Physics of Earth’s Climate
Physics of Earth's Climate

Tapio Schneider and Lenka Novak
This is a draft of what eventually is to become a book. You are welcome to use it. I would appreciate hearing about any comments and suggestions you have. (Some figures will only render correctly in Adobe Acrobat but not, for example, in Preview.)

— Tapio Schneider (tapio@caltech.edu)
The fact that we live at the bottom of a deep gravity well, on the surface of a gas covered planet going around a nuclear fireball 90 million miles away and think this to be normal is obviously some indication of how skewed our perspective tends to be.

Contents

Preface xi

1 Phenomena 1
  1.1 Temperature ............................................... 2
  1.2 Wind ....................................................... 20
  1.3 Water ..................................................... 27
  1.4 Oceans .................................................... 42
  1.5 Cryosphere ............................................... 46
  1.6 Land ...................................................... 47
  1.7 Greenhouse gases and aerosols ......................... 47

2 Radiative Energy 50
  2.1 Solar Radiation ........................................... 50
  2.2 Bond Albedo ............................................... 54
  2.3 Terrestrial Radiation .................................... 56
  2.4 Blackbody Radiation .................................... 58
  2.5 Effective Temperature ................................... 69
  2.6 Greenhouse Effect: Arrhenius Model .................... 72

3 Earth’s Orbit and Insolation 78
  3.1 Orbital Geometry ......................................... 78
  3.2 Orbital Motion ............................................ 81
  3.3 Solar Radiative Energy Flux ............................. 86
  3.4 Insolation ................................................ 88
  3.5 Milankovitch Cycles ..................................... 95
  3.6 Ice Ages and Orbitally Paced Climate Variations .... 103

4 Radiation and Matter 108
  4.1 Absorption and Emission ................................ 108
  4.2 Earth’s Primary Absorbers ............................... 111
  4.3 Absorption Line Shapes .................................. 113
  4.4 Absorption Spectrum ..................................... 116
  4.5 Scattering ................................................ 117
  4.6 Earth’s Albedo ............................................ 120

5 Radiative Transfer 126
## CONTENTS

5.1 Radiance ............................................. 126
5.2 Exponential Attenuation Law ...................... 126
5.3 Radiative Heating Rate ............................. 128
5.4 Schwarzschild Equation ............................ 130
5.5 Two-Stream Approximation ......................... 132
5.6 Radiative Equilibrium .............................. 133
5.7 Arrhenius Model With Clouds ....................... 135
5.8 Cloud Radiative Effects on Earth ................. 137

6 Radiative-Convective Equilibrium ................. 141
6.1 Static Stability ..................................... 141
6.2 RCE Models ......................................... 144
6.3 Atmospheric Thermal Structure .................... 146
6.4 Greenhouse Effect .................................. 146
6.5 Tropopause Height .................................. 148
6.6 Simpson's Paradox ................................ 150

7 Energy Balances and Temperatures ............... 152
7.1 Global Energy Balance .............................. 152
7.2 Surface Energy Balance ............................ 154
7.3 Top of Atmosphere Energy Balance ................ 160
7.4 Forms of Energy .................................... 163
7.5 Mean Flows and Eddies ............................. 165
7.6 Atmospheric Energy Transport ...................... 166
7.7 Oceanic Energy Transport .......................... 170
7.8 Energy Balance Models ............................. 170

8 Winds and Angular Momentum Balance ........... 174
8.1 Total Angular Momentum ........................... 174
8.2 Planetary Angular Momentum ....................... 176
8.3 Relative Angular Momentum ......................... 177
8.4 Regional Balances in the Atmosphere ............... 183
8.5 Regional Balances in the Ocean ...................... 195
8.6 Global Momentum Transport Cycle ................. 197
8.7 Appendix / Supplementary Material (move to the end of the book) ............... 197
8.8 Ocean .................................................. 200

9 Hydrologic Cycle, Precipitation, and Evaporation . 202
9.1 Water Reservoirs ................................... 202
9.2 Basic Equation of Hydrology ....................... 203
9.3 Water Transport ...................................... 203
9.4 Clausius-Clapeyron Relation ......................... 203
9.5 Humidity Variables .................................. 203
9.6 Climate Changes ..................................... 203
9.7 Water Isotopes ....................................... 208
CONTENTS

10 Clouds, Convection, and Boundary Layers 209
Bibliography 213
Index 219
Index 219
Preface

The climate system is first and foremost a physical system. The oldest questions that have been asked about it are of physical nature: What controls the surface temperatures and winds? What shapes rainfall patterns? Where and when do clouds form? How are ocean waters set into motion? How do glaciers and ice sheets advance and retreat? Such questions are among the first scientific questions humans have asked, going back at least to Aristotle’s *Meteorology* written 2300 years ago. Given the astounding progress the physical sciences have made on the smallest and largest scales—from quantum mechanics that gave us the transistor and digital technologies, to the theory of relativity that fixed our vantage point in an expanding universe—it may come as a surprise that we still do not have complete answers to such basic questions.

We do know the fundamental physical laws governing climate. Quantum mechanics and classical electrodynamics govern how radiative energy propagates from the Sun to the Earth and how it is absorbed, scattered, and re-emitted within the climate system, eventually making its way back to space. Classical mechanics and thermodynamics govern how radiative energy is converted into kinetic energy of air and water motion, and how it drives evaporation, water vapor transport in the atmosphere, and rainfall. We do know the macroscopic physical balances the climate system must satisfy so that its energy, angular momentum, and water balances are maintained. What we do not always know is what other climate states satisfy these balances in circumstances other than those we currently observe on Earth. Were winds weaker or stronger during the ice ages? Will there be more or fewer clouds as the atmospheric concentration of greenhouse gases increases and the climate warms? We do not have complete answers to such questions because answering them requires an understanding of complex interactions within the climate system as a whole. Its subcomponents interact nonlinearly through multiple physical processes and across a vast range of scales, from the micrometers of rain droplets to the thousands of kilometers of atmosphere and ocean circulations spanning the globe. Lack of detailed data that is continuous across this range of scales, and our inability to perform controlled experiments with the climate system, have inhibited scientific progress so far. However, thanks primarily to space-based remote sensing and increasingly also to autonomous vehicles exploring the oceans, we now are blessed with a trove of detailed data on the global climate system. And even though simulating all relevant scales and their interactions globally will remain out of reach, our ability to perform numerical experiments with com-
puter models of the climate system and its subcomponents is maturing with continually increasing computing power. Ours likely is the age in which the laws that govern climate as an aggregate system will be discovered.

This book is meant as a guide to what currently is known about the physics of the climate system. It begins with a survey of the climate phenomena that call for an explanation, based on the best data that are available. It goes on to present an overview of the physical processes that shape the climate system: from radiative transfer to the dynamics of clouds and global atmosphere and ocean circulations. The emphasis throughout is on the physical balances—of energy, angular momentum, and water vapor—that must be satisfied, and on illustrative models and order-of-magnitude reasoning of how such balances can be maintained. What we hope readers will take from it is an appreciation of how far one can go in making inferences about climate and possible climate changes without running computer models, and how far we still have to go before we know the laws governing climate as an aggregate system.

While the focus is on Earth’s climate, the physical balances discussed apply equally well to other planetary climates. We mention other planetary climates along the way where that can broaden the perspective, for example, on what kind of wind belts can occur on planets of different sizes and different rotation rates.

Readers are expected to have a solid grounding in standard undergraduate physics but are not expected to have knowledge of more specialized areas needed to understand Earth’s climate, such as fluid dynamics. We envision this book as the first in a series, as a survey that introduces themes to be developed in greater depth in subsequent books, for example, on the dynamics of global atmospheric circulations and of clouds and boundary layers.

Acknowledgments

Momme Hell and Davide Panosetti produced most figures in this book, and Thomas Müller, Philip Nelson, and Clare Singer provided helpful comments on the manuscript.
Chapter One

Phenomena

We begin with the observed phenomena—the collection of facts about Earth’s climate system that call for an explanation. We focus on those phenomena that have a direct bearing on Earth’s near-surface climate, which is what we experience and commonly refer to as “the climate.”

Climate is average weather, or more generally, a collection of long-term, slowly varying weather statistics. These statistics include mean values, variances, and covariances, for example, of winds, temperatures, and humidities. Usually, weather statistics are accumulated over a few decades, thirty years being the traditional standard. But they can and do fluctuate on longer timescales: that is, climate varies. For the past several decades, an abundance of weather data is available from which detailed statistics can be accumulated. The atmosphere is the best observed planetary fluid system, thanks to globe-spanning initiatives that began in the midst of the cold war. During the International Geophysical Year (IGY) in 1957/58, countries in east and west came together in a coordinated campaign to measure the state of the Earth and its nearby surroundings in space in unprecedented detail. The measurements included meteorological observations that were coordinated globally. Upper-air observations with radiosondes—weather balloons equipped with radio transmitters—were synchronized to the standard times 00:00h and 12:00h Coordinated Universal Time (UTC). These hours have become etched in as the standard radiosonde observation times ever since. IGY led to a dramatic expansion of meteorological observations, and the countries involved agreed on long-term plans to archive and share the data. IGY was also the beginning of Earth observations from space: the Soviet Union launched the first two Earth-orbiting satellites in 1957, Sputnik 1 and Sputnik 2, and the United States—led by Caltech, this being before NASA was founded—followed suit with Explorer 1 in 1958. The first measurements of Earth’s radiation balance from space were made soon thereafter, in 1959 with Explorer 7. Since then, ground-based observational networks have expanded, and our capabilities of observing Earth from space have exploded. What also lives on from IGY is the spirit of international cooperation and data-sharing that now puts a plethora of climate data at our fingertips, easily findable and accessible on the internet.

It is convenient to work with data that are homogeneous in space and time, but this is generally not the case for ground- or space-based measurements. Observing stations on the ground are spaced irregularly. For example, most stations are on continents, and there are few in the southern hemisphere. Stations
rarely have continuous records over long time spans. Satellite measurements generally provide better spatial coverage. But their temporal resolution can be limited. For polar-orbiting satellites, for example, the period over which the ground track of a measurement footprint repeats can exceed 10 days. This is longer than the evolution timescale of weather systems. Additionally, satellites and their instruments become obsolete and are replaced by new ones, posing challenges to ensure data continuity. Climate analysts therefore often resort to reanalysis data: statistically optimized blends of short-term weather forecasts with as many historical observations as possible. The analysis step of a weather forecast determines the current state of the atmosphere globally, filling gaps between observations with data assimilation techniques that leverage previous forecasts of the atmospheric evolution. Reanalysis repeats this process with a modern weather prediction model and archived historical data, to obtain the best estimate of the state of the atmosphere at times in the past.

Since 1979, microwave radiometers aboard polar-orbiting satellites have measured with global coverage the intensity of microwave radiation emitted by atmospheric oxygen. Temperatures at different levels in the atmosphere can be inferred from microwave measurements at different wavelengths. Microwave measurements led to marked improvements in atmospheric analyses: They provided the first temperature estimates with three-dimensional global coverage. Therefore, for many purposes and unless longer records are necessary, it is common to focus on reanalysis data post 1979. In what follows and unless otherwise noted, we show data from the European Centre for Medium-Range Weather Forecasting (ECMWF) Interim Reanalysis for the 30 years from 1981 through 2010. It is common to refer to both direct instrumental measurements and reanalysis data as observations.

1.1 TEMPERATURE

1.1.1 Current Distribution and Seasonal Variations at Surface

The annual-mean surface air temperature, conventionally defined as the temperature 2 m above the surface, exhibits many features that should be broadly familiar (Fig. 1.1, top). Temperatures are greatest around the equator (~300 K or ~27°C) and are relatively uniform within the latitude belt between around 30°N to 30°S. This latitude belt constitutes the meteorological tropics, which have no sharp boundary but extend to about 30°N/S, that is, beyond the astronomical Tropic of Cancer (23.4°N) and Tropic of Capricorn (23.4°S). In the extratropics, temperatures generally decrease poleward, at a rate of ~4 K/(1000 km), which is much greater than tropical temperature gradients. For example, the average temperatures on the Galapagos Islands at the equator and on the Hawaiian Islands at 21°N latitude are almost the same (about 24°C). By contrast, over the same meridional distance of 22° latitude, or about 2400 km in the north-south
direction in the extratropics, the average temperatures drop from about 16°C in Los Angeles at 34°N to about 7°C in coastal southern Alaska at 56°N. Additionally, higher elevations such as the Tibetan Plateau (mean elevation around 5000 m) and the ice sheets of Greenland and Antarctica (mean elevations around 2000 m) have lower surface air temperatures than surrounding lower elevations. Annual-mean temperatures below the freezing point of water (273.15 K) are found poleward of ~60° latitude. This is the region of permafrost. The annual temperature cycle in the ground is gradually attenuated with depth, and beneath ~7 m depth, the temperature is approximately equal to the annual-mean surface temperature. Where the annual-mean surface temperature is below freezing, soil beneath a top layer that may thaw seasonally remains frozen year-round. The global- and annual-mean surface air temperature is about 289 K.

