrier jets meeting that definition are given in Table 20-5 and in Fig. 20-21. Such barrier jets are common around the world and influence local climate and meridional heat transport. Jets from pole to equator bring cold air equatorward. Jets toward the pole bring warm air poleward.

In the simplest model of a barrier jet, a geostrophic flow approaches a high mountain barrier and is slowed down by an adverse pressure gradient (Overland 1984; Pierrehumbert and Wyman 1985; Colle and Mass 1996). With slower flow, the Coriolis force is reduced and the flow turns to the left (in the Northern Hemisphere) and flows along the terrain slope. The mountain-induced pressure gradient is balanced by the Coriolis force. The width of the barrier jet is related to \( RR \) in (20-49).

Probably the best-studied barrier jet is along the coasts of British Columbia and southeast Alaska, where the Canadian Rockies rise sharply from the sea. One case there is shown in Fig. 20-22. This jet blows northward along the coast (Winstead et al. 2006; Olson et al. 2007; Olson and Colle 2009).

There are a few examples of “barrier jets” to which the above definition does not apply. Two cases are the cold northerly airflow along the California coast (Burk and Thompson 1996) and a cold southerly flow along the northern coast of Chile (Garreaud and Munoz 2005). These two examples are “mirror image” flows across the equator. They are not simple blocking flows because the Coriolis force pushes air away from the ridge. Rather, they move under the influence of an independent regional pressure gradient.

e. Mountain wakes

A mountain wake is defined as a region of slower airflow downwind of a mountain obstacle. Wakes occur on a small scale of a kilometer or less (Voigt and Wirth 2013), but here we focus on wakes that extend for tens or hundreds of kilometers downwind. While they can be identified over land, they are much easier to see over the uniform ocean (Deardorff 1976; Etling 1989).

There is general agreement that the occurrence of wakes relates closely to the generation and advection of vertical vorticity in the atmosphere:

\[
\xi(x, y) = \frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}, \tag{20-50}\]

For a fluid dynamicist, the first question relates to the origin of this vorticity. One useful vorticity theorem for steady flow is

\[
U \times \xi = \nabla B, \tag{20-51}\]

relating vorticity to gradients in the Bernoulli function (20-48). In the wake of an obstacle, the pressure field tends to even out so if there are gradients in velocity (i.e., in the wake), there must be gradients in \( B \) as well. Gradients in the Bernoulli function are caused by dissipative processes along the streamline. The cause of Bernoulli loss near the mountain could be friction, turbulence, hydraulic jumps, or even convection. In other

### Table 20-5. Some barrier jets.

<table>
<thead>
<tr>
<th>Location</th>
<th>Direction of flow</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sierra Nevada</td>
<td>Northward</td>
<td>Kingsmill et al. (2013)</td>
</tr>
<tr>
<td>East Greenland</td>
<td>Southward</td>
<td>Petersen et al. (2009)</td>
</tr>
<tr>
<td>Alaska</td>
<td>Northward</td>
<td>Olson et al. (2007)</td>
</tr>
<tr>
<td>Zagros</td>
<td>Northward</td>
<td>Dezfuli et al. (2017)</td>
</tr>
<tr>
<td>New Zealand</td>
<td>Southward</td>
<td>Yang et al. (2017)</td>
</tr>
<tr>
<td>Australia</td>
<td>Northward</td>
<td>Colquhoun et al. (1985)</td>
</tr>
<tr>
<td>Rocky Mountains</td>
<td>Southward</td>
<td>Colle and Mass (1995)</td>
</tr>
<tr>
<td>Norway</td>
<td>Northward</td>
<td>Barstad and Grønås (2005)</td>
</tr>
<tr>
<td>Taiwan</td>
<td>Northward</td>
<td>Li and Chen (1998)</td>
</tr>
<tr>
<td>Appalachians</td>
<td>Southward</td>
<td>Bell and Bosart (1988)</td>
</tr>
<tr>
<td>Mexico</td>
<td>Southward</td>
<td>Luna-Niño and Cavazos (2018)</td>
</tr>
</tbody>
</table>

![Fig. 20-21. Sixteen barrier jets around the world. In each case, the Coriolis force pushes the air against the associated mountain range (NOAA background). [Image source: NOAA/NCEI ETOPO1 Ice Surface Global Relief Model (arrows added by author).]](image-url)
words, mountain wakes are an expression of dissipation near the mountain (Schär and Smith 1993a, b; Schär and Durran 1997; Epifanio and Durran 2002a, b). Non-dissipative theories of wake generation are discussed by Smolarkiewicz and Rotunno (1989) and Rotunno et al. (1999).

The second key question is about the advection of the vorticity. Once created, vertical vorticity may be approximately conserved following the flow obeying

$$\frac{D\xi}{Dt} = 0.$$ (20-52)

Vorticity can be advected by the mean flow or by vortical eddies within the wake. If the mean flow dominates, the wake will just trail off downstream as a straight region of reduced wind speed. Stronger wake vorticity will self-advect, causing steady return flows or oscillating wakes. Oscillating wakes are well known in fluid dynamics; for example, high-frequency shedding eddies from telephone wires cause an audible hum.

To illustrate these wake possibilities, we give three real examples. In Fig. 20-23, we show a sunglint image of a long wake extending westward from the Windward Islands in the eastern Caribbean (Smith et al. 1997). In Fig. 20-24, we show the wake of Madeira as seen in a shallow layer of stratocumulus clouds (Grubišić et al. 2015). The eddies of alternating spin move slowly downwind while self-adveecting nearby vortices. In Fig. 20-25, we see the quasi-steady wake of the Big Island of Hawaii, as mapped by a research aircraft repeatedly crossing the wake (Smith and Grubišić 1993). In this wake, two large quasi-steady eddies form the wake. In environmental easterly flow, they cause a westerly return flow pushing air back toward the island.

Larger-scale Alpine wakes have been discussed by Aebischer and Schär (1998) and Flamant et al. (2004).

f. Thermally driven winds

An important local effect of sloping terrain is the thermally driven slope winds (Barry 2008; Whiteman 1990, 2000). These winds arise whenever there is sensible heat flux between the sloping Earth’s surface (e.g., soil, forest, snow) and the atmosphere. When the heat flux is negative, as during a cold clear night, a dark winter season, or on the shady side of a mountain range, the lowest few meters of the atmosphere lose heat and become cooler and denser. Under the influence of gravity, this dense air layer begins to slide downslope.

---

Fig. 20-22. A southeasterly barrier jet along the Alaska coast seen in wind-induced ocean roughness. Arrows show the wind direction. See Fig. 20-21 for location. [From Loescher et al. (2006).]

---

Fig. 20-23. A sunglint image of a long wake extending westward from the Windward Islands in the eastern Caribbean (Smith et al. 1997). In this wake, two large quasi-steady eddies form the wake. In environmental easterly flow, they cause a westerly return flow pushing air back toward the island.
steady state is often reached where the continuous loss of heat to the surface is balanced by continuous compressive heating. This “drainage flow layer” moves silently downslope, following the terrain and converging in valleys and valley junctions, not so different than stream tributaries converging into rivers. Once joined into a “valley wind,” these cold airstreams continue to move away from the high terrain, even if the slope of the valley floor is small. At mountain peaks and along ridge crestlines, where the downslope flows first originate, they draw their mass from the free atmosphere. At lower elevations, these flows can spread into an open plain where they mix into the environment or collect in a depression, forming a cold-air pool (CAP; see next section).

When these cold drainage flows are sufficiently large and deep they are called katabatic winds. Katabatic winds may not need continual cooling; their momentum may be sufficient to keep them moving. The largest, most persistent, and well-studied katabatic flows occur around the perimeter of the Antarctic Ice Sheet (e.g., Davis and McNider 1997). This air has lost heat over several days during its stay on the high ice sheet. As it spills down the steep slopes to the sea, this layer of air can reach speeds of 50 m s\(^{-1}\) or greater. This high-momentum flow can disperse out over the Southern Ocean or decelerate suddenly in a turbulent hydraulic jump.

When the heat flux from the surface is positive, as on a sunny slope, the first few meters of air warms and tries to rise. Often, because the free atmosphere is statically stable, the air slides up along the slope instead of rising vertically. This “upslope layer” of air can also reach an approximate steady state where continued positive sensible heating is balanced by adiabatic cooling or by existing vertical gradients of potential temperature in the free atmosphere. When this layer reaches a ridgeline or a mountain peak, it must leave the surface, either forming an elevated horizontally moving gravity current or rising vertically to form cumulus clouds aloft. The midmorning occurrence of cumulus clouds over the
Rockies and the Alps in summer is caused by the upslope winds rising off the surface at mountain peaks (e.g., Banta and Schaaf 1987; Damiani et al. 2008; Bennett et al. 2011; Smith et al. 2012). Figure 20-26 shows a GOES image of cumulus clouds forming in this way over the Rincon Mountains in Arizona from Damiani et al. (2008). If there is a prevailing wind aloft, the elevated warmed air can be advected downwind. In the case of the Rocky Mountains, the impact of this diurnally warmed midtroposphere air has been linked to severe weather events hundreds of kilometers to the east (Carbone and Tuttle 2008; Li and Smith 2010).

g. **Cold-air pool**

One of the most important impacts of terrain on surface winds is the cold-air pool (CAP). A CAP will form in a terrain depression when dense, cold air collects and pools under the influence of gravity. The cold air often comes from a negative surface heat budget with strong thermal infrared cooling to space. Equally important, however, is the advection of warmer air in the middle troposphere. The CAP is usually of a diurnal or a synoptic type. In the diurnal variety, the nighttime radiative clear-sky cooling dominates and the CAP may dissipate...
the next day. In the synoptic type, lasting many days, several processes may contribute. The cold air in the basin may come from cold air advection or a negative surface heat flux arising from low sun angle or high albedo. Warm air advection aloft could be important too.

As long as the CAP remains in place, the surface winds are nearly calm. Winds aloft will not be able to mix their momentum down into the CAP due to its stable stratification. Furthermore, any weak pressure gradient could tilt the stable inversion but probably not push the cold air out of the basin.

Many occurrences of CAP have been studied around the world. Examples include Salt Lake Valley (Crosman and Horel 2017; Wei et al. 2013; Lu and Zhong 2014; Lareau et al. 2013), the Columbia basin (Whiteman et al. 2001), western U.S. canyons (Whiteman et al. 1999a,b; Billings et al. 2006), Alpine valleys (Zängl 2005), interior Alaska (Mölders and Kramm 2010), the Appalachians (Allwine et al. 1992), and Aizu basin, Japan (Kondo et al. 1989). An example from the Salt Lake Valley is shown in Fig. 20-27. During a 10-day period from 31 December to 10 January, warm air exceeding \( \theta = 290 \text{ K} \) moved in above colder surface air with \( \theta = 280 \text{ K} \). Halfway through the event on 4 January cold air advection weakened the CAP, but it strengthened again on next day. Not until 9 January did cold air advection aloft eliminate the CAP.

Two CAP examples from April and May in the Aizu basin in Japan are shown in Fig. 20-28. Both cases have westerly winds aloft. In the stippled CAP layer, the winds are calm. Only after the CAP disappeared did the surface winds approximate the winds aloft.

The occurrence of CAP is considered a forecasting challenge of the first order. CAP is closely associated with pollution events with human health impacts. CAP forecasting is complex because of the multiple processes involved such as the surface heat budget, shear-induced turbulence, pressure gradient force, and temperature advection aloft. Some success using numerical mesoscale models is reported by Zhong et al. (2001) and Zängl (2005).

5. Orographic precipitation

Orographic precipitation (OP) is generally defined as the enhancement or spatial redistribution of precipitation by mountains. In some cases, mountains will cause precipitation where there would otherwise be little or none. More commonly, the mountains will enhance or reduce precipitation in already wet regions. Mountains usually enhance precipitation on the windward slopes where air is rising. They reduce precipitation on the lee slopes where air is sinking (Fig. 20-29). The precipitation enhancement ratio (PER) is the ratio of precipitation \( P \) at the mountain peak, or some other selected point, to a reference point or region \( P_{\text{REF}} \):

\[
\text{PER} = \frac{P}{P_{\text{REF}}}. \quad (20-53)
\]

The reference region should be far enough removed that it is unaffected by the mountain, but still in the same climate zone. Often, the reference region is difficult to define (Dettinger et al. 2004). A site with OP enhancement would have \( \text{PER} > 1 \); in the lee of mountains, in the so-called “rain shadow”, \( \text{PER} < 1 \). Both the numerator and denominator of (20-53) could be the instantaneous precipitation rates, storm event totals, or
long-term averages. Another useful integral measure of OP, the drying ratio, will be defined later.

Orographic precipitation is very important in hydropower, water resources for cities and farms, and in the formation of glaciers and large ice sheets. The idea of mountains as the "water towers of the world" arises from the combination of enhanced precipitation by forced air ascent and increased water storage at high elevation due to lower temperatures. In the drier climates of the world, large cities such as Los Angeles, Denver, Santiago, Cairo, Cape Town, and Sydney obtain most of their water resources from mountain-fed rivers. Likewise, many of the world’s most productive farmlands need to be irrigated from mountain-fed rivers, for example, the Tigris and Euphrates in the Fertile Crescent and California’s productive Central Valley. For hydropower, the elevation plays another role. By capturing precipitation before it falls to sea level, it provides the potential energy that can be converted to electricity in a hydropower turbine.

On high mountains at high latitudes, annual snowpack accumulates until balanced by slow gravitational glacial sliding down the slopes. For the large ice sheets of Greenland and Antarctica, the thick ice sheet provides its own terrain. The highest annual precipitation occurs near the windward edges of these plateaus where the moist air in frontal cyclones is forced to rise.

Even the early books on meteorology mention the orographic enhancement of precipitation caused by mountains and the attendant rain shadow on the lee side. This wet–dry contrast was explained by the thermodynamic theory of moist adiabatic lifting and dry descent, put forward by the American scientist James Espy and others in the nineteenth century. Air that is lifted by sloping terrain cools by adiabatic expansion and the saturation vapor pressure \( e_{\text{SAT}}(T) \) decreases. Correspondingly, the relative humidity (RH) increases toward unity (section 2d). Above the LCL, where RH = 1, the excess water vapor condenses onto microscopic cloud condensation nuclei (CCN) to form cloud droplets. Under certain conditions, the cloud droplets can combine to form hydrometeors (i.e., raindrops, snowflakes, or graupel) that fall to Earth’s surface. When the air reaches the lee side and descends, if most of the condensed water has been removed, it will warm at the dry adiabatic rate, resulting in warmer, drier air. If the wind direction is persistent, the lee side will experience a drier “rain shadow” climate due in part to the descent and in part to the upstream removal of water vapor.

These basic ideas of moist thermodynamics are so fundamental to atmospheric science that they are taught in virtually every college-level course on meteorology and climate. This subject is very well presented in standard textbooks such as Wallace and Hobbs (2016), Yau and Rogers (1996), Mason (2010), and Lamb and Verlinde (2011). Because of its intuitive nature, these basic principles of moist thermodynamics are often illustrated with mountain upslope and downslope flows (i.e., Fig. 20-29).

The history of research on orographic precipitation can be traced using the review articles in Table 20-6.
a. Measuring orographic precipitation

The measurement of precipitation in complex terrain has benefited from several new technologies, but it is still difficult and prone to error. Due to strong winds carrying rain and snow particles, precipitation gauges are subject to “undercatch.” Snow is particularly difficult to record due to its variable packing density and its redistribution from wind transport (Wayand et al. 2015). Furthermore, due to difficulty in gauge installation and maintenance, gauges are sparse in regions of steep terrain with a large bias toward easy-to-reach valley sites. These two problems, undercatch and valley siting bias, cause an underestimation in regional precipitation amounts.

Another serious difficulty is the interpolation of precipitation data between gauges. A widely used interpolation scheme is the Parameter-Elevation Regressions on Independent Slopes Model (PRISM) algorithm by Daly et al. (1994), Lundquist et al. (2010), and Henn et al. (2018). This method uses data from each terrain facet to interpolate precipitation amounts. While the overall accuracy of the PRISM interpolation is unknown, it provides very useful spatial patterns and regional totals. A few authors have used stream gauge data to verify direct precipitation measurements, but runoff delays, infiltration, storage, and evaporation introduce new uncertainties (e.g., Lundquist et al. 2010).

