

Introduction

ABSTRACT

This opening chapter defines the discipline known as geophysical fluid dynamics, stresses its importance, and highlights its most distinctive attributes. A brief history of numerical simulations in meteorology and oceanography is also presented. Scale analysis and its relationship with finite differences are introduced to show how discrete numerical grids depend on the scales under investigation and how finite differences permit the approximation of derivatives at those scales. The problem of unresolved scales is introduced as an aliasing problem in discretization.

1.1 OBJECTIVE

The object of geophysical fluid dynamics is the study of naturally occurring, large-scale flows on Earth and elsewhere, but mostly on Earth. Although the discipline encompasses the motions of both fluid phases – liquids (waters in the ocean, molten rock in the outer core) and gases (air in our atmosphere, atmospheres of other planets, ionized gases in stars) – a restriction is placed on the *scale* of these motions. Only the large-scale motions fall within the scope of geophysical fluid dynamics. For example, problems related to river flow, micro-turbulence in the upper ocean, and convection in clouds are traditionally viewed as topics specific to hydrology, oceanography, and meteorology, respectively. Geophysical fluid dynamics deals exclusively with those motions observed in various systems and under different guises but nonetheless governed by similar dynamics. For example, large anticyclones of our weather are dynamically germane to vortices spun off by the Gulf Stream and to Jupiter's Great Red Spot. Most of these problems, it turns out, are at the large-scale end, where either the ambient rotation (of Earth, planet, or star) or density differences (warm and cold air masses, fresh and saline waters), or both assume some importance. In this respect, geophysical fluid dynamics comprises rotating-stratified fluid dynamics.

Typical problems in geophysical fluid dynamics concern the variability of the atmosphere (weather and climate dynamics), ocean (waves, vortices, and currents), and, to a lesser extent, the motions in the earth's interior responsible for the dynamo effect, vortices on other planets (such as Jupiter's Great

Red Spot and Neptune's Great Dark Spot), and convection in stars (the sun, in particular).

1.2 IMPORTANCE OF GEOPHYSICAL FLUID DYNAMICS

Without its atmosphere and oceans, it is certain that our planet would not sustain life. The natural fluid motions occurring in these systems are therefore of vital importance to us, and their understanding extends beyond intellectual curiosity—it is a necessity. Historically, weather vagaries have baffled scientists and laypersons alike since times immemorial. Likewise, conditions at sea have long influenced a wide range of human activities, from exploration to commerce, tourism, fisheries, and even wars.

Thanks in large part to advances in geophysical fluid dynamics, the ability to predict with some confidence the paths of hurricanes (Figs. 1.1 and 1.2) has led to the establishment of a warning system that, no doubt, has saved numerous lives at sea and in coastal areas (Abbott, 2004). However, warning systems are only useful if sufficiently dense observing systems are implemented, fast prediction capabilities are available, and efficient flow of information is ensured. A dreadful example of a situation in which a warning system was not yet adequate to save lives was the earthquake off Indonesia's Sumatra Island on

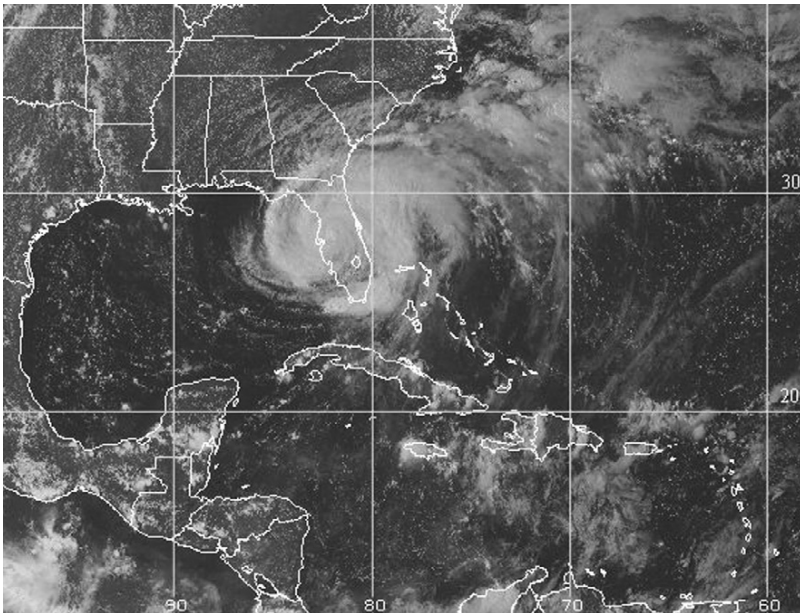


FIGURE 1.1 Hurricane Frances during her passage over Florida on 5 September 2004. The diameter of the storm was about 830 km, and its top wind speed approached 200 km per hour. (Courtesy of NOAA, Department of Commerce, Washington, D.C.)

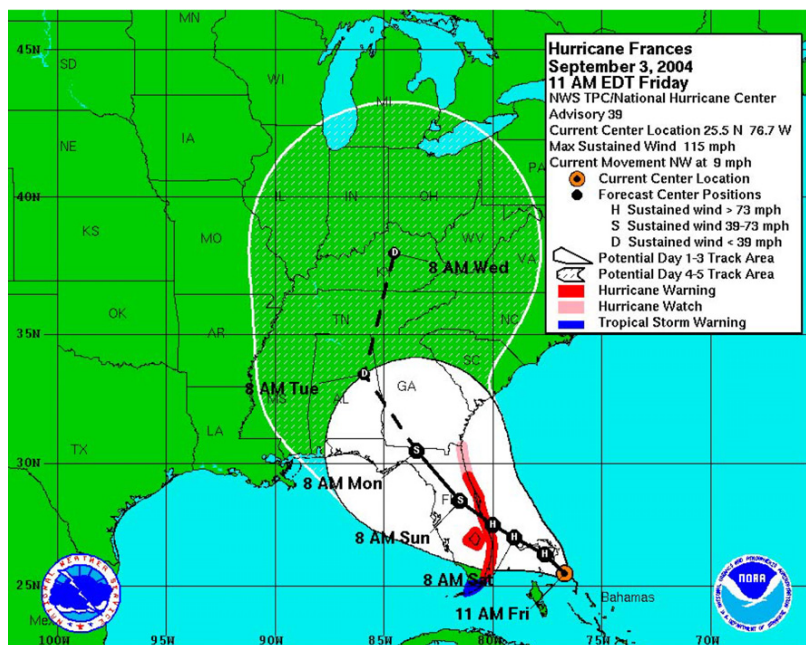


FIGURE 1.2 Computer prediction of the path of Hurricane Frances. The calculations were performed on Friday, 3 September 2004, to predict the hurricane path and characteristics over the next 5 days (until Wednesday, 8 September). The outline surrounding the trajectory indicates the level of uncertainty. Compare the position predicted for Sunday, 5 September, with the actual position shown on Fig. 1.1. (Courtesy of NOAA, Department of Commerce, Washington, D.C.)

26 December 2004. The tsunami generated by the earthquake was not detected, its consequences not assessed, and authorities not alerted within the 2 h needed for the wave to reach beaches in the region. On a larger scale, the passage every 3–5 years of an anomalously warm water mass along the tropical Pacific Ocean and the western coast of South America, known as the El-Niño event, has long been blamed for serious ecological damage and disastrous economical consequences in some countries (Glantz, 2001; O’Brien, 1978). Now, thanks to increased understanding of long oceanic waves, atmospheric convection, and natural oscillations in air–sea interactions (D’Aleo, 2002; Philander, 1990), scientists have successfully removed the veil of mystery on this complex event, and numerical models (e.g., Chen, Cane, Kaplan, Zebiak & Huang, 2004) offer reliable predictions with at least one year of *lead time*, that is, there is a year between the moment the prediction is made and the time to which it applies.

Having acknowledged that our industrial society is placing a tremendous burden on the planetary atmosphere and consequently on all of us, scientists, engineers, and the public are becoming increasingly concerned about the fate of pollutants and greenhouse gases dispersed in the environment and especially

about their cumulative effect. Will the accumulation of greenhouse gases in the atmosphere lead to global climatic changes that, in turn, will affect our lives and societies? What are the various roles played by the oceans in maintaining our present climate? Is it possible to reverse the trend toward depletion of the ozone in the upper atmosphere? Is it safe to deposit hazardous wastes on the ocean floor? Such pressing questions cannot find answers without, first, an in-depth understanding of atmospheric and oceanic dynamics and, second, the development of predictive models. In this twin endeavor, geophysical fluid dynamics assumes an essential role, and the numerical aspects should not be underestimated in view of the required predictive tools.

1.3 DISTINGUISHING ATTRIBUTES OF GEOPHYSICAL FLOWS

Two main ingredients distinguish the discipline from traditional fluid mechanics: the effects of rotation and those of stratification. The controlling influence of one, the other, or both leads to peculiarities exhibited only by geophysical flows. In a nutshell, this book can be viewed as an account of these peculiarities.

The presence of an ambient rotation, such as that due to the earth's spin about its axis, introduces in the equations of motion two acceleration terms that, in the rotating framework, can be interpreted as forces. They are the Coriolis force and the centrifugal force. Although the latter is the more palpable of the two, it plays no role in geophysical flows; however, surprising this may be.¹ The former and less intuitive of the two turns out to be a crucial factor in geophysical motions. For a detailed explanation of the Coriolis force, the reader is referred to the following chapter in this book or to the book by Stommel and Moore (1989). A more intuitive explanation and laboratory illustrations can be found in Chapter 6 of Marshall and Plumb (2008).

In anticipation of the following chapters, it can be mentioned here (without explanation) that a major effect of the Coriolis force is to impart a certain vertical rigidity to the fluid. In rapidly rotating, homogeneous fluids, this effect can be so strong that the flow displays strict columnar motions; that is, all particles along the same vertical evolve in concert, thus retaining their vertical alignment over long periods of time. The discovery of this property is attributed to Geoffrey I. Taylor, a British physicist famous for his varied contributions to fluid dynamics. (See the short biography at the end of Chapter 7.) It is said that Taylor first arrived at the rigidity property with mathematical arguments alone. Not believing that this could be correct, he then performed laboratory experiments that revealed, much to his amazement, that the theoretical prediction was indeed correct. Drops of dye released in such rapidly rotating, homogeneous

¹ Here we speak about the centrifugal force associated with the earth's planetary rotation, not to be confused with the centrifugal force associated with the strong rotation of eddies or hurricanes.

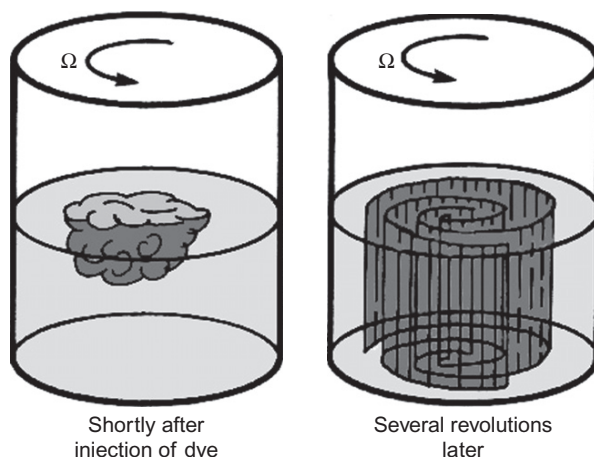


FIGURE 1.3 Experimental evidence of the rigidity of a rapidly rotating, homogeneous fluid. In a spinning vessel filled with clear water, an initially amorphous cloud of aqueous dye is transformed in the course of several rotations into perfectly vertical sheets, known as *Taylor curtains*.

fluids form vertical streaks, which, within a few rotations, shear laterally to form spiral sheets of dyed fluid (Fig. 1.3). The vertical coherence of these sheets is truly fascinating!