The annual-mean surface air temperatures are generally similar to those during the equinox seasons (spring and fall). Because for many climate statistics, the annual mean is quite similar to the means over the equinox seasons, we will not show equinox seasons separately in what follows. However, annual or equinox means, for example, of surface air temperature mask large variations between the solstice seasons (summer and winter), especially in middle and higher latitudes. Winters are colder than summers, of course, but more so over continents than over oceans. In December, January, February (DJF, northern hemisphere winter), mean surface air temperatures are lowest over the north Asian continental interior (Siberia) and over the high elevations of the Greenland ice sheet, where they drop to ~40°C (Fig. 1.1, middle). Generally, eastern continental boundaries in mid-latitudes are colder than western continental boundaries. For example, northeast Asia and northeastern North America are colder than northwestern North America and western Europe at similar latitudes. The eastern continental boundaries are even colder than continental interiors at similar latitudes; for example, eastern Asia is colder than central Asia. In June, July, August (JJA, northern hemisphere summer), mean surface air temperatures are highest over continents in the subtropical latitude belt between around 10° and 35°N, the transition zone between the warm, tropical climates and temperate, midlatitude climates. For example, in the Sahara Desert in northern Africa or in the Gobi Desert in Asia, JJA mean temperatures reach 40°C (Fig. 1.1, bottom). The coldest temperatures on Earth have been recorded over the Antarctic ice sheet in winter (JJA), with seasonal mean values below ~60°C. Wintertime extremes there have reached ~89°C or 184 K—a temperature at which carbon dioxide freezes to solid dry ice.

Taking the difference between summer and winter temperatures—that is, JJA minus DJF in the northern hemisphere, and JJA minus DJF in the southern hemisphere—clearly reveals the role of continentality in controlling the annual temperature range (Fig. 1.2). The annual temperature range is lowest (~5 K) in low latitudes, within about 15° latitude of the equator. The weak annual temperature variations in the tropics make it possible to produce a plot like Fig. 1.2 without an evident seam (with the rather coarse contour interval of
Figure 1.1: Surface air temperature in the annual mean (top), DJF mean (middle), and JJA mean (bottom). As is common practice in meteorology, surface temperature here is defined as the temperature 2 m above the surface.

5 K) at the equator, where what is meant by summer and winter switches discontinuously. The continents in middle and high latitudes stand out as regions of large annual temperature ranges (≥15 K). The annual temperature range is smaller over oceans at similar latitudes. Their thermal inertia is greater because turbulent vertical mixing of water masses in the upper ocean means that between $h \approx 30$ m of the water column in the tropics and up to more than $h \approx 2000$ m in polar regions in winter are affected by seasonal atmospheric temperature variations; by contrast only the upper $h \approx 7$ m of solid ground are affected by the seasonal cycle. The effective heat capacity per unit area of the portion of the surface participating in seasonal heat exchange with the atmosphere is $\rho c_s h$, where $\rho$ is the density of the surface, $c_s$ is its specific heat
Figure 1.2: Annual range of surface air temperatures. The annual temperature range is the difference between mean summer and winter temperatures, that is, the difference between JJA and DJF temperatures in the northern hemisphere, and the difference between DJF and JJA temperatures in the southern hemisphere. (This looks continuous across the equator because the annual temperature range near the equator is small.)

capacity (heat capacity per unit mass), and \( h \) is the depth of the surface layer affected by the seasonal cycle. The density \( \rho \) of rocks or soil is about a factor 2–3 greater than that of water, but their specific heat capacity \( c_s \) is about a factor of 4 smaller. Together with the much thinner layer of solid ground that participates in seasonal exchange with the atmosphere, this implies that per unit area, continents have a much smaller effective heat capacity than oceans. Hence, the seasonal cycle over oceans is damped more strongly than over land. Additionally, lack of moisture availability and reduced evaporation can also enhance the seasonal cycle over land relative to oceans, as we will discuss in chapter 7.

1.1.2 Recent Changes at the Surface

Annual- and seasonal-mean temperatures are not constant but vary on longer timescales. In the past 160 years, surface temperatures have generally increased, in the global and annual mean by about 1.2 K between 1850 and 2021, with a warming trend of 0.18 K per decade over the past 40 years (Fig. 1.3). Most of the increase in mean surface temperatures occurred in two stages, the first from the 1910s to the 1940s, the second from the 1970s onward and continuing through today. In between were three decades of variable temperatures, with no clear global-mean trend.

The regional surface temperature variations that give rise to the global-mean variations show that the warming of the past 160 years indeed has been global, albeit not globally and temporally uniform (Fig. 1.4). Between 1850 and the 1910s, there was no coherent large-scale pattern of temperature changes but regional variations over different ocean and land areas (Fig. 1.4a). Around 1910,
the world was globally slightly cooler than between 1850 and 1900 (Figs. 1.3, 1.4). Between the 1910s and 1940s, however, Earth’s surface warmed globally, and most strongly in high northern latitudes (Fig. 1.4b). The North American prairies experienced the Dust Bowl, a period of drought and relative warmth during the 1930s. Extensive farming and deep plowing in the Great Plains had exposed previously grass-covered top soil, which fed severe dust storms and led to soil erosion during the drought. In the decades following the 1940s, much of the Earth’s surface continued to warm, but some regions cooled. For example, the North Atlantic cooled through the 1960s and 70s, the eastern South Pacific cooled or did not warm much through the 1940s and 50s, and some land areas in the eastern United States, Central Africa, and Southeast Asia also cooled or warmed only weakly (Fig. 1.4c). These land areas are regions in which, for example, industrial processes and burning of biomass have led to increasing emissions of pollutants from which aerosols (tiny particles suspended in air, forming smog) can form. Aerosols reflect sunlight and so cool the surface underneath. It thus appears that the relatively constant global-mean temperatures between the 1940s and 70s arose because widespread warming was counteracted by aerosol-induced cooling as well as transient cooling over the North Atlantic and South Pacific. The ocean cooling is likely a result of natural cycles in the Atlantic and Pacific, such as the Pacific Decadal Oscillation, a pattern of climate variability that recurs on timescales of several decades.

Since the 1970s, however, Earth’s surface has warmed globally. Land areas
Figure 1.4: Decadal regional surface air temperature changes relative to 1850–1899 for selected years between 1910 and 2010. Grey areas have insufficient data to infer decadal temperature variations over the past century. The temperature variations are filtered with a filter that maximizes slower (decadal) variations relative to faster (interannual) variations. Thus, each figure panel is to be taken as representative of several years centered on the indicated year. [update]
have warmed faster than ocean areas, and high northern latitudes have warmed fastest (Fig. 1.4d). Relative to the 19th century, the warming is truly global, with only a few small areas (e.g., over the North Atlantic) that have not warmed much. It is difficult to imagine any process of internal re-distribution of energy within the climate system that could give rise to such global warming. Instead, this warming is just what is expected to occur when the atmospheric concentrations of greenhouse gases, such as carbon dioxide (CO$_2$), increase. They have continuously risen since the industrial revolution in the 19th century. After hovering around 270 ppm (parts per million) for the past 10,000 years before the industrial revolution, carbon dioxide concentrations began to climb substantially late in the 19th century, reaching 295 ppm by 1900, 320 ppm by 1970, and 415 ppm by 2020 (Fig. [add Keeling’s co2 figure]). The atmospheric concentration of carbon dioxide rose because of human emission from burning of fossil fuels, production of cement, and deforestation. The global warming over the past 160 years has occurred because of the accumulation of greenhouse gases in the atmosphere owing to human emissions (primarily carbon dioxide but also other trace gases such as methane (CH$_4$), nitrous oxide (N$_2$O), and fluorinated gases).

How can increased concentrations of a trace constituent of the atmosphere such as carbon dioxide lead to substantial global warming? In chapter 4, we will analyze how carbon dioxide and other gases interact with radiative energy fluxes and come to be important for Earth’s energy balance. In chapters 5 and 6, we will study how increasing atmospheric greenhouse gas concentrations lead to global warming.

## 1.1.3 Long-Term Variations

On yet longer timescales, direct temperature measurements are scant, and we rely on indirect methods, for example, using geochemical temperature proxies, to infer temperature variations.

### Holocene

The Holocene spans the past 11,700 years (11.7 kyr) and represents the geological epoch that extends from the end of the last glacial period (“ice age”) to the present. Temperature changes over the Holocene can be inferred from multiple paleo-temperature proxies. For example, oxygen has three naturally occurring stable isotopes: oxygen-16 ($^{16}$O, with 8 protons and 8 neutrons in the nucleus), oxygen-17 ($^{17}$O, with 8 protons and 9 neutrons), and oxygen-18 ($^{18}$O, with 8 protons and 10 neutrons). The primary isotope $^{16}$O is the most common, with a natural abundance of 99.76%; the heavier isotopes $^{17}$O and $^{18}$O are much rarer, with natural abundances of 0.04% and 0.20%, respectively. How frequently the heavier and rarer $^{18}$O (or the even rarer $^{17}$O) replace the common $^{16}$O in the calcium carbonate (CaCO$_3$) shells of sea organisms depends on temperature. The same is true for the frequency with which magnesium impurities are incorpo-
Figure 1.5: Estimated global- and annual-mean surface temperatures over the past 3 million years. The temperatures are expressed as anomalies (deviations) relative to the mean for 1961–1990. (a) Temperatures over the past 11.3 kyr, since the early Holocene, inferred from a variety of globally distributed temperature proxies. LIA marks the Little Ice Age. The orange line on the far right shows the smoothed direct temperature measurements from Fig. 1.3. [update instrument T] (b) Temperatures over the past 420 kyr, inferred from the relative abundance of deuterium (²H) to hydrogen (¹H) in an Antarctic ice core. (c) Temperatures over the past 3 Myr, inferred from the abundance of the heavy oxygen isotope ¹⁸O relative to the common ¹⁶O in the calcium carbonate shells of benthic (bottom dwelling) foraminifera, recovered from ocean sediments. Shading in all panels indicates rough estimates of 95% confidence bands. Note the changing temperature axes from panel to panel.
rated in the carbonate shells. Recovering the fossil shells of sea organisms such as corals or foraminifera (unicellular microorganisms the size of sand grains) and measuring their oxygen isotopes or magnesium impurities allows us to infer the water temperature where they lived. The age of these fossils can be determined by dating the shells radiometrically: one measures the abundance of radioactively decaying (e.g., carbon or uranium) isotopes relative to their decay products to establish how long ago they were incorporated into the shells. Similar dating techniques are also applied to land-based proxies of temperature (e.g., pollen in lake deposits, speleothems in caves, and tree rings).

From a combination of multiple such paleo-temperature proxies, it was estimated that Earth warmed by ~0.6 K in the global mean over the first 2–3 kyr of the Holocene, a period over which the glacial ice sheets in the northern hemisphere continued to retreat (Fig. 1.5a). Temperatures were relatively high and stable thereafter, during the Holocene Thermal Maximum between 9 and 5 kyr before present (BP, here referring to years before the year 2000 of the Common Era, CE).9 After that, Earth cooled gradually by ~0.7 K, until the end of the Little Ice Age about 150 BP, or 1850 CE (Fig. 1.5a). At least regionally, however, this cooling was not monotonic but was punctuated by intervals of warming or little temperature change, such as the Medieval Warm Period in the North Atlantic region around 900 BP, or 1100 CE. Because some of the proxies on which this temperature reconstruction is based are most sensitive to the summer (growing) season, it is possible that this temperature reconstruction is more representative of temperature variations in summer than in the annual mean. But irrespective of the precise interpretation of the Holocene record, the past century of warming (Fig. 1.3 and orange line in Fig. 1.5a) represents an unprecedented spike: the temperature rise over the past century is very rapid and large compared with the temperature variations over the Holocene.

The precise magnitude, seasonality, and geography of temperature changes over the Holocene and of epochs deeper in the past are uncertain. Likewise uncertain are the mechanisms responsible for these climate changes. Variations in Earth's orbit around the sun and the concomitant variations in Earth's incoming solar energy are certainly implicated in many of the climate changes with timescales of millennia and longer; we will discuss them in chapter 3. How subtle variations in incoming solar energy work their way through the climate system to give rise to large climate changes remains insufficiently understood.

The Holocene represents one of the climatically most stable periods of the last 3 million years, with relatively small global temperature fluctuations. This stability allowed nomadic human hunter-gatherers to settle. Tribes in the Middle East, Central America, and East Asia began farming, apparently independently of each other. They invented metal tools, such as the plow, and formed the civilizations in Mesopotamia, Egypt, and China that produced the first written documents. Human populations flourished and cultures as we know them emerged in the Holocene—thanks to its climatic stability.
Ice Ages

Zooming out in time, the temporal resolution and spatial coverage of our temperature proxies decrease, as does the accuracy with which temperature variations can be inferred from them. But up to 800 kyr back, Antarctic ice cores provide detailed records with a temporal resolution that gradually degrades from $O(10\ \text{yr})$ (“order 10 years”) in the recent past to $O(1\ \text{kyr})$ in the more distant past. Greenland ice cores provide similar records but only extend ~120 kyr back because the Greenland ice sheet is younger than the Antarctic ice sheets.