The PRISM-derived annual precipitation distribution for the United States in Fig. 20-30 illustrates several aspects of OP. The intense orographic precipitation along the U.S. West Coast in Fig. 20-30 is caused by landfalling Pacific winter frontal cyclones (Fig. 20-31) and atmospheric rivers. The OP influence of the coastal ranges, the Sierra Nevada Range, and the Cascades in Washington, Oregon, and California is well studied (e.g., Pandey et al. 1999; Dettinger et al. 2004; Reeves et al. 2008). Just to the east, the intermountain states of Idaho, Utah, Wyoming, and Colorado are mostly in a rain shadow of the coastal ranges, but they still have some local OP maxima. In the southwestern United States, Arizona and New Mexico receive OP in the summer monsoon season, especially along the Mogollon Rim. In the eastern United States, the Appalachians enhance precipitation from the Carolinas northward to New York and Vermont.

In flat terrain, meteorological radar has revolutionized quantitative precipitation measurement. Radar has great advantages on mountaious terrain too, but it has some limitations (Hill et al. 1981; Houze et al. 2001; Houze 2012). First is beam blocking. Most radars in high terrain have limited sectors in which they see. In addition to sector blocking, radars have trouble seeing the first kilometer or so above terrain. This is problematic because in steep terrain the rain rate may increase over the last few hundred meters due to raindrop scavenging of cloud droplets. Thus radars, like gauges, tend to underestimate precipitation. A recent valuable comparison between radar and other observing systems was the Olympic Mountain Experiment (OLYMPEX) project (Houze et al. 2017).

An ideal situation for radar is the monitoring of precipitation on the Caribbean island of Dominica (Fig. 20-32; Smith et al. 2009a,b). In this case, the Météo-France radars are located on the adjacent islands of Guadeloupe and Martinique, giving nearly unobstructed views of Dominica clouds. It found a precipitation enhancement ratio (PER) of about four. Even so, the off-island radar probably missed some low-level enhancement. Airborne radar can also avoid beam-blocking issues (Geerts et al. 2011).

Time-resolved radar images have provided some keen insights into the mechanisms of orographic precipitation. Consider the well-known time–distance diagrams of radar reflectivity over Wales in Fig. 20-33 (Hill et al. 1981). At the lower altitudes (500–900 m MSL), the radar detects incoming precipitating rain cells, probably embedded convection in a frontal system. When they reach the coast, the rain rate detected...
at that level increases and merges into a nearly continuous downpour. Meanwhile, aloft (2700–3300 m MSL) the cells are less disturbed. Such observations can be interpreted two different ways. First, it might support the important “seeder–feeder” theory of orographic precipitation. In that theory, preexisting mid-level clouds (i.e., the “seeder cloud”) rain down into low-level cloud caused by forced orographic ascent. Scavenging of cloud droplets in the lower “feeder cloud” enlarges the hydrometeors and increases the low-level precipitation. A second interpretation is that the forced ascent over the coast has invigorated the convective cells. Such questions are fundamental to the science of orographic precipitation.

20.26 METEOROLOGICAL MONOGRAPHS VOLUME 59

30-yr Normal Precipitation: Annual
Period: 1991-2010

FIG. 20-30. U.S. annual precipitation (in.) using the PRISM interpolation method. (Copyright 2015 PRISM Climate Group, Oregon State University.)

FIG. 20-31. GOES-West TIR image of a Pacific frontal cyclone hitting the west coast of North America (source: NOAA).

FIG. 20-32. Average daily rainfall (mm day$^{-1}$) over mountainous Dominica for 2007, derived from the Météo-France radar on Guadeloupe (TFFR). Trade wind flow is from right to left. Note the rainfall maxima over the hills and dry rain shadow. [From Smith et al. (2012).]
FIG. 20-33. Time–distance diagram showing radar echoes moving eastward over the Bristol Channel and the Glamorgan Hills in Wales. Two altitudes are shown. (a) 500–900 m and (b) 2700–3300 m. Low- and midlevel precipitating cells drift over the coastline and amplify. [From Hill et al. (1981), copyright Royal Meteorological Society.]
Satellites have revolutionized precipitation by continuous monitoring of weather systems over terrain through passive thermal infrared (TIR) imagery. The example image in Fig. 20-31 shows a Pacific frontal cyclone entering North America and passing over the coastal mountain ranges. As seen in this image, however, the observed cloud patterns appear to be only slightly modified by the terrain. Note three small “foehn gap” clearings just east of the Cascades. Strong orographic precipitation enhancement is mostly occurring at low levels, blocked from view by the higher clouds.

The more recent development of satellite-borne radar has revolutionized precipitation estimates over rough terrain (Shige et al. 2006; Sapiano and Arkin 2009; Anders and Nesbitt 2015; Houze et al. 2017). With a top-down radar view, beam blocking is largely eliminated. Still, due to low-level enhancement from scavenging, errors may arise from nearby ground clutter. Today, satellite radar systems such as Tropical Rainfall Measuring Mission (TRMM) and Global Precipitation Measurement (GPM) provide nearly global monitoring. An example of TRMM data over Southeast Asia is shown in Fig. 20-34. Strong orographic enhancement of monsoon precipitation is seen over the Western Ghats (WG) of India and the Arakan Yoma (AY) mountains of Myanmar (Shige et al. 2017).

A most provocative and interesting hypothesis has been put forward by Jessica Lundquist (2018, personal communication) and Beck et al. (2019), that modern well-tuned mesoscale models such as the Weather Research and Forecasting (WRF) Model may be more accurate than interpolated rain gauge data, and in some cases more accurate than the spot rain gauge data. This claim recognizes 1) errors in observing technology, 2) interpolation errors, and 3) recent advances in modeling skill. Objections to the idea arise, however. With no accepted objective truth, it leaves no clear path forward for model improvement (i.e., Stoelinga et al. 2003; Garvert et al. 2007). The hydrology community seems poised to accept the supremacy of models for annual average precipitation. However, on a storm-by-storm basis, the models may have less skill. The chaotic nature of the atmosphere leaves single nested mesoscale storm analysis with large uncertainty. In principle, this problem might be overcome with a probabilistic forecast using an ensemble approach.

b. Orographic precipitation concepts

1) WATER VAPOR FLUX

An important quantity in orographic precipitation is the horizontal water vapor flux (WVF) vector given by the vertical integrals

$$ WVF = \int_0^\rho_w U \, dz = \int_0^\rho q_v U \, dz = -\frac{1}{g} \int_0^{q_v} q_v U \, dp, $$

(20-54)

where $\rho$ is the air density, $\rho_w$ is the water vapor mass density, $q_v$ is the specific humidity, and $U(z)$ is the horizontal wind vector. This vector quantity WVF has
typical magnitudes of 50–1000 kg m\(^{-1}\) s\(^{-1}\). Large values are found in warm trade winds, monsoon inflows, hurricanes, and “atmospheric rivers” carrying moist air from the tropics to the midlatitudes (Hu et al. 2017). If one imagines two tall posts spaced a distance \(L\) apart, with the wind blowing between them, the included rate of water vapor transport is the product of \(WVF\) and \(L\). To be more precise, the water vapor flux is the dot product of the \(WVF\) vector and the normal vector \(ds\) describing the length and orientation of a line segment on Earth’s surface, so

\[
\text{Transport} = WVF \cdot ds \tag{20-55}
\]

in units of kilograms per second.

The water vapor flux vector (20-54) is easily computed from a balloon sounding in which humidity, wind speed, and wind direction are measured with altitude. Alternatively, if the column density of water vapor (i.e., precipitable water; \(PW\)) is known from a GPS time delay (Bevis et al. 1994),

\[
PW = \int_0^h \rho q_V \, dz = -\left(\frac{1}{g}\right) \int_{p_s}^0 q_V \, dp, \tag{20-56}
\]

the water vapor flux could be estimated as the product

\[
WVF = p_w \cdot U_w,
\]

where \(U_w\) is a water vapor weighted mean wind vector. If the water vapor profile is exponential (20-7), then \(PW = p_w h w\). For example, if the water vapor density at Earth’s surface is \(p_w = 0.01\) kg m\(^{-3}\), \(h w = 2000\) m, and the average wind speed is \(|U| = 20\) m s\(^{-1}\), then the magnitude of the vapor flux vector is \(|WVF| = 400\) kg m\(^{-1}\) s\(^{-1}\).

The conservation of water relates the difference in surface evaporation (\(E\)) and precipitation (\(P\)) to the divergence in the horizontal water vapor flux vector field:

\[
E - P = \nabla \cdot WVF, \tag{20-57}
\]

neglecting any condensed water storage in the atmosphere in the form of cloud liquid or ice or hydrometeors.

In this article, we emphasize the use of water vapor flux (20-54) to quantify the bulk amount of orographic precipitation. In a simple unidirectional wind field, we consider a flow path crossing a mountain range from point A on the windward side to point B on the leeward side. In steady-state 2D flow, with no evaporation, the change in \(WVF\) along this path is the downstream integral of (20-57):

\[
WVF_A - WVF_B = \int_A^B P \, dx, \tag{20-58}
\]

where \(dx\) is an increment of distance along the path. The drying ratio (\(DR\)) is defined as the fraction of the water vapor flux that is removed by precipitation:

\[
DR = \frac{WVF_A - WVF_B}{WVF_A} = \int_A^B P \, ds \quad \text{with} \quad 0 < DR < 1. \tag{20-59}
\]

This simple definition of \(DR\) may be difficult to apply to the real world due to three-dimensionality and unsteadiness.

2) SIMPLE UPSLOPE MODELS OF OROGRAPHIC PRECIPITATION

A number of authors (e.g., Rhea and Grant 1974; Smith 1979; Alpert and Shafir 1989; Sinclair 1994) have proposed simple upslope models of orographic precipitation. These models clarify aspects of the physics and provide rough estimates of precipitation amount.

In the simplest upslope model of orographic precipitation, we assume that the airflow at each altitude rises along the same slope as the underlying terrain. This gives the vertical velocity field

\[
w(x,y,z) = U(z) \cdot \nabla h(x,y) \tag{20-60}
\]

from (20-14) and (20-15). If the air is saturated, the rate of condensate formation aloft is given by the product of the vertical velocity and the slope of the moist-adiabatic saturation water vapor density \(dp_{SAT}/dz\). The logarithmic derivative defined as

\[
H_{SAT}^{-1} = -\frac{1}{\rho_{SAT}} \frac{dp_{SAT}}{dz} \tag{20-61}
\]

is used to define the “saturation scale height” for water vapor for air parcels ascending moist adiabatically. This scale height (20-61) is distinct from (20-8) as it represents a thermodynamic process rather than an observed state of the atmosphere. They are equal in a saturated atmosphere with a moist adiabatic profile.

If the air column has a moist-adiabatic lapse rate and is saturated at all levels, the vertical integral of water condensation rate (\(CR\)) is related to the vertical air velocity (20-60) by

\[
CR(x,y) = \int_0^h H_{SAT}^{-1} \rho_{SAT} w(x,y,z) \, dz = \int_0^h H_{SAT}^{-1} \rho_{SAT} U \cdot \nabla h(x,y) \, dz. \tag{20-62}
\]

In such an atmosphere, if \(H_{SAT}\) is constant and if the condensed water falls to the ground instantaneously, then the precipitation rate [using (20-54)] is

\[
P(x,y) = H_{SAT}^{-1} WVF \cdot \nabla h(x,y) \tag{20-63}
\]
with units kg m$^{-2}$s$^{-1}$. This formula (20-63) is the “up-slope model” of orographic precipitation. It states that the precipitation rate is proportional to the dot product of the horizontal water vapor flux and the terrain slope. For example, if the flux and terrain slope vectors are parallel, with $|WVF| = 300$ kg m$^{-1}$s$^{-1}$ and $|\nabla h| = 0.1$ and $H_{SAT} = 3000$ m, then $P = 0.01$ kg m$^{-2}$s$^{-1} = 36$ mm h$^{-1}$.

The coefficient in (20-63), the inverse saturation scale height, is approximated using the Clausius–Clapeyron equation (20-4), giving

$$H_{SAT}^1 = -\frac{1}{\rho_{SAT}} \frac{d\rho_{SAT}}{dz} \approx \left[ \left( \frac{1}{e_s} \right) \frac{d e_s}{dT} \right]_{d z} = \frac{L\gamma}{RT^2},$$

(20-64)

where the moist adiabatic lapse rate is $\gamma = [dT/dz]_{moist}$. In the special case when little latent heat is released, rising parcels cool dry adiabatically so $\gamma = -(g/CP) = -9.8^\circ C$ km$^{-1}$ and (20-64) becomes

$$H_{SAT}^1 \approx \frac{Lg}{RT^2 CP} or H_{SAT} = T^*/52.9.$$

(20-65)

This special case might be valid in a cold climate where there is so little water vapor in a saturated atmosphere that the lifting is nearly dry adiabatic. In that case, the $H_{SAT}$ values are small (e.g., $H_{SAT} \approx 1500$ m). In the general case, when latent heat of condensation is added to a rising air parcel, the parcel cools more slowly and $H_{SAT}$ is much larger than (20-65). For example, if the moist lapse rate is $\gamma = -0.0065^\circ C$ m$^{-1}$, then from (20-64) $H_{SAT} \approx 2600$ m.

The upslope model is generally understood to be an oversimplified view of orographic precipitation. First, the streamline slope would probably decrease aloft rather than remaining parallel to Earth’s terrain several kilometers below. Second, the conversion from cloud droplets to hydrometers would probably be delayed and generally incomplete. Third, further delays and possibly evaporative losses would occur as the hydrometeors fall to Earth. Fourth, the air approaching a mountain range may not be fully saturated or have dry layers within it. All four of these errors would cause an overprediction of precipitation rate by (20-63). Note also that (20-63) predicts negative precipitation in downslope regions; an unphysical result. In addition, the use of (20-63) is quite sensitive to the amount of smoothing applied to the terrain. Extensions to the upslope model to resolve these issues are discussed in a later section.

3) WATER VAPOR FLUX DEPLETION

As air rises over higher and higher terrain and precipitation occurs, water vapor will be removed from the airstream and $WVF$ (20-54) will diminish. With a smaller $WVF$, the same incremental terrain rise would yield less rain. Recall that the $WVF$ in (20-63) is the local value, so this reduction in $WVF$ must be accounted for. To analyze this depletion, consider a unidirectional airflow over a ridge. The conservation of water (20-57) or (20-58) requires

$$\frac{dWVF}{dx} = -P.$$  

(20-66)

Combining (20-63) with the 1D version of (20-57) gives

$$\frac{1}{WVF} \frac{dWVF}{dx} = -\frac{1}{H_{SAT}} \frac{dh}{dx}.$$  

(20-67)

which integrates to

$$WVF(x) = WVF_0 \exp \left[ \frac{-h(x)}{H_{SAT}} \right].$$  

(20-68)

The water vapor flux decreases exponentially as the air climbs to higher elevation. Expressed as a total drying ratio (20-59), this is

$$DR = 1 - \exp \left( \frac{h_{MAX}}{H_{SAT}} \right).$$  

(20-69)

After the air crosses the ridge and precipitation ends, the DR is locked in with a value set by the height $h_{MAX}$ of the highest ridge. For example, a ridge with $h_{MAX} = 1000$ m and $H_{SAT} = 2000$ m, (20-69) gives $DR = 0.39 = 39\%$. As $H_{SAT}$ is greater in a warmer atmosphere, (20-69) suggests that warmer seasons or climates may have reduced DR. While admiring the simplicity of (20-69), the reader should recall the numerous strong assumptions that were used to derive it.

For an initially unsaturated atmosphere, air must ascend to the LCL before condensation can begin. This delay could be accounted for by using an effective maximum mountain height in (20-69). Using (20-9),

$$h_{MAX} = h - Z_{LCL} = h - 25[100 - RH(\%)].$$  

(20-70)

For the example above, with $h = 1000$ m, but with RH = 90%, then $h_{MAX} = 750$ m and DR = 31%. If the mountain height does not exceed the LCL height, (20-70) gives a negative value and we assume that DR = 0.