In large-scale atmospheric and oceanic flows, such state of perfect vertical rigidity is not realized chiefly because the rotation rate is not sufficiently fast and the density is not sufficiently uniform to mask other, ongoing processes. Nonetheless, motions in the atmosphere, in the oceans, and on other planets manifest a tendency toward columnar behavior. For example, currents in the western North Atlantic have been observed to extend vertically over 4000 m without significant change in amplitude and direction (Schmitz, 1980).

Stratification, the other distinguishing attribute of geophysical fluid dynamics, arises because naturally occurring flows typically involve fluids of different densities (e.g., warm and cold air masses, fresh and saline waters). Here, the gravitational force is of great importance, for it tends to lower the heaviest fluid and to raise the lightest. Under equilibrium conditions, the fluid is stably stratified, consisting of vertically stacked horizontal layers. However, fluid motions disturb this equilibrium, in which gravity systematically strives to restore. Small perturbations generate internal waves, the three-dimensional analogue of surface waves, with which we are all familiar. Large perturbations, especially those maintained over time, may cause mixing and convection. For example, the prevailing winds in our atmosphere are manifestations of the planetary convection driven by the pole-to-equator temperature difference.

It is worth mentioning the perplexing situation in which a boat may experience strong resistance to forward motion while sailing under apparently calm conditions. This phenomenon, called *dead waters* by mariners, was first

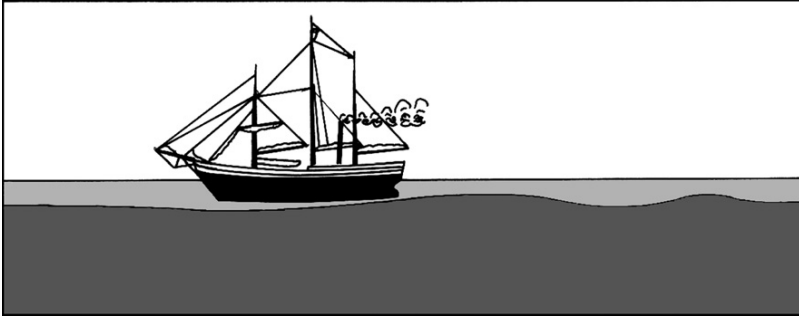


FIGURE 1.4 A laboratory experiment by Ekman (1904) showing internal waves generated by a model ship in a tank filled with two fluids of different densities. The heavier fluid at the bottom has been colored to make the interface visible. The model ship (the superstructure of which was drawn onto the original picture to depict Fridtjof Nansen's *Fram*) is towed from right to left, causing a wake of waves on the interface. The energy consumed by the generation of these waves produces a drag that, for a real ship, would translate into a resistance to forward motion. The absence of any significant surface wave has prompted sailors to call such situations *dead waters*. (From Ekman, 1904, and adapted by Gill, 1982)

documented by the Norwegian oceanographer Fridtjof Nansen, famous for his epic expedition on the *Fram* through the Arctic Ocean, begun in 1893. Nansen reported the problem to his Swedish colleague Vagn Walfrid Ekman who, after performing laboratory simulations (Ekman, 1904), affirmed that internal waves were to blame. The scenario is as follows: During times of dead waters, Nansen must have been sailing in a layer of relatively fresh water capping the more saline oceanic waters and of thickness, coincidentally, comparable to the ship draft; the ship created a wake of internal waves along the interface (Fig. 1.4), unseen at the surface but radiating considerable energy and causing the noted resistance to the forward motion of the ship.

1.4 SCALES OF MOTIONS

To discern whether a physical process is dynamically important in any particular situation, geophysical fluid dynamicists introduce *scales of motion*. These are dimensional quantities expressing the overall magnitude of the variables under consideration. They are estimates rather than precisely defined quantities and are understood solely as *orders of magnitude* of physical variables. In most situations, the key scales are those for time, length, and velocity. For example, in the dead-water situation investigated by V.W. Ekman (Fig. 1.4), fluid motions comprise a series of waves whose dominant wavelength is about the length of the submerged ship hull; this length is the natural choice for the length scale L of the problem; likewise, the ship speed provides a reference velocity that can be taken as the velocity scale U ; finally, the time taken for the ship to travel the distance L at its speed U is the natural choice of time scale: $T = L/U$.

As a second example, consider Hurricane Frances during her course over the southeastern United States in early September 2004 (Fig. 1.1). The satellite picture reveals a nearly circular feature spanning approximately 7.5° of latitude (830 km). Sustained surface wind speeds of a category-4 hurricane such as Frances range from 59 to 69 m/s. In general, hurricane tracks display appreciable change in direction and speed of propagation over 2-day intervals. Altogether, these elements suggest the following choice of scales for a hurricane: $L = 800$ km, $U = 60$ m/s, and $T = 2 \times 10^5$ s ($= 55.6$ h).

As a third example, consider the famous Great Red Spot in Jupiter's atmosphere (Fig. 1.5), which is known to have existed at least several hundred years. The structure is an elliptical vortex centered at 22° S and spanning approximately 12° in latitude and 25° in longitude; its highest wind speeds exceed 110 m/s, and the entire feature slowly drifts zonally at a speed of 3 m/s (Dowling & Ingersoll, 1988; Ingersoll et al., 1979). Knowing that the planet's equatorial radius is 71,400 km, we determine the vortex semi-major and semi-minor axes (14,400 km and 7,500 km, respectively) and deem $L = 10,000$ km to be an appropriate length scale. A natural velocity scale for the fluid is $U = 100$ m/s. The selection of a timescale is somewhat problematic in view of the nearly

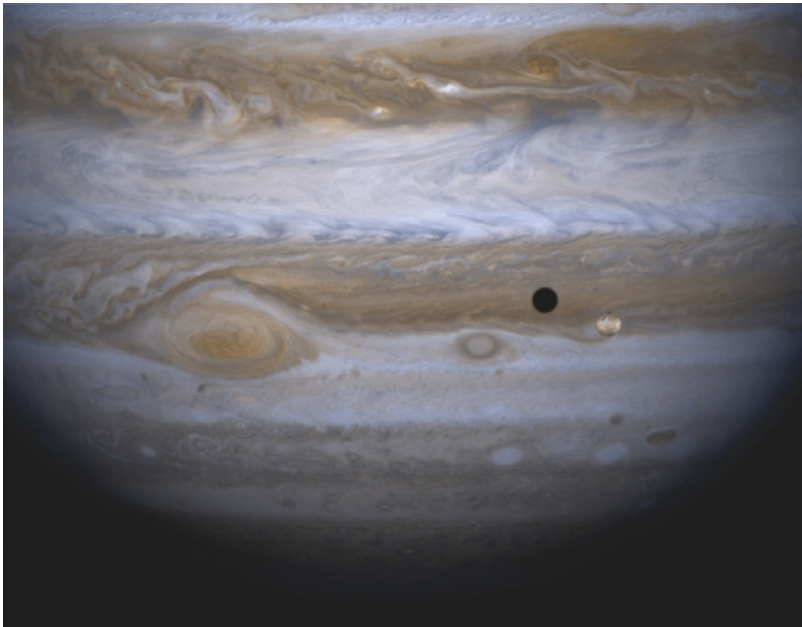


FIGURE 1.5 Southern hemisphere of Jupiter as seen by the spacecraft *Cassini* in 2000. The Jupiter moon Io, of size comparable to our moon, projects its shadow onto the zonal jets between which the Great Red Spot of Jupiter is located (on the left). For further images, visit <http://photojournal.jpl.nasa.gov/target/Jupiter>. (Image courtesy of NASA/JPL/University of Arizona)

steady state of the vortex; one choice is the time taken by a fluid particle to cover the distance L at the speed U ($T = L/U = 10^5$ s), whereas another is the time taken by the vortex to drift zonally over a distance equal to its longitudinal extent ($T = 10^7$ s). Additional information on the physics of the problem is clearly needed before selecting a timescale. Such ambiguity is not uncommon because many natural phenomena vary on different temporal scales (e.g., the terrestrial atmosphere exhibits daily weather variation as well as decadal climatic variations, among others). The selection of a timescale then reflects the particular choice of physical processes being investigated in the system.

There are three additional scales that play important roles in analyzing geophysical fluid problems. As we mentioned earlier, geophysical fluids generally exhibit a certain degree of density heterogeneity, called stratification. The important parameters are then the average density ρ_0 , the range of density variations $\Delta\rho$, and the height H over which such density variations occur. In the ocean, the weak compressibility of water under changes of pressure, temperature, and salinity translates into values of $\Delta\rho$ always much less than ρ_0 , whereas the compressibility of air renders the selection of $\Delta\rho$ in atmospheric flows somewhat delicate. Because geophysical flows are generally bounded in the vertical direction, the total depth of the fluid may be substituted for the height scale H . Usually, the smaller of the two height scales is selected.

As an example, the density and height scales in the dead-water problem (Fig. 1.4) can be chosen as follows: $\rho_0 = 1025 \text{ kg/m}^3$, the density of either fluid layer (almost the same); $\Delta\rho = 1 \text{ kg/m}^3$, the density difference between lower and upper layers (much smaller than ρ_0), and $H = 5 \text{ m}$, the depth of the upper layer.

As the person new to geophysical fluid dynamics has already realized, the selection of scales for any given problem is more an art than a science. Choices are rather subjective. The trick is to choose quantities that are relevant to the problem, yet simple to establish. There is freedom. Fortunately, small inaccuracies are inconsequential because the scales are meant only to guide in the clarification of the problem, whereas grossly inappropriate scales will usually lead to flagrant contradictions. Practice, which forms intuition, is necessary to build confidence.

1.5 IMPORTANCE OF ROTATION

Naturally, we may wonder at which scales the ambient rotation becomes an important factor in controlling the fluid motions. To answer this question, we must first know the ambient rotation rate, which we denote by Ω and define as

$$\Omega = \frac{2\pi \text{ radians}}{\text{time of one revolution}}. \quad (1.1)$$

Since our planet Earth actually rotates in two ways simultaneously, once per day about itself and once a year around the sun, the terrestrial value of Ω consists of two terms, $2\pi/24 \text{ hours} + 2\pi/365.24 \text{ days} = 2\pi/1 \text{ sidereal day} = 7.2921 \times 10^{-5} \text{ s}^{-1}$. The *sidereal day*, equal to 23 h 56 min and 4.1 s, is the

period of time spanning the moment when a fixed (distant) star is seen one day and the moment on the next day when it is seen at the same angle from the same point on Earth. It is slightly shorter than the 24-hour solar day, the time elapsed between the sun reaching its highest point in the sky two consecutive times, because the earth's orbital motion about the sun makes the earth rotate slightly more than one full turn with respect to distant stars before reaching the same Earth–Sun orientation.

If fluid motions evolve on a timescale comparable to or longer than the time of one rotation, we anticipate that the fluid does feel the effect of the ambient rotation. We thus define the dimensionless quantity

$$\omega = \frac{\text{time of one revolution}}{\text{motion timescale}} = \frac{2\pi/\Omega}{T} = \frac{2\pi}{\Omega T}, \quad (1.2)$$

where T is used to denote the timescale of the flow. Our criterion is as follows: If ω is on the order of or less than unity ($\omega \lesssim 1$), rotation effects should be considered. On Earth, this occurs when T exceeds 24 h.

Yet, motions with shorter timescales ($\omega \gtrsim 1$) but sufficiently large spatial extent could also be influenced by rotation. A second and usually more useful criterion results from considering the velocity and length scales of the motion. Let us denote these by U and L , respectively. Naturally, if a particle traveling at the speed U covers the distance L in a time longer than or comparable to a rotation period, we expect the trajectory to be influenced by the ambient rotation, so we write

$$\begin{aligned} \epsilon &= \frac{\text{time of one revolution}}{\text{time taken by particle to cover distance } L \text{ at speed } U} \\ &= \frac{2\pi/\Omega}{L/U} = \frac{2\pi U}{\Omega L}. \end{aligned} \quad (1.3)$$

If ϵ is on the order of or less than unity ($\epsilon \lesssim 1$), we conclude that rotation is important.