Temperatures of the past can be inferred from the isotopic composition of ice that formed in the past. The relative rates at which heavier and rare water isotopologues (water with heavier isotopes of H and O) are incorporated into the ice depends on temperature. The heavier water isotopologues with $^2\text{H}$ (also known as deuterium, D) or $^{18}\text{O}$ in place of the common $^1\text{H}$ or $^{16}\text{O}$ condense and freeze more easily, and the lighter isotopologues evaporate more easily. The fractionation rate—the relative partitioning of the heavier and lighter isotopes—during evaporation and condensation/freezing depends on temperature. The end result is that the heavier isotopologues are rarer in ice than in the oceans from which the water originated that, after evaporation, transport, and condensation/freezing, is deposited on the ice sheet as snow. How rare they are depends on the temperature along the path the water takes from its origin in lower-latitude oceans to its resting place on the high-latitude ice sheet (chapter 9). Thus, analyzing how the isotopic composition of the ice varies with age allows us to infer temperature variations on geologic timescales by linking global temperatures through physics-based assumptions to the path-dependent temperatures that control the isotopic composition of ice.

Focusing on the past 420 kyr of global-mean temperature variations, we can discern a cycle of prolonged, cold glacial periods, interrupted by shorter, warm interglacial periods (Fig. 1.5b). Interglacial periods are spaced about 100 kyr apart. The current interglacial period, the Holocene, began with the end of the last glacial period continues today (Fig. 1.5b). The inception of glacial periods is marked by an initially relatively rapid cooling of up to about 5 K over 20 kyr, which over the ensuing tens of thousands of years continues more gradually. The glacial periods terminate relatively suddenly, with a warming spike that brings the climate from glacial to interglacial temperatures within a few thousand years. (Although geologically this represents a warming spike, the rate of warming of around 1 K/kyr during the deglaciations is an order of magnitude smaller than the rate of the recent human-induced global warming.)

The last glacial period began about 110 kyr BP and ended about 15 kyr BP. Glaciers and ice sheets advanced in this period and, at the Last Glacial Maximum (LGM) around 18 kyr BP, covered large areas of the northern hemisphere continents, down to about 40° latitude. In North America, the Laurentide ice sheet reached the modern sites of New York City and Chicago, and in Europe, the Fennoscanidian ice sheet reached France, Germany, and Poland (Fig. 1.6). The Holocene can be compared to the previous interglacial period, the Eemian,
Figure 1.6: Reconstructed elevation of northern-hemisphere ice sheets during the Last Glacial Maximum (LGM).\textsuperscript{10} The Laurentide Ice Sheet over North America ended near modern-day New York. Its remnants are still visible there today: giant boulders dropped by the ice in Central Park, and rubble carried by the ice and deposited in long lines of hills called moraines in Brooklyn and Queens. The Fennoscandian Ice Sheet reached southern England, France, Germany, and Poland, likewise leaving its traces in European landscapes.

between about 130 and 115 kyr BP. Climate in the Eemian was broadly similar to how it is today, with temperatures at least in high latitudes a few kelvin above today’s (Fig. 1.5b). Sea level was ~6 m (20 ft) higher than today. The ice sheets in Greenland and Antarctica were smaller, so less water was locked up as ice on land. It is remarkable that a global warming of a few kelvin—well within what may occur this century—sustained over a few thousand years produced such a large sea level rise. A comparable sea level rise today would inundate the many coastal cities that house much of the world’s population.

The global temperature variations on these glacial-interglacial timescales are much greater than those during the Holocene (compare Figs. 1.5a and b). It is clear that the glacial-interglacial cycles are paced by variations in Earth’s orbit around the sun, which have similar periodicities in the range of
20 kyr to 100–400 kyr (chapter 3). But how orbital variations generate glacial-interglacial cycles is one of the great unresolved questions in the geosciences: What kind of orbital variations are essential, and how are they transmitted and amplified in the climate system to produce the enormous and (compared with the sinusoidal orbital variations) abrupt climate changes between glacial and interglacial periods? From bubbles of ancient air trapped in ice, we do know that atmospheric concentrations of greenhouse gases such as carbon dioxide vary together with temperature on timescales of glacial-interglacial cycles (Fig. 1.7). However, on these long timescales, the greenhouse gas variations not only drive temperature changes but also respond to them, amplifying each other in a feedback loop that is incompletely understood. Temperature increases drive greenhouse gas concentration increases through both physical and biological processes in oceans and on land which, for example, link the large carbon reservoir in the deep oceans to the atmospheric composition on timescales of $O(10^4 \text{ yr})$. The complicated interlocking dance of the processes involved remains to be elucidated.

Embedded in the glacial-interglacial cycles are large-amplitude, higher-frequency temperature variations, with timescales of millennia and shorter. The origin of these millennial events likewise is not clear. They are likely triggered by internal instabilities within the climate system, such as an abrupt alteration of the ocean circulation that might occur when an ice sheet collapses and iceberg discharge and melting freshens salty surface waters. For example, changes in meltwater fluxes from the Laurentide ice sheet have been implicated in the sudden return to glacial temperatures in most of the Northern Hemisphere known as the Younger Dryas. The Younger Dryas lasted from about 12.9 kyr BP to the onset of the Holocene 11.7 kyr BP.

**Past 3 Million Years**

Zooming further out in time, more glacial-interglacial cycles come into view, but the mystery of how they are generated deepens (Fig. 1.5c). On these timescales, we can no longer rely on ice cores to infer temperatures, but we must use ocean sediment cores, which have lower temporal resolution because sedimentation occurs slowly on the ocean floor. In these cores, one can find shells of benthic (bottom-dwelling) foraminifera, and from the abundance of $^{18}$O relative to $^{16}$O in their carbonate shells, it is in principle possible to infer the temperature of the deep-ocean water in which the foraminifera lived—provided one knows the relative $^{18}$O abundance of the sea water at the time. Unfortunately, the oxygen isotopic composition of the sea water itself also changes with time, as the volume of land ice changes: Polar ice sheets contain a disproportionate share of the lighter isotope $^{16}$O, leaving behind sea water enriched in the heavier isotope $^{18}$O as continental ice sheets grow. (This arises because the snow that accumulates
on ice sheets is depleted in $^{18}$O, much of which condenses earlier along the path the water takes through the atmosphere from lower latitudes to the poles; we will discuss the processes involved in chapter 9.) But it is possible to deconvolve the effects of ice volume and temperature on the relative $^{18}$O abundance using ice models that couple land ice volume to air temperature, and using physically based assumptions about how air and deep-ocean temperatures covary.

The resulting estimate of global-mean temperature variations over the past 3 million years (3 Myr) shows many glacial-interglacial cycles, superimposed on a general cooling trend of 3–4 K (Fig. 1.5c). Intriguingly, the dominant period of the glacial-interglacial cycles changed from a periodicity of ~40 kyr before 1 Myr to a periodicity of ~100 kyr thereafter, though neither cycle is exactly periodic. Both ~40 kyr and ~100 kyr are close to periods of variation in the tilt (obliquity) of Earth’s rotational axis and in the eccentricity of Earth’s orbit around the Sun, respectively (chapter 3). But the change in the dominant period of the glacial-interglacial cycles occurred without an evident change in Earth’s orbital variations. Why did the dominant period of the glacial-interglacial cycles change? And what drove the long-term cooling trend? The answers to these and many other questions about the climate system’s past are presently unknown. It is once again clear that the cooling trend coincides with a downward trend in the concentration of greenhouse gases such as carbon dioxide (Fig. xx). But to what degree the temperature trend drove the carbon dioxide trend, versus the other way around, is unclear.

The transition in the dominant period of glacial-interglacial cycles occurred in the middle of the geological epoch known as the Pleistocene, which lasted from 2.588 Myr until the beginning of the Holocene 11.7 kyr. The Pleistocene is the epoch of extensive and repeated glaciations in the northern hemisphere. It was the time of the Paleolithic Age, in which humans first developed primitive stone tools, and later began to use fire, bury their dead, and produce bracelets, beads, cave paintings and other early forms of art. Large mammals such as mammoths, mastodons, saber-toothed cats, and giant ground sloths roamed the Earth. They began to die out at the end of the Pleistocene, in a major extinction of megafauna, which coincided region-by-region with the spread of humans over the continents. The topic is still debated, but hunting by the expanding human populations—perhaps hunting primarily of the large herbivores, on which the large carnivores depended as their prey—almost certainly was a factor in the megafauna extinction, as may have been climate change.11

The epoch preceding the Pleistocene, the Pliocene, extended from 5.333 Myr to 2.58 Myr BP and was 2–3 K warmer than today. It is the most recent time in geological history that carbon dioxide concentrations were as high as they are now (over 415 ppm). At the time, northern hemisphere ice sheets were ephemeral, and sea level was ~20 m higher than today. The Pliocene saw the first hominins, as well as geomorphological changes, such as the closing of the Isthmus of Panama that joined the two Americas and may have contributed to changes in ocean circulation and climate.
Deeper in Time

Looking yet deeper in time, climate reconstructions become even more uncertain, and their temporal resolution degrades further, down to $\mathcal{O}(100$ kyr) about 50 Myr before present (BP). As for the Pleistocene, climate reconstructions on timescales of millions of years primarily use the oxygen isotopic composition of benthic foraminifera shells recovered from ocean sediment cores, combined with physical models and assumptions to relate variations of the relative abundance of $^{18}$O to temperature variations. Estimating surface temperature variations in this way becomes uncertain because, for example, variations of ice volume and sea level, which affect the oxygen isotopic composition of ocean water, are not well constrained independently of the measured $^{18}$O variations in the foraminifera shells. However, while surface temperature variations are only known to within a few kelvin on such timescales, qualitative trends can be inferred.

Over the past 50 Myr, climate has generally cooled. During the Early Eocene Climatic Optimum around 50 Myr BP, Earth may have been $\sim 15$ K warmer in the global mean than it is today (Fig. 1.8). Crocodilians and palm trees thrived in the Arctic, and atmospheric concentrations of greenhouse gases such as carbon dioxide and methane were much higher than today. It is unclear precisely how much higher they were, but current evidence suggests carbon dioxide concentrations did not exceed 2,000 ppm even during the warmest periods of the Eocene (Fig.xx). The general cooling since the early Eocene was not monotonic but was punctuated, for example, by geologically brief warm periods such as the Middle Eocene Climatic Optimum, during which Earth warmed 4–8 K for a few hundred thousand years. Similar geologically brief warm spikes (hyperthermals) occurred earlier, for example, at the Paleocene-Eocene Thermal Maximum at the beginning of the Eocene about 55.8 Myr BP. At the beginning of the Oligocene around 33 Myr BP, a geologically rapid temperature drop occurred, associated with growth of Antarctic ice sheets, followed by warming and Antarctic thawing around 26 Myr BP. Geological processes such as plate tectonics are almost certainly implicated in climate changes on such long timescales. They were likely modulated by more rapid feedback processes within the climate system. Geological processes control the composition of the atmosphere and thus Earth’s energy balance (chapters 4 and 7) through emission of trace gases such as carbon dioxide by volcanism, and drawdown by the weathering of silicate rocks, which fixes carbon dioxide from the atmosphere into carbonate that eventually forms sediments. Climate changes on timescales of millions of years are driven by changes in such geological processes, rather than by the more rapid and quasi-periodic variations in Earth’s orbit (timescales of up to 400 kyr), or by the much slower and gradual changes in the sun’s luminosity over its lifetime (chapter 2). But it is not understood how geological processes such as plate tectonics lead to changes in the atmospheric composition and in climate, and how they are modulated by feedbacks on shorter than tectonic timescales. What the deeper-time perspective does illustrate is that the climate system is a complex dynamical system whose behavior cannot be
understood by examining subcomponents in isolation from another. A holistic view is required. On timescales of millions of years, this must include geological processes.

What zooming out in time has revealed clearly is that the magnitude of temperature variations increases with timescale: witness the increasingly stretched temperature axes as timescales increase from Fig. 1.5a–c to Fig. 1.8. One speaks of the climate system as having a red spectrum, meaning it has greater power at lower frequencies (“redder” frequencies, in analogy with the visible light spectrum, where red occupies the low-frequency end). We will discuss the spectrum in more quantitative detail in chapter 3.6. The temperature reconstructions here show that on geological timescales, the climate system is capable of changes
that are very large compared with those experienced during the evolution of humans over the past several hundred thousand years.