4) EXTENSIONS OF THE UPSLOPE MODEL

Several extensions to the upslope model have been proposed (Kunz and Kottmeier 2006; Roe and Baker 2006; Gutmann et al. 2016). These important extensions are designed to account for the shift and decay of vertical air motion $w(x, y, z)$ with height above terrain (see section 6) and the microphysical time delays of...
hydrometeor formation and fallout. One of the most widely used extensions is the linear theory of orographic precipitation (LTOP) proposed by Smith and Barstad (2004). In this model, the spatial pattern of precipitation \( P(x, y) \) is related to the spatial terrain function \( h(x, y) \) through their Fourier transforms (see section 3). In the simplest case of a moist adiabatic profile, 

\[
\hat{P}(k, l) = \frac{i C_W \sigma \hat{h}(k, l)}{(1 - i m H_W)(1 + i \sigma \tau_C)(1 + i \sigma \tau_F)}. \tag{20-71}
\]

In (20-71), \( h(k, l) \) and \( \hat{P}(k, l) \) are the double Fourier transforms of the terrain and the predicted precipitation field, respectively. Symbols \( k \) and \( l \) are the zonal and meridional wavenumbers. The intrinsic frequency is \( \sigma = U k + V l \), where \( U \) and \( V \) are the zonal and meridional wind components. The use of Fourier methods in (20-71) arises because of the scale dependence of airflow dynamics and cloud physics.

The coefficient \( C_W \) in (20-71) represents the column rate of water condensation for a unit of base uplift. In a saturated moist neutral atmosphere \( C_W \approx \rho_W (\sigma = 0) \), that is, the water vapor density at the surface. Other parameters in (20-71) are the vertical wavenumber of the flow disturbance \( m = (N^2/|U|); \text{ section 6} \) and the characteristic delay times for conversion of cloud droplet to hydrometeors (\( \tau_C \)) and fallout (\( \tau_F \)). The first factor in the denominator accounts for the upward decay and/or upwind shift of the ascent field. The second and third factors account for the downwind drift of condensed water by the wind. Delay times of \( \tau_C = \tau_F = 1000 \) s give reasonable results in some regions (see references in Table 20-7).

In cases with fast winds and long delay times, the product of intrinsic frequency and the delay time grow large (i.e., \( \sigma \tau \gg 1 \)) and the magnitude of \( P \) from (20-71) diminishes. Thus, this formula captures the basic characteristic of orographic precipitation; namely, that it is a

<table>
<thead>
<tr>
<th>Location</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oregon</td>
<td>Smith et al. (2005)</td>
</tr>
<tr>
<td>Olympics</td>
<td>Anders et al. (2007)</td>
</tr>
<tr>
<td>Southern Chile</td>
<td>Smith and Evans (2007)</td>
</tr>
<tr>
<td>Iceland</td>
<td>Crochet et al. (2007)</td>
</tr>
<tr>
<td>Norway</td>
<td>Schuler et al. (2008)</td>
</tr>
<tr>
<td>Southern California</td>
<td>Hughes et al. (2009)</td>
</tr>
<tr>
<td>Sierra Nevada</td>
<td>Lundquist et al. (2010)</td>
</tr>
<tr>
<td>British Columbia</td>
<td>Jarosch et al. (2012)</td>
</tr>
<tr>
<td>Mars</td>
<td>Scanlon et al. (2013)</td>
</tr>
<tr>
<td>Arizona</td>
<td>Hughes et al. (2014)</td>
</tr>
<tr>
<td>Northern Chile</td>
<td>Garreaud et al. (2016)</td>
</tr>
<tr>
<td>Alaska</td>
<td>Roth et al. (2018)</td>
</tr>
<tr>
<td>Spain</td>
<td>Durán and Barstad (2018)</td>
</tr>
</tbody>
</table>

“race against time.” If cloud water is not quickly converted to hydrometeors and if the hydrometeors do not quickly fall to the ground, the chance to precipitate will be lost. The condensed water will move into the leeside region where it will quickly evaporate in descending air.

In the special case of \( m = \tau_C = \tau_F = 0 \), the LTOP model (20-71) reduces to

\[
\hat{P}(k, l) = i C_w \sigma \hat{h}(k, l), \tag{20-72}
\]

identical to the upslope model (20-63). This happens because the forced ascent \( U \cdot \nabla h \) in (20-63) is equivalent to \( i \sigma h \) in Fourier space (section 3).

After \( \hat{P}(k, l) \) is computed, the rainrate distribution is recovered from an inverse Fourier transform according to

\[
P(x, y) = \text{Max} \left\{ \left| \int \hat{P}(k, l) \exp[-i(kx + ly)] dk dl + P_{\infty} \right| \right\}. \tag{20-73}
\]

The introduction of the background rainrate \( P_{\infty} \) and the Max function eliminates negative values. An example of LTOP application is shown in Fig. 20-35 for the Olympic Mountains in Washington State during southwesterly airflow (Smith and Barstad 2004; Anders et al. 2007).

The resulting precipitation patterns have maxima on the southwest-facing ridges and a dry rain shadow on the lee slopes to the northeast.
LTOP has the advantage of very fast run time, often less than 1 s, because of the efficiency of the fast Fourier transform. The LTOP model has been applied to mountains and glaciers around the world and even to the planet Mars (Table 20-7). The LTOP model has been extended by Barstad and Smith (2005), Barstad and Schüller (2011), and Siler and Durran (2015).

The most serious limitations of all upslope models, including extended upslope models such as LTOP, is that they cannot be run continuously. In practice, such models should be run only when rain in the region is predicted or is already occurring. This is so because the models assume that the incoming airflow is near saturation and/or that some regional precipitation is already occurring. Thus, upslope models only predict the spatial distribution of precipitation, not whether it will precipitate or not. This limitation on (20-71) might be corrected using a modified terrain as in (20-70). In general, however, if a robust prediction is needed, a full physics mesoscale model should be used (e.g., Durran and Klemp 1986). The NCAR community WRF Model can be nested into a global model to capture the varying nature of the environment. It will then give a complete description of the precipitation timing and spatial distribution.

The partial success of extended upslope models may suggest a simplicity to orographic precipitation, but a deeper look reveals many important complexities that will always limit their accuracy and utility. Several of these complexities are described in the next section.

c. Classification of orographic precipitation

Progress on understanding OP will be slow until it is recognized that OP comes in many forms. Any attempt to combine different types into a single generic OP category will cause confusion and false comparisons. Unfortunately, there are many potential control parameters, so we are faced with a multidimensional parameter space. We organize this section by discussing several factors and processes that control the physics of OP.

1) Nature of the Ambient Disturbance

While the textbook description of OP (Fig. 20-29) shows an undisturbed incoming flow lifted by terrain, the more common situation is that the ambient flow is already strongly disturbed. Thus, we must categorize OP by the nature and degree of preexisting ambient disturbance. A common occurrence of ambient disturbance is the midlatitude frontal cyclone moving eastward against the west coasts of North or South America from the Pacific Ocean (Fig. 20-31) or England, Wales, Scotland, or South Africa from the Atlantic Ocean. Frontal cyclones moving through the Mediterranean Sea impacting the Alps are discussed by Rotunno and Ferretti (2001) and Smith et al. (2003).

These cool-season weather systems are precipitating all by themselves, but their precipitation patterns can be strongly modulated by terrain. Existing storms set the local environment for OP in many ways. First they usually bring strong water vapor flux against the terrain, often in the form of an atmospheric river (Kingsmill et al. 2006, 2013; Hughes et al. 2014; Hu et al. 2017).

Second, the storm system will control the stability profile and the relative humidity of the inflow airstream. Unless the airflow has an RH > 80% or so, mountain lifting may not bring the air to saturation (20-70). The stability will influence upstream blocking and deflection and mountain wave generation. Third, preexisting precipitation in the ambient disturbance will accelerate the conversion of cloud water to rain or snow. According to the well-known seeder–feeder mechanism, droplet scavenging or snowflake riming from existing precipitation will reduce the time delay associated with “autoconversion,” the conversion of cloud droplets to hydrometeors.

So important is the ambient frontal cyclone to OP that investigators subdivide the cyclone environment into subenvironments, for example, the warm front (WF), the warm sector (WS), and the post front (PF) (e.g., Zagrodnik et al. 2018; Purnell and Kirshbaum 2018). The WF environment has veering winds, stable layers, and widespread stratiform precipitation. The WS environment has strong water vapor flux and moist neutral profiles. The PF environment has colder air, unstable profiles, and little background precipitation. Each of these environments responds differently to the terrain.

Another highly disturbed environment is the tropical cyclone. Like frontal cyclones, warm-season tropical cyclones bring very large water vapor fluxes (exceeding 1000 kg m\(^{-2}\) s\(^{-1}\)) against terrain. OP within these systems have been studied in Taiwan (Fang et al. 2011; Hong et al. 2015), Japan (Wang et al. 2009), the Philippines (Bagtasa 2017), and the Caribbean (Smith et al. 2009a).

It would be interesting and useful to identify cases of OP in simple undisturbed flow, but such cases are difficult to find. Perhaps we should look at steady easterly trade winds in the subtropics (e.g., Smith et al. 2012) or steady westerly monsoon winds (e.g., Zhang and Smith 2018). However, even these low-latitude wind systems are not fully steady and undisturbed. Weak tropical disturbances embedded in these flows cause precipitation rates to fluctuate significantly from day to day. The reason may be subtle changes in wind speed, humidity, large-scale ascent, or static stability.
2) MICROPHYSICAL PROCESSES

Orographic precipitation is sensitive to cloud microphysical processes, and thus OP can be used as a natural laboratory for cloud studies. It is convenient for these studies to have a fixed upslope location where instruments and numerical simulations can be focused. OP events include a wide variety of factors that control precipitation: aerosol type and concentration, updraft speed, embedded convection, and temperature. Examples of OP field projects include HaRP, TAMEX, CALJET, IMPRoVE, DOMEX, and OLYMPEX (Table 20-1).

A standard subdivision in cloud physics is between “warm cloud” and “cold cloud” microphysical processes. In warm clouds, with $T > 0^\circ C$, collision–coalescence and raindrop scavenging of cloud droplets dominate the hydrometeor formation. When $T < 0^\circ C$, the ice-phase process along with graupel and snow may take over (Colle et al. 2005). Below the melting level, the snow or graupel melt to form raindrops (section 2). This melting is almost always visible as a radar “brightband” signature due to partial snowflake melting and clumping.

A fascinating special category is the “non-bright-band” precipitation described by Kingsmill et al. (2016) and others in Southern California. In this situation, collision–coalescence dominates above the freezing level (i.e., $T < 0^\circ C$) and the mixed phase processes dominate. In the lower panel, the lifting occurs below the freezing level (i.e., $T > 0^\circ C$) and warm rain processes may dominate.

A nice illustration is the schematic in Fig. 20-36 (Zagrodnik et al. 2018). In the upper panel, with cold airflow, the strong orographic lifting occurs above the freezing level (i.e., $T < 0^\circ C$) and the mixed phase processes dominate. In the lower panel, the lifting occurs below the freezing level (i.e., $T > 0^\circ C$) and warm rain processes may dominate.

One important difference between warm and cold clouds is the fall speed of the hydrometeors (Hobbs et al. 1973). Due to their compact shape, raindrops fall with a terminal speed of about 6 m s$^{-1}$ while intricate snowflakes are much slower (say 2 m s$^{-1}$). Given the importance of downstream hydrometeor drift in OP, this difference in fall speed is significant. With a wind speed of 10 m s$^{-1}$ and a fall distance of 2000 meters, snow will drift 10 km while rain will only drift 3.3 km.

Another control parameter is the number, size, and composition of aerosols in the incoming flow. Rosenfeld and Givati (2006) proposed that a shift in the aerosol size distribution could explain a reduction in the precipitation enhancement factor over the western United States. Aerosols act as CCN upon which vapor can condense. Greater aerosol concentration would increase competition between growing cloud droplets, giving more but smaller droplets. This change would suppress hydrometeor growth. These mechanisms were investigated by Muhlbauer and Lohmann (2008) and Pousse-Nottelmann et al. (2015) using a numerical model. For a warm rain scenario, they found that increasing aerosol load delayed rain drop formation by collision–coalescence.

In an extensive set of simulations, Moore et al. (2016) varied air temperature, mountain height, and aerosol concentration in mixed-phase orographic precipitation. In addition to precipitation amount, they computed the isotope fractionation (section 5d). Similar to Muhlbauer and Lohmann (2008), they found that higher aerosol loading gave less precipitation and isotope fractionation.

Smith et al. (2009b), Minder et al. (2013), Watson et al. (2015), and Nugent et al. (2016) studied warm rain precipitation over the tropical island of Dominica in the eastern Caribbean. With weak trade winds, thermal convection over the island carried surface-derived aerosols into the clouds, giving more but smaller cloud droplets (Fig. 20-37). While the aerosol impact on the droplet size was striking, the impact on precipitation was far less significant.

3) CONVECTIVE VERSUS STRATIFORM CLOUDS

The distinction between OP from convective versus stratiform clouds is a fundamental one (Houze 2012). In cold climates with little latent heat release and/or with a
moist statically stable lapse rate, the uplifted air will remain laminar and the clouds will have a stratiform nature, almost like that of a lenticular cloud in the cold stable middle troposphere. In warmer or less stable climates, orographic lifting will trigger shallow or deep convection. The vertical motion field will consist of turbulent cloudy ascent and dry descent in addition to the orographic lifting.

There are many ways that convection will alter the physics of precipitation. First, it concentrates the vertical motion and may increase the liquid water concentration and the rate at which air parcels rise through the LCL. The faster rise rate can increase the degree of supersaturation and thus increase the number of activated CCN. Second, at the cloud edge, dry air may be entrained into the cloud, causing droplet evaporation and loss of buoyancy. Convection will localize precipitation and make it intermittent as cells drift over the terrain. Convection may have the added effect of mixing the atmosphere vertically. With all this complexity, it is not easy to say whether convection will enhance or suppress the total amount of OP.

Convection in OP may be deep or shallow (Stein 2004). So-called “embedded convection” is shallow convection within a layered cloud. It contributes a variance to the vertical motion and cloud water density that is missing in a purely stratiform cloud. At the other extreme is deep convection in which rising air exists in tall discrete towers reaching to the tropopause. Vertical wind speeds may reach several meters per second. An early discussion of deep orographic convection is Sakakibara (1979) and later by Lin et al. (2001).

Some of the deepest moist convection events on planet Earth are found adjacent to mountains (Zipser et al. 2006). Perhaps the most extreme example is on the east side of the Andes in South America (Romatschke and Houze 2010; Rasmussen and Houze 2011; Rasmussen et al. 2016). The cause of this deep convection involves a moist low-level jet slanting up the Andean slopes. Similar dynamics may be responsible for deep convection along the steep southern slopes on the Himalayas (Houze et al. 2007).

The first theoretical treatments of dynamically triggered convection were the contemporaneous papers by Fuhrer and Schär (2005) and Kirshbaum and Durran (2004, 2005). Fuhrer and Schär focused on shallow cellular convection while Kirshbaum and Durran focused on shallow banded convection. Chen and Lin (2005) included 3D hills in their analysis of orographic convection.

Four mechanisms of convection triggering by terrain will be mentioned here. First is a homogenous airstream that is conditionally unstable and with high relative humidity. When this air is lifted to its LCL by terrain, the local uplifted region of cloudy air will begin to accelerate upward due to its buoyancy. To the left and right, mass conserving downdrafts will suppress adjacent cloud formation. After a few minutes, this latent heat-driven circulation will drift away downwind and the up slope orographic ascent will recover and the process will begin again. The result is a periodic generation of convective clouds. An early attempt to model this process was Smolarkiewicz et al. (1988).

Second, consider an incoming airflow with existing scattered cumulus or layered stratocumulus clouds. As this heterogeneous air is smoothly lifted by the terrain, the cloudy and dry air segments will cool at different rates. The cloudy air will quickly become buoyant relative to the dry air and the convection will invigorate. This mechanism can be quantified using the “slice method” (Kirshbaum and Smith 2008, 2009; Kirshbaum and Grant 2012).