Let us now consider a variety of possible length scales, using the value Ω for Earth. The corresponding velocity criteria are listed in [Table 1.1](#).

Obviously, in most engineering applications, such as the flow of water at a speed of 5 m/s in a turbine 1 m in diameter ($\epsilon \sim 4 \times 10^5$) or the air flow past a 5-m wing on an airplane flying at 100 m/s ($\epsilon \sim 2 \times 10^6$), the inequality is not met, and the effects of rotation can be ignored. Likewise, the common task of emptying a bathtub (horizontal scale of 1 m, draining speed in the order of 0.01 m/s and a lapse of about 1000 s, giving $\omega \sim 90$ and $\epsilon \sim 900$) does not fall under the scope of geophysical fluid dynamics. On the contrary, geophysical flows (such as an ocean current flowing at 10 cm/s and meandering over a distance of 10 km or a wind blowing at 10 m/s in a 1000-km-wide anticyclonic formation) do meet the inequality. This demonstrates that rotation is usually important in geophysical flows.

TABLE 1.1 Length and Velocity Scales of Motions in Which Rotation Effects are Important

$L = 1 \text{ m}$	$U \leq 0.012 \text{ mm/s}$
$L = 10 \text{ m}$	$U \leq 0.12 \text{ mm/s}$
$L = 100 \text{ m}$	$U \leq 1.2 \text{ mm/s}$
$L = 1 \text{ km}$	$U \leq 1.2 \text{ cm/s}$
$L = 10 \text{ km}$	$U \leq 12 \text{ cm/s}$
$L = 100 \text{ km}$	$U \leq 1.2 \text{ m/s}$
$L = 1000 \text{ km}$	$U \leq 12 \text{ m/s}$
$L = \text{Earth radius} = 6371 \text{ km}$	$U \leq 74 \text{ m/s}$

1.6 IMPORTANCE OF STRATIFICATION

The next question concerns the condition under which stratification effects are expected to play an important dynamical role. Geophysical fluids typically consist of fluid masses of different densities, which under gravitational action tend to arrange themselves in vertical stacks (Fig. 1.6), corresponding to a state of minimal potential energy. But, motions continuously disturb this equilibrium, tending to raise dense fluid and lower light fluid. The corresponding increase of potential energy is at the expense of kinetic energy, thereby slowing the flow. On occasions, the opposite happens: Previously disturbed stratification returns toward equilibrium, potential energy converts into kinetic energy, and the flow gains momentum. In sum, the dynamical importance of stratification can be evaluated by comparing potential and kinetic energies.

If $\Delta\rho$ is the scale of density variations in the fluid and H is its height scale, a prototypical perturbation to the stratification consists in raising a fluid element of density $\rho_0 + \Delta\rho$ over the height H and, in order to conserve volume, lowering a lighter fluid element of density ρ_0 over the same height. The corresponding change in potential energy, per unit volume, is $(\rho_0 + \Delta\rho)gH - \rho_0gH = \Delta\rho gH$. With a typical fluid velocity U , the kinetic energy available per unit volume is $\frac{1}{2}\rho_0 U^2$. Accordingly, we construct the comparative energy ratio

$$\sigma = \frac{\frac{1}{2}\rho_0 U^2}{\Delta\rho gH}, \quad (1.4)$$

to which we can give the following interpretation. If σ is on the order of unity ($\sigma \sim 1$), a typical potential-energy increase necessary to perturb the stratification consumes a sizable portion of the available kinetic energy, thereby modifying the flow field substantially. Stratification is then important. If σ is much less

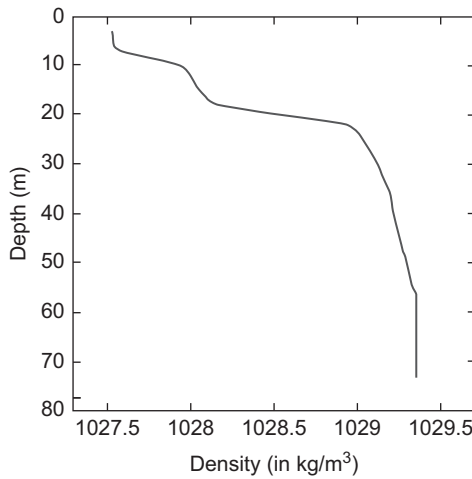


FIGURE 1.6 Vertical profile of density in the northern Adriatic Sea ($43^{\circ}32'N$, $14^{\circ}03'E$) on 27 May 2003. Density increases downward by leaps and bounds, revealing the presence of different water masses stacked on top of one another in such a way that lighter waters float above denser waters. A region where the density increases significantly faster than above and below, marking the transition from one water mass to the next, is called a *pycnocline*. (Data courtesy of Drs. Hartmut Peters and Mirko Orlić)

than unity ($\sigma \ll 1$), there is insufficient kinetic energy to perturb significantly the stratification, and the latter greatly constrains the flow. Finally, if σ is much greater than unity ($\sigma \gg 1$), potential-energy modifications occur at very little cost to the kinetic energy, and stratification hardly affects the flow. In conclusion, stratification effects cannot be ignored in the first two cases—that is, when the dimensionless ratio defined in Eq. (1.4) is on the order of or much less than unity ($\sigma \lesssim 1$). In other words, σ is to stratification what the number ϵ , defined in Eq. (1.3), is to rotation.

A most interesting situation arises in geophysical fluids when rotation and stratification effects are simultaneously important, yet neither dominates over the other. Mathematically, this occurs when $\epsilon \sim 1$ and $\sigma \sim 1$ and yields the following relations among the various scales:

$$L \sim \frac{U}{\Omega} \quad \text{and} \quad U \sim \sqrt{\frac{\Delta\rho}{\rho_0} gH}. \quad (1.5)$$

(The factors 2π and $\frac{1}{2}$ have been omitted because they are secondary in a scale analysis.) Elimination of the velocity U yields a fundamental length scale:

$$L \sim \frac{1}{\Omega} \sqrt{\frac{\Delta\rho}{\rho_0} gH}. \quad (1.6)$$

In a given fluid, of mean density ρ_0 and density variation $\Delta\rho$, occupying a height H on a planet rotating at rate Ω and exerting a gravitational acceleration g , the scale L arises as a preferential length over which motions take place. On Earth ($\Omega = 7.29 \times 10^{-5} \text{ s}^{-1}$ and $g = 9.81 \text{ m/s}^2$), typical conditions in the atmosphere ($\rho_0 = 1.2 \text{ kg/m}^3$, $\Delta\rho = 0.03 \text{ kg/m}^3$, $H = 5000 \text{ m}$) and in the ocean ($\rho_0 = 1028 \text{ kg/m}^3$, $\Delta\rho = 2 \text{ kg/m}^3$, $H = 1000 \text{ m}$) yield the following natural length and velocity scales:

$$\begin{aligned} L_{\text{atmosphere}} &\sim 500 \text{ km} & U_{\text{atmosphere}} &\sim 30 \text{ m/s} \\ L_{\text{ocean}} &\sim 60 \text{ km} & U_{\text{ocean}} &\sim 4 \text{ m/s}. \end{aligned}$$

Although these estimates are relatively crude, we can easily recognize here the typical size and wind speed of weather patterns in the lower atmosphere and the typical width and speed of major currents in the upper ocean.

1.7 DISTINCTION BETWEEN THE ATMOSPHERE AND OCEANS

Generally, motions of the air in our atmosphere and of seawater in the oceans that fall under the scope of geophysical fluid dynamics occur on scales of several kilometers up to the size of the earth. Atmospheric phenomena comprise the coastal sea breeze, local to regional processes associated with topography, the cyclones, anticyclones, and fronts that form our daily weather, the general atmospheric circulation, and the climatic variations. Oceanic phenomena of interest include estuarine flow, coastal upwelling and other processes associated with the presence of a coast, large eddies and fronts, major ocean currents such as the Gulf Stream, and the large-scale circulation. [Table 1.2](#) lists the typical velocity, length and time scales of these motions, whereas [Fig. 1.7](#) ranks a sample of atmospheric and oceanic processes according to their spatial and temporal scales. As we can readily see, the general rule is that oceanic motions are slower and slightly more confined than their atmospheric counterparts. Also, the ocean tends to evolve more slowly than the atmosphere.

Besides notable scale disparities, the atmosphere and oceans also have their own peculiarities. For example, a number of oceanic processes are caused by the presence of lateral boundaries (continents, islands), a constraint practically nonexistent in the atmosphere, except in stratified flows where mountain ridges can sometimes play such a role, exactly as do mid-ocean ridges for stratified ocean currents. On the other hand, atmospheric motions are sometimes strongly dependent on the moisture content of the air (clouds, precipitation), a characteristic without oceanic counterpart.

Flow patterns in the atmosphere and oceans are generated by vastly different mechanisms. By and large, the atmosphere is thermodynamically driven, that is, its primary source of energy is the solar radiation. Briefly, this shortwave solar

TABLE 1.2 Length, Velocity and Time Scales in the Earth’s Atmosphere and Oceans

Phenomenon	Length Scale <i>L</i>	Velocity Scale <i>U</i>	Timescale <i>T</i>
Atmosphere			
Microturbulence	10–100 cm	5–50 cm/s	few seconds
Thunderstorms	few km	1–10 m/s	few hours
Sea breeze	5–50 km	1–10 m/s	6 h
Tornado	10–500 m	30–100 m/s	10–60 min
Hurricane	300–500 km	30–60 m/s	Days to weeks
Mountain waves	10–100 km	1–20 m/s	Days
Weather patterns	100–5000 km	1–50 m/s	Days to weeks
Prevailing winds	Global	5–50 m/s	Seasons to years
Climatic variations	Global	1–50 m/s	Decades and beyond
Ocean			
Microturbulence	1–100 cm	1–10 cm/s	10–100 s
Internal waves	1–20 km	0.05–0.5 m/s	Minutes to hours
Tides	Basin scale	1–100 m/s	Hours
Coastal upwelling	1–10 km	0.1–1 m/s	Several days
Fronts	1–20 km	0.5–5 m/s	Few days
Eddies	5–100 km	0.1–1 m/s	Days to weeks
Major currents	50–500 km	0.5–2 m/s	Weeks to seasons
Large-scale gyres	Basin scale	0.01–0.1 m/s	Decades and beyond

radiation traverses the air layer to be partially absorbed by the continents and oceans, which in turn re-emit a radiation at longer wavelengths. This second-hand radiation effectively heats the atmosphere from below, and the resulting convection drives the winds.