1.1.4 Vertical Structure

While surface temperatures can be inferred indirectly on geological timescales, information about the third dimension, the vertical structure of the atmosphere, only comes from direct measurements in the recent past. The first measurements were carried out late in the 18th century by scientists who ascended mountains. The Swiss scientist and alpinist Horace-Bénédict de Saussure climbed Mont Blanc in 1787, measured pressure and temperature along the way, and discovered that air temperature decreases with a lapse rate of roughly 7 K per kilometer height. Sufficiently accurate measurements with sounding balloons that reached greater altitudes became available at the end of the 19th century. They established that the temperature decrease with height does not continue indefinitely, which would be impossible in any case, as temperatures near absolute zero would be reached around 40 km altitude. Rather, on the basis of balloon measurements in France and Germany, the meteorologists Léon Teisserenc de Bort and Richard Assmann reported in 1902 that there is a layer in which temperatures are nearly constant with height between around 11 and 20 km altitude; above, temperatures increase with height. They had discovered the stratosphere, which lies above the troposphere and is separated from it by the tropopause (Fig. 1.9). The troposphere takes its name from the Greek tropos, “turning,” because in the troposphere, overturning air masses mix vigorously. The stratosphere derives its name from the Latin stratum, because it is more stratified and air masses mix less vigorously. Temperatures increase with height through the stratosphere, primarily because solar radiation is absorbed by ozone, heating the air, as we will discuss in chapter 5. The temperature increase with height continues up to the stratopause around 50 km altitude. Above lies the mesosphere, in which temperatures again decrease with height (Fig. 1.9). Collectively, the stratosphere and mesosphere are referred to as the middle atmosphere, which reaches up to the mesopause near 85 km altitude. Pressure is often used as a vertical coordinate in atmospheric sciences, because it serves as a mass coordinate (discussed in chapter XX).

The vertical temperature structure of the atmosphere is complex, and more layers with distinct thermal properties lie above the mesosphere: for example, the thermosphere, where the dance of auroras in polar regions occurs. For understanding the surface climate, however, it is most important to understand the properties and the dynamics of the troposphere, because it contains about 80% of the mass of the atmosphere. Layers above play secondary or altogether negligible roles for the transport of energy, momentum, and water that shapes the surface climate. Our focus will therefore be on the troposphere.

Modern reanalysis data incorporating measurements with radiosondes and satellite data reveal the vertical temperature structure of the troposphere and lower stratosphere with global coverage (Fig. 1.10). Averaged zonally (over lon-
Figure 1.9: Vertical temperature structure of the atmosphere up to ~85 km altitude. (This is based on the U.S. standard atmosphere and is representative of annual-mean conditions in midlatitudes.)

...
Figure 1.10: Vertical structure of temperatures in the zonal mean: annual mean (top), DJF mean (middle), and JJA mean (bottom). The thick red line marks the tropopause, determined according to the World Meteorological Organization’s definition as the lowest level at which the temperature lapse rate decreases to 2 K km\(^{-1}\) or less and stays as low for at least another 2 km.
~200 K, which is the lowest temperature in Earth’s lower atmosphere. Because water vapor has to go through this “cold trap” to reach the stratosphere from the tropics, the tropical tropopause temperature has important implications for the amount of water vapor that can reach the stratosphere.

Why is there a tropopause? What controls its height and how can it change with climate? What sets the tropospheric temperature lapse rate? Interactions of the radiative energy balance with dynamical energy fluxes will turn out to be the key to answering such questions and will be discussed in chapter 6.

1.2 WIND

1.2.1 Surface

The surface wind is conventionally defined as the wind at 10-m height. In the mean, the zonal (east-west) winds on Earth are about an order of magnitude stronger than the meridional (south-north) winds, a consequence of Earth’s rotation around its spin axis (chapter 8). The mean zonal surface winds are easterly (from the east) in the tropics, westerly (from the west) in midlatitudes, and weakly easterly in polar latitudes, especially in the southern hemisphere (Fig. 1.11). The strength of the mean zonal surface winds varies seasonally. In midlatitudes, winds are strongest in winter, in the regions centered around 45°N/S known as the storm tracks: over the Pacific and Atlantic in the northern hemisphere, and over the Southern Ocean encircling Antarctica in the southern hemisphere. In the southern-hemisphere latitude belt centered around 45°S, the zonal winds circling the globe are unimpeded by continental landmasses (except for the tip of South America) and are especially strong, with average values around 10 m s⁻¹. The region is known among seafarers as the “Roaring Forties,” because it spans the forties of degrees southern latitude. Strong surface winds are also found during the boreal summer months (JJA) in the Asian monsoon region, off the Horn of Africa: there, the southwesterly Somali jet blows diagonally across the Indian Ocean, hugging the coasts of Somalia and Oman, with mean wind speeds that exceed 10 m s⁻¹. Otherwise, however, between 30°N and 30°S, the gentler and steady easterly winds known as the trade winds, or trades for short, prevail. The subtropical transition latitudes around 30°N/S, where mean zonal winds change sign from tropical easterlies to midlatitude westerlies, are known as the horse latitudes. There winds are variable but calm in the mean. This pattern of alternating easterlies in the tropics, westerlies in midlatitudes, and weak easterlies again in polar regions is present throughout the year. Interestingly and curiously, despite the strong seasonal variations of temperature and wind speeds, the latitudes at which the mean zonal surface wind changes sign from tropical easterlies to midlatitude westerlies shift by only ~5° (600 km) between summer and winter.

That the mean zonal surface winds have prevalent directions has been
Figure 1.11: Surface winds in the annual mean (top), DJF mean (middle), and JJA mean (bottom). The color shading indicates the strength and direction of the zonal surface wind. As is common practice in meteorology, the surface wind here is defined as the wind at 10-m height above the surface.

known and exploited by mariners for centuries, as is reflected by the archaic names of the wind belts. The term trade winds derives from the Middle Low German word *trade*, meaning track, because winds “blowing trade” provided sailors with a steady pull on a definite track or course. (Originally, the term trade winds was used for all winds that blow steadily in the same direction but nowadays is used more restrictively for the tropical easterlies.) When Christopher Columbus set out on his first New World voyage in 1492, he knew from Portuguese explorers before him, who had sailed southward along Africa’s west coast, that he had to drop south to exploit the easterly trades for his planned westward passage. He sailed from Spain south toward the Canary Islands...
Figure 1.12: Route of Christopher Columbus’ first voyage in 1492/93 (green line), overlaid over surface wind vectors for 1981–2000 averaged over the time of year of Columbus’ voyage (September–February). Columbus crossed the Atlantic going westward from September through early October 1492, and returned going eastward from January through early March 1493. The latitude at which mean easterlies transition to mean westerlies was likely similar in Columbus’ age as it is now, given how little that latitude varies seasonally. As in Fig. 1.11, the color shading indicates the strength and direction of the zonal surface wind.

(28°N), from where the trades carried him to the Bahamas and eventually to the Caribbean (Fig. 1.12). To return to the Old World, he headed northeast diagonally across the trades, until he reached the midlatitude westerlies, which carried him back eastward to the Azores and eventually to Europe. In the following centuries, when international trade was dominated by sailing ships, mariners similarly used the easterly trades to sail from Europe westward to the Americas and the midlatitude westerlies for the return trip. They spoke of “running the easting down” when they used the westerly Roaring Forties for a swift eastward passage from Europe around the Cape of Good Hope (tip of South Africa) toward Australasia.

Because the mean meridional surface winds are weaker than the mean zonal surface winds, their directions became known only later. The trade winds have a northerly component in the northern hemisphere, where they are called the northeasterly trades, and a southerly component in the southern hemisphere, where they are called the southeasterly trades (Fig. 1.11). The northeasterly and southeasterly trades converge in the Intertropical Convergence Zone (ITCZ).
Midlatitude westerlies, on the other hand, have a weak poleward component, so surface winds diverge in the subtropical horse latitudes. Polar easterlies, where they exist, have a weak equatorward component in the zonal mean, although this is difficult to discern locally because of the strong zonal variations of the polar easterlies, e.g., around Antarctica. Stated more generally, mean meridional surface winds are directed equatorward in regions of mean surface easterlies, and poleward in regions of mean surface westerlies. This statement also holds, for example, during the Asian summer monsoon, when the mean northward surface wind across the equator near the Somali jet is accompanied by surface westerlies. That is, when surface winds change from southward to northward at the beginning of the summer season, the usually easterly trade winds reverse direction to westerly—a wind reversal that is a hallmark of large-scale monsoons (Fig. 1.11).

What controls the direction and strength of surface winds and the width of the wind belts? How are zonal and meridional surface winds related? It may be surprising that we still do not have complete answers to such age-old questions. However, it is clear that for understanding surface winds, it is essential to understand the angular momentum balance of the atmosphere and how turbulent fluxes enter in it. This is the topic of chapter 8.

1.2.2 Vertical Structure of Zonal Winds

The vertical structure of Earth’s winds only became known in some detail with the advent of aviation and the need for improved aviation weather forecasts in the 20th century. Radar tracking of ascending radiosondes to infer the speed and direction of winds became common during World War II. These rawinsondes (“radar-wind sondes”) continue to be used today. Atmospheric reanalyses that assimilate rawinsonde data and data collected, for example, by commercial aircraft and satellites give us a quantitative picture of how winds vary in the vertical.

In the zonal mean, zonal winds generally have an eastward shear with height within the troposphere, that is, they become more westerly with height (Fig. 1.13). There are regional exceptions to this rule, most prominently during the Asian summer monsoon, when winds near the equator have a westward shear with height, which leads to upper-tropospheric easterlies appearing in the JJA zonal mean. But otherwise, because of the general eastward shear with height, mean westerlies occupy a broader range of latitudes in the upper troposphere than near the surface. Their strengths reach local maxima in the jet streams in the upper troposphere and lower stratosphere, with mean zonal winds around 30 m s⁻¹ in the annual mean, with higher values (~40 m s⁻¹) in winter and lower values in summer (~25 m s⁻¹). The cruising altitude of commercial aircraft lies in the upper troposphere. The strong westerlies and jet streams there are responsible for the fact that flying west in midlatitudes takes longer (headwind) than flying east (tailwind).

While the angular momentum balance controls the direction and strength...
Figure 1.13: Vertical structure of zonal winds in the zonal mean: annual mean (top), DJF mean (middle), and JJA mean (bottom).
of surface winds, the temperature distribution of the atmosphere and hence the energy balance turn out to be crucial for the vertical structure of zonal winds and features such as the strength of jets. The reason is that meridional temperature gradients are related to the shear of the zonal wind with height through what is known as thermal wind balance, to be discussed in chapter 8.

1.2.3 Meridional Overturning Circulation

We have already seen that the mean meridional surface winds are directed equatorward in the region of tropical easterlies, and poleward in the region of midlatitude westerlies (Fig. 1.11). In the zonal mean and averaged over a sufficiently long time so that flow statistics can be approximated as stationary, this means that there must be compensating meridional return flows of air masses aloft. Otherwise mass in atmospheric columns would be transported across latitude belts, which would have to change the surface pressure (a measure of the mass per unit area of atmospheric columns); a statistically steady state with constant mean surface pressure would not be possible. Mass conservation also implies that air masses in the zonal mean must rise in the ITCZ, where surface winds converge, and sink in the subtropics, where surface winds diverge. Taken together, this implies that a zonal-mean overturning circulation must exist in the meridional (latitude-height) plane.

The meridional overturning circulation closes within the troposphere, just below the tropopause (Fig. 1.14). Air masses converge in the ITCZ, which in the annual and zonal mean lies just north of the equator, at around 5°N. The converging air masses rise through the troposphere, diverge aloft, and flow toward the subtropics of both hemispheres. They sink again in the subtropics and diverge near the surface near 30°N/S. This tropical overturning circulation is called the Hadley circulation, after the 18th-century English lawyer and meteorologist George Hadley, who provided the first mechanistic account of the circulation that still has some validity today. The Hadley circulation consists of one cell in each hemisphere. The two cells share a rising branch at the ITCZ and have separate subtropical sinking branches at their respective poleward edges. The ITCZ and rising branch of the Hadley circulation shift seasonally toward the summer hemisphere. But just like the subtropical dividing line between the mean surface easterlies and westerlies, the sinking branches shift only slightly around 30°N/S year-round. This is the case although the strength of the Hadley cells changes dramatically with seasons: The winter Hadley cell—in which air masses rise in the summer hemisphere, cross the equator, and sink in the winter hemisphere—transports more than 200 Sv of mass, with 1 Sv (Sverdrup) = \(10^9\) kg s\(^{-1}\); the summer Hadley cell—in which air masses rise and sink in the summer hemisphere—transports less than 55 Sv of mass (Fig. 1.14).

The Hadley cells share their sinking branches with midlatitude overturning cells called the Ferrel cells, after the 19th-century American meteorologist William Ferrel, who proposed to account for the poleward component of the
Figure 1.14: Streamfunction of meridional mass circulation in the annual mean (top), DJF mean (middle), and JJA mean (bottom). Positive contours (warm colors) for clockwise rotation; negative contours (cold colors) for counterclockwise rotation. The mass flux is given in units of Sverdrup, $1 \text{ Sv} = 10^9 \text{ kg s}^{-1}$. The streamfunction extrema are labelled with their mass transport values in Sverdrup.
surface westerlies through midlatitude circulations cells with the opposite sense of rotation as the Hadley cells. However, the circulation cells and mechanisms Ferrel proposed for them have little to do with how we conceptualize the cells that bear his name today, fundamentally because he and his contemporaries were insufficiently aware of the role large-scale turbulence plays in maintaining them. The Ferrel cells are considerably weaker than the Hadley cells. They are stronger in winter than in summer, with typical mass transports around 25–40 Sv. The rising branches of the Ferrel cells lie at the poleward edge of the region of midlatitude westerlies. They are shared with the rising branches of usually very weak polar circulation cells (not recognizable with the contouring in Fig. 1.14). The polar circulation cells, to the extent they are manifest, overlie regions of polar easterlies (compare the zonal surface winds in Fig 1.13 with the overturning circulations in Fig. 1.14).