A third method of convection initiation occurs when heterogeneous noncloudy air rises over terrain. In the incoming airstream, we suppose that there are humidity fluctuations even if these variations are “buoyancy adjusted” so that the virtual temperature is uniform. As the air is lifted, the moister air parcels will reach their LCL first and will rise along a moist adiabat. With sufficient lifting, the drier air parcels will eventually reach their LCL too, but by then the moist “seeds” will be warmer and more buoyant (Fig. 20-38). In a conditionally stable atmosphere the convection will accelerate. This mechanism was proposed by Woodcock (1960) and tested by Nugent and Smith (2014).
The fourth source of orographic convection is thermally driven by solar heating of high or sloping terrain (see section 4f). This mechanism is widely seen and well documented by time-lapse cameras, geostationary thermal infrared (TIR) imagery, and scanning radars. It is most easily appreciated by viewing TIR movie loops from a geostationary satellite (Banta and Schaaf 1987; Schaaf et al. 1988; Damiani et al. 2008). In the middle of each day, convective clouds appear over high terrain (e.g., Rockies, Andes, Alps) and large islands (e.g., Puerto Rico) and drift slowly away. A good example of thermally driven convection is found on the Santa Catalina Mountains in Arizona (Demko and Geerts 2010). Such convection is easily classified as thermally driven because of its clear diurnal cycle.

In a conditionally stable atmosphere, mechanical and thermal convective generation mechanisms may compete with each other. The nature of this competition over the small tropical island of Dominica was studied by Nugent et al. (2014). With trade winds in excess of 5 m s⁻¹, the island was well “ventilated” and the diurnal temperature cycle of the island surface was limited to a few degrees Celsius. The ambient airstream quickly carried off the excess surface heat. As the air lifted over the island’s mountains, the “dynamical” lifting triggered convection by one of the mechanisms mentioned above. With a trade wind less than 5 m s⁻¹, there was less “ventilation.” The diurnal cycle of surface temperature on the island was increased to 5°C or more and diurnal thermal convection ensued.

A subtle but important difference between the two types of convection, dynamical and thermal, is the property of air entering the cloud. In the former case, this “cloud-base” air has never touched Earth’s surface. It was only forced to rise by the mountain-induced pressure field. In the latter case, the air entering the cloud has “touched” Earth’s surface in the sense that it has gained its excess heat from the surface. Nugent et al. (2014) showed that this surface-warmed air had gained aerosol and lost carbon dioxide from contact with the forest canopy. In both types of convection, dry air entrainment above the trade wind inversion prevented deep cloud development.

For a more comprehensive discussion of orographic convection see Houze (2012).

4) WIDE OR NARROW MOUNTAINS

While it is obvious that mountain height will be a dominant factor in OP (e.g., 20-69), mountain width may be equally important. The reason for this is that air crossing a mountain range has a finite “advection time” (AT = a/U) to reach the lee side. As an example, a mountain with width a = 30 km with a wind speed of U = 15 m s⁻¹ would have an advection time of AT = 2000 s = 0.55 h. Air parcels or convective cells that have not generated precipitation in that interval will start to descend with all chances of precipitation lost. Thus, OP is fundamentally a “race against time,” as the microphysical processes of precipitation have a limited time in which to act.

A good example of the advective time limitation is shown by Eidhammer et al. (2018). They looked at cool-season precipitation in subranges of the Rocky Mountains. In Fig. 20-39 they show the influence of wind speed on the DR for several mountains within the Colorado Rockies, ordered by decreasing width. For the wide ranges, the DR is nearly independent of wind speed, suggesting that the advection time is not a limiting factor on DR. For narrow ridges, however, the DR decreases quickly with increasing wind speed, suggesting that there is not sufficient advection time to produce precipitation. The apparent delay time in precipitation generation could be caused by convective development, hydrometeor generation, or gravitational fall out.

Colle and Zeng (2004a,b) used a numerical simulation to clarify how microphysical pathways and mountain width interact. They varied the ridge half-width from a = 50 to 10 km and the freezing level from 1000 to 500 hPa. For each situation, they computed the windward precipitation efficiency (PE) as ratio of precipitation on the windward slopes to the condensation rate. As
FIG. 20-39. DR for winter precipitation across several mountain ranges in the central Rocky Mountains as a function of wind speed. Diagrams are arranged in order of decreasing mountain width. The DRs for narrower mountain are more sensitive to wind speed. [From Eidhammer et al. (2018).]
expected, the PE decreased for the narrower hills (e.g., 84%–50%) as leeside descent began before the hydrometer formation and fallout were completed. When they lowered the freezing level from 500 to 1000 hPa, the PE also decreased (e.g., 88%–76%). Apparently, the warm rain mechanisms were more efficient than the cold cloud mechanisms. See also Kingsmill et al. (2016).

The roles of mountain width and advection time are qualitatively captured in the LTOP model as it includes a parameterized time delay for hydrometeor formation.

5) **UPSTREAM BLOCKING AND AIRFLOW DEFLECTION**

If orographic precipitation is caused by upslope flow, then it follows that airflow blocking or deflection will reduce precipitation. The reason for airflow deflection of moist airstreams is the same as for dry airstreams: the hydrostatic development of high pressure over the upslope region. Jiang and Smith (2003) showed that the same flow-splitting criterion based on nondimensional mountain height

\[ h = \frac{N_M h}{U} > 1, \]  

which is (20-2) with dry stability $N$ replaced by moist stability $N_M$, can be used to predict flow deflection in simple terrain geometries. As $N_M < N$, moist air has an easier time ascending over high hills, but it still may be blocked or deflected (Durran and Klemp 1982, 1983; Jiang 2003; Hughes et al. 2009). In realistic situations, detailed observations and full-physics mesoscale models can help understand and predict blocking and deflection. Three examples will be given here.

As frontal cyclones and atmospheric rivers over the Pacific Ocean hit the coast of California, the moist airstreams may easily pass over the lower coastal range but are strongly deflected by the Sierra Nevada range forming a type of barrier jet (Fig. 20-40). This deflection shifts some of the heaviest precipitation northward (Hughes et al. 2009; Lundquist et al. 2010; Kingsmill et al. 2013). When frontal cyclones pass through the Mediterranean Sea, strong moist southerly flows push against the south facing slopes of the Alps. These flows are deflected westward, forming a barrier jet through the Po Valley and focusing the heaviest precipitation in the Gulf of Genoa region (Rotunno and Houze 2007). As cyclones move farther east from the Mediterranean Sea, a barrier jet brings moisture from the Arabian Sea into the Zagros Mountain of Turkey and Iran. The upslope rains supply water for the Tigris and Euphrates Rivers and allow rainfed agriculture in the Fertile Crescent (Evans and Alsamawi 2011).

6) **COASTAL VERSUS CONTINENTAL OP**

While the importance of the five factors discussed above (ambient disturbance, cloud microphysics, convection, width, and blocking) is well accepted, other factors may also be important. One potential factor is the difference between coastal and interior continental OP. For example, a landfalling Pacific cyclone may be modulated by the California coastal ranges differently than by the interior Rocky Mountains a thousand kilometers inland. Before reaching the Rockies, the cyclone will have lost much of its water vapor; lost its distinct fronts and sectors; and altered its temperature, convective available potential energy (CAPE), boundary layer, and aerosol characteristics. In Europe, for example, one can imagine that OP in coastal sites such Wales and Norway would differ from interior sites such as the Carpathians or Urals, although these differences have not been systematically studied. In monsoonal India, precipitation in the coastal Western Ghats may differ from the interior sites along the Himalayas. In general, the classification of OP at different sites and seasons around the world has not been attempted.

In summary, different cases of orographic precipitation should be classified using the factors discussed above: ambient disturbance, microphysical process, convective versus stratiform, wide or narrow mountain, upstream blocking, continentality, etc. In this way, the wide variety of OP in Earth’s atmosphere can be appreciated and understood.

**d. Rain shadow, spillover, drying ratio, and isotope fractionation**

As an airstream passes over a mountain range, its net water vapor flux will be reduced by orographic precipitation (20-69). In addition, its water vapor profile...
may be altered by the layer-wise condensation, hydrometeor evaporation, convective detrainment of cloud droplets, or general turbulent mixing. The temperature profile may be altered in a similar way. The whole process should be considered to be an “orographic airmass transformation,” not just airmass drying (Smith et al. 2003).

The concept of the “rain shadow” is central to the study of orographic precipitation. In its simplest form, it invokes the reduced water vapor concentration and the warming by descent on the lee slopes to explain reduced leeside cloudiness and precipitation. The former of these processes could have a long-lasting effect, reducing humidity and precipitation for hundreds of kilometers downwind. In addition, the alteration of the vertical profile of temperature and humidity will have a long-lasting effect, perhaps by increasing the convection inhibition (CI) or reducing the CAPE. Thus, the concept of rain shadow should be partitioned into short-range and long-range effects.

As air parcels reach their maximum height over a mountain range and begin to descend, the small cloud droplets will be the first to evaporate. The evaporation of hydrometeors (i.e., snowflakes, graupel, and raindrops) will be significantly delayed due to their larger size. Strong ridge-top winds may carry this precipitation a few kilometers or even tens of kilometers downwind of the ridge crest. This precipitation on the lee slope is called “spillover” (Sinclair et al. 1997; Chater and Sturman 1998; Smith et al. 2012; Siler et al. 2013; Siler and Durran 2016). Spillover is essential in providing water to leeside glaciers, watersheds, farmlands, and reservoirs.

The total fraction of the water vapor flux removed as air flows over a mountain is called the drying ratio, defined by (20-79). In principle, DR can be computed using balloon soundings or GPS-integrated water vapor measurements to determine upwind and downwind WVF. If sufficient rain gauge, snow gauge, or radar stations are available, the integrated precipitation in (20-79) may be estimated.

It has been postulated that the DR can be determined using the stable isotope fractionation of water vapor crossing a mountain range. As air passes over a mountain and loses water vapor, the heavier isotope (e.g., with deuterium versus hydrogen or $^{18}$O versus $^{16}$O) condenses and precipitates more readily than the lighter isotope (Dansgaard 1964; Merlivat and Jouzel 1979; Gat 1996). As a result, each successive increment of precipitation is isotopically lighter than the previous one, reflecting the amount of water vapor already removed. Using the famous Rayleigh fractionation equation gives an expression for the DR (20-79):

\[
DR(x) = 1 - \left[ \frac{R(x)}{R_0} \right]^{\frac{1}{\alpha-1}}.
\]

In (20-75), the $R(x)$ values are the ratio of heavy to light isotopes in the precipitation while $R_0$ is its value in the first precipitation on the upwind side. The factor $\alpha$ is the fractionation factor that describes the difference in saturation vapor pressure of the two isotopes. It generally increases as temperature decreases, so some estimate of cloud temperature is required to use (20-75). By choosing $R(x)$ at the farthest downwind point with measurable precipitation (20-75) gives the DR for the entire range.

This isotope method was used by Smith et al. (2005) for the Cascade Range in Oregon using streamwater samples. The method has been applied more recently to the Andes (Smith and Evans 2007) and the Southern Alps (Kerr et al. 2015). Kerr’s results for New Zealand are shown in Fig. 20-41. The DR increases to nearly 40% at the top of the ridge. These results for DR must be taken with caution as the validity of (20-75) requires strong assumptions. Isotope-enabled mesoscale models have been used to test the accuracy of (20-75) (Moore et al. 2016).

6. Mountain waves

The idea that mountains will disturb the airflow patterns near the surface of Earth is quite intuitive (section 4), but the insight that this distortion could be a vertically propagating gravity wave, carrying momentum and energy to remote regions of the atmosphere, was a more subtle and advanced discovery (Lyra 1943; Queney 1948; Queney et al. 1960; Scorer 1949; Eliassen and Palm 1960). This theoretical idea was tested in some of the first airborne mountain wave observations (see Grubišić and Lewis 2004). The role that mountain wave transport of momentum might play in the general circulation of the
atmosphere and climate was suggested by Sawyer (1959), Blumen (1965), Bretherton (1969), Lilly (1972), Holton (1982), and others. Mountain wave parameterizations were introduced into global models by Palmer et al. (1986) and McFarlane (1987), with positive results. Recognizing their importance, there have been many theoretical, observational, and numerical studies of mountain wave momentum fluxes, including Lilly and Kennedy (1973), Kim and Mahrt (1992), Alexander et al. (2009), Geller et al. (2013), Vosper et al. (2016), Sandu et al. (2016), and Smith and Kruse (2018).

There are several other reasons for wishing to understand and predict mountain waves. The launching of waves by terrain has an influence on the surface wind field (section 4). For example, wave launching modifies upstream airflow blocking, gap and barrier jet, and severe downslope winds. Visually, mountain waves generate beautiful displays of lenticular clouds, lee wave clouds, rotor clouds, and broader arch clouds. They also produce an upstream shift in orographic clouds that modifies precipitation patterns. In the troposphere and low stratosphere, smooth wave motion and sudden wave-induced clear air turbulence (CAT) can disturb commercial air transport. As waves reach the stratosphere and mesosphere, they amplify inversely with the square root of air density, and thus can become almost catastrophic in amplitude there. When these waves break at high altitude, they heat and mix the air (Fritts and Alexander 2003).

The literature on mountain waves is large. Table 20-8 includes a list of recommended books and review articles that trace the history of the study of gravity waves and mountain waves. In this section, we briefly review the physical principles needed to understand mountain waves.

a. Dispersion relation for internal gravity waves

Internal gravity waves (IGW) are oscillations in a stably stratified fluid with gravity (or buoyancy) as the restoring force (see references in Table 20-8). When a fluid parcel is displaced vertically, buoyancy forces pull it back to its neutral position. As it reaches its neutral position, inertia makes it overshoot and it must be restored again. This oscillation does not remain in a fixed region but propagates through the fluid from one region to another. As we will see below, propagation is an inherent property of IGW. Mountain waves are special type of IGW.

To begin, we again define again the buoyancy frequency \( N \) in terms of the vertical gradient of potential temperature \( \theta \), (20-3):

\[
N^2 = \frac{g}{C_p} \frac{d\theta}{dz} = \left( \frac{g}{C_p} \right) \left( \frac{dT}{dz} \right)
\]  

(20-76)

for a compressible atmosphere. This parameter is a measure of the buoyancy-restoring force when an air parcel is lifted or depressed. Typical values for \( N \) from (20-76) are \( N \approx 0.01 \text{ s}^{-1} \) in Earth’s troposphere or \( N \approx 0.02 \text{ s}^{-1} \) in the stratosphere. Values of \( N \) can be reexpressed as buoyancy periods using \( T_N = \frac{2\pi}{N} \) with \( T_N = 628 \text{ s} \approx 10 \text{ min} \) and \( T_N = 314 \text{ s} \approx 5 \text{ min} \) for the

Table 20-8. Books and review articles on internal gravity waves and mountain waves.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eckart (1960)</td>
<td><em>Hydrodynamics of Oceans and Atmospheres</em></td>
</tr>
<tr>
<td>Queney et al. (1960)</td>
<td>The airflow over mountains</td>
</tr>
<tr>
<td>Turner (1973)</td>
<td><em>Buoyancy Effects in Fluids</em></td>
</tr>
<tr>
<td>Gossard and Hooke (1975)</td>
<td><em>Waves in the Atmosphere: Atmospheric Infrasound and Gravity Waves, Their Generation and Propagation</em></td>
</tr>
<tr>
<td>Scorer (1978)</td>
<td><em>Environmental Aerodynamics</em></td>
</tr>
<tr>
<td>Gill (1982)</td>
<td><em>Atmosphere-Ocean Dynamics</em></td>
</tr>
<tr>
<td>Smith (1989a)</td>
<td>Hydrostatic airflow over mountains</td>
</tr>
<tr>
<td>Durrant (1990)</td>
<td>Mountain waves and downslope winds</td>
</tr>
<tr>
<td>Hamilton (1997)</td>
<td><em>Gravity Wave Processes: Their Parameterization in Global Climate Models</em></td>
</tr>
<tr>
<td>Lighthill (2001)</td>
<td><em>Waves in Fluids</em></td>
</tr>
<tr>
<td>Smith (2002)</td>
<td>Stratified flow over topography</td>
</tr>
<tr>
<td>Nappo (2002)</td>
<td><em>An Introduction to Atmospheric Gravity Waves</em></td>
</tr>
<tr>
<td>Fritts and Alexander (2003)</td>
<td>Gravity wave dynamics and effects in the middle atmosphere</td>
</tr>
<tr>
<td>Bühler (2014)</td>
<td><em>Waves and Mean Flows</em></td>
</tr>
<tr>
<td>Sutherland (2010)</td>
<td><em>Internal Gravity Waves</em></td>
</tr>
<tr>
<td>Jackson et al. (2013)</td>
<td>Dynamically-driven winds</td>
</tr>
<tr>
<td>Teixeira (2014)</td>
<td>The physics of orographic gravity wave drag</td>
</tr>
<tr>
<td>Durrant (2015a)</td>
<td>Mountain meteorology: Downslope winds</td>
</tr>
<tr>
<td>Durrant (2015b)</td>
<td>Mountain meteorology: Lee waves and mountain waves</td>
</tr>
</tbody>
</table>
troposphere and stratosphere, respectively. Using this definition for $N^2$ and the linear Boussinesq equations of motion, one can derive the dispersion relation for IGW in two dimensions:

$$\omega^2 = \frac{N^2 k^2}{(k^2 + m^2)}, \quad (20-77)$$

where $\omega$ is the frequency of fluid oscillation and $k$ and $m$ are the horizontal and vertical wavenumbers. This important formula (20-77) describes the relationship between wave scale, wave phase line tilt, and wave frequency for plane wave disturbances of the form

$$f(x, z, t) = f_0 \exp[i(kx + mz + \omega t)], \quad (20-78)$$

where the function $f(x, z, t)$ describes any of the local physical properties involved in the wave motion such as vertical or horizontal velocity, perturbation pressure, temperature, or air density. In (20-77) and (20-78), the two components of the wavenumber vector $\mathbf{k} = (k, m)$ are related to the horizontal and vertical wavelengths by

$$k = 2\pi/\lambda_X, \quad m = 2\pi/\lambda_Z.$$

The first lesson from (20-77) is that $|\omega| \leq N$, so that only disturbances with periods $T = 2\pi/\omega$ or greater than 10 or 5 min can propagate as IGW in the troposphere and stratosphere, respectively. Disturbances with shorter periods ($T < T_N$) can exist in the vicinity of local forcing but they cannot propagate through the fluid as an IGW. The second lesson from (20-77) and (20-78) is that the slope of the phase lines ($m/k$) is related to the frequency ratio ($\omega/N$). For low-frequency waves, as $\omega/N \to 0$, the phase lines approach horizontal.