In contrast, the oceans are forced by a variety of mechanisms. In addition to the periodic gravitational forces of the moon and sun that generate the tides, the ocean surface is subjected to a wind stress that drives most ocean currents. Finally, local differences between air and sea temperatures generate heat fluxes,

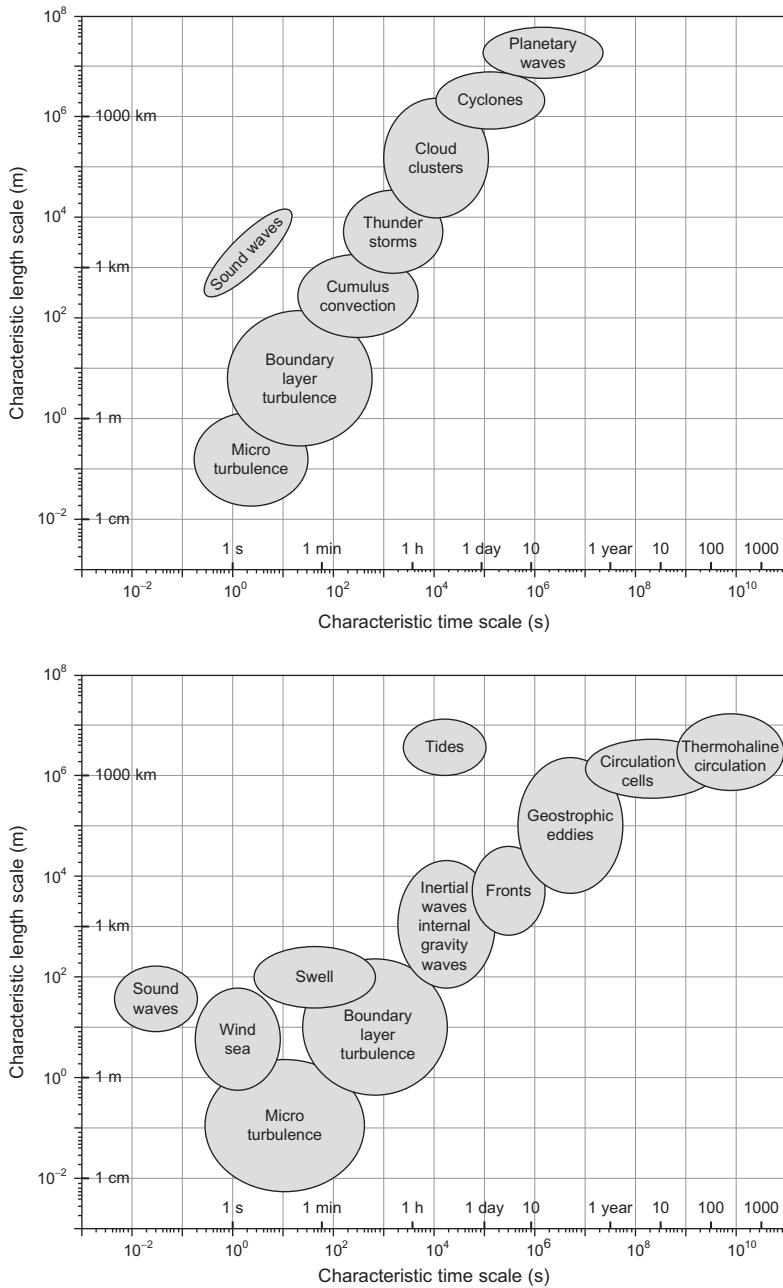


FIGURE 1.7 Various types of processes and structures in the atmosphere (top panel) and oceans (bottom panel), ranked according to their respective length and time scales. (Diagram courtesy of Hans von Storch)

evaporation, and precipitation, which in turn act as thermodynamical forcings capable of modifying the wind-driven currents or producing additional currents.

In passing, while we are contrasting the atmosphere with the oceans, it is appropriate to mention an enduring difference in terminology. Because meteorologists and laypeople alike are generally interested in knowing from where the winds are blowing, it is common in meteorology to refer to air velocities by their direction of origin, such as easterly (from the east—that is, toward the west). On the contrary, sailors and navigators are interested in knowing where ocean currents may take them. Hence, oceanographers designate currents by their downstream direction, such as westward (from the east or to the west). However, meteorologists and oceanographers agree on the terminology for vertical motions: upward or downward.

1.8 DATA ACQUISITION

Because geophysical fluid dynamics deals exclusively with naturally occurring flows and, moreover, those of rather sizable proportions, full-scale experimentation must be ruled out. Indeed, how could one conceive of changing the weather, even locally, for the sake of a scientific inquiry? Also, the Gulf Stream determines its own fancy path, irrespective of what oceanographers wish to study about it. In that respect, the situation is somewhat analogous to that of the economist who may not ask the government to prompt a disastrous recession for the sake of determining some parameters of the national economy. The inability to control the system under study is greatly alleviated by simulations. In geophysical fluid dynamics, these investigations are conducted via laboratory experiments and numerical models.

As well as being reduced to noting the whims of nature, observers of geophysical flows also face length and timescales that can be impractically large. A typical challenge is the survey of an oceanic feature several hundred kilometers wide. With a single ship (which is already quite expensive, especially if the feature is far away from the home shore), a typical survey can take several weeks, a time interval during which the feature might translate, distort, or otherwise evolve substantially. A faster survey might not reveal details with a sufficiently fine horizontal representation. Advances in satellite imagery and other methods of remote sensing (Conway & the Maryland Space Grant Consortium, 1997; Marzano & Visconti, 2002) do provide synoptic (i.e., quasi-instantaneous) fields, but those are usually restricted to specific levels in the vertical (e.g., cloud tops and ocean surface) or provide vertically integrated quantities. Also, some quantities simply defy measurement, such as the heat flux and vorticity. Those quantities can only be derived by an analysis on sets of proxy observations.

Finally, there are processes for which the timescale is well beyond the span of human life if not the age of civilization. For example, climate studies require a certain understanding of glaciation cycles. Our only recourse here is to be clever and to identify today some traces of past glaciation events, such as geological

records. Such an indirect approach usually requires a number of assumptions, some of which may never be adequately tested. Finally, exploration of other planets and of the sun is even more arduous.

At this point one may ask: What can we actually measure in the atmosphere and oceans with a reasonable degree of confidence? First and foremost, a number of scalar properties can be measured directly and with conventional instruments. For both the atmosphere and ocean, it is generally not difficult to measure the pressure and temperature. In fact, in the ocean, the pressure can be measured so much more accurately than depth that, typically, depth is calculated from measured pressure on instruments that are gradually lowered into the sea. In the atmosphere, one can also accurately measure the water vapor, rainfall, and some radiative heat fluxes (Marzano & Visconti, 2002; Rao, Holmes, Anderson, Winston & Lehr, 1990). Similarly, the salinity of seawater can be either determined directly or inferred from electrical conductivity (Pickard & Emery, 1990). Also, the sea level can be monitored at shore stations. The typical problem, however, is that the measured quantities are not necessarily those preferred from a physical perspective. For example, one would prefer direct measurements of the vorticity field, Bernoulli function, diffusion coefficients, and turbulent correlation quantities.

Vectorial quantities are usually more difficult to obtain than scalars. Horizontal winds and currents can now be determined routinely by anemometers and current meters of various types, including some without rotating components (Lutgens & Tarbuck, 1986; Pickard & Emery, 1990) although usually not with the desired degree of spatial resolution. Fixed instruments, such as anemometers atop buildings and oceanic current meters at specific depths along a mooring line, offer fine temporal coverage, but adequate spatial coverage typically requires a prohibitive number of such instruments. To remedy the situation, instruments on drifting platforms (e.g., balloons in the atmosphere and drifters or floats in the ocean) are routinely deployed. However, these instruments provide information that is mixed in time and space and thus is not ideally suited to most purposes. A persistent problem is the measurements of the vertical velocity. Although vertical speeds can be measured with acoustic Doppler current profilers, the meaningful signal is often buried below the level of ambient turbulence and instrumental error (position and sensitivity). Measuring the vector vorticity, so dear to theoreticians, is out of the question as is the three-dimensional heat flux.

Also, some uncertainty resides in the interpretation of the measured quantities. For example, can the wind measured in the vicinity of a building be taken as representative of the prevailing wind over the city and so be used in weather forecasting, or is it more representative of a small-scale flow pattern resulting from the obstruction of the wind by the building?

Finally, sampling frequencies might not always permit the unambiguous identification of a process. Measuring quantities at a given location every week

might well lead to a data set that includes also residual information on faster processes than weekly variations or a slower signal that we would like to capture with our measurements. For example, if we measure temperature on Monday at 3 o'clock in the afternoon one week and Monday at 7 o'clock in the morning the next week, the measurement will include a diurnal heating component superimposed on the weekly variation. The measurements are thus not necessarily representative of the process of interest.

1.9 THE EMERGENCE OF NUMERICAL SIMULATIONS

Given the complexity of weather patterns and ocean currents, one can easily anticipate that the equations governing geophysical fluid motions, which we are going to establish in this book, are formidable and not amenable to analytical solution except in rare instances and after much simplification. Thus, one faces the tall challenge of having to solve the apparently unsolvable. The advent of the electronic computer has come to the rescue, but at a definite cost. Indeed, computers cannot solve differential equations but can only perform the most basic arithmetic operations. The partial differential equations (PDEs) of geophysical fluid dynamics (GFD) need therefore to be transformed into a sequence of arithmetic operations. This process requires careful transformations and attention to details.

The purpose of numerical simulations of GFD flows is not limited to weather prediction, operational ocean forecasting, and climate studies. There are situations when one desires to gain insight and understanding of a specific process, such as a particular form of instability or the role of friction under particular conditions. Computer simulations are our only way to experiment with the planet. Also, there is the occasional need to experiment with a novel numerical technique in order to assess its speed and accuracy. Simulations increasingly go hand in hand with observations in the sense that the latter can point to places in which the model needs refinements while model results can suggest optimal placing of observational platforms or help to define sampling strategies. Finally, simulations can be a retracing of the past (*hindcasting*) or a smart interpolation of scattered data (*nowcasting*), as well as the prediction of future states (*forecasting*).

Models of GFD flows in meteorology, oceanography, and climate studies come in all types and sizes, depending on the geographical domain of interest (local, regional, continental, basinwide, or global) and the desired level of physical realism. Regional models are far too numerous to list here, and we only mention the existence of Atmospheric General Circulations Models (AGCMs), Oceanic General Circulation Models (OGCMs) and coupled General Circulation Models (GCMs). A truly comprehensive model does not exist because the coupling of air, sea, ice, and land physics over the entire planet is always open to the inclusion of yet one more process heretofore excluded from the model. In developing a numerical model of some GFD system, the question immediately

arises as to what actually needs to be simulated. The answer largely dictates the level of details necessary and, therefore also, the level of physical approximation and the degree of numerical resolution.

Geophysical flows are governed by a set of coupled, nonlinear equations in four-dimensional space-time and exhibit a high sensitivity to details. In mathematical terms, it is said that the system possesses chaotic properties, and the consequence is that geophysical flows are inherently unpredictable as Lorenz demonstrated for the atmosphere several decades ago (Lorenz, 1963). The physical reality is that geophysical fluid systems are replete with instabilities, which amplify in a finite time minor details into significant structures (the butterfly-causing-a-tempest syndrome). The cyclones and anticyclones of midlatitude weather and the meandering of the coastal currents are but a couple of examples among many. Needless to say, the simulation of atmospheric and oceanographic fluid motions is a highly challenging task.

The initial impetus for geophysical fluid simulations was, not surprisingly, weather prediction, an aspiration as old as mankind. More recently, climate studies have become another leading force in model development because of their need for extremely large and complex models.

The first decisive step in the quest for weather prediction was made by Vilhelm Bjerknes (1904) in a paper titled “The problem of weather prediction considered from the point of view of mechanics and physics.” He was the first to pose the problem as a set of time-dependent equations derived from physics and to be solved from a given, and hopefully complete, set of initial conditions. Bjerknes immediately faced the daunting task of integrating complicated partial differential equations, and, because this was well before electronic computers, resorted to graphical methods of solution. Unfortunately, these had little practical value and never surpassed the art of subjective forecasting by a trained person pouring over weather charts.