How are the mean meridional overturning circulations and zonal surface winds linked? What controls the strength of the meridional overturning circulations? Such questions will also be discussed in chapter 8.

1.3 WATER

Water plays multiple important roles in the climate system. Not only is water the fluid of the oceans and so the carrier of the energy the oceans transport. Water vapor is also a major carrier of energy in the atmosphere. On its way through the atmosphere, it transports energy in the form of latent heat, which it takes up when liquid water evaporates, and which it releases when water vapor condenses. By evaporating at the surface and condensing in the atmosphere, water thus carries energy from the surface into the atmosphere. Additionally, water in the atmosphere is a principal absorber of radiative energy. Especially infrared (thermal) radiation is absorbed effectively by both water vapor and condensed water in clouds. Thus, water regulates Earth’s energy balance and surface temperature. Beyond that, of course, water is essential for life on Earth. The relative distribution of precipitation and evaporation, along with temperature, regulate the habitability of Earth’s surface and the distribution of ecosystems.

1.3.1 Evaporation

Water vapor can enter the atmosphere through several pathways. The dominant pathway is direct evaporation, primarily from the oceans but also from land surfaces. Over land, evaporation is augmented by transpiration: the process in which plants take up water through their roots, move it upward, and evaporate it through their stomata (minute pores) in stems and leaves. Another, globally less important pathway by which water vapor enters the atmosphere is sublimation of snow and ice, for example, over glaciers. We will subsume all
Figure 1.15: Evaporation rate ($E$) in the annual mean (top), DJF mean (middle), and JJA mean (bottom). The global- and annual-mean evaporation rate is $\sim 3.4 \text{ mm day}^{-1}$, or $\sim 1.2 \text{ m yr}^{-1}$.

Some aspects of the distribution of mean evaporation rates ($E$) at Earth’s surface resemble the distribution of mean surface temperatures (Fig. 1.15). Mean evaporation rates are highest in the tropics, with typical values over oceans around $5 \text{ mm day}^{-1}$, with $1 \text{ day} = 86400 \text{ s}$. (That is, $5 \text{ mm of water evaporate from tropical ocean surfaces per day},$ or $1.8 \text{ m evaporate per year}.$) Near the equator, the mean evaporation rate has a local minimum, which arises principally because upwelling of colder ocean waters from depth cools the surface and reduces the energy available to evaporate water (chapter 7). High evaporation rates are found over the Kuroshio Current in the western North Pacific and...
over the Gulf Stream in the western North Atlantic, especially in winter. The Kuroshio and Gulf Stream are the Pacific’s and Atlantic’s western boundary currents, which carry warm waters from the subtropics poleward. In winter, the westerlies blow cold and relatively dry air from the continents over these warm ocean currents, leading to high evaporation rates. Otherwise, evaporation rates over oceans generally decrease going poleward, to values around 1 mm d$^{-1}$ and below in high latitudes. Over continents, evaporation rates are often limited by the availability of surface or near-surface water. For example, they are close to zero over the subtropical deserts in summer.

Seasonally, mean evaporation rates vary relatively little in the tropics. However, they vary more strongly in the extratropics, where their seasonal variations do not always resemble temperature variations (compare Figs. 1.1 and 1.15). Over the extratropical continents, evaporation rates are generally higher in summer than in winter, as are the surface temperatures. Over the extratropical oceans, however, evaporation rates are usually higher in winter than in summer, despite the lower surface temperatures.

What controls the evaporation rates and their spatial and temporal variations? It turns out that the surface energy balance and the wind speed and humidity of the air immediately above the surface are crucial, as we will discuss in chapter 7.

In the global and annual mean, the evaporation rate is $\approx3.4$ mm d$^{-1}$, or $\approx1.2$ m yr$^{-1}$. The mean evaporation rate over oceans alone is similar. Given that the mean depth of the oceans is about 4 km, this implies that water molecules reside on average over 3000 years in the oceans before being cycled through the atmosphere. It is to be kept in mind, though, that this residence time of water in the oceans is an average. Ocean waters mix to the abyss only on timescales of $O(1000$ yr), so that individual water molecules can reside in the oceans for much longer or shorter periods, for example, when they evaporate soon after precipitating onto the ocean surface.

### 1.3.2 Precipitation

Because in a statistically steady state, the rate at which water vapor enters the atmosphere through evaporation must equal the rate at which it leaves the atmosphere through precipitation, the evaporation rate must equal the precipitation rate when averaged globally and over a sufficiently long time (e.g., a year). What water goes up must come down.$^{17}$ Like the evaporation rate, the precipitation rate in the global and annual mean therefore is also $\approx3.4$ mm d$^{-1}$, or $\approx1.2$ m yr$^{-1}$.

However, the distribution of mean precipitation rates ($P$) at Earth’s surface has more pronounced spatial and seasonal variations than the distribution of mean evaporation rates (Fig. 1.16). Mean precipitation rates assume their global maxima in the ITCZ and in the South Pacific Convergence Zone (SPCZ). The SPCZ is most evident in austral (southern hemisphere) summer (DJF) and extends from the maritime continent in the equatorial western Pacific south-
Figure 1.16: Precipitation rate ($P$) in the annual mean (top), DJF mean (middle), and JJA mean (bottom). The contouring is logarithmic, with factors of $\sqrt{2}$ between adjacent contour levels. The Intertropical Convergence Zone (ITCZ) and South Pacific Convergence Zone (SPCZ) stand out as bands of maximal mean precipitation in the tropics. In the global and annual mean, the precipitation rate is equal to the evaporation rate: $\sim 3.4$ mm d$^{-1}$, or $\sim 1.2$ m yr$^{-1}$. 
eastward toward French Polynesia. It arises in part because of zonal variations in the strength of easterlies (Fig. 1.11). In these tropical convergence zones, warm and moist air masses converge near the surface, rise, and cool along their way upward through the troposphere. The water vapor in the rising and cooling air masses reaches saturation, condenses, and falls out as precipitation. Mean precipitation rates in the tropical convergence zones reach values around 10 mm d$^{-1}$. From there, they drop toward local minima in the subtropics, in the sinking branches of the Hadley circulation, with precipitation rates around 2 mm d$^{-1}$ in the zonal mean and considerably lower values regionally, for example, over the deserts on continents or over eastern portions of ocean basins. Going farther poleward, mean precipitation rates increase again, reaching secondary maxima with values around 5 mm d$^{-1}$ in the midlatitude storm tracks. While mean precipitation rates over extratropical oceans are generally higher in winter than in summer, the opposite holds over extratropical continents: mean precipitation rates are generally higher in summer than in winter, because of the intense summertime convection (thunderstorms) that is typical over continents in summer.

### 1.3.3 Net precipitation

What controls the humidity or aridity of the surface climate is not the precipitation rate alone. More important is the net precipitation rate, the difference between precipitation and evaporation rates, $P - E$. Where the net precipitation rate is positive, precipitation exceeds evaporation. Over continents, the excess precipitation that does not infiltrate the soil runs off, eventually to flow through rivers into the oceans. Where the net precipitation rate is negative, evaporation exceeds precipitation. Over continents, negative net precipitation (positive net evaporation) is only possible on regional scales, where net evaporation is balanced by surface or subsurface transport of water from nearby regions with net precipitation. But on continental scales, net evaporation would require a steady continental water source, which is implausible given that continents lie above mean sea level, preventing oceans from resupplying a continental water source.

The distribution of mean net precipitation rates bears the imprint of the strong variations of mean precipitation rates onto the weaker variations of mean evaporation rates (Fig. 1.17). Precipitation exceeds evaporation in the tropical convergence zones and in the extratropics. Evaporation exceeds precipitation over the subtropical oceans. This means that the atmosphere must transport water from the subtropical oceans both into the tropical convergence zones and into the extratropics. Because ocean water that evaporates leaves its sea salt behind, continued net evaporation renders the subtropical oceans the most saline of the world. For a statistically steady state with a steady ocean salinity to be possible, the oceans must transport freshwater from the extratropics and the tropical convergence zones into the subtropics. In the driest parts of the subtropics, the net evaporation rate ($E - P$) amounts to $\geq 6$ mm d$^{-1}$ in the annual
mean, or ≥2 m yr⁻¹; this amount of freshwater must be replenished by transport of water in the oceans. Positive net precipitation in the annual mean occurs over continents. It feeds continental runoff and rivers, which transport the excess precipitation back toward the oceans, where net evaporation can occur.

Although mean precipitation rates over extratropical continents are generally higher in summer than in winter, the corresponding net precipitation rates are generally higher in winter and can even be negative in summer, indicating net evaporation of water that was seasonally stored at or near the surface (e.g., in snow packs that form in winter and melt in spring and summer). This means that in summer, a larger fraction of the precipitation over extratropical
continents comes from local sources; in winter, more is transported there from remote locations, because any net transport of water into a region by the atmosphere leaves a positive signature in the net precipitation rate. In other words, more of the rain falling in summertime convection over extratropical continents has evaporated nearby, whereas more of the wintertime precipitation has evaporated farther away.

What controls the net precipitation rates and their spatial and temporal variations? Precipitation data with global coverage are only available since 1979, when we began measurements with microwave radiometers aboard satellites. We know much less about what controls precipitation and net precipitation than what controls surface temperature. We do know that the distribution of net precipitation has varied over Earth’s history. For example, the Sahara desert had lakes in the middle Holocene, during what is known as the African Humid Period. Paintings in Saharan caves from that period depict lush vegetation, elephants, giraffes, and hippopotami, next to human swimmers in lakes. The African Humid Period ended about 5 kyr BP, the lakes dried out, and the Sahara became the desert it currently is. A rearrangement of the atmospheric circulation that led to changes in water transport must have led to these dramatic changes in African hydroclimate. But the precise circulation dynamics involved remain unclear. In chapter 9, we will discuss in greater detail factors that can change atmospheric water transport and net precipitation.

1.3.4 Precipitable Water

We have seen that in the global and annual mean, \( \sim 3.4 \text{ mm d}^{-1} \) or \( \sim 1.2 \text{ m yr}^{-1} \) of water enter the atmosphere by evaporation and leave it by precipitation. How much of it is retained in the atmosphere? The answer is, very little. The discussion of the water content of the atmosphere can first focus on water vapor because almost all (99.7%) of the water residing in the atmosphere is in the gas phase (as opposed to the liquid or ice that form clouds). The amount of water vapor in an atmospheric column is called the precipitable water; it is usually measured as the depth of the liquid layer that would result if all water vapor in the column were condensed onto the cross-sectional area of the column. In the global and annual mean, the precipitable water in Earth’s atmosphere amounts to a liquid layer on Earth’s surface only about 24 mm (about an inch) thin. The mean residence time of water vapor in the atmosphere thus is only \( \sim 24 \text{ mm}/(3.4 \text{ mm d}^{-1}) \approx 7 \text{ d} \), a much shorter timescale than the 3000-year timescale of water molecules in the oceans inferred above.

While precipitable water in the atmosphere has a small global mean, it varies strongly spatially (Fig. 1.18). It is greatest in the tropical convergence zones, where it reaches values around 50 mm. From there, it generally decreases going poleward, to lower values in the winter hemisphere than in the summer hemisphere. For example, over extratropical continents in summer, there are typically \( \sim 20 \text{ mm} \) precipitable water. Yet rainfall rates in intense summertime convective storms over continents regularly exceed 20 mm h\(^{-1}\). This would
Figure 1.18: Precipitable water in the annual mean (top), DJF mean (middle), and JJA mean (bottom). The precipitable water is expressed as the thickness of the equivalent liquid layer if all water vapor in an atmospheric column were condensed onto the cross-sectional area of the column. The global- and annual-mean precipitable water is 24 mm (about an inch).

delepte the precipitable water where it rains within ~1 h if the storm could only draw upon the precipitable water in the rainy region. Clearly, this is not the case (storms can last longer, and the atmosphere is not completely dry in their wake). Evaporation is about two orders of magnitude too small to resupply precipitable water in the rainy region. Therefore, such convective storms and atmospheric storms generally converge atmospheric water vapor from neighboring regions into the precipitation region. The rain water that falls when it pours comes from afar. And, hence, calling the amount of water vapor in an atmospheric column “precipitable water” is a bit misleading: The precipitable water can neither all
precipitate out, leaving dry air in the wake, nor does it limit the amount of rain that falls in a given storm. However, it is limiting the amount of rain that can fall averaged over regions much larger than storm scales.