The propagation of a "packet" of waves is described by the group velocity

$$C_g = \frac{m}{k} = \frac{\pm Nm}{(k^2 + m^2)^{3/2}}, \quad (20-79a)$$

$$C_g = \frac{\partial \omega}{\partial m} = \frac{\pm Nkm}{(k^2 + m^2)^{3/2}}, \quad (20-79b)$$

The mechanism of wave propagation is "pressure–velocity work," where oscillatory fluid motion covaries with a pressure oscillation causing work to be done on one region of fluid by an adjacent region. The corresponding energy flux has two components:

$$\text{EF}_x = \langle u'p' \rangle, \quad (20-80a)$$

$$\text{EF}_z = \langle w'p' \rangle. \quad (20-80b)$$

The reader will be familiar with (20-80) as it is the same pressure–velocity work mechanism that sound waves use to propagate from speaker to listener.

A clearest illustration of IGW propagation comes from the famous laboratory experiment where a vertically oscillating object is placed in the middle of a quiescent stratified fluid (e.g., Turner 1973; Nappo 2002). The resulting disturbance in the fluid (Fig. 20-42) includes four slanting beams of wave energy in a pattern reminiscent of a St. Andrew’s cross (i.e., a national symbol of Scotland). The angle of the ray paths, given by the ratio of the two group velocity components (20-79), is controlled by the relative frequency of the oscillator ($\omega/N$). For the upper-left beam, the $\text{EF}_x < 0$ while $\text{EF}_z > 0$.

**b. Steady mountain waves**

To apply IGW theory to a steady-state mountain wave situation, one imagines a constant mean flow $U$ passing over a sinusoidal lower boundary

$$h(x) = h_m \cos(kx) \quad (20-81)$$

with wavelength $\lambda = 2\pi/k$. Under the assumptions of steady flow, Boussinesq fluid, and linear dynamics, the terrain-induced flow disturbance satisfies the famous Scorer equation

$$\frac{d^2 \phi}{dz^2} + [S(z)^2 - k^2] \phi = 0, \quad (20-82)$$

where $\phi(k, z)$ is the Fourier transform of the vertical velocity field $w(x, z)$ and the Scorer parameter is

$$S^2(z) = \frac{N^2(z)}{U^2(z)} - \frac{U_{zz}}{U(z)}. \quad (20-83)$$
Scorer’s equation (20-82) is discussed at length in many articles and books (e.g., Table 20-8). To derive (20-82) the horizontal vorticity equation is derived from the linearized equations of motion and Fourier transformed in the \( x \) direction. The linearized lower boundary condition is

\[
w(x, z = 0) = U(0)\frac{dh}{dx} = -U(0)h_0 km \sin(kx). \tag{20-84}
\]

The simplest case is when the wind speed and stability are constant with height so \( S^2 = N^2/U^2 \) is constant. The solution to (20-82) is then of the form

\[
\hat{w}(z) = \text{Re}[A \exp(iz)]. \tag{20-85}
\]

The pattern of vertical velocity above the terrain depends on whether \( |k| \) is greater than or less than \( |S| \). In the former case, \( m = \pm i(k^2 - S^2)^{1/2} \) and

\[
w(z; k) = -Uh_0 km \exp(-|m|z) \sin(kx). \tag{20-86a}
\]

This disturbance decays with height and has no phase tilt. In the latter case \( m = \text{sgn}(Uk)[S^2 - k^2]^{1/2} \) and

\[
w(z; k) = -Uh_0 km \sin(kx - mz). \tag{20-86b}
\]

This disturbance does not decay with height and the phase lines tilt upward. This is the case where the stability is strong enough, the wind is weak enough, and the terrain wavelength is long enough to generate mountain waves. In other words, the intrinsic frequency \( \sigma_f = Uk \) for a moving air parcel is less than the buoyancy frequency \( N \). These two behaviors are illustrated in Fig. 20-43.

For an isolated ridge, the mountain waves form a compact beam of wave energy, making it easier to see the direction of wave propagation. With a mean flow \( (U) \) added, the \( x \) component of the group velocity \( C_{g_x} \) in Earth coordinates becomes

\[
C_{g_x} = U - \frac{\partial \omega}{\partial k} = U \mp \frac{Nm^2}{(k^2 + m^2)^{3/2}}.
\]

To understand the propagation path of mountain waves, we distinguish the \( C_{g_F} \) relative to the ambient fluid from the \( C_{g_E} \) in Earth-fixed coordinates. The latter includes the advection of wave energy by the mean flow. In Earth coordinates, the wave energy moves along a ray path that tilts downstream (Fig. 20-44), while the phase lines still tilt upstream.

In the real atmosphere, the Scorer parameter (20-83) varies with height. For a given wavenumber \( k = 2\pi/\lambda \), we evaluate (20-83) at each altitude to determine if the disturbance is propagating \((|S| > |k|)\) or evanescent \((|S| < |k|)\). To illustrate this procedure (see Fig. 20-45), we consider a typical structured atmosphere with surface winds of 15 m s\(^{-1}\) and positive wind shear in the troposphere with a jet maximum of \( U = 32 \text{ m s}^{-1} \) at \( z = 12 \text{ km} \). Above the jet, the wind speed decreases to \( U = 16 \text{ m s}^{-1} \) at \( z = 19 \text{ km} \) and then increases upward through the middle stratosphere. The stability frequency increases from \( N = 0.01 \text{ s}^{-1} \) in the troposphere to \( N = 0.02 \text{ s}^{-1} \) in the stratosphere.

In Fig. 20-45 we consider four wavelengths \( \lambda = 5, 10, 25, 200 \text{ km} \) (Table 20-9). For \( \lambda = 5 \text{ km} \), \( S(z) < k \) at all
heights and the disturbance will decay upward as in Fig. 20-43a. The two longest waves in Fig. 20-45, with \( \lambda = 25 \) and 200 km, satisfy \( S(z) > k \) at all altitudes and thus waves propagate vertically as in Fig. 20-43b.

An important intermediate case in Fig. 20-45 is \( \lambda = 10 \) km, for which the disturbance will propagate in some altitudes and decay in others. In this case, the two wind minima near \( z = 0 \) km and \( z = 19 \) km, with \( S > k \) locally, can act as “wave ducts” (e.g., Wang and Lin 1999). Waves can propagate horizontally in these ducts, trapped by the low \( S \) values above and below. The wave duct in the lower troposphere could support trapped lee waves generated by mountains (see Fig. 20-46). The occurrence of trapped lee waves in the troposphere is generally well predicted from the “Scorer criterion”: that the Scorer parameter decreases strongly with altitude (Scorer 1949). In this situation, mountain waves in the low troposphere reflect downward when they try to propagate into the jet stream in the upper troposphere.

Another important property of mountain waves is the ratio \( \left( \frac{w'}{u'} \right) \) of vertical to horizontal wind perturbations. For long wavelengths \( (S \gg k) \), this ratio is small. For wavelengths such that \( S \approx k \), this ratio is large. The air parcels move up and down, barely changing their horizontal speed (Table 20-9).

The subject of trapped lee waves is too broad to discuss here. The reader should consult Scorer (1949), Smith (1976), Gjevik and Marthinsen (1978), Nance and Durran (1997, 1998), and more recent studies such as Smith (2002), Durran et al. (2015), Hills et al. (2016), and Portele et al. (2018).

c. The hydrostatic limit

In mountain wave theory, the hydrostatic limit provides a valuable simplification and is widely used in wave theory and wave drag parameterization (e.g., Smith 1989a; Zängl 2003). In the special case of long waves, strong stratification, and/or weak winds (i.e., \( k \ll N/U \)), the vertical wavenumber simplifies to its hydrostatic form

\[
m \approx \left( \frac{N}{U} \right) \text{sgn}(Uk).
\]

The factor \( \text{sgn}(Uk) \) selects the upwind tilting phase lines corresponding to upward energy flux (\( EFz > 0 \)). It is easy to show that this case corresponds to the hydrostatic limit where vertical accelerations within the wave field are small. With parameters \( U = 10 \text{ m s}^{-1} \) and \( N = 0.01 \text{ s}^{-1} \), the vertical wavelength from (20-87) is \( \lambda Z = 2\pi U/N \approx 6.28 \) km. We will test this prediction in section 6e.

<table>
<thead>
<tr>
<th>Wavelength</th>
<th>Regime</th>
<th>( m )</th>
<th>Hydrostatic</th>
<th>Ratio ( w'/u' )</th>
<th>( C_g )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 km</td>
<td>Evanescent</td>
<td>Imaginary</td>
<td>No</td>
<td>Moderate</td>
<td>—</td>
</tr>
<tr>
<td>10 km</td>
<td>Variable</td>
<td>Both</td>
<td>No</td>
<td>Large</td>
<td>—</td>
</tr>
<tr>
<td>25 km</td>
<td>Propagating</td>
<td>Real</td>
<td>Nearly</td>
<td>Moderate</td>
<td>Fast</td>
</tr>
<tr>
<td>200 km</td>
<td>Propagating</td>
<td>Real</td>
<td>Yes</td>
<td>Small</td>
<td>Slow</td>
</tr>
</tbody>
</table>
The vertical fluxes of energy and horizontal momentum become

\[ EF_z = \langle p'w' \rangle = \left( \frac{1}{2} \right) \rho U^2 N k A^2, \] (20-88a)

\[ MF = \rho \langle u'w' \rangle \approx -\left( \frac{1}{2} \right) \rho U N k A^2. \] (20-88b)

Note that for a given amplitude of the vertical velocity perturbations \( A \), the shorter waves carry more flux.

Probably the most remarkable feature of the hydrostatic limit is that the horizontal component of the fluid-relative group velocity is exactly cancelled by the advection of wave energy by the mean wind so that in Earth coordinates \( C_{gX} = 0 \). This important result predicts that hydrostatic mountain waves will be found directly above their generating terrain. In Fig. 20-44, the energy flux vector \( \mathbf{C}_{gE} \) would be vertical. The vertical component of group velocity (20-79b) reduces to

\[ C_{gZ} = kU^2/N. \] (20-89)

This quantity is important in estimating how long it would take a field of mountain waves to develop vertically. For example, if \( U = 10 \text{ m s}^{-1} \) and \( N = 0.01 \text{ s}^{-1} \), horizontal wavelengths of \( \lambda = 2\pi/k \) of 20 or 200 km, give \( C_{gZ} = 3.14 \) and 0.31 m s\(^{-1}\), respectively. The elapsed time needed for these short and long mountain waves to reach tropopause (say \( z = 10 \text{ km} \)) would be then 0.9 and 9 h, respectively.

The most famous closed form solution for mountain waves is the hydrostatic Boussinesq solution derived by Queney (1948). He examined airflow over the Witch-of-Agnesi hill shape (20-29),

\[ h(x) = \frac{h_m a^2}{x^2 + a^2}, \] (20-90)

with constant \( N \) and \( U \) with height. In (20-90) \( h_m \) and \( a \) are the ridge height and half-width. For this special ridge shape, Queney was able to solve the governing equations in Fourier space including the inverse Fourier transform (section 3). He carefully selected the upward propagating wave modes with positive energy flux and thus satisfied physical causality. The displacement field is given by

\[ \eta(x, z) = \frac{h_m a [a \times \cos(mz) - x \times \sin(mz)]}{x^2 + a^2} \] (20-91)

as shown in Fig. 20-47. Other variables such as \( u'(x, z) \), \( w'(x, z) \), and \( p'(x, z) \) can be readily derived from (20-91). In Fig. 20-47, where the streamlines move apart, the wind speed is reduced and the pressure is high as expected from Bernoulli’s law. Where the streamlines are closely spaced, the wind is fast and the pressure is low. The pressure difference across the ridge gives a pressure drag. The phase lines of this solution tilt upstream with height as they did in Figs. 20-42–20-44. The stability of this steady mountain wave solution was analyzed by Laprise and Peltier (1989) and Lee et al. (2006).
The pressure drag on the hill is

$$\text{Drag} = \int_{-\infty}^{\infty} p \frac{dh}{dx} \, dx = \left( \frac{\pi}{4} \right) \rho N U_h^2 \quad (20-92)$$

with the same sign as the mean flow ($U$). The horizontally integrated vertical flux of horizontal momentum by the waves is equal to the pressure drag but with the opposite sign,

$$\text{MF} = \int_{-\infty}^{\infty} p u' w' \, dx = - \left( \frac{\pi}{4} \right) \rho N U_h^2 \quad (20-93a)$$

at every altitude. Thus, Newton’s first law of action and reaction is satisfied. The units for (20-92) and (20-93) are newtons per meter. The vertical flux of energy is also independent of height and given by

$$\text{EF}_z = \int_{-\infty}^{\infty} p w' \, dx = \left( \frac{\pi}{4} \right) \rho N U_h^2 \quad (20-93b)$$

with units of watts per meter. Note that the momentum and energy fluxes in (20-93) are related by

$$\text{EF}_z = - U \times \text{MF}. \quad (20-94)$$

Queney’s wonderfully simple solution (20-91) is often used to answer basic questions about the structure of mountain waves. A surprising result is the complexity of the pointwise momentum ($u' w'$) and energy fluxes ($p' w'$). While the field is composed only of Fourier modes with negative momentum flux (MF) and positive EF$_z$, the pointwise fluxes have both positive and negative regions. This result implies that pointwise energy fluxes are not representative of the wave field as a whole.

While the Queney solution for steady gravity waves (20-91) is useful for illustrating basic mechanics, it does not account for the important influence of variable atmospheric properties on wave propagation. In the next section, we consider the effects of vertical variation in air density, wind speed, and static stability.

d. Hydrostatic waves in a variable atmosphere

In a real atmosphere, the ambient air density $\rho(z)$, wind speed $U(z)$, and static stability $N(z)$ vary with altitude. These variations in the background state will influence mountain waves in many ways (e.g., Eliassen and Palm 1960; Bretherton 1966). In this section we assume that the wavelength is long enough to be hydrostatic at all levels and thereby to propagate continuously upward. Using the group velocity formula (20-89), the waves will propagate quickly up through layers of fast wind but slowly through layer with slow wind. Assuming that the wave stays in the hydrostatic regime, the elapsed time (ET) for the wave to reach a given altitude $z$ is

$$\text{ET}(z) = \int_0^z C_{GZ}(z')^{-1} \, dz' = k^{-1} \int_0^z \frac{N(z')}{U(z')^2} \, dz' \quad (20-95)$$

(e.g., Kruse and Smith 2018). An application of (20-95) is given in Fig. 20-48. Here we use the same atmospheric structure shown in Fig. 20-45 with two example wavelengths of 25 and 200 km. The shorter wave propagates faster and reaches the top of the domain ($z = 60$ km) within 1 h. The longer wave takes nearly 10 h to reach that altitude.