Taking a different tack, Lewis Fry Richardson (1922; see biography at the end of Chapter 14) decided that it would be better to reduce the differential equations to a set of arithmetic operations (additions, subtractions, multiplications, and divisions exclusively) so that a step-by-step method of solution may be followed and performed by people not necessarily trained in meteorology. Such reduction could be accomplished, he reasoned, by seeking the solution at only selected points in the domain and by approximating spatial derivatives of the unknown variables by finite differences across those points. Likewise, time could be divided into finite intervals, and temporal derivatives be approximated as differences across those time intervals, and thus was born numerical analysis. Richardson’s work culminated in his 1922 book entitled *Weather Prediction by Numerical Process*. His first grid, to forecast weather over western Europe, is reproduced here as [Fig. 1.8](#). After the equations of motion had been dissected into a sequence of individual arithmetic operations, the first algorithm before the word existed, computations were performed by a large group of people,

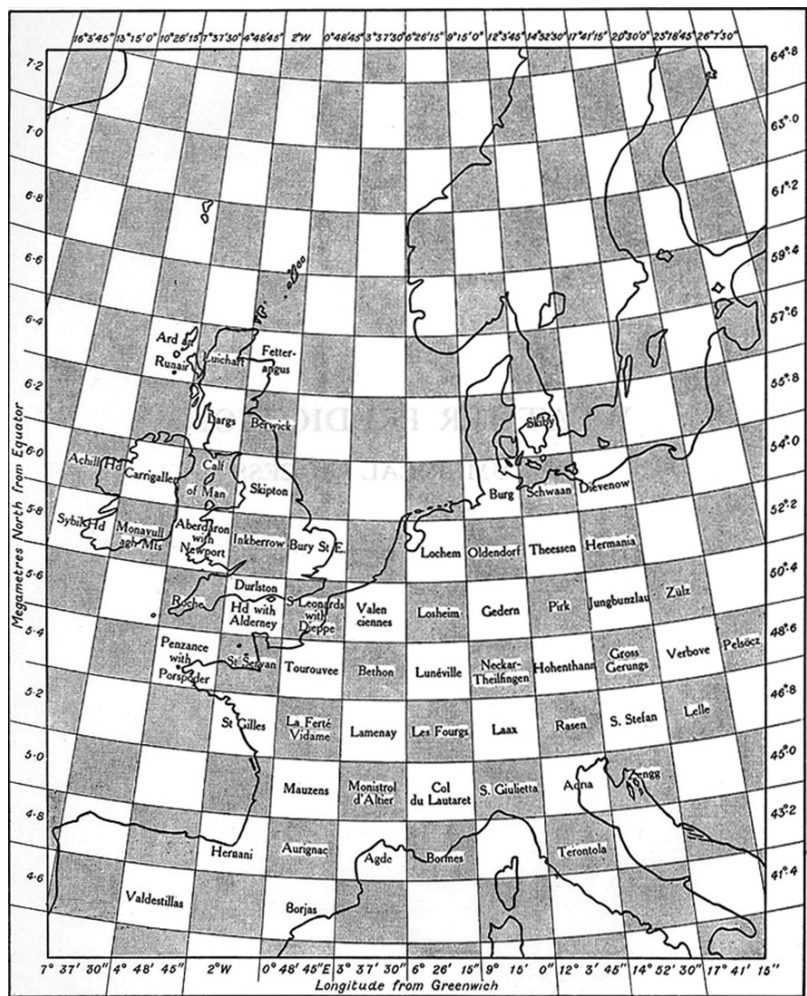


FIGURE 1.8 Model grid used by Lewis Fry Richardson as reported in his 1922 book *Weather Prediction by Numerical Process*. The grid was designed to optimize the fit between cells and existing meteorological stations, with observed surface pressures being used at the center of every shaded cell and winds at the center of every white cell.

called *computers*, sitting around an auditorium equipped with slide rules and passing their results to their neighbors. Synchronization was accomplished by a leader in the pit of the auditorium as a conductor leads an orchestra. Needless to say, the work was tedious and slow, requiring an impractically large number of people to conduct the calculations quickly enough so that a 24-h forecast could be obtained in less than 24 h.

Despite an enormous effort on Richardson's part, the enterprise was a failure, with predicted pressure variations rapidly drifting away from meteorologically acceptable values. In retrospective, we now know that Richardson's model was improperly initiated for lack of upper level data and that its 6-h time step was exceeding the limit required by numerical stability, of which, of course, he was not aware. The concept of numerical stability was not known until 1928 when it was elucidated by Richard Courant, Karl Friedrichs and Hans Lewy.

The work of Richardson was abandoned and relegated to the status of a curiosity or, as he put it himself, "a dream," only to be picked up again seriously at the advent of electronic computers. In the 1940s, the mathematician John von Neumann (see biography at end of Chapter 5) became interested in hydrodynamics and was seeking mathematical aids to solve nonlinear differential equations. Contact with Alan Turing, the inventor of the electronic computer, gave him the idea to build an automated electronic machine that could perform sequential calculations at a speed greatly surpassing that of humans. He collaborated with Howard Aiken at Harvard University, who built the first electronic calculator, named the Automatic Sequence Controlled Calculator (ASCC). In 1943, von Neumann helped build the Electronic Numerical Integrator and Computer (ENIAC) at the University of Pennsylvania and, in 1945, the Electronic Discrete Variable Calculator (EDVAC) at Princeton University. Primarily because of the wartime need for improved weather forecasts and also out of personal challenge, von Neumann paired with Jule Charney (see biography at end of Chapter 16) and selected weather forecasting as the scientific challenge. But, unlike Richardson before them, von Neumann and Charney started humbly with a much reduced set of dynamics, a single equation to predict the pressure at mid-level in the troposphere. The results (Charney, Fjörtoft & von Neumann, 1950) exceeded expectations.

Success with a much reduced set of dynamics only encouraged further development. Phillips (1956) developed a two-layer quasi-geostrophic² model over a hemispheric domain. The results did not predict actual weather but did behave like weather, with realistic cyclones generated at the wrong places and times. This was nonetheless quite encouraging. A major limitation of the quasi-geostrophic simplification is that it fails near the equator, and the only remedy was a return to the full equations (called *primitive equations*), back to where Richardson started. The main problem, it was found by then, is that primitive equations retain fast-moving gravity waves, and although these hold only a small amount of energy, their resolution demands both a much shorter time step of integration and a far better set of initial conditions than were available at the time.

From then on, the major intellectual challenges were overcome, and steady progress (Fig. 1.9) has been achieved, thanks to ever-faster and larger computers

² Quasi-geostrophic dynamics are described in Chapter 16. It suffices here to say that the formalism eliminates the velocity components under the assumption that rotational effects are very strong. The result is a drastic reduction in the number of equations.

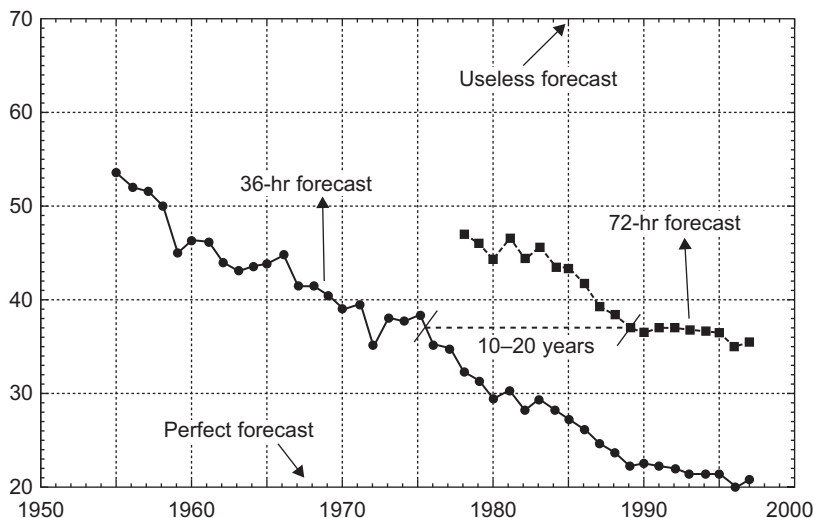


FIGURE 1.9 Historical improvement of weather forecasting skill over North America. The S1 score shown here is a measure of the relative error in the pressure gradient predictions at mid-height in the troposphere. (From Kalnay, Lord & McPherson, 1998, reproduction with the kind permission of the American Meteorological Society)

(Fig. 1.10) and to the gathering of an ever-denser array of data around the globe. The reader interested in the historical developments of weather forecasting will find an excellent book-length account in Nebeker (1995).

1.10 SCALES ANALYSIS AND FINITE DIFFERENCES

In the preceding section, we saw that computers are used to solve numerically equations otherwise difficult to apprehend. Yet, even with the latest supercomputers and unchanged physical laws, scientists are requesting more computer power than ever, and we may rightfully ask what is the root cause of this unquenchable demand. To answer, we introduce a simple numerical technique (*finite differences*) that shows the strong relationship between scale analysis and numerical requirement. It is a prototypical example foreshowing a characteristic of more elaborate numerical methods that will be introduced in later chapters for more realistic problems.

When performing a timescale analysis, we assume that a physical variable u changes significantly over a timescale T by a typical value U (Fig. 1.11). With this definition of scales, the time derivative is on the order of

$$\frac{du}{dt} \sim \frac{U}{T}. \quad (1.7)$$

If we then assume that the timescale over which the function u changes is also the one over which its derivative changes (in other words, we assume the

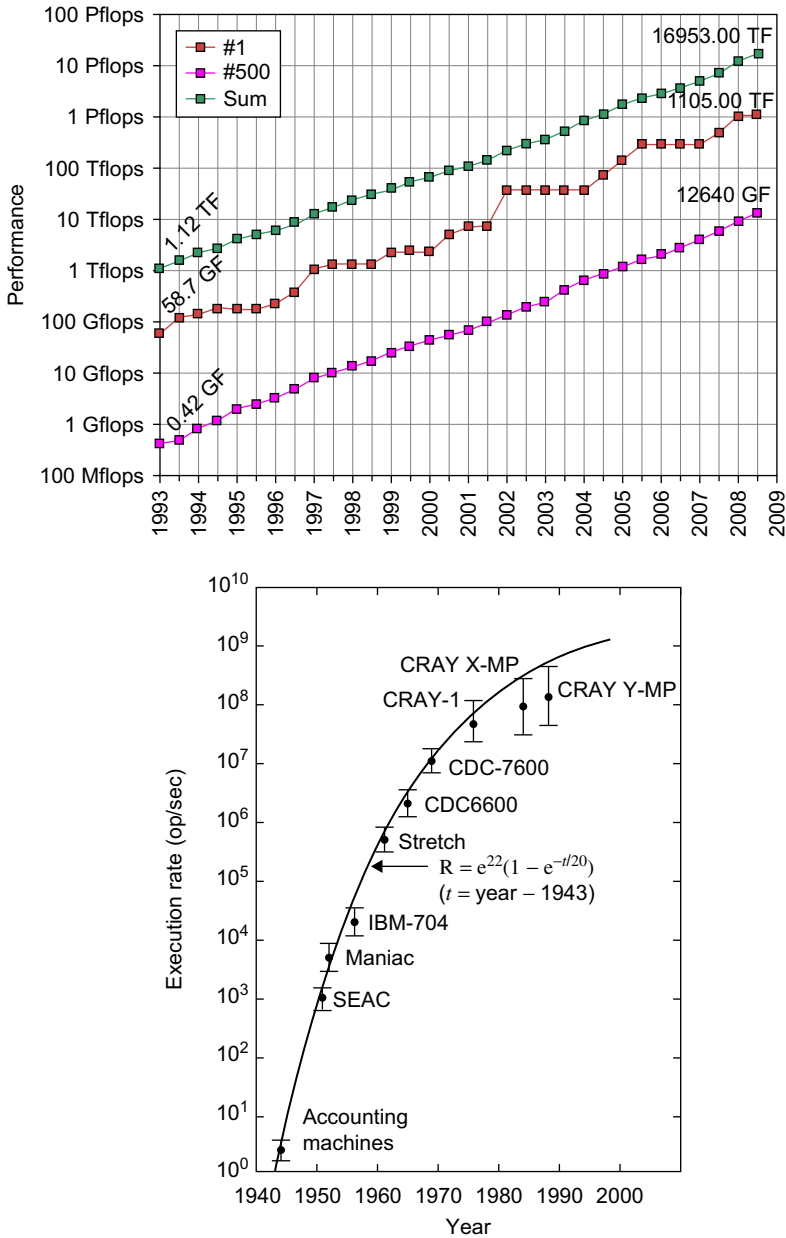


FIGURE 1.10 Historical increase of computational speed, as measured by the number of operations performed per second. (Adapted and supplemented from Hack (1992), who gives credit to a 1987 personal communication with Worlton (left panel) and with recent data from <http://www.top500.org>, upper panel)

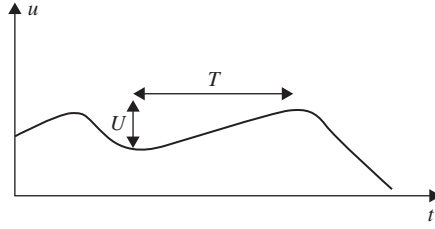


FIGURE 1.11 Timescale analysis of a variable u . The timescale T is the time interval over which the variable u exhibits variations comparable to its standard deviation U .

timescale T to be representative of all types of variabilities, including differentiated fields), we can also estimate the order of magnitude of variations of the second derivative

$$\frac{d^2u}{dt^2} = \frac{d}{dt} \left(\frac{du}{dt} \right) \sim \frac{U/T}{T} = \frac{U}{T^2}, \quad (1.8)$$

and so on for higher-order derivatives. This approach is the basis for estimating the relative importance of different terms in time-marching equations, an exercise we will repeat several times in the following chapters.