1.3.5 Humidity

The concentration of water vapor in the atmosphere varies not only in the horizontal but also in the vertical. A measure of water vapor concentration in the atmosphere is the specific humidity. It is defined as the ratio \( q = \frac{\rho_v}{\rho} \) of the density of water vapor \( \rho_v \) to the density of (moist) air \( \rho \), or of the mass of water vapor per unit mass of moist air.\(^{19}\) The zonal-mean specific humidity is largest near the surface around the equator and rapidly decreases from there going poleward or upward (Fig. 1.19). Near the surface around the equator, the specific humidity reaches values exceeding 15 g kg\(^{-1}\) (i.e., 15 grams of water vapor per kilogram of air, or \( 15 \times 10^{-3} \)). It drops by three orders of magnitude to values below \( \sim 5 \times 10^{-2} \) g kg\(^{-1}\) near the tropopause. Near the surface in polar regions, it assumes values of order 1 g kg\(^{-1}\) and lower. That is, even where water vapor is most abundant in the atmosphere, it remains an atmospheric trace constituent, accounting for no more than a few percent of the mass of the atmosphere locally. In the global mean, it accounts for 0.2% of the mass of the atmosphere.

The reason water vapor concentrations in the atmosphere are so low is that they are limited by water vapor reaching saturation. Where water vapor reaches saturation, it generally condenses. Although supersaturation of water vapor can in principle occur when an insufficient number of condensation nuclei (e.g., sea salt crystals or dust particles) are available, this only occurs occasionally in isolated regions such as the upper troposphere. Therefore, water vapor concentrations in the atmosphere are generally limited by saturation.

A measure of the degree to which water vapor in the atmosphere reaches saturation is the relative humidity, \( H = e/e^* \), the ratio of the partial pressure of water vapor \( e \) to the saturation vapor pressure \( e^* \). The zonal-mean relative humidity varies far less than the specific humidity and is everywhere below 100% (Fig. 1.20). This shows that the strong poleward and upward decrease of specific humidity in the troposphere principally arises from the poleward and upward decrease of temperature (Fig. 1.10), as we will discuss in chapter 9. The zonal-mean relative humidity is generally large near the surface, where it reaches 80–90%. It assumes similar values in the interior troposphere in polar regions and near the ITCZ, in the rising branch of the Hadley circulation, which transports moist air masses from the surface upward. The relative humidity has pronounced subtropical minima in the subtropical interior troposphere, in the sinking branches of the Hadley circulation (cf. Fig. 1.14). These minima are deepest, with zonal-mean relative humidities below 20%, in the sinking branches of the strong winter Hadley cells that cross the equator. One reason for the relative dryness of the subtropical interior troposphere is that the sinking branches of the Hadley cells bring very dry air from near the tropopause.
Figure 1.19: Specific humidity in the zonal mean: annual mean (top), DJF mean (middle), and JJA mean (bottom). The contouring is logarithmic, with factors of 2 between adjacent contour levels.
Figure 1.20: Relative humidity in the zonal mean: annual mean (top), DJF mean (middle), and JJA mean (bottom).
38
CHAPTER 1

downward, with specific humidities three orders of magnitude smaller than near the surface.

It is clear that the distribution of water vapor in the atmosphere is controlled both by thermodynamics, through the temperature structure and the saturation constraint, and by circulation dynamics, which affect features such as the subtropical minima of the relative humidity and which we will study further in chapter 9. Because water vapor is a potent greenhouse gas, its distribution in the atmosphere is important for Earth’s radiative energy balance. In turn, the energy balance of Earth’s surface imparts important constraints on how the humidity of near-surface air can change as the climate changes, as we will discuss in chapter 7.

1.3.6 Clouds

Given how ubiquitous clouds are when one looks up into the sky, it may come as a surprise how little water they contain. The white appearance of clouds is due to sunlight scattered by water in its condensed phases: by liquid droplets or ice crystals, which remain suspended in clouds as long as their fall velocity is balanced by upward air motion. On average, clouds cover about 70% of the sky. But the amount of condensed water they contain is tiny. If one takes all liquid water in clouds and spreads it as a liquid layer on Earth’s surface, one would get a film ~60 \( \mu \text{m} \) thin (1 \( \mu \text{m} = 10^{-6} \text{ m} \)). This a factor ~400 less than the amount of water vapor in the atmosphere. There is even less ice: all ice in clouds when melted would give a liquid layer on Earth’s surface that is ~30 \( \mu \text{m} \) thin. That is, taking all condensed water in clouds in the sky and spreading it as a liquid layer on Earth’s surface would give a layer ~100 \( \mu \text{m} \) thin—roughly the thickness of a human hair or of a coat of paint. And yet, clouds and their condensate are tremendously important for Earth’s radiative energy balance, as we will see in chapters 4 and 7.

The amount of condensate in clouds is measured, for example, in situ by radiosondes and remotely by ground-based and space-based microwave radiometers. Each measurement technique comes with distinct errors and biases, and the amount of condensate in clouds is only known to within 15–30%. Nonetheless, atmospheric reanalyses that assimilate measurements with different sensors allow us to compile a global view of the cloud water path—the amount of cloud condensate (liquid and ice) in atmospheric columns. Expressed as the thickness of the equivalent liquid layer on the cross-sectional area of an atmospheric column, the cloud water path ranges from \( \lesssim 10 \mu \text{m} \) over subtropical deserts to \( \gtrsim 200 \mu \text{m} \) in regions of strong precipitation, such as the tropical convergence zones and the Asian summer monsoon region (Fig. 1.21). Overall, regions of large cloud water paths broadly correspond to regions of strong (convective) precipitation (Fig. 1.16).

However, the tiny global- and annual-mean cloud water path shows that it does not take much condensate to make a visible cloud. Cloud water paths \( \gtrsim 20 \mu \text{m} \) suffice to make a dense marine stratocumulus deck, like those com-
Figure 1.21: Cloud water path in atmospheric columns in the annual mean (top), DJF mean (middle), and JJA mean (bottom). The cloud water path contains the contributions from liquid water (liquid water path) and ice (ice path) and is expressed as the depth of the equivalent liquid layer (µm) on the cross-sectional area of an atmospheric column. It is also common to express the cloud water path as a mass per unit area in g m⁻². Because the density of liquid water is 10³ kg m⁻³, this has the same numerical value as the depth of the equivalent liquid layer in µm. The colorscale saturates in the areas of continental convection over South America (where cloud water paths reach ~900 µm) and in the South Asian monsoon (where they reach ~550 µm in JJA). The global- and annual-mean cloud water path is ~90 µm.
Figure 1.22: Cloud fraction (fraction of air volumes occupied by clouds) in the annual and zonal mean, averaged over the years 2006–2011. The cloud fractions are obtained from combined space-based radar and lidar data from CloudSat and CALIPSO.\textsuperscript{22}

Commonly found over eastern subtropical oceans, for example, off the coasts of California or off Chile and Peru. Only since 2006 do we have global data that reveal the three-dimensional structure of clouds, thanks to space-based radar and lidar instruments aboard the CloudSat and CALIPSO satellites. The radar instrument aboard CloudSat sends a pulse of microwave radiation downward, which can penetrate even thick clouds and allows us to infer their vertical structure from the backscatter measured by the satellite. The lidar (the word is a blend of “light” and “radar”) instrument aboard CALIPSO shines a laser downward and allows us to infer the vertical structure particularly of thin clouds from the backscatter. Combined they yield a global three-dimensional view of clouds.

In the annual and zonal mean, deep clouds reaching the tropopause are found at the ITCZ and in the extratropical storm tracks (Fig. 1.22). The ITCZ clouds regularly occupy around 30% of air volumes at altitudes \(~15\) km—higher than the cruising altitude of commercial aircraft, which therefore have to fly around rather than over the deep ITCZ clouds. Through these deep clouds, the mass in the rising branch of the Hadley circulation is transported upward. Low clouds (below \(~3\) km altitude or \(700\) hPa pressure) occur at all latitudes. Even in the subtropics, where cloud water paths are small (Fig. 1.21), low clouds occupy \(~20\)% of air volumes in the lower troposphere (Fig. 1.22). In the extratropical storm tracks, low clouds are ubiquitous, occupying \(~40\)% of air volumes in the lower troposphere.

From the vertical integral of the total cloud amount, we can derive the cloud fraction per unit area: the area fraction of the sky occupied by clouds (Fig. 1.23, top). Earth’s global-mean cloud fraction is \(~70\)%, meaning that most
of Earth is shrouded in clouds. In the tropical convergence zones and in the extratropical storm tracks, cloud fractions exceed ~80%. In the subtropics, cloud fractions are at their minimum, with values below 20% over the continental deserts. However, even over the eastern subtropical oceans, where cloud water paths are are only ~40 μm (Fig. 1.21), cloud fractions still exceed 60%. The clouds that are prevalent over the eastern subtropical oceans are primarily low clouds (Fig. 1.23, middle). For example, off the coasts of California and off Chile and Peru, vast decks of stratocumulus clouds blanket the ocean. But
low clouds are also found in most other regions, with the exception of some subtropical regions, for example, over continental deserts and over warm ocean waters (cf. Fig. 1.1). Low clouds reflect sunlight and cool the underlying ocean surface—notwithstanding the relatively small amounts of cloud condensate in them. Because of this effect on the energy balance of the surface, they are very important for the global climate. We do not fully understand how they respond to climate changes, which is the cause of the greatest uncertainties in climate change projections.

By contrast, thick clouds occur along the tropical convergence zones, convective regions over land, and in the extratropical storm tracks (Fig. 1.23, bottom panel). High fractional cover by thick clouds coincides with high precipitation rates (Fig. 1.16) and high cloud water path (Fig. 1.21). That is, strong precipitation occurs in deep convective clouds with a high cloud water path.

The separation into low and thick clouds is made for the purposes of demonstrating the spatial distribution of the two opposing ends of the cloud type spectrum. In reality, there is a seamless spectrum of cloud types, with continuous transitions between them. Capturing this continuous spectrum of cloud types remains a challenge for climate models and our understanding of the climate system.

How do low clouds arise? What controls where there are deep convective clouds? These and other questions of cloud dynamics and convection will be addressed in chapter 10.

1.4 OCEANS

What goes on in the atmosphere is highly dependent its underlying surface, two thirds of which are covered by the ocean. Through evaporation, the ocean directly provides the vast majority of atmospheric moisture that results in the vapour, cloud and precipitation distributions that we saw earlier. Oceans also absorb, store and transport energy, which is released into the atmosphere as latent and sensible heat. Additionally, by transporting and storing chemical tracers it indirectly controls the radiative properties of the atmosphere aloft. Mechanically, the ocean retards the surface atmospheric winds via momentum transfer that is dependent on its surface roughness and current speeds.

Observations of the ocean state are substantially more limited, compared to the atmosphere. They are more discontinuous in time, spatially sparse, and often suffer from large calibration and method errors. Despite these challenges, reanalysis datasets for the ocean exist, though one still needs to be aware of their shortcomings when reproducing Earth’s ocean history, especially times before the satellite era. Here we discuss the ocean properties that are most important for our climate.
1.4.1 Temperature

Ship-based observations of sea surface temperatures (SST) go back more than a century, but with these being measured rather haphazardly with no consistent protocol, constraining them is challenging. It wasn’t until the satellite era in 1970s, when a global coverage became possible with visible, infrared and microwave radiometers. International in-situ measurement programs could later be used to calibrate the satellite measurements and provide additional detail, as well as information at depth. Examples of particularly impactful programs are the NOAA’s Global Drifter Program that started in 1990s, or the international Argo float program that started providing depth-profile measurements a decade later.

On average SSTs decrease with latitude with zonal asymmetries arising due to continents (Fig. XX). As is physically justified in sections XX, the wind-driven circulation in the top ocean layer favors upwelling of deep cold waters near eastern edges of continents and downwelling of warm surface waters in the center of the basins, which results in the cold and warm anomalies on the SST map.

The depth-latitude profiles of the world’s oceans reveal some common features. In the first few hundred meters, there is a relatively homogeneous turbulent layer dominated by convection and mechanical stirring from the surface winds aloft. This "mixed layer" is separated from the deep ocean by a sharp temperature gradient, called the thermocline. The thermocline is particularly shallow and sharp in regions of upwelling and surface heating (e.g., the Equator) and very weak at polar latitudes where convection of dense (i.e. cold and saline) waters can deepen the mixed layer throughout the water column. Over time, all types of SST observations agree unequivocally that SSTs have been increasing since pre-industrial times, a very similar trend to the surface temperature plot.

1.4.2 Sea surface height

As with SST, sea surface height (SSH) has been measured for more than a century, using tide gauges. Similar problems with calibration and the resulting large measurement errors pertain, with global high-precision altimetry data from satellites with microwave radars and/or gravity sensors only becoming available from 1992.