As the wave propagates upward, its amplitudes will also vary. We say “amplitudes” (i.e., plural) because we can monitor the amplitude of any of the oscillating variables: vertical velocity ($w$), horizontal velocity ($u$), displacement ($\eta$), and perturbation temperature ($T$). The simplest way to estimate the impact of the environmental variations on the properties of the wave is to use the theorem of Eliassen and Palm (1960) that the MF is independent of height in steady nondissipative flow. For convenience, we write the MF in terms of each amplitude:

$$\text{MF} = \rho \langle u' w' \rangle = - \frac{\rho N \langle w \rangle^2}{U k}, \quad (20-96a)$$

$$\text{MF} = \rho \langle u' w' \rangle = - \rho U N \langle \eta \rangle^2, \quad (20-96b)$$
MF = \rho \langle u'w' \rangle = -\frac{\rho U k \langle u \rangle^2}{N}, \quad (20-96c)

MF = \rho \langle u'w' \rangle = -\frac{\rho U N k}{(\frac{\partial T}{\partial z})^2} \langle T \rangle^2, \quad (20-96d)

where \( k = 2\pi/\lambda \) is the horizontal wavenumber. For the example, if the rms \( \langle \eta \rangle = 1 \) m in a layer with \( \rho = 1.2 \) kg m\(^{-3}\), \( U = 10 \) m s\(^{-1}\), and \( N = 0.01 \) s\(^{-1}\), then from (20-96b), MF = 38 and 3.8 mPa for wavelengths of 20 and 200 km, respectively.

The validity of (20-96) depends on the additional assumption that there are upward propagating waves only. The lack of downgoing waves requires that the atmospheric properties vary so slowly with altitude that no wave reflection occurs. This assumption is formalized in the WKB method described by Bretherton (1966), Lighthill (2001), Bühler (2014), and others.

With MF constant with height, these expressions (20-96) predict how a hydrostatic wave with fixed horizontal wavenumber \( k \) will react as it propagates upward through a variable atmosphere. We begin with the effect of density decreasing systematically with height as given approximately by isothermal atmosphere formula (20-1):

\[
\rho(z) = \rho_0 \exp\left(-\frac{z}{H_s}\right). \quad (20-97)
\]

where \( H_s = RT/g \approx 8400 \) m. According to (20-96) and (20-97), all the wave amplitudes \( \langle w \rangle, \langle \eta \rangle, \langle T \rangle, \) and \( \langle u \rangle \) will grow exponentially with height proportional to \( \exp(z/2H_s) \). A wave propagating from sea level to \( z = 50 \) km will amplify by a factor of about 20. Even small-amplitude waves in the troposphere will eventually become large as they propagate far into the upper atmosphere. In practice, \( N(z) \) and \( U(z) \) vary also. In layers with reduced wind speed \( U(z) \), \( \langle w \rangle \) will drop while \( \langle \eta \rangle, \langle T \rangle, \) and \( \langle u \rangle \) will grow (20-96).

To estimate where the wave might break down, we define the nonlinearity ratio (NLR) \( \langle u' \rangle \). If the fluctuations in horizontal wind in the mountain wave field \( \langle u' \rangle \) grow to be as large as the mean flow \( \langle U \rangle \), this NLR approaches unity. In such a wave field, the airflow could stagnate locally at a point where \( u' = -U \). To predict this occurrence, we use (20-96c) to obtain

\[
NLR = \frac{\langle u \rangle}{U} = \frac{N \times \left| \text{MF} \right|}{\rho U^3 k}. \quad (20-98)
\]

According to (20-98), atmospheric layers with low wind speed, large \( N \), and low air density may experience wave breaking. In the extreme case of a “critical level” where \( U(z) = 0 \), all waves must break no matter how small their momentum flux or how large their wavenumber (Booker and Bretherton 1967; Breeding 1971; Grubišić 1997).

If we take \( NLR = 1 \) as the wave breaking criterion, the maximum (negative) momentum flux MF in a given layer is

\[
\text{MF}_{\text{MAX}} \approx -\frac{\rho U^2 k}{N}. \quad (20-99)
\]

Noting wavenumber \( k \) in the numerator of (20-99), it seems that shorter waves can carry more MF without breaking than longer waves because their wind speed perturbation amplitude \( \langle u \rangle \) is less. According to the “saturation hypothesis” when \( \text{MF}_{\text{MAX}} \) decreases below MF, wave breaking will destroy the excess MF and \( \text{MF}(z) \) will follow (20-99) (Lindzen 1981; Palmer et al. 1986; McFarlane 1987). As air density decreases toward zero in the upper atmosphere, eventually all the wave momentum flux must be deposited into the atmosphere as \( \text{MF}_{\text{MAX}} \rightarrow 0 \).

If the terrain-generated momentum flux (MF) and wavenumber \( k \) for the wave are known, (20-98) allows us to predict where the mountain waves will break. For illustration, we use the environment shown in Fig. 20-45. Important features of our ideal atmosphere are the jet at 10 km and the wind minimum at 18 km. We imagine two hydrostatic vertically propagating mountain waves with wavelengths 25 and 200 km each carrying 0.2 Pa of momentum flux. As shown in Fig. 20-48, the shorter waves have a faster group velocity and reach \( z = 60 \) km within 1 h. The longer waves require nearly 8 h to reach that altitude. Both waves become more nonlinear as they propagate through the slow wind layer at \( z = 18 \) km. Kruse et al. (2016) call this layer a “valve layer” as the ambient wind speed there controls wave penetration. When wind speed is high the valve is “open.” When wind speed is low, the valve is “closed” because the wave breaks and may not penetrate farther upward.

For the short wave in Fig. 20-48, with its smaller \( \langle u \rangle \), the NLR remains less than NLR = 0.3 and the wave penetrates without breaking. Having passed through the “valve layer” it quickly propagates to the top of the stratosphere where decreasing air density increases NLR, eventually forcing it to break. The longer waves \( (L = 200 \text{ km}) \) have a larger NLR and probably break in the valve layer.

Combining the insights from Figs. 20-48b and 20-48c, we can predict a descending nature of wave breaking during a mountain wave event. The shorter waves start to break aloft at \( z = 60 \) km within an hour of wave launching by the terrain. The longer waves start to break at \( z = 18 \) km about 5 h after launch. Thus, the wave breaking appears to descend with time, even though the
waves are all propagating upward (Lott and Teitelbaum 1993; Smith and Kruse 2017; Kruse and Smith 2018).

A graphical example of the wave dissipation in a valve layer is shown in Fig. 20-49. It shows two wave-resolving numerical simulations of mountain waves over New Zealand during the DEEPWAVE project (Kruse et al. 2016). Shown is the EF$_z = 1$ W m$^{-2}$ isosurface of the EF$_z(x, y, z)$ field. During RF04, waves penetrate the top of the domain near $z = 40$ km. During RF09, waves become nonlinear and break in the “valve layer” with reduced wind speed at $z = 18$ km. [After Kruse and Smith (2015); courtesy of C. Kruse.]

The example in Fig. 20-46 shows a beautiful pattern of trapped lee wave clouds in the lower troposphere generated by a small island.

2) AIRCRAFT IN SITU OBSERVATIONS

Instrumented research aircraft are well suited for mountain wave detection (e.g., Lilly and Kennedy 1973; Lilly 1978; Lilly et al. 1974, 1982; Hoinka 1984, 1985; Bacmeister 1996; Bougeault et al. 1997; Smith et al. 2007, 2008, 2016; Smith and Broad 2003; Duck and Whiteway 2005). The aircraft gust probe consists of a forward boom or nose cone with differential pressure probes to measure the angle of flow relative to the aircraft. Laser beam backscatter can also be used (Cooper et al. 2014). The aircraft motion vector and orientation are known from an inertial platform with frequent updates from GPS location information. Air temperature is measured using thermistors, carefully corrected for dynamical heating. Static pressure is measured on the fuselage and corrected for local airflow distortion. With differential GPS altitude measurement accurate to 10 cm, static pressure can be reduced to a standard level (Parish et al. 2007; Smith et al. 2008, 2016). Together, these systems allow measurements of wave variables $u$, $v$, $w$, $T$, and $p$ at 25 Hz (e.g., $\Delta x \approx 10$ m) resolution along a level flight leg. These quantities can be used to compute momentum and energy fluxes (e.g., Lilly and Kennedy 1973) and other useful diagnostic measures.

A simple way to visualize the aircraft-observed mountain waves is to compute the vertical displacement by integrating the measured vertical velocity along the track; assumed parallel to the wind direction,

$$\eta(x) \approx \int_{u(x')} w(x') dx'.$$

(20-100)

An example is given in Fig. 20-50, showing nine aircraft legs over New Zealand from the DEEPWAVE project in 2014.

Unauthenticated | Downloaded 07/30/22 08:08 PM UTC
When an aircraft experiences variable winds, temperature, and pressures along a flight leg, it is important to verify that this disturbance is indeed a mountain wave. A valuable check on theory and observation is the relationship between energy flux and momentum flux. In steady, linear waves Eliassen and Palm (1960) predicted

\[ EF_z = \frac{-U \cdot MF}{C_1} \]

a more general form of (20-94). In Fig. 20-51 we plot the two sides of (20-101) for 94 cross-mountain aircraft legs in the DEEPWAVE project. In Fig. 20-51, all the EFz values are positive, all the MF values are negative, and the two sides of (20-101) generally agree. This agreement verifies that the observed disturbances were steady, linear, vertically propagating mountain waves.

3) BALLOON MEASUREMENTS

Gravity waves are easily detected by buoyant rising balloons (Dönrbrack et al. 1999; Geller et al. 2013) or by drifting constant-level (e.g., Plougonven et al. 2008; Jewtoukoff et al. 2015) balloons. Both types of balloon are tracked with onboard GPS and they detect waves, mostly through observed perturbation wind speed \( u' \). With more sophisticated signal processing, they may be able to detect perturbations to the vertical air motion \( w' \) (Vincent and Hertzog 2014). Air temperature fluctuations can be also detected from rising balloons, while poor sensor ventilation can lead to temperature errors in drifting constant level balloons.

4) GROUND-BASED AND SATELLITE-BASED PASSIVE TIR AND MICROWAVE

A significant development is the remote TIR detection of gravity waves by local fluctuation of temperature (Hoffmann et al. 2013). This technology allows global maps and regional statistics of gravity amplitude in the stratosphere especially (Fig. 20-52). This method has revolutionized our understanding of mountain wave climatology. There are limitations, however. First, by observing temperature only, they bias the observations toward longer waves with larger temperature perturbations [see (20-96d)]. Second, as the measurement “footprint” is large, it further biases the observations toward waves with large vertical and horizontal wavelengths.

A clever new technique is the Mesosphere Temperature Mapper (MTM; Pautet et al. 2014; Fritts et al. 2016). This system images infrared emissions from OH radicals in a thin layer at \( z = 80 \) km near the mesopause. Ground-based, airborne, or satellite-based MTM can resolve the crest and trough structure of waves with wavelength of 10 km.

5) RAYLEIGH LIDAR

Recent developments of pulsed lidar technology have allowed detection of gravity waves in the stratosphere and mesosphere. At these altitudes, lidar backscatter is dominated by molecular Rayleigh scattering proportional to air density. Knowledge of air density \( \rho(z) \) allows the hydrostatic equation to be integrated to obtain pressure and temperature perturbations (Kaufier et al. 2015; Ehard et al. 2016). The example in Fig. 20-53
from New Zealand shows stationary waves in the stratosphere with a vertical wavelength of about 6 km [see (20-87)]. In the mesosphere, above 50 km, the waves are still detectable, but they are irregular and unsteady. Lidar systems cannot be used to determine air density in the troposphere where scattering is dominated by aerosol. Instead they could be used to track the vertical displacement of aerosol layers and lenticular clouds (Smith et al. 2002).

A similar Rayleigh lidar system was carried on an aircraft during the recent DEEPWAVE campaign (Fritts et al. 2016). This airborne lidar can map out the horizontal structure of a stationary wave, instead of waiting for gravity waves to propagate or drift over a fixed land-based lidar station. In Fig. 20-54 we show two transects over New Zealand. The observed temperature anomalies exhibit a vertical periodicity and upstream phase line tilt expected in mountain waves (e.g., Figs. 20-42–44 and 20-47). Superposed on this diagram are temperature contours from the ECMWF global model. The agreement is surprisingly good. Smaller-scale waves may be present too, but their temperature amplitude is small (20-96d) and they would not be resolved in the global model.

f. Nonlinear wave generation, severe winds, and rotors

While the linear wave theory described above gives a consistent picture of the mountain wave generation, actual mountain waves are often nonlinear (section 4b).
In general, we use the nondimensional mountain height \( \tilde{h} = Nh / U \) (20-2) to judge the importance of nonlinear effects during wave generation. When \( \tilde{h} \) exceeds about 0.25 nonlinear effects begin to be important but when it exceeds unity, nonlinear effects will dominate. Severe downslope winds are an example of nonlinear wave generation (Durran 1992). Trapped lee waves can also be generated by a nonlinear mechanism (Smith 1976).

The first attempt to attack the finite amplitude problem was the development of “Long’s equation” (e.g., Long 1955; Huppert and Miles 1969). Several authors have used this elegant 2D formulation, but it is mostly limited to cases with uniform wind speed and stability. These special cases are less susceptible to nonlinearity, so the Long’s equation approach can underestimate the influence of nonlinearity (see Smith 1976; Durran 1992; Sachseperger et al. 2017). Another approach is that of Lott (2016), showing that a substantial part of the nonlinearity comes from the lower boundary condition alone. When numerical simulations using the full nonlinear governing equations in 2D and 3D became possible in the 1970s, studies of nonlinear effects advanced rapidly. The primary goal of these early studies was an understanding of severe downslope winds (see section 4b).

One remarkable nonlinear mountain wave phenomenon is the rotor. A “rotor” is a long leeside overturning eddy, parallel to the mountain ridge, rotating “away” from the ridge (Doyle and Durran 2002, 2004; Hertenstein and Kuettner 2005; Grubišić et al. 2008; Strauss et al. 2016). It usually resides 1 or 2 km above Earth’s surface. If strong enough, it can generate reverse winds at the ground. Most often, the rotor appears to sit under the first crest of a trapped lee wave. Its turbulence and dissipation seems to reduce the lee wave amplitude farther downwind. An example is shown in Fig. 20-55 from Doyle and Durran (2002). Some of the concentrated vorticity in the rotor may come from frictional boundary layer vorticity, lifted off the surface and concentrated in the rotor.

g. 3D mountain waves

Our understanding of mountain waves comes partly from the early 2D formulations, but many new interesting and important features arise when the hills and wave fields are fully three-dimensional. The simplest way to pose a mountain problem in 3D is to consider a simple isolated hill in a unidirectional flow (e.g., Wurtele 1957; Crapper 1962; Blumen and McGregor 1976). In this case, waves with many different wavenumber orientations will be generated. In the hydrostatic limit, the wave fronts perpendicular to the wind will propagate vertically, as in the 2D case. The oblique waves, however, will follow ray paths that are tilted laterally and downstream. The wave energy from a lone isolated peak spreads laterally along a horizontal parabola at each altitude \( z \), given by

\[ x = U y^2 / N z a, \]

where \( a \) is the radius of the axisymmetric terrain and \( x \) and \( y \) are the downwind and crosswind distance from the peak (Smith 1980). This lateral dispersion weakens the local wave amplitude with height and postpones wave breaking. The implications of this wave spreading are discussed by Eckermann et al. (2015).

When the Scorer condition is met (section 6b), an isolated hill may generate a beautiful pattern of trapped lee waves (Fig. 20-46). This pattern is sometimes called a “ship wave,” because it resembles the pattern of surface waves behind ships at sea. Two families of waves are possible: the diverging waves and the transverse waves (Gjevik and Marthinsen 1978; Smith 2002).

When the terrain is anisotropic with a long and short axis (see section 3), the nature of the wave disturbance depends on the wind direction. Smith (1989b) showed that airflow directly along a ridge is less disturbed. Low-level flow splitting is more likely and wave breaking is delayed as the mountain waves disperse more quickly aloft.