We now turn our attention to the question of estimating derivatives with more accuracy than by a mere order of magnitude. Typically, this problem arises on discretizing equations, a process by which all derivatives are replaced by algebraic approximations based on a few discrete values of the function u (Fig. 1.12). Such *discretization* is necessary because computers possess a finite memory and are incapable of manipulating derivatives. We then face the following problem: Having stored only a few values of the function, how can we retrieve the value of the function's derivatives that appear in the equations?

First, it is necessary to discretize the independent variable time t , since the first dynamical equations that we shall solve numerically are time-evolving equations. For simplicity, we shall suppose that the discrete time moments t^n , at which the function values are to be known, are uniformly distributed with a constant *time step* Δt

$$t^n = t^0 + n\Delta t, \quad n = 1, 2, \dots \quad (1.9)$$

where the superscript index (not an exponent) n identifies the discrete time. Then, we note by u^n the value of u at time t^n , that is, $u^n = u(t^n)$. We now would like to determine the value of the derivative du/dt at time t^n knowing only the discrete values u^n . From the definition of a derivative

$$\frac{du}{dt} = \lim_{\Delta t \rightarrow 0} \frac{u(t + \Delta t) - u(t)}{\Delta t}, \quad (1.10)$$

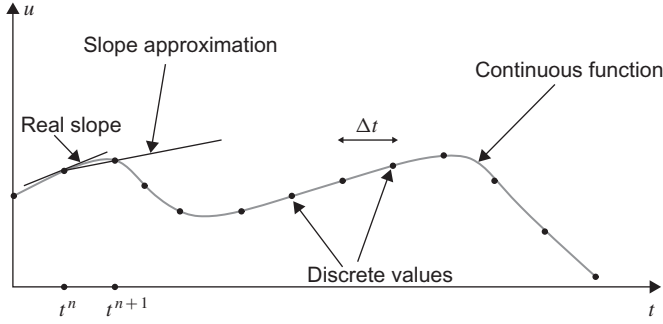


FIGURE 1.12 Representation of a function by a finite number of sampled values and approximation of a first derivative by a finite difference over Δt .

we could directly deduce an approximation by allowing Δt to remain the finite time step

$$\frac{du}{dt} \simeq \frac{u(t + \Delta t) - u(t)}{\Delta t} \rightarrow \left. \frac{du}{dt} \right|_{t^n} \simeq \frac{u^{n+1} - u^n}{\Delta t}. \quad (1.11)$$

The accuracy of this approximation can be determined with the help of a Taylor series:

$$u(t + \Delta t) = u(t) + \Delta t \left. \frac{du}{dt} \right|_t + \underbrace{\frac{\Delta t^2}{2} \left. \frac{d^2u}{dt^2} \right|_t}_{\Delta t^2 \frac{U}{T^2}} + \underbrace{\frac{\Delta t^3}{6} \left. \frac{d^3u}{dt^3} \right|_t}_{\Delta t^3 \frac{U}{T^3}} + \underbrace{\mathcal{O}(\Delta t^4)}_{\Delta t^4 \frac{U}{T^4}}. \quad (1.12)$$

To the leading order for small Δt , we obtain the following estimate

$$\frac{du}{dt} = \frac{u(t + \Delta t) - u(t)}{\Delta t} + \mathcal{O}\left(\frac{\Delta t}{T} \frac{U}{T}\right). \quad (1.13)$$

The *relative* error on the derivative (the difference between the finite-difference approximation and the actual derivative, divided by the scale U/T) is therefore of the order $\Delta t/T$. For the approximation to be acceptable, this relative error should be much smaller than 1, which demands that the time step Δt be sufficiently short compared to the timescale at hand:

$$\Delta t \ll T. \quad (1.14)$$

This condition can be visualized graphically by considering the effect of various values of Δt on the resulting estimation of the time derivative (Fig. 1.13). In the following, we write the formal approximation as

$$\left. \frac{du}{dt} \right|_{t^n} = \frac{u^{n+1} - u^n}{\Delta t} + \mathcal{O}(\Delta t), \quad (1.15)$$

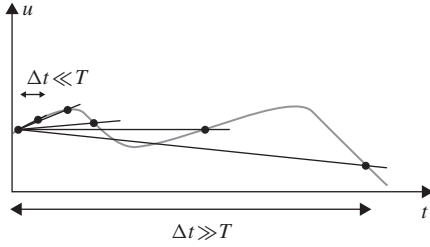


FIGURE 1.13 Finite differencing with various Δt values. Only when the time step is sufficiently short compared to the time scale, $\Delta t \ll T$, is the finite-difference slope close to the derivative, that is, the true slope.

where it is understood that the measure of whether or not Δt is “small enough” must be based on the timescale T of the variability of the variable u . Since in the simple finite difference (1.15), the error, called *truncation error*, is proportional to Δt , the approximation is said to be of first order. For an error proportional to Δt^2 , the approximation is said of second order and so on.

For spatial derivatives, the preceding analysis is easily applicable, and we obtain a condition on the horizontal grid size Δx relatively to the horizontal length scale L , while the vertical grid space Δz is constrained by the vertical length scale H of the variable under investigation:

$$\Delta x \ll L, \quad \Delta z \ll H. \quad (1.16)$$

With these constraints on time steps and grid sizes, we can begin to understand the need for significant computer resources in GFD simulations: The number of grid points M in a 3D domain of surface \mathcal{S} and height H is

$$M = \frac{H}{\Delta z} \frac{\mathcal{S}}{\Delta x^2}, \quad (1.17)$$

while the total number of time steps N needed to cover a time period P is

$$N = \frac{P}{\Delta t}. \quad (1.18)$$

For a model covering the Atlantic Ocean ($\mathcal{S} \sim 10^{14} \text{ m}^2$), resolving geostrophic eddies (see Fig. 1.7: $\Delta x \sim \Delta y \leq 10^4 \text{ m}$) and stratified water masses ($H/\Delta z \sim 50$), the number of grid points is about $M \sim 5 \times 10^7$. Then, at each of these points, several variables need to be stored and calculated (three-dimensional velocity, pressure, temperature, etc.). Since each variable takes 4 or 8 bytes of memory depending on the desired number of significant digits, 2 Gigabytes of RAM is required. The number of floating point operations to be executed to simulate a single year can be estimated by taking a time step resolving the rotational period of Earth $\Delta t \sim 10^3 \text{ s}$, leading to $N \sim 30000$ time steps. The total number of operations to simulate a full year can then be estimated by observing that for every grid point and time step, a series of calculations must be performed (typically several hundreds) so that the total number of calculations

amounts to 10^{14} – 10^{15} . Therefore, on a contemporary supercomputer (one of the top 500 machines) with 1 Teraflops = 10^{12} floating operations per second exclusively dedicated to the simulation, less than half an hour would pass before the response is available, whereas on a PC delivering 1–2 Gigaflips, we would need to wait several days before getting our results. And yet, even with such a large model, we can only resolve the largest scales of motion (see Fig. 1.7) while motions on shorter spatial and temporal scales simply cannot be simulated with this level of grid resolution. However, this does not mean that those shorter scale motions may altogether be neglected and, as we will see (e.g., Chapter 14), one of the problems of large-scale oceanic and atmospheric models is the need for appropriate *parameterization* of shorter scale motions so that they may properly bear their effects onto the larger scale motions.

Should we dream to avoid such a parameterization by explicitly calculating all scales, we would need about $M \sim 10^{24}$ grid points demanding 5×10^{16} Gigabytes of computer memory and $N \sim 3 \times 10^7$ time steps, for a total number of operations on the order of 10^{34} . Willing to wait only for 10^6 s before obtaining the results, we would need a computer delivering 10^{28} flops. This is a factor $10^{16} = 2^{53}$ higher than the present capabilities, both for speed and memory requirements. Using Moore's Law, the celebrated rule that forecasts a factor 2 in gain of computing power every 18 months, we would have to wait 53 times 18 months, that is, for about 80 years before computers could handle such a task.

Increasing resolution will therefore continue to call for the most powerful computers available, and models will need to include parameterization of turbulence or other unresolved motions for quite some time. Grid spacing will thus remain a crucial aspect of all GFD models, simply because of the large domain sizes and broad range of scales.

1.11 HIGHER-ORDER METHODS

Rather than to increase resolution to better represent structures, we may wonder whether using other approximations for derivatives than our simple finite difference (1.11) would allow larger time steps or higher quality approximations and improved model results. Based on a Taylor series

$$u^{n+1} = u^n + \Delta t \left. \frac{du}{dt} \right|_{t^n} + \frac{\Delta t^2}{2} \left. \frac{d^2u}{dt^2} \right|_{t^n} + \frac{\Delta t^3}{6} \left. \frac{d^3u}{dt^3} \right|_{t^n} + \mathcal{O}(\Delta t^4) \quad (1.19)$$

$$u^{n-1} = u^n - \Delta t \left. \frac{du}{dt} \right|_{t^n} + \frac{\Delta t^2}{2} \left. \frac{d^2u}{dt^2} \right|_{t^n} - \frac{\Delta t^3}{6} \left. \frac{d^3u}{dt^3} \right|_{t^n} + \mathcal{O}(\Delta t^4), \quad (1.20)$$

we can imagine that instead of using a *forward-difference* approximation of the time derivative (1.11), we try a backward Taylor series (1.20) to design

a *backward-difference* approximation. This approximation is obviously still of first order because of its truncation error:

$$\left. \frac{du}{dt} \right|_{t^n} = \frac{u^n - u^{n-1}}{\Delta t} + \mathcal{O}(\Delta t). \quad (1.21)$$

Comparing Eq. (1.19) with Eq. (1.20), we observe that the truncation errors of the first-order forward and backward finite differences are the same but have opposite signs so that by averaging both, we obtain a second-order truncation error (you can verify this statement by taking the difference between Eqs. (1.19) and (1.20)):

$$\left. \frac{du}{dt} \right|_{t^n} = \frac{u^{n+1} - u^{n-1}}{2\Delta t} + \mathcal{O}(\Delta t^2). \quad (1.22)$$

Before considering higher-order approximations, let us first check whether the increase in order of approximation actually leads to improved approximations of the derivatives. To do so, consider the sinusoidal function of period T (and associated frequency ω)

$$u = U \sin\left(2\pi \frac{t}{T}\right) = U \sin(\omega t), \quad \omega = \frac{2\pi}{T}. \quad (1.23)$$

Knowing that the exact derivative is $\omega U \cos(\omega t)$, we can calculate the errors made by the various finite-difference approximations (Fig. 1.14). Both the forward and backward finite differences converge toward the exact value for $\omega \Delta t \rightarrow 0$, with errors decreasing proportionally to Δt . As expected, the second-order approximation (1.22) exhibits a second-order convergence (the slope is 2 in a log–log graph).