SSH decreases with latitude as is expected from the geoidal shape due to earth’s rotation and gravity (Fig. 1.24). Elevated bulges can also be observed in the western parts of the basins, as water is in essence piled up against the continents, due to drag from surface winds and Earth’s rotation (Ekman transport, discussed in Sec. XX). The historical global record is a measure of global sea-level rise. It indicates 8 inch sea-level rise since pre-industrial times.
1.4.3 Salinity

The ship-based observation record of sea surface salinity (SSS) is shorter than for SST, around 50 years, and the satellite-based SSS record is shorter than that for SSH. Because salinity alters the emission properties of seawater by altering its electrical conductivity, satellite interferometric radiometers have recently been used to obtain a global observational coverage of surface salinity (e.g., NASA’s Aquarius and SMAP operating since 2011). Argo floats can again provide targeted depth profiles down to mid levels. The global ocean salinity is 34.8 ‰ (parts per thousand), and its spatial distribution is closely related to that of the net precipitation (last chapter). Large-scale distillation over dry subtropical “ocean deserts” produces anomalously salty waters, and the ITCZ and storm track regions receive surplus fresher water from heavy precipitation. The Atlantic is generally saltier than the Pacific largely due to the net fresh water transport into the Pacific via the equatorial winds. A freshwater imprint from the major rivers can also be observed, for example, at the outflow of the Amazon. At depth, salinity is generally more spatially uniform, with a halocline separating the well mixed layer. The halocline is disturbed around 30°N in the Atlantic, where an extremely salty tongue extends from the depths of the Mediterranean sea. Although the magnitude of these spatial salinity variations may seem small, it is crucial (along with temperature and pressure)
for determining seawater density and thus buoyancy, especially in the polar regions.

add SSS, depth, (time?)

1.4.4 Currents

1.4.4.1 Wind-driven circulation

Near-surface ocean currents are difficult to observe directly, but they can be deduced from measurements of SSH, SST and wind stress. These estimates can then be verified by in-situ measurements from the aforementioned drifters and Argo floats or ship-based measurements.

The world’s main currents are schematically overlaid on the SSH map (Fig. 1.24). The bulges in SSH due to wind drag induce an outward pressure gradient that has to be balanced by flow along the pressure contours (geostrophic balance, discussed in section XX), thus forming almost closed basin-wide circulations, called gyres. Gyres are clockwise in the northern hemisphere and anti-clockwise in the southern hemisphere. The gyres are analogous to atmospheric high pressure systems, which are also controlled by the geostrophic balance.

The western branches of the gyres are narrow and strong, and are referred to as the western boundary currents. These include the Kurushio and Gulf streams, whose heat transport is responsible for the mild climates of the northwestern coasts of North America and Europe. The returning equatorward flows are weaker and more diffuse. This east-west asymmetry is necessary, due to the sphericity of the Earth and conservation of absolute vorticity, as will be discussed in chapter XX on Sverdrup balance.

The circumpolar winds in the southern hemisphere give rise to the Antarctic Circumpolar current, which is associated with upwelling on its poleward flank and downwelling on its equatorward flank (due to Ekman balance, also discussed in chapter XX).

Sea surface height as streamfunction of near surface flow

1.4.4.2 Thermohaline circulation

Water movement at depth is induced from chemical properties of waters. Because the deep ocean experiences minimal mixing, it is assumed that water masses of similar properties have a similar origin. These properties include density (derived from temperature, salinity and pressure) and chemical composition, such as oxygen saturation. The latter is assumed to be inversely proportional to the age of the water since it was last in contact with the surface, because dissolved oxygen becomes depleted by fauna and chemical reactions once submerged.

Homogeneous properties of individual water bodies (referred to as modes) helps infer deep water movement (Fig XX). Downwelling into the depths from
the polar North Atlantic is the North Atlantic Deep Water (NADW), which is formed mainly by cooling the salty waters supplied by the Gulf Stream. This water extends southward and upwells at high latitudes of the southern hemisphere. This upwelling can be partially counteracted by the Ekman-induced overturning circulation in the wind-driven mixed layer. The southern hemisphere has its own deep water formation, the Antarctic Bottom Water (ABW), which mainly arises from sea ice formation rejecting brine and thus increasing surface water density. The ABW is the densest water mode and is located below the NADW. The North Pacific does not have a region of extensive deep water formation, and so it contains some of the oldest and most oxygen-depleted seawater on Earth. Because the deep water circulation is driven by buoyancy generation due to thermal and saline contrasts, it is called the thermohaline circulation. Although it is the wind-driven circulation that transports most of the oceanic energy, the thermohaline circulation is especially important in polar areas where it can affect sea-ice melt and in upwelling/downwelling regions where it can affect ocean heat intake, SSTs and chemical sequestration, all of which can be impactful on the global climate.

3D plot of currents

1.5 CRYOSPHERE

Cryosphere, derived from the Greek word Kryos (meaning ice), refers to all solid water bodies on Earth, including ice sheets, glaciers, snow, sea ice, icebergs and frozen ground water. The cryosphere stores approximately 2% of world’s water and three quarters of the world’s freshwater. It affects the atmosphere directly by deflecting winds by high topography and regulating the amount of sunlight that is reflected back to space, and indirectly by changing the thermohaline profiles of the polar oceans and, in turn, the global oceanic circulation. Sea ice also acts as an insulating cap, impacting ocean stability, and limiting heat and momentum exchanges between the ocean and atmosphere, with potential large-scale effects on the atmospheric temperature gradients. Permafrost, meaning permanently frozen groundwater, has also been linked to underground storage of methane (a potent greenhouse gas), though the scale of this is currently being researched.

Ice coverage can be measured with some accuracy with satellite observations. The thickness (and volume) of ice can be measured with high-precision laser and radar altimeters from the early 1990s, and gravimetry from 2002. Figure XX shows the global extent of the cryosphere. Most of cryosphere resides in land ice as ice sheets and glaciers. Polar regions contain the vast majority, with the Antarctic storing 58m of sea level rise and Greenland ice sheet 6m. But both the Antarctic and Greenland ice sheets have been losing volume, as evidenced by all satellite measurements, with the Greenland ice sheet contributing the most.
Sea ice is formed by sea water freeze. Some Arctic sea ice persists year-round, and almost all Antarctic sea ice is seasonal (remember, ice shelves are not sea ice; they are extensions of land-based ice sheets and their glaciers).

Permafrost covers 15-20% of Earth’s land and it is measured by probes in situ.

1.6 LAND
1.6.1 Temperature
Globally warming much faster than the ocean, as we saw in the atmospheric temperatures. Carbon

1.7 GREENHOUSE GASES AND AEROSOLS
Unlike H2O, well mixed

FURTHER READING

Notes
1. The 30-year time frame over which climate statistics are traditionally aggregated led to the quip that a climatologist is someone who adds 30 numbers and divides by 30. What we want to achieve here is go beyond that and train climate scientists who can answer “why?” and “whereto?” questions about climate.
2. The data are available at data-portal.ecmwf.int and are described in Dee et al. (2011).
3. If we approximate Earth as a sphere with Earth’s mean radius $r_e = 6.371 \times 10^6$ m, the meridional distance corresponding to 1° latitude is $(2\pi/360)r_e = 111$ km everywhere on Earth. Therefore, the distance from the equator to the poles is $(\pi/2)r_e = 10000$ km—a round number that is no coincidence: the meter was originally defined, by the French Academy of Sciences in 1791, as $10^{-7}$ of the length of the meridian from the equator to the pole (through Paris).
4. At greater depth, owing to Earth’s internal heat, temperatures increase at the geothermal gradient of $\sim 25$ K km$^{-1}$.
5. Figure 1.3 is based on the Berkeley Earth surface temperature dataset. The data are available at http://berkeleyearth.org.
6. The filtering of the temperature data in 1.4 is done with the technique described in Schneider
7. The most recent report of Working Group I of the Intergovernmental Panel on Climate Change summarizes the evidence that the recent warming is anthropogenic (Intergovernmental Panel on Climate Change, 2021).

8. In Fig. 1.5, the temperature reconstruction in panel (a) is from Marcott et al. (2013). The temperature reconstruction in panel (b) is based on the European Project for Ice Coring in Antarctica (EPICA) Dome C ice core (Jouzel et al., 2007). The temperature anomaly of snow formation that Jouzel et al. estimated from the relative deuterium abundance in the ice is divided by a factor 2 to obtain an approximate global-mean surface temperature anomaly. The factor 2 accounts for the polar amplification of global climate change and is estimated from climate model simulations (Masson-Delmotte et al., 2010). The temperature reconstruction in panel (c) is based on the relative $^{18}$O abundance in the shells of benthic foraminifera, recovered from ocean sediments at sites distributed around the globe (Lisiecki and Raymo, 2005). Bintanja and van de Wal (2008) used a land ice model to deconvolve the effects of ice volume and temperature on the relative abundance of $^{18}$O in foraminifera shells. The resulting deep-ocean temperature anomaly inferred by Bintanja and van de Wal (2008) was multiplied by a factor 1.59 to obtain an approximate global-mean surface temperature anomaly that matches the temperature anomaly inferred from the EPICA Dome C record in the period of overlap in a least-squares sense (cf. Masson-Delmotte et al., 2010). The confidence bands are approximations based on published uncertainty estimates: in panel (a), they are from Marcott et al. (2013); in panels (b) and (c), they are 2 K wide, roughly taking into account measurement uncertainties and the uncertainties in obtaining the estimated global-mean surface temperatures from the inferred Antarctic and deep-ocean temperatures (Masson-Delmotte et al., 2010).

9. Throughout this book, we use the year 2000 as the reference year for the $^\ominus$ scale. This makes for easier conversion to calendar years $^\ominus$ than the commonly used reference year 1950, which was adopted when radiocarbon dating with the unstable carbon isotope $^{14}$C first arose.

10. The ice sheet topography data are described in Peltier (1994) and are available from the NOAA National Climatic Data Center at ncdc.noaa.gov.

11. See Barnosky et al. (2004) and Koch and Barnosky (2006) for reviews of current knowledge about what caused the megafauna extinction.

12. The temperature reconstruction in Fig. 1.8 is from Hansen et al. (2013), who used simple physical assumptions to reconstruct global surface temperatures from the isotope abundances of $^{18}$O relative to $^{16}$O in the benthic foraminifera samples compiled by Zachos et al. (2001, 2008). The temperature reconstructions in Fig. 1.5 and 1.8 do not coincide exactly in their period of overlap, but they are consistent with each other within the estimated error bounds. The ice sheet indicators in Fig. 1.8 follow Zachos et al. (2001, 2008).

13. See Hoinka (1997) for a summary of the history of how the tropopause was discovered.

14. The term boreal, from the Greek god of the north winds Boreas, is used to refer to the northern hemisphere, as in boreal summer (JJA). It contrasts with the term austral, from the Latin term australis for southern, which is used to refer to the southern hemisphere, as in austral summer (DJF).

15. Originally, the Sverdrup is a unit of volume transport named in honor of the Norwegian oceanographer Harald Sverdrup, with 1 Sv = $10^6$ m$^3$ s$^{-1}$. With the approximate density of water of $10^3$ kg m$^{-3}$, this corresponds to a water mass transport of $10^9$ kg s$^{-1}$. As is increasingly common in the atmospheric sciences, we use this corresponding mass transport as the definition of 1 Sv. For water, this re-definition of the Sverdrup as a mass transport is essentially equivalent to the traditional definition of a volume transport; additionally, it is also suitable for a mass transport in the atmosphere, in which density varies strongly with height.

16. See Thomson (1892) and Lorenz (1983) for a history of ideas about winds and meridional overturning circulations.

17. A tiny amount of photodissociated water escapes from the top of the atmosphere to space in the form of hydrogen and oxygen ions. But because water vapor concentrations in the upper atmosphere are very small, this water loss is not significant for climate. It may amount to $O(10^{-9}$ m yr$^{-1}$) in liquid water equivalent based on measurements of O$^+$ escape (Kistler et al., 2010).

18. We can think of the atmosphere as being composed of vertical columns. In principle, these
should be spherical cones radiating out from the center of Earth. But because the atmosphere is so thin relative to Earth’s radius, it is adequate to speak of columns.

19. The modifier “specific” generally indicates quantities that are referenced to a unit mass. So specific humidity is the water vapor mass per unit mass of moist air, and the specific energy is the energy per unit mass of moist air.

20. These cloud condensate amounts are global and annual means from the ECMWF Interim Reanalysis for the years 1981–2010.

21. See O’Dell et al. (2008) for a discussion of errors in measurements of cloud condensate.

22. The combined radar (CloudSat) and lidar (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations, CALIPSO) data in Figs. 1.22 and 1.23 were compiled by Jennifer Kay (National Center for Atmospheric Research and University of Colorado) and are available at www.cgd.ucar.edu/staff/jenkay/cloudsatcaliop/. See Kay and Gettelman (2009) for a description of the analysis method used.

23. The sea surface height data in Fig. 1.24 are described in Maximenko et al. (2009) and were provided by the late Nikolai Maximenko (International Pacific Research Center) and Peter Niiler (Scripps Institution of Oceanography).
Chapter Two

Radiative Energy

Sunlight provides the energy source to essentially all atmospheric and oceanic motions. It is the ultimate driver of climate. Some of the sunlight incident at the top of Earth’s atmosphere is scattered back into space, some is absorbed in the atmosphere, and some is absorbed at the surface. The absorbed solar energy drives energy transformations within the climate system and is eventually, to close the energy balance, radiated back into space as terrestrial infrared radiation, which balances the incoming energy at the top of the atmosphere.

This chapter surveys the fundamental physical laws governing the radiative energy balance from a global perspective. It describes the characteristics of the radiation emitted by the Sun and the Earth, discusses how those characteristics depend on the temperature of the emitter, and conceptualizes how the presence of an absorbing atmosphere generally increases surface temperatures through the greenhouse effect.