The case when the wind hits the terrain obliquely is more complex. In this case, Phillips (1984) showed that the wave drag vector is not oriented parallel to the wind vector. The drag component perpendicular to the wind vector is called the “transverse drag.” Several authors
have used Phillips’ solutions for elliptical hills to estimate wave drag on complex terrain (e.g., Scinocca and McFarlane 2000; Teixeira and Miranda 2006).

Garner (2005) and Smith and Kruse (2018) showed that in the hydrostatic limit, the wave drag vector $\mathbf{D}$ acting on any complex terrain can be written as the product of the wind vector $\mathbf{U}$ and a $2 \times 2$ symmetric positive-definite drag matrix $\mathbf{D}$ that is

$$\mathbf{D} = \mathbf{U} \cdot \mathbf{D}.$$  \hspace{1cm} (20-103)

The positive-definite property of $\mathbf{D}$ satisfies the causality condition $\varepsilon F > 0$. As with terrain matrices (20-22) and (20-24), the ratio of eigenvalues of $\mathbf{D}$ is a measure of terrain anisotropy.

A successful application of this scheme to wave drag on New Zealand is shown in Fig. 20-56. In this analysis, the “actual” wave drag is estimated from a high resolution numerical model using a filtering procedure (Kruse and Smith 2015). Due to the variation of $\mathbf{U}(t)$ in passing frontal cyclones (e.g., Fig. 20-31), the wave drag is highly episodic.

Another aspect of wave three-dimensionality is the effect of wind turning with height. For an isolated hill, the different oblique waves will react differently to the wind turning. Waves for which the wind turns along the phase lines will experience a decreasing intrinsic frequency

$$\sigma = \mathbf{U} \cdot \mathbf{k}.$$ \hspace{1cm} (20-104)

In the extreme case, these waves may cease to propagate, as they would at a critical level. (Broad 1995; Shutts 1995; Doyle and Jiang 2006).

Another interesting 3D mountain wave situation is when there is lateral wind shear. Examples were given by Blumen and McGregor (1976) and Jiang et al. (2013). If the zonal wind varies with latitude, waves with initial north–south phase lines will be refracted into the meridional direction and giving so-called “trailing waves.”

**h. Mountain wave breaking and PV generation**

In section 6d, we hypothesized that mountain waves would break down to turbulence when the nonlinearity ratio NLR $\approx 1$. This hypothesis arises from the idea that if the local horizontal flow can be decelerated to zero speed, then the streamlines can be locally vertical and fluid overturning can begin. Once colder air has been put atop warmer air, turbulent convection will begin. While there is a good deal of circumstantial evidence to support this idea, the reality is probably more complicated (Dörnbrack et al. 1995; Whiteway et al. 2003; Smith et al. 2016).

The second conventional explanation for wave breaking is that the vertical shear inherent in gravity wave motion will trigger Kelvin–Helmholtz (KH) instability.

$$\mathbf{D} = \mathbf{U} \cdot \mathbf{D}.$$  \hspace{1cm} (20-103)

The positive-definite property of $\mathbf{D}$ satisfies the causality condition $\varepsilon F > 0$. As with terrain matrices (20-22) and (20-24), the ratio of eigenvalues of $\mathbf{D}$ is a measure of terrain anisotropy.

A successful application of this scheme to wave drag on New Zealand is shown in Fig. 20-56. In this analysis, the “actual” wave drag is estimated from a high resolution numerical model using a filtering procedure (Kruse and Smith 2015). Due to the variation of $\mathbf{U}(t)$ in passing frontal cyclones (e.g., Fig. 20-31), the wave drag is highly episodic.

Another aspect of wave three-dimensionality is the effect of wind turning with height. For an isolated hill, the different oblique waves will react differently to the wind turning. Waves for which the wind turns along the phase lines will experience a decreasing intrinsic frequency

$$\sigma = \mathbf{U} \cdot \mathbf{k}.$$ \hspace{1cm} (20-104)

In the extreme case, these waves may cease to propagate, as they would at a critical level. (Broad 1995; Shutts 1995; Doyle and Jiang 2006).

Another interesting 3D mountain wave situation is when there is lateral wind shear. Examples were given by Blumen and McGregor (1976) and Jiang et al. (2013). If the zonal wind varies with latitude, waves with initial north–south phase lines will be refracted into the meridional direction and giving so-called “trailing waves.”

In a macroscopic point of view, the key question about wave breaking is how the wave is dissipated and how momentum flux is deposited into the fluid. A good indicator of turbulent wave dissipation is the Ertel potential vorticity

$$\mathbf{PV} = \frac{\mathbf{j}}{\rho} \cdot \nabla \vartheta,$$ \hspace{1cm} (20-106)

where $\mathbf{j} = \nabla \times \mathbf{u}$ is the fluid vorticity. PV satisfies the conservation law

$$\frac{D\mathbf{PV}}{Dt} = \frac{1}{\rho} \nabla \vartheta \cdot \nabla \times \mathbf{F} + \frac{1}{\rho} \mathbf{\xi} \cdot \nabla \vartheta,$$ \hspace{1cm} (20-107)

where $\mathbf{F}$ is a vector body force and $\vartheta$ is the rate of change of potential temperature from heating (Haynes and McIntyre 1987; Schär 1993; Schär and Durran 1997; Aebischer and Schär 1998). The body force could be associated with a subgrid-scale turbulent flux of
momentum. The heating could be radiative, latent heat, or subgrid turbulent transport of heat. In laminar nondissipating gravity waves, PV (20-106) is conserved because the baroclinic vorticity generation lies parallel to the isentropic surfaces ($\theta = \text{constant}$). In wave breaking, however, the small-scale turbulence [right side of (20-107)] creates PV pairs of equal magnitude and opposite sign that advect downstream from their source regions. These streaks of opposite PV are called "PV banners." This is shown in Fig. 20-57 for flow in the valve layer over New Zealand. After creation, these vortices may self advect each other into a complex wake structure (see section 4e).

These PV banners in the stratosphere nicely illustrate the "action at a distance" principle of mountain wave momentum transport. In aerodynamics, pairs of vortices would be shed from any object immersed in an airstream associated with the drag force. The vorticity would come from separated boundary layer vorticity. In the case of mountain waves, the drag force is transported upward by wave motion and then deposited in the fluid. The vorticity associated with mountain drag is thus generated far above the mountain.

i. Gravity wave drag and wave–mean flow interaction

The question of how the ambient flow responds to mountain wave drag is a broad and challenging one. We start by defining the gravity wave drag (GWD) as the divergence of the wave momentum flux, divided by air density:

$$\text{GWD} = \rho^{-1} \frac{dMF}{dz} \quad (20-108)$$

with units of acceleration. For example, of the MF decreased from $MF = -100 \text{ mPa}$ at $z = 18 \text{ km}$ to zero at $z = 19 \text{ km}$, where the air density is $\rho = 0.14 \text{ kg m}^{-3}$, the GWD $= 0.000714 \text{ m s}^{-2} = 2.6 \text{ m s}^{-1} \text{ h}^{-1}$. While GWD has units of acceleration, the ambient flow may not decelerate locally. An unbounded geostrophic flow, subject to an applied force, has several degrees of freedom in response to GWD (Durran 1995a; Chen et al. 2005, 2007; Bölni et al. 2016; Menchaca and Durran 2017, 2018).

GWD (20-108) may be caused by wave dissipation or by wave transience. In a brief nondissipative wave event, GWD will tend to slow the flow as the wave arrives and then release it as the wave departs. A more compact way to describe this process of "reversible deceleration" is to imagine that the reduced flow speed is tied to the wave packet itself: its so-called "pseudo momentum" (PM). In 2D flow with a uniform environment

$$\text{PM} = -\overline{\eta \xi}, \quad (20-109)$$

where $\eta(x, z, t)$ and $\xi(x, z, t) = (\partial w/\partial x) - (\partial u/\partial z)$ are the wave induced vertical displacement and the horizontal vorticity. As the waves enter a region of the atmosphere, the mean flow will be altered according to

$$\frac{\partial U}{\partial t} = -\frac{\partial \text{PM}}{\partial t} + F, \quad (20-110)$$

where $F$ is the influence of dissipative wave breaking. This idea has broad implications for other types of wave phenomena (Bretherton 1966; Andrews and McIntyre 1978 a,b; Bretherton and Garrett 1968; Scinocca and Shepherd 1992; Durran 1995b; Bühler 2014). Kruse and Smith (2018) used (20-109) and (20-110) to distinguish between dissipative and nondissipative GWD in numerical simulations. They found that dissipative wave breaking dominated for all but the weakest waves.

On broad scales, persistent dissipative GWD from mountain waves will break the geostrophic balance in Earth’s zonal jets and produce a persistent meridional overturning in Earth’s atmosphere. This meridional circulation has been studied by many authors (e.g.,
Haynes et al. 1991; Garcia and Boville 1994; Shaw and Boos 2012; Cohen and Boos 2017). A key property of the meridional circulation is the principle of “downward control” (Haynes et al. 1991) whereby most of the vertical motion caused by GWD occurs below the level of momentum deposition (Fig. 20-58). The meridional overturning below the GWD level may be responsible for warming the polar vortex, ozone transport across the tropopause, and even rainfall enhancement or suppression.

7. Global climate impacts

The previous sections have described the local influence of particular mountain ranges on surface winds, precipitation, and gravity waves. Furthermore, we have mostly considered transient phenomena lasting only a few hours, days, or weeks. In this chapter we expand the space and time scale of our review to include regional, global, and historical climate impacts.

a. The polar vortex

We consider here how the polar vortex is influenced by global distribution of mountains and other variations in surface properties. By “polar vortex” we mean the global-scale westerly winds circling the pole in mid- and high latitudes of both hemispheres. The term is used to describe both the tropospheric and stratospheric circumpolar jet streams. These meandering jets are important parts of the full general circulation of the atmosphere. As discussed by Waugh et al. (2017), these upper and lower air currents differ significantly in their location and characteristics, so we must be careful not to confuse them (Fig. 20-59). The stratospheric vortex is smaller in radius but stronger in wind strength. The tropospheric vortex is larger in radius, weaker in wind speed, and often more highly disturbed. Both wind systems satisfy the thermal wind relationships between vertical wind shear and meridional temperature gradient. Because of this relationship, the jets reach maximum strength in the winter months when there is strong differential heating between the pole and equator. Both vortices are disturbed with stationary and transient waves. The coupling between the two polar vortices is examined by Gerber and Polvani (2009) and Domeisen et al. (2013), among others.

b. Stationary waves and eddies

One of the most surprising and important impacts of Earth’s surface irregularities on the atmosphere is the pattern of stationary waves distorting the polar vortex, especially in the Northern Hemisphere and especially in the winter season (Charney and Eliassen 1949; Saltzman and Irsch 1972; Held et al. 2002). While we might expect the circumpolar westerlies to be disturbed by transient baroclinic wave and eddies, we find instead that a significant fraction of the disturbance is stationary. Stationary disturbances are usually defined as the residual after time averages of months or seasons have been computed. If there were no surface inhomogeneities, such as the fictional “aquaplanet” (Smith et al. 2006; Marshall et al. 2007), every airflow disturbance would be transient, randomly distributed, and/or migrating. Long-term climate averaging would eliminate these transient perturbations, leaving a simple zonally uniform atmospheric structure. Instead, monthly averaging reveals a
fixed winter pattern of meridional flows in mid- and high northern latitudes (Fig. 20-60). Southward flows are seen at 100°E and 120°W longitudes and compensating northward flows at 40° and 150°W. To a first approximation, high pressure “anticyclones” are located over the higher terrain.

Most investigators agree that these stationary atmospheric features must be caused by stationary surface forcing, probably airflow uplift or deflection by broad mountains (e.g., Rockies or Tibet/Mongolian highlands). The rough coincidence between high terrain, anticyclonic vorticity, and high pressure might suggest a simple vortex compression mechanism (e.g., Smith 1979, 1984; Gill 1982), but this is likely an oversimplification. Many investigators have suggested that stationary sources of heat may be equally important (e.g., Held 2002). Zonal variations in atmospheric heating could be caused by the mountain albedo differences or by latent heating from persistent orographic precipitation, but just as easily by the contrast between ocean and continent surface properties and sensible heat fluxes. Two well-known stationary hot spots of surface heat flux to the atmosphere are the warm Kuroshio and Gulf Stream in winter. In principle, these heat sources could generate stationary waves without any topography. Thus, we cannot confidently assign the cause of the stationary waves to the mountains themselves.

One important clue to the source of planetary waves is their wintertime preference. In winter, the high latitudes have larger static stability (20-3), a shifted and stronger polar jet, and strengthened local oceanic heat sources, any of which could be important in stationary wave generation.

A second clue is the stronger stationary waves in the Northern versus the Southern Hemisphere. It is generally recognized that the Northern Hemisphere has a

---

**Fig. 20-60.** Longitude–height distribution of mean meridional component of wind (m s$^{-1}$) along the 45°N latitude circle. Averages are computed over five Januaries. (top) Observed, (middle) M-model, and (bottom) NM-model. [From Manabe and Terpstra (1974).]
greater fraction of continental area and higher broader mountains. The Northern Hemisphere has the Rockies, Alps, and the Himalayas/Tibetan Plateau (Fig. 20-5). The Southern Hemisphere has the impressive but narrow Andes, and less continental area. In most other respects, the two hemispheres are identical, if one takes into account the 6-month time offset between the seasons. Only the small ellipticity of Earth’s orbit breaks this north–south symmetry with regard to the seasonal forcing by solar radiation.

c. Mountain impacts in the troposphere

The generation of planetary waves by large-scale terrain has received much attention (e.g., Charney and Eliassen 1949; Charney and Drazin 1961; Saltzman and Irsch 1972; Manabe and Terpstra 1974; Karoly and Hoskins 1983; Valdes and Hoskins 1991; Broccoli and Manabe 1992; Cook and Held 1992; Ringler and Cook 1997; Wang and Ting 1999; Held et al. 2002; Kaspi and Schneider 2013; Wills and Schneider 2018). One of the earliest global simulations of mountain influence on the midlatitude westerlies was Manabe and Terpstra (1974). Using an early-generation general circulation model (GCM), they carried one of the first mountain/no-mountain (i.e., M and NM) comparisons of Earth’s general circulation. Their GCM was rather primitive compared to today’s models, as the stratosphere was very poorly resolved and it lacked gravity wave drag parameterizations. Radiative heating and cooling, surface fluxes, and cloud properties were treated with simple algorithms. They compared their two numerical simulations with interpolated global radiosonde data (Fig. 20-60).

As expected, they found the biggest impact of mountains in the Northern Hemisphere. By averaging over several Januaries, they identified a stationary wavenumber 2 or 3 pattern in the winter troposphere. Stationary troughs were seen to the lee of the Rockies and the Tibetan Plateau (Fig. 20-60). This wave pattern nearly vanished when the mountains were removed from the numerical model. The residual wave was due to other surface inhomogeneities, for example, the difference in friction and heat storage between the continents and oceans. Importantly, the M run (Fig. 20-60b) agreed better with observations (Fig. 20-60a) than the NM run (Fig. 20-60c).

An idealized model of stationary wave generation was considered by Cook and Held (1992). They inserted a 0.7-km-high idealized mountain at 45°N latitude in a GCM (Fig. 20-61). As midlatitude westerlies passed over this terrain, a stationary Rossby wave was generated with anticyclonic flow to the west and cyclonic flow to the east of the high ground. The Rossby wave train propagates quickly equatorward and seems to be absorbed in the tropics. Valdes and Hoskins (1991) argued that the wave generation process might be associated either with orographic forced ascent or forced meridional deflection of the winds.

Another aspect of midlatitude weather impacted by terrain is the storm tracks (e.g., Trenberth 1991; Bengtsson et al. 2006; Brayshaw et al. 2009; Saulière et al. 2012; Kaspi and Schneider 2013; Chang et al. 2013; Lehmann et al. 2014). The level of storminess may also be affected. Chang et al. (2013) used a variety of GCMs to compute the strength of the midlatitude storm tracks as represented in the “one-day” variance of the meridional velocity (Fig. 20-62). Both hemispheres show a storm track at 50° of latitude, but the Southern Hemisphere track is stronger and more compact due to fewer mountains. One explanation is that without stationary waves, the transient “storm” disturbances must be amplified to carry the necessary heat. This postulate supports the sailor’s experience that the Southern Ocean is the stormiest place on Earth.