The convergence rate obeys our theoretical estimate for $\omega \Delta t \ll 1$. However, when the time step is relatively large (Fig. 1.15), the error associated with the finite-difference approximations can be as large as the derivative itself. For coarse resolution, $\omega \Delta t \sim \mathcal{O}(1)$, the relative error is of order 1 so that we expect a 100% error on the finite-difference approximation. Obviously, even with a second-order finite difference, we need at least $\omega \Delta t \leq 0.8$ to keep the relative error below 10%. In terms of the period of the signal $T = (2\pi)/\omega$, we need a time step not larger than $\Delta t \lesssim T/8$, which implies that 8 points are needed along one period to resolve its derivatives within a 10% error level. Even a fourth-order method (to be shown shortly) cannot reconstruct derivatives correctly from a function sampled with fewer than several points per period.

The design of the second-order difference was accomplished simply by inspection of a Taylor series, a technique that cannot be extended to obtain higher-order approximations. An alternate method exists to obtain in a systematic way finite-difference approximations to any desired order, and it can be

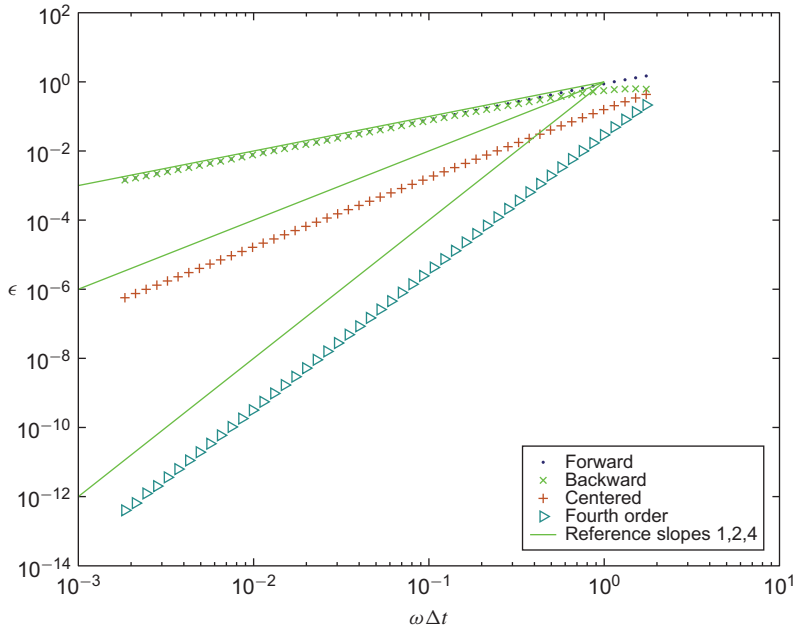


FIGURE 1.14 Relative error ϵ of various finite-difference approximations of the first derivative of the sinusoidal function as function of $\omega\Delta t$ when $\omega t = 1$. Scales are logarithmic and continuous lines of slope 1, 2, and 4 are added. First-order methods have a slope of 1, and the second-order method a slope of 2. The error behaves as expected for decreasing Δt .

illustrated with the design of a fourth-order centered finite-difference approximation of the first derivative. Expecting that higher-order approximations need more information about a function in order to estimate its derivative at time t^n , we will combine values over a longer time interval, including t^{n-2} , t^{n-1} , t^n , t^{n+1} , and t^{n+2} :

$$\left. \frac{du}{dt} \right|_{t^n} \simeq a_{-2}u^{n-2} + a_{-1}u^{n-1} + a_0u^n + a_1u^{n+1} + a_2u^{n+2}. \quad (1.24)$$

Expanding u^{n+2} and the other values around t^n by Taylor series, we can write

$$\begin{aligned} \left. \frac{du}{dt} \right|_{t^n} &= (a_{-2} + a_{-1} + a_0 + a_1 + a_2) u^n \\ &\quad + (-2a_{-2} - a_{-1} + a_1 + 2a_2) \Delta t \left. \frac{du}{dt} \right|_{t^n} \\ &\quad + (4a_{-2} + a_{-1} + a_1 + 4a_2) \frac{\Delta t^2}{2} \left. \frac{d^2u}{dt^2} \right|_{t^n} \end{aligned}$$

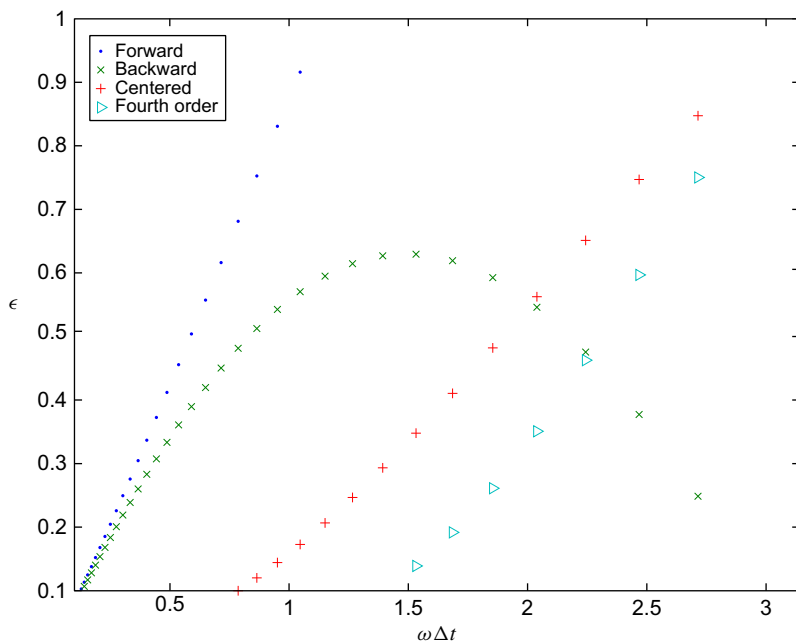


FIGURE 1.15 Relative error ϵ of various finite-difference approximations of the first derivative of the sinusoidal function as function of $\omega \Delta t$ when $\omega t = 1$. For coarse resolution $\omega \Delta t \sim \mathcal{O}(1)$, the relative error is of order 1 so that we expect a 100% error on the finite-difference approximation.

$$\begin{aligned}
 & + (-8a_{-2} - a_{-1} + a_1 + 8a_2) \frac{\Delta t^3}{6} \left. \frac{d^3 u}{dt^3} \right|_t^n \\
 & + (16a_{-2} + a_{-1} + a_1 + 16a_2) \frac{\Delta t^4}{24} \left. \frac{d^4 u}{dt^4} \right|_t^n \\
 & + (-32a_{-2} - a_{-1} + a_1 + 32a_2) \frac{\Delta t^5}{120} \left. \frac{d^5 u}{dt^5} \right|_t^n \\
 & + \mathcal{O}(\Delta t^6).
 \end{aligned} \tag{1.25}$$

There are five coefficients, a_{-2} to a_2 , to be determined. Two conditions must be satisfied to obtain an approximation that tends to be the first derivative as $\Delta t \rightarrow 0$:

$$\begin{aligned}
 a_{-2} + a_{-1} + a_0 + a_1 + a_2 &= 0, \\
 (-2a_{-2} - a_{-1} + a_1 + 2a_2) \Delta t &= 1.
 \end{aligned}$$

After satisfying these two necessary conditions, we have three parameters that can be freely chosen so as to obtain the highest possible level of accuracy. This

is achieved by imposing that the coefficients of the next three truncation errors be zero:

$$\begin{aligned} 4a_{-2} + a_{-1} + a_1 + 4a_2 &= 0 \\ -8a_{-2} - a_{-1} + a_1 + 8a_2 &= 0 \\ 16a_{-2} + a_{-1} + a_1 + 16a_2 &= 0. \end{aligned}$$

Equipped with five equations for five unknowns, we can proceed with the solution:

$$-a_{-1} = a_1 = \frac{8}{12\Delta t}, \quad a_0 = 0, \quad -a_{-2} = a_2 = -\frac{1}{12\Delta t},$$

so that the fourth-order finite-difference approximation of the first derivative is

$$\left. \frac{du}{dt} \right|_{t^n} \simeq \frac{4}{3} \left(\frac{u^{n+1} - u^{n-1}}{2\Delta t} \right) - \frac{1}{3} \left(\frac{u^{n+2} - u^{n-2}}{4\Delta t} \right). \quad (1.26)$$

This formula can be interpreted as a linear combination of two centered differences, one across $2\Delta t$ and the other across $4\Delta t$. The truncation error can be assessed by looking at the next term in the series (1.25)

$$(-32a_{-2} - a_{-1} + a_1 + 32a_2) \frac{\Delta t^5}{120} \left. \frac{d^5 u}{dt^5} \right|_{t^n} = -\frac{\Delta t^4}{30} \left. \frac{d^5 u}{dt^5} \right|_{t^n}, \quad (1.27)$$

which shows that the approximation is indeed of fourth order.

The method can be generalized to approximate a derivative of any order p at time t^n using the current value u^n , m points in the past (before t^n) and m points in the future (after t^n):

$$\left. \frac{d^p u}{dt^p} \right|_{t_n} = a_{-m} u^{n-m} + \dots + a_{-1} u^{n-1} + a_0 u^n + a_1 u^{n+1} + \dots + a_m u^{n+m}. \quad (1.28)$$

The discrete points $n-m$ to $n+m$ involved in the approximation define the so-called *numerical stencil* of the operator. Using a Taylor expansion for each term

$$u^{n+q} = u^n + q\Delta t \left. \frac{du}{dt} \right|_{t^n} + q^2 \frac{\Delta t^2}{2} \left. \frac{d^2 u}{dt^2} \right|_{t^n} + \dots + q^p \frac{\Delta t^p}{p!} \left. \frac{d^p u}{dt^p} \right|_{t^n} + \mathcal{O}(\Delta t^{p+1}) \quad (1.29)$$

and injecting Eq. (1.29) for $q = -m, \dots, m$ into the approximation (1.28), we have on the left-hand side the derivative we want to approximate and on the right a sum of derivatives. We impose that the sum of coefficients multiplying a derivative lower than order p be zero, whereas the sum of the coefficients multiplying the p th derivative be 1. This forms a set of $p+1$ equations for the $2m+1$ unknown coefficients a_q ($q = -m, \dots, m$). All constraints can be satisfied simultaneously only if we use a number $2m+1$ of points equal to or greater than

$p+1$, that is, $2m \geq p$. When there are more points than necessary, we can take advantage of the remaining degrees of freedom to cancel the next few terms in the truncation errors. With $2m+1$ points, we can then obtain a finite difference of order $2m-p+1$. For example, with $m=1$ and $p=1$, we obtained Eq. (1.22), a second-order approximation of the first derivative, and with $m=2$ and $p=1$, Eq. (1.26), a fourth-order approximation.

Let us now turn to the second derivative, a very common occurrence, at least when considering spatial derivatives. With $p=2$, m must be at least 1, that is, three values of the function are required as a minimum: one old, one current, and one future values. Applying the preceding method, we immediately obtain

$$\left. \frac{d^2 u}{dt^2} \right|_{t_n} \simeq \left(\frac{u^{n-1} - 2u^n + u^{n+1}}{\Delta t^2} \right), \quad (1.30)$$

a result we could also have obtained by direct inspection of Eqs. (1.19) and (1.20).

Appendix C recapitulates a variety of discretization schemes for different orders of derivatives and various levels of accuracy. It also includes skewed schemes, which are not symmetric between past and future values but can be constructed in a way similar to the fourth-order finite-difference approximation of the first derivative.

1.12 ALIASING

We learned that the accuracy of a finite-difference approximation of the first derivative degrades rapidly when the time step Δt is not kept much shorter than the timescale T of the variable, and we might wonder what would happen if Δt should by some misfortune be larger than T . To answer this question, we return to a physical signal u of period T ,

$$u = U \sin(\omega t + \phi), \quad \omega = \frac{2\pi}{T}, \quad (1.31)$$

sampled on a uniform grid of time step Δt ,

$$u^n = U \sin(n \omega \Delta t + \phi), \quad (1.32)$$

and assume that there exists another signal v of higher frequency $\tilde{\omega}$ given by

$$v = U \sin(\tilde{\omega} t + \phi), \quad \tilde{\omega} = \omega + \frac{2\pi}{\Delta t}. \quad (1.33)$$

The sampling of this other function at the same time intervals yields a discrete set of values

$$v^n = U \sin(n \tilde{\omega} \Delta t + \phi) = U \sin(n \omega \Delta t + 2n\pi + \phi) = u^n, \quad (1.34)$$

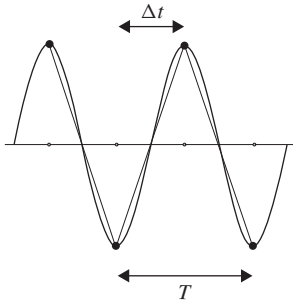


FIGURE 1.16 Shortest wave (at cutoff frequency $\pi/\Delta t$ or period $2\Delta t$) resolved by uniform grid in time.

which cannot be distinguished from the discrete values u^n of the first signal although the two signals are clearly not equal to each other. Thus, frequencies ω and $\omega + 2\pi/\Delta t$ cannot be separated in a sampling with time interval Δt because the higher-frequency signal masquerades as the lower-frequency signal. This unavoidable consequence of sampling is called *aliasing*.

Since signals of frequency $\omega + 2\pi/\Delta t$ and ω cannot be distinguished from each other, it appears that only frequencies within the following range

$$-\frac{\pi}{\Delta t} \leq \omega \leq \frac{\pi}{\Delta t} \quad (1.35)$$

can be recognized with a sampling interval Δt , and all other frequencies should preferably be absent, lest they contaminate the sampling process.

Since a negative frequency corresponds to a 180° phase shift, because $\sin(-\omega t + \phi) = \sin(\omega t - \phi + \pi)$, the useful range is actually $0 \leq \omega \leq \pi/\Delta t$, and to sample a wave of frequency ω , the time step Δt may not exceed $\Delta t_{\max} = \pi/\omega = T/2$, which implies that at least two samples of the signal must be taken per period. This minimum required sampling frequency is called the *Nyquist frequency*. Looking at the problem in a different way, with a given sampling interval Δt (rather than a given frequency), we recognize that the highest resolved frequency is $\pi/\Delta t$, called the *cutoff frequency* (Fig. 1.16).

Should higher frequencies be present and sampled, aliasing inevitably occurs, as illustrated by a sinusoidal function sampled with increasingly fewer points per period (Fig. 1.17). The reader is invited to experiment with MATLAB™ script `aliasanim.m`. Up to $\Delta t = T/2$, the signal is recognizable, but, beyond that, lines connecting consecutive sampled values appear to tunnel through crests and troughs, giving the impression of a signal with longer period.

Aliasing is a major concern, and the danger it poses is often underestimated. This is because we do not know whether the signal being represented by the discretization scheme contains frequencies higher than the cutoff frequency, precisely because variability at those frequencies is not retained and computed. In geophysical situations, the time step and grid spacing are most often set not by the physics of the problem but by computer-hardware limits. This forces the modeler to discard variability at unresolved frequencies and wavelengths and

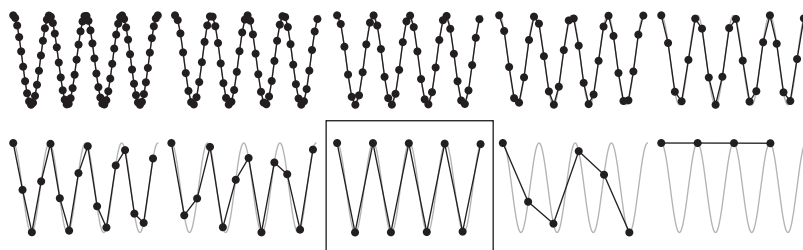


FIGURE 1.17 Aliasing illustrated by sampling a given signal (gray sinusoidal curve) with an increasing time interval. A high sampling rate (top row of images) resolves the signal properly. The boxed image on the bottom row corresponds to the cutoff frequency, and the sampled signal appears as a seesaw. The last two images correspond to excessively long time intervals that alias the signal, making it appear as if it had a longer period than it actually has.

creates aliasing. Methods to overcome the undesired effects of aliasing will be presented in subsequent chapters.

ANALYTICAL PROBLEMS

- 1.1. Name three naturally occurring flows in the atmosphere.
- 1.2. How did geophysical flows contribute to Christopher Columbus' discovery of the New World and to the subsequent exploration of the eastern shore of North America? (Think of both large-scale winds and major ocean currents.)
- 1.3. The sea breeze is a light wind blowing from the sea as the result of a temperature difference between land and sea. As this temperature difference reverses from day to night, the daytime sea breeze turns into a nighttime land breeze. If you were to construct a numerical model of the sea–land breeze, should you include the effects of the planetary rotation?
- 1.4. The Great Red Spot of Jupiter, centered at 22°S and spanning 12° in latitude and 25° in longitude, exhibits wind speeds of about 100 m/s. The planet's equatorial radius and rotation rates are, respectively, 71,400 km and $1.763 \times 10^{-4} \text{ s}^{-1}$. Is the Great Red Spot influenced by planetary rotation?
- 1.5. Can you think of a technique for measuring wind speeds and ocean velocities with an instrument that has no rotating component? (*Hint:* Think of measurable quantities whose values are affected by translation.)

NUMERICAL EXERCISES

- 1.1. Using the temperature measurements of Nansen (Fig. 1.18 left), estimate typical vertical temperature gradient values and typical temperature values.

Compare those values to the estimates based on a Mediterranean profile (Fig. 1.18 right). Which temperature scale do you need for the estimation of gradients in each case?

- 1.2. Perform a numerical differentiation of $e^{-\omega(t+|t|)}$ using $\omega\Delta t = 0.1, 0.01, 0.001$. Compare the first-order forward, first-order backward, and second-order centered schemes at $t = 1/\omega$. Then, repeat the derivation with $t = 0.000001/\omega$ and compare. What do you conclude?
- 1.3. Apply forward, backward, second-order, and fourth-order discretizations to $\sinh(kx)$ at $x = 1/k$ for values of $k\Delta x$ covering the range $[10^{-4}, 1]$. Plot errors on a logarithmic scale and verify the convergence rates. Repeat the exercise for $x = 0$. What strange effect do you observe and why?
- 1.4. Establish a purely forward, finite-difference approximation of a first derivative that is of second or higher-order in accuracy. How many sampling points are required as a function of the order of accuracy?
- 1.5. Suppose you need to evaluate $\partial u/\partial x$ not at grid node i , but at mid-distance between nodes $x_i = i\Delta x$ and $x_{i+1} = (i+1)\Delta x$. Establish second-order and fourth-order finite-difference approximations to do so and compare the truncation errors to the corresponding discretizations centered on the nodal point i . What does this analysis suggest?
- 1.6. Assume that a spatial two-dimensional domain is covered by a uniform grid with spacing Δx in the x -direction and Δy in the y -direction. How can you discretize $\partial^2 u/\partial x \partial y$ to second order? Does the approximation satisfy a similar property as its mathematical counterpart $\partial^2 u/\partial x \partial y = \partial^2 u/\partial y \partial x$?
- 1.7. Determine how a wave of wavelength $\frac{4}{3}\Delta x$ and period $\frac{5}{3}\Delta t$ is interpreted in a uniform grid of mesh Δx and time step Δt . How does the computed propagation speed resulting from the discrete sampling compare to the true speed?

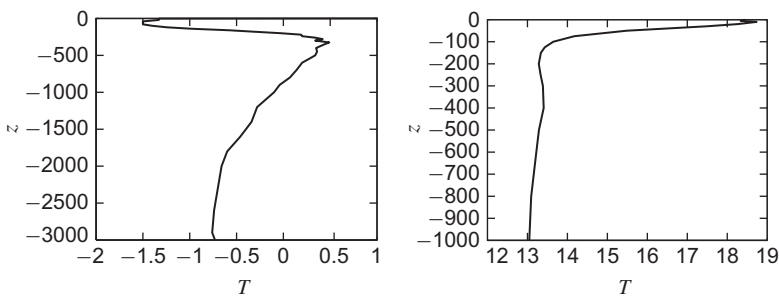


FIGURE 1.18 Temperature values measured by Nansen during his 1894 North Pole expedition (left) and a typical temperature profile in the Mediterranean Sea (from Medar© data base, right).

- 1.8.** Suppose you use numerical finite differencing to estimate derivatives of a function that was sampled with some noise. Assuming the noise is uncorrelated (i.e., purely randomly distributed, independently of the sampling interval), what do you expect would happen during the finite differencing of first-order derivatives, second derivatives etc.? Devise a numerical program that verifies your assertion, by adding a random noise of intensity $10^{-5}A$ to the function $A \sin(\omega t)$, where the frequency ω is well resolved by the numerical sampling ($\omega \Delta t = 0.05$). Did you correctly guess what would happen? Now plot the convergence as a function of $\omega \Delta t$ for noise levels of $10^{-5}A$ and $10^{-4}A$.

Walsh Cottage, Woods Hole, Massachusetts 1962–present



Every summer since 1962, this unassuming building of the Woods Hole Oceanographic Institution (Falmouth, Massachusetts, USA) has been home to the Geophysical Fluid Dynamics Summer Program, which has gathered oceanographers, meteorologists, physicists, and mathematicians from around the world. This program (begun in 1959) has single-handedly been responsible for many of the developments of geophysical fluid dynamics, from its humble beginnings to its present status as a recognized discipline in physical sciences. *(Drawing by Ryuji Kimura, reproduced with permission)*

UK Meteorological Office, Exeter, England 1854–present



The Meteorological Office in the United Kingdom was established in 1854 to provide meteorological information and sea currents by telegraph to people at sea and, within a few years, began also to issue storm warnings to seaports and weather forecasts to the press. In 1920, separate meteorological military services established during World War I were merged with the civilian office under the Air Ministry. World War II led to another large increase in both personnel and resources, including the use of balloons. The “Met Office” has an outstanding record of capitalizing on new technologies, beginning broadcasts on radio (in 1922) and then on television (in 1936 by means of simple captions and live broadcasts in 1954). The first electronic computer was installed in 1962, and satellite imagery was incorporated in 1964.

As weather forecasts began to depend less on trained meteorologists drawing weather maps and more on computational models, the need for the latest and best performing computer platform became a driving force, leading to the acquisition of a Cyber supercomputer in 1981 and a series of ever faster Cray supercomputers in the 1990s.

The impact of the Met Office can hardly be underestimated: Its numerical activities have contributed enormously not only to the field of meteorology but also to the development of computational fluid dynamics and physical oceanography, while the scope of its data analyses and forecasts has spread well beyond tomorrow’s weather to other areas such as the impact of the weather on the environment and human health. (*For additional information, see <http://www.metoffice.gov.uk/about-us/who/our-history>*)