A key property of the stationary wave is its transport of heat from the pole to equator to balance the radiative imbalance. This meridional heat transport can be computed across any latitude (φ) circle using

\[ F(\phi) = R \cos(\phi) \int \rho v M d\lambda dz, \]  

(20-111)

where \( \rho \) is air density, \( v \) is the meridional velocity, \( R \) is Earth’s radius, and the moist static energy is \( M = C_p T + g z + L q_v \). In (20-111) we integrate in altitude and longitude, giving units of watts. Several authors have suggested splitting the heat flux into three components: the mean meridional overturning, the stationary waves/eddies,
and the transient waves/eddies (e.g., Nakamura and Oort 1988; Serreze et al. 2007; Fan et al. 2015). Due to the decrease in air density with height, most of this heat transport occurs in the troposphere. A surprising result of Nakamura and Oort and Serreze et al. is that in the Northern Hemisphere winter a significant fraction of the heat flux is carried by stationary waves; waves that are probably caused by Earth’s terrain. This important result is shown in Fig. 20-63 in a longitude–time diagram. In January, for example, there are strong inflows and outflows of heat from stationary waves. The stationary wave contribution is partly from shallow barrier jets near Greenland, Asia, and North America and partly due to deeper waves filling the depth of the troposphere. By modulating northward heat flux, terrain influences the Arctic climate. By contrast, the stationary wave contribution is nearly negligible in the Southern Hemisphere due to fewer mountains there (Nakamura and Oort 1988).

Another important impact of mountains on the troposphere is lee cyclogenesis (Smith 1984; Speranza et al. 1985; Buzzi and Speranza 1986; Egger 1988; Tafferner and Egger 1990; Schär 1990; Bannon 1992; Aebischer and Schär 1998; Schultz and Doswell 2000; McTaggart-Cowan et al. 2010). It has been noted that the formation of midlatitude frontal cyclones is not uniform around the world but concentrated to the lee (eastern or equatorward flanks) of high mountains. The Alps and the Rockies have been well studied in this regard. Once formed, lee cyclones produce severe weather locally and to the east where these cyclones may drift in the westerlies. Some theories of lee cyclogenesis simply rely on vortex stretching as fluid columns depart from high plateaus and return to lower elevation. Some rely on dissipative PV banners and eddies as discussed in section 4. Others are true baroclinic models in which baroclinic instability is triggered by terrain distortion of the temperature field.

d. Mountain impacts in the stratosphere

It might be supposed that the stratosphere would be less affected by mountains than the troposphere as it is higher and lies above all the mountaintops, but this is not the case. As described in section 6, mountain waves carry momentum into the stratosphere. Even more important is the vertical propagation of long planetary waves generated by the large-scale terrain (e.g., Charney and Drazin 1961). Both types of waves amplify as they enter stratosphere with its reduced air density.

As stratospheric datasets improved, stationary and transient waves could be better distinguished. An interesting approach was taken by Waugh and Randel (1999). They used spatially interpolated data from the period 1978 to 1998 to identify stationary wave patterns using “elliptical diagnostics” (ED); essentially fitting ellipses to the circumpolar geopotential and potential vorticity patterns. This method, they argued, gives a better representation of the polar vortex than the traditional spectral decomposition as the vortex center was often shifted away from the pole.
The results of their ED approach are shown in Fig. 20-64 for the winter months in both hemispheres. Figure 20-64 shows a best-fit ellipse to the stratospheric polar vortex at the 850-K potential temperature surface. Equivalent months are paired in this diagram. As winter advances, differential heating generates the polar vortex, reaching a maximum in July/January. In the Southern Hemisphere, the vortex is larger, nearly circular, and nearly centered on the pole. In the Northern Hemisphere, the vortex is smaller, elliptical, and offset from the pole. This difference between the two hemispheres is almost certainly due to the different terrain in the two hemispheres, but the exact mechanism is still unclear. It could be planetary wave generation from terrain or from fixed foci of precipitation and latent heat release. It could also reflect the global distribution of the GWD from the small-scale terrain.

Occasionally the impact of mountains on Northern Hemisphere stratospheric polar vortex is even greater. About six times per decade, the vortex distortion increases dramatically and the cold core is eliminated altogether. This is the sudden stratospheric warming (SSW) phenomenon discussed by Matsuno (1971), Charlton and Polvani (2007), Gerber and Polvani (2009), Palmeiro et al. (2015), and others. According to Charlton and Polvani (2007), the breakdown of the polar vortex can be of two types: a vortex displacement and vortex splitting (Fig. 20-65). The fact that SSW is common in the Northern Hemisphere but rare in the Southern Hemisphere illustrates the importance of the Northern Hemisphere terrain.

Another well-known difference between the Northern and Southern Hemisphere stratosphere is the occurrence of a smooth, oval ozone hole over the South Pole (Fig. 20-66). Due to the larger orographic disturbances in the Northern Hemisphere, a well-formed ozone hole is rare there.

The connection between terrain and disturbances on the polar vortex is shown more explicitly by Gerber and Polvani (2009). In an idealized analysis of polar vortex asymmetry, they introduced a smooth periodic
midlatitude terrain and examined its impact on the stratospheric polar vortex. They found that wave-number 2 terrain with an amplitude to 3000 m gave the best prediction of vortex structure and SSW frequency.

e. Mountain building and decay over geologic time

In this final section, we review how Earth’s mountains have changed over geologic time and the potential impact on climate (Ruddiman 1997). As shown in Table 20-10, ice sheets can appear and disappear in just a few thousand years by the accumulation of snow, melting, or gravitational sliding. Mountains come and go on a longer time scales associated with tectonic plate collisions and bedrock erosion. According to the theme of this review, these terrain changes will alter local and global climate, although admittedly they may not be the dominant influence on these long time scales. Orbital variations (e.g., the Milankovitch cycles) and changes in atmospheric composition (e.g., carbon dioxide concentration) are currently thought to have the dominant effect on climate change over millennia. Nevertheless, several recent studies have investigated the role of terrain variation on climate change.

1) THE LAURENTIDE ICE SHEET AND CLIMATE DURING THE LAST GLACIAL MAXIMUM

A well-studied example of changing mountain influence on climate is impact of the Laurentide Ice sheet on climate during the Last Glacial Maximum (LGM) about 20 000 years before present. Most investigators agree that the Pleistocene glacial cycles leading to the LGM were primarily caused by the Milankovitch orbital forcing and greenhouse gas feedbacks combined with a long term Cenozoic cooling. In addition, however, the development of large continental ice sheets causes a large feedback to regional climate change (Roe and Lindzen 2001; Kageyama and Valdes 2000; Justino et al. 2005; Liakka 2012; Lora et al. 2016; Löfverström and Lora 2017; Sherriff-Tadano et al. 2018; Roberts et al. 2019). A key question is the “self-sustaining” property of the ice sheet. That is, does the ice sheet increase the approaching water vapor transport and orographic
precipitation to maintain itself? If so, how does it begin and how does it end its life?

In Fig. 20-67 from Löfverström and Lora (2017), we see their assumed shrinkage of the Laurentide Ice Sheet from 20,000 to 12,000 years ago. Using a numerical climate model, they estimated the change in wind patterns caused by these terrain changes. To a good approximation, the dynamical impact of the ice sheet was similar to earlier results described in section 7c. A stationary Rossby wave was generated causing widespread changes, including upstream over the Pacific Ocean and downstream over the Atlantic Ocean. Some hint of flow splitting around the ice sheet is seen in Fig. 20-67.

2) PLATE TECTONICS, TIBETAN PLATEAU, AND ASIAN CLIMATE

The largest single terrain feature on Earth is the Tibetan Plateau, with an area of 2.5 million km² and height exceeding 3 km. Compared to the age of Earth (4.5 billion years) the plateau is relatively recent, probably about 50 million years old (Table 20-10). Its uplift was associated with the collision of the Indian and Eurasian plates. Numerous authors have investigated how this large uplift might have modified the climate of Asia (see Molnar et al. 1993, 2010; Liu and Dong 2013).

Liu and Yin (2002) emphasize the impact of Tibet on the wintertime circulation of East Asia (Fig. 20-68). From numerical simulations, they show that the fully elevated plateau generates a strong anticyclone over central China. This high pressure region brings aridity to central China and associated northerly winds over eastern China.

<table>
<thead>
<tr>
<th>Mountain range</th>
<th>Age (million years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laurentide Ice Sheet</td>
<td>2 to 0.015 (variable)</td>
</tr>
<tr>
<td>Greenland Ice Sheet</td>
<td>3 to present</td>
</tr>
<tr>
<td>Taiwan CMR</td>
<td>8 to present</td>
</tr>
<tr>
<td>Southern Alps (NZ)</td>
<td>10 to present</td>
</tr>
<tr>
<td>Antarctic Ice Sheet</td>
<td>23 to present</td>
</tr>
<tr>
<td>Alps (Europe)</td>
<td>50 to present</td>
</tr>
<tr>
<td>Tibetan Plateau</td>
<td>50 to present</td>
</tr>
<tr>
<td>Andes</td>
<td>80 to present</td>
</tr>
<tr>
<td>Rockies</td>
<td>80 to present</td>
</tr>
<tr>
<td>Sierra Nevada</td>
<td>100 to present</td>
</tr>
<tr>
<td>Appalachians</td>
<td>500 to present</td>
</tr>
</tbody>
</table>

FIG. 20-66. 1 Oct 2016 ozone hole in the Southern Hemisphere. Due to mountains in the Northern Hemisphere, such a stable polar vortex and associated ozone hole rarely occur over the North Pole. [Source: NASA Earth Observatory image by Jesse Allen, using Suomi NPP OMPS data provided courtesy of Colin Seftor (SSAI) and Aura OMI data provided courtesy of the Aura OMI science team.]

FIG. 20-67. Impact of the decaying Laurentide Ice Sheet on regional flow patterns. Estimates of the North American ice sheet topography (colors) from 20 to 12 ky BP are used to drive a global atmospheric model. Shown are model-computed low-level wind patterns (arrows) in winter (DJF) at the end of the Pleistocene period. [From Löfverström and Lora (2017).]
The impact of Tibet on the summer monsoon has received intense study (e.g., Liu and Yin 2002; Molnar et al. 2010; Boos and Kuang 2010, 2013; Park et al. 2012; Tang et al. 2013; Sha et al. 2015; Fallah et al. 2016; Tada et al. 2016; White et al. 2017; Shi et al. 2017; Naiman et al. 2017). Many physical processes might be at play as Tibet modifies regional climate, for example, uplift of the westerly winds, lateral deflection of the westerlies, vertical mixing, elevated heating, blocking of cold air from the north, and mountain wave drag. For many years, the leading hypothesis for Tibet’s impact on the monsoon has been the role of elevated heating on the plateau surface. According to this idea, the solar heating of the elevated plateau surface creates a large meridional temperature difference with the free atmosphere at the same elevation but farther south. This temperature gradient hydrostatically generates a pressure gradient that draws in moist air from the Indian Ocean.

This “elevated heating” hypothesis has recently been questioned by Boos and Kuang (2010, 2013). They begin their counterargument by noting that both the hottest equivalent potential temperature at Earth’s surface and the hottest air temperature near the tropopause is found over the northern plains of India, not over the Tibetan Plateau (Fig. 20-69). Thus, the center of the thermal circulation is low-elevation northern India, not over the elevated plateau.

To examine the sensitivity to surface temperature, they compared a control simulation with two runs with altered surface heat budgets. First, they reduced the sensible heat flux over the elevated plateau. Second, they reduced the sensible heat flux over northern India. The latter had the greater impact.

From this result and other evidence, they conclude that the most likely cause of a reduced Indian monsoon is the advection of cold air from the north by a mountain barrier. They argue that the terrain blocking of cold air from central China may have a larger role than the elevated plateau heating. If this conclusion is true, only when the Himalayas grew to block cold air from central China.
northern India rather than the Tibetan Plateau. [From Boos and Kuang (2013); reprinted by permission from Springer Nature: Scientific Reports, Sensitivity of the South Asian monsoon to elevated and non-elevated heating, William R. Boos and Zhiming Kuang, 2013. https://creativecommons.org/licenses/by-nc-nd/3.0/]

China from reaching northern India could the Southeast Asian monsoon establish its full strength.

8. Final remarks

In the early twentieth century, only the most basic qualitative concepts of mountain meteorology were described and understood, for example, ideas such as pressure and temperature decreasing with mountain elevation, windy mountain peaks and gaps, and rainy windward and dry leeward mountain slopes. Today, 100 years later, we have a much wider appreciation of mountain influences on the atmosphere and sophisticated theories for many such occurrences. Our ability to accurately numerically simulate orographic disturbances has advanced far faster than expected. This progress is due to an aggressive application of theory, observation, and modeling by many scientists. A few ideas have come and gone, but many older ideas are still valuable, contributing to our knowledge of the complex Earth–atmosphere system.

As with many areas of science, our field has become specialized. There are few of us today that work in all areas of mountain meteorology, and probably even fewer tomorrow. Nevertheless, the subject has a nice coherence that we can appreciate.

In spite of our great progress, our new understanding is partly illusory. Many of our theories are oversimplified and some are probably wrong. Future workers must ask hard questions and demand stronger evidence for basic ideas. We should raise our standards for scientific proof. The motivation for this enterprise is strong. We seek nothing less than an understanding of the weather, the climate, and the geography of our planet.

Acknowledgments. The subject of mountain meteorology is geographically diverse because orographic phenomena can occur in numerous different mountainous terrains around the world. It is thematically diverse as it includes many physical processes, length scales, and scientific strategies. Due to space limitations, only a few examples of each phenomenon could be discussed here. I apologize to the authors of many excellent and relevant research papers that could not be mentioned in this short summary. I have benefited from discussing mountain meteorology with many colleagues; too numerous to mention. A few scientists with an early impact on my ideas include Robert Long, Francis Bretherton, Douglas Lilly, William Blumen, Arnt Eliassen, Morton Wurtele, Arne Foldvik, and Joachim Kuetttner. Among my contemporaries, I owe a strong debt to Larry Armi, Peter Baines, Robert Banta, Philippe Bourgeault, Rit Carbone, Brian Colle, William Cooper, Andreas Dörrnbrack, James Doyle, Dale Durran, Huw Davies, Steven Eckermann, Dave Fritts, Rene Garreaud, Klaus-Peter Hoinka, Robert Houze, Dave Kingsmill, Joseph Klemp, Francois Lott, Steven Mobbs, Richard Peltier, Ray Pierrehumbert, Mike Revell, Piotr Smolarkiewicz, Richard Rotunno, Reinhold Steinacker, Ignaz Vergeiner, Hans Volkert, Simon Vosper, David Whiteman, and others. Their enthusiasm inspired me. A large influence also came from my graduate students: Yuh-Lang Lin, Jielun Sun, Eric Salathe, Vanda Grubisic, Qingfeng Jiang, Hui He, Benjamin Zaichik, Yanping Li, Bryan K. Woods, Alison Nugent, Christopher Kruse and post-docs Wen-dar Chen, Christoph Schärf, Steven Skubis, Daniel Kirshbaum, Jason Evans, Idar Barstad, Justin Minder, Campbell Watson, and Gang Zhang. My work was mostly supported by the Physical and Dynamical Meteorology program and the Aeronomy program at the National Science Foundation. Thanks also to the Yale Department of Geology and Geophysics for welcoming this research over the years. Sigrid R-P Smith helped with the references, editing, and morale. Mark Brandon helped with Table 20-10, Dale Durran and two other reviewers helped to correct many deficiencies in the early drafts. Finally, thanks to the American Meteorological Society for sponsoring this review.

REFERENCES

REFERENCES

REFERENCES

REFERENCES


Crapper, G. D., 1962: Waves in the lee of a mountain with elliptical
Crosman, E. T., and J. D. Horel, 2017: Large-eddy simulations of a
Crouch, P., T. J. Órnsson, and I. Barstad, 2007: Estimating the
Colman, B. R., and C. F. Dierking, 1992: The Taku wind of


Roe, G. H., 2005: Orographic precipitation.


Schulz, T., P. Crochet, R. Hock, M. Jackson, I. Barstad, and T. Johannesson, 2008: Distribution of snow accumulation on the


Shi, Z., Y. Sha, and X. Liu, 2017: Effect of Yunnan–Guizhou to-