2.1 SOLAR RADIATION

2.1.1 Luminosity and Radiative Energy Flux

The Sun in its core converts matter into energy by fusion of hydrogen nuclei into helium. The energy released in the fusion reaction is transported outward by radiation and convection, and is eventually emitted into space as radiative energy. The total amount of radiative energy the Sun emits per unit time is called the solar luminosity. It currently is \( L_\odot = 3.83 \times 10^{26} \text{ W} \) (the subscript \( \odot \) denotes properties of the Sun; see Table 2.1 for a list). However, the luminosity

<table>
<thead>
<tr>
<th>Table 2.1: Properties of the Sun</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luminosity ( L_\odot )</td>
</tr>
<tr>
<td>Mean radius ( r_\odot )</td>
</tr>
<tr>
<td>Mass</td>
</tr>
<tr>
<td>Effective temperature</td>
</tr>
<tr>
<td>Mean distance from Earth ( d_0 )</td>
</tr>
</tbody>
</table>
Figure 2.1: The radiative energy passing through spherical surfaces concentric with the Sun is constant and equal to the solar luminosity $L_\odot$. Because the surface area of a sphere with radius $d$ is $4\pi d^2$, the radiative energy flux $S$ that passes through a fixed area $A$ decreases with the square of the distance $d$ from the Sun, $S \propto d^{-2}$.

The radiation emitted by the Sun propagates outward into space. As it propagates through the vacuum of space, its energy is conserved. Thus, the same radiative energy per unit time, or power, passes through spherical surfaces changes over time. It fluctuates, for example, by ±0.06% over the 11-year cycle over which the number of sunspots waxes and wanes. Interestingly, the luminosity is larger when more dark and relatively cool sunspots are present, because bright hot spots called faculae co-occur with the sunspots and emit more additional radiative energy than is held back by the sunspots. On longer timescales, the solar luminosity changes as the Sun undergoes the evolution characteristic of hydrogen-fusing main-sequence stars: Nuclear fusion leads to helium accumulation in the core, which increases its mean molecular weight and density. As the core gets denser, it contracts through the action of gravity. The pressure increases, and the fusion reaction accelerates, leading to a gradual increase of the solar luminosity, in the present epoch by about 9% per billion years (Gyr). This luminosity increase has been going on since shortly after the Sun’s formation 4.57 Gyr ago; it will continue for the remaining ~5.5 Gyr of the Sun’s lifetime as a main-sequence star. Early in Earth’s history, ~4.5 Gyr ago, the solar luminosity was 30% lower than today. By the end of the Sun’s lifetime as a main-sequence star, before the hydrogen supply in its core runs out and the Sun becomes a red giant, the solar luminosity will be about 70% higher than today. This will suffice to extinguish life on Earth and burn off all its water.
Figure 2.2: The part of the electromagnetic spectrum containing the bulk of solar and terrestrial radiative energy. Frequencies $\nu$ increase and wavelengths $\lambda$ decrease from left to right. Visible light occupies the wavelength range between about 0.4 and 0.7 $\mu$m.

concentric with the Sun, irrespective of their distance from the Sun (Fig. 2.1). The radiative energy passing per unit time through each unit surface area of a sphere at a distance $d$ from the Sun is

$$S = \frac{L_\odot}{4\pi d^2},$$

(2.1)

which is essentially independent of the direction of emission because the Sun is very nearly spherical and emits radiation isotropically.\textsuperscript{3} The quantity $S$ is called the radiative energy flux density, or simply the radiative energy flux. If it is incident on a surface, it is also called the irradiance of the surface, or the insolation of the surface if the radiative energy flux in question originates at the Sun, as it does here. The solar radiative energy flux at a distance of 1 astronomical unit, $1 \text{ au} = 1.496 \times 10^{11} \text{ m} = d_0$, which is approximately the mean Earth-Sun distance,\textsuperscript{4} is known as the total solar irradiance, or sometimes also as the solar constant,

$$S_0 = \frac{L_\odot}{4\pi d_0^2} = 1362 \text{ W m}^{-2}.$$

(2.2)

The total solar irradiance varies with the solar luminosity $L_\odot$, for example, by $\pm 0.8 \text{ W m}^{-2}$ over the 11-year sunspot cycle (which is why calling it the “solar constant” is something of a misnomer).
2.1.2 Spectrum

Solar radiation is made up of electromagnetic waves, which consist of a magnetic field and an electric field that mutually support each other in a self-sustaining oscillation: a changing electric field produces a changing magnetic field, which in turn produces a changing electric field, and so on. The electromagnetic waves span a broad spectrum of frequencies and wavelengths. In order of increasing frequency $\nu$ or decreasing wavelength $\lambda$, the bulk of the solar radiative energy resides in the near-infrared, visible, and ultraviolet bands (Fig. 2.2). Frequency and wavelength are linked through the speed of light,

$$c = \nu \lambda,$$

which is a universal physical constant in a vacuum ($c \approx 3 \times 10^8$ m s$^{-1}$); in air, it is only about 0.03% lower. Hence, we can express the energy content of solar radiation as a function of either frequency or wavelength. Frequencies are often expressed as wavenumbers (inverse wavelengths) $\tilde{\nu} = \lambda^{-1} = \nu / c$, particularly in the infrared portion of the spectrum, where they are measured in inverse centimeters (1 cm$^{-1} = 10^2$ m$^{-1}$). Wavelengths are measured in nanometers (1 nm = 10$^{-9}$ m) or micrometers (also called microns, with 1 $\mu$m = 10$^{-6}$ m). For example, visible light occupies the wavelength range of about 400 to 700 nm, or 0.4 to 0.7 $\mu$m (Fig. 2.2). Throughout this book, we preferentially characterize electromagnetic radiation by its wavelength, because length scales are more intuitive and more easily linked to scales of matter than inverse time scales.

Figure 2.3 (yellow shading) shows the energy spectrum of solar radiation at the top of Earth’s atmosphere as a function of wavelength. More precisely, it shows the spectral radiative energy flux, or the spectral irradiance, incident on a surface normal to the solar beam and located at the mean Earth-Sun distance $1 \text{ au}$. (The “spectral” modifier indicates that the flux is a function of wavelength or frequency.) The integral of this spectral irradiance over wavelength is the total solar irradiance $S_0$. The peak of the spectrum occurs in the visible wavelength band, at green light, when viewed, as we do here, as a function of wavelength. (The peak occurs in the near-infrared when the spectral irradiance is viewed as a function of frequency; we will get to how to transform spectral radiative energy fluxes from wavelengths to frequencies in section 2.4.2.)

Figure 2.3 (red shading) shows the spectral irradiance that remains at Earth’s surface after the sunlight has passed through a cloudless atmosphere, under typical conditions in midlatitudes. Atmospheric constituents such as ozone, molecular oxygen, water vapor, and carbon dioxide absorb solar radiation, preferentially in distinct wavelength bands. Such absorption in distinct wavelength bands makes the solar energy spectrum at Earth’s surface more jagged than it is at the top of the atmosphere. For example, water vapor strongly absorbs solar radiation in the near-infrared wavelength band centered on 1.38 $\mu$m. In this wavelength band, the irradiance at Earth’s surface is close to zero: the atmosphere is effectively opaque (Fig. 2.3). The opacity in this wavelength band
Figure 2.3: Solar spectral irradiance at the top of the atmosphere at the mean Earth-Sun distance (yellow) and at Earth’s surface (red). The irradiances are for surfaces normal to the solar beam. The spectral irradiance at the surface is typical for conditions in midlatitudes, under cloudless skies and with an atmospheric aerosol loading representative of rural locations. Some absorption bands of atmospheric constituents are identified: ozone (O$_3$), molecular oxygen (O$_2$), water vapor (H$_2$O), and carbon dioxide (CO$_2$). The blue line shows a corresponding blackbody spectrum for the effective temperature of the Sun, $T_e = 5775$ K.$^5$

is used by satellites to detect high and thin ice clouds (cirrus clouds), because any solar radiation scattered back into space in this wavelength band must have been scattered high in the atmosphere, because otherwise it would have been absorbed by water vapor in the lower atmosphere.

2.2 BOND ALBEDO

Earth’s surface and atmosphere not only absorb solar radiation. They also scatter it, reflecting some of the solar radiation incident on Earth back into space. The fraction of radiative energy that is reflected by a body is called its albedo—a word that derives from the Latin albus, “white.” A body’s albedo (“whiteness”) measures its reflectivity: the larger the albedo, the more of the incident radiative energy the body reflects. An albedo of 0 indicates a perfect absorber, an albedo of 1, or 100%, a perfect reflector. A body’s albedo depends
Table 2.2: Typical albedos in percent

<table>
<thead>
<tr>
<th>Material</th>
<th>Albedo (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fresh snow</td>
<td>70–90</td>
</tr>
<tr>
<td>Sea ice</td>
<td>50–70</td>
</tr>
<tr>
<td>Clouds</td>
<td>20–80</td>
</tr>
<tr>
<td>Desert</td>
<td>40</td>
</tr>
<tr>
<td>Deciduous forest</td>
<td>20</td>
</tr>
<tr>
<td>Evergreen forest</td>
<td>10</td>
</tr>
<tr>
<td>Ocean</td>
<td>3–40</td>
</tr>
</tbody>
</table>

on its material properties but is not determined by them alone. The shape of the body and its surface texture play a role, too, as does the wavelength and angle of incidence of the incoming radiation. For example, an albedo that depends on wavelength gives an object its color: Leaves are green because they absorb red and blue light for photosynthesis, preferentially reflecting the green portion of the visible band. The sky is blue because air preferentially scatters the blue portion of the visible band (for reasons to be discussed in chapter 4). Clouds are white because they reflect visible light relatively independently of wavelength. Table 2.2 lists a few representative albedo values integrated over the solar spectrum.

The Bond albedo of an astronomical object refers to the fraction of the total solar radiative energy incident on the object that is reflected back into space. It is the albedo of the object as a whole for solar radiative energy integrated over all wavelengths. Earth’s Bond albedo is produced by scattering of solar radiation in the atmosphere and at Earth’s surface—processes that we will discuss in greater detail in chapter 4. Clouds are efficient scatterers of solar radiation and exert the dominant control on Earth’s Bond albedo. The highly reflective snow- and ice-covered surfaces in the polar regions also contribute substantially. Because their extent shrinks as the climate warms, which exposes darker ocean or land surfaces underneath when sea ice or land ice retreat, they provide an important feedback mechanism: the ice-albedo feedback, which can amplify climate changes.

Before the advent of satellites, Earth’s Bond albedo was estimated by measuring the albedo of Earth’s surface and of clouds from mountains or balloons, and by attempting to combine such measurements in a weighted mean to form a Bond albedo estimate. The resulting albedo estimates were in error by as much as 20%. Since the early 20th century, Earth’s Bond albedo has also been measured astronomically by measuring Earthshine—the sunlight reflected from the Earth that faintly illuminates the dark portion of the Moon’s face. Earth’s Bond albedo can be inferred by measuring the relative brightness of directly sunlit and earth-lit portions of the Moon during different lunar phases. However,
albedo estimates obtained in this way by the middle of the 20th century still had errors of at least 6%.8

Nowadays, Earth’s Bond albedo is determined by measuring the solar radiation incident on Earth and that reflected by it with broadband spectrometers on satellites in space. The Bond albedo currently is 29%, with a standard error of about 0.3%. This means that, on average, 29% of the solar energy incident at the top of Earth’s atmosphere is reflected back into space. The remaining 71% are absorbed within the atmosphere and at Earth’s surface. This is the portion that drives the climate system.

2.3 TERRESTRIAL RADIATION

The solar energy that is absorbed in the atmosphere and at Earth’s surface drives energy transformations in the climate system. Some of the absorbed solar energy heats Earth’s surface. Some is converted into kinetic energy of atmospheric and oceanic motions, giving rise to winds and circulations that eventually convert their kinetic energy to heat by turbulent dissipation. And some is used to evaporate water, which eventually condenses, precipitates, and releases its latent heat to the atmosphere. Thus, the absorbed solar energy changes the state of the absorbing matter and eventually becomes heat, often
Figure 2.5: Terrestrial spectral radiative energy flux emanating from the top of Earth’s atmosphere. The spectrum is calculated with a radiative transfer model for the U.S. standard atmosphere (Fig. 1.9), which is representative of midlatitudes. Effects of clouds are ignored in this spectrum, so it represents clear-sky conditions. Absorption bands of a few atmospheric trace constituents are identified: water vapor (H₂O), methane (CH₄), ozone (O₃), and carbon dioxide (CO₂). The effective temperature of the spectrum shown is 260 K, which is close to Earth’s effective temperature of 255 K. The blue lines show blackbody spectra for temperatures of 255 K (Earth’s effective temperature) and 289 K (Earth’s global-mean surface temperature and the surface temperature of the U.S. standard atmosphere).  

Far away from where in the climate system it was absorbed. Ultimately, to close Earth’s energy balance, Earth’s surface and atmosphere radiate the amount of energy they absorb from the Sun back into space as terrestrial radiation. While scattering and reflection do not substantially change the wavelength of the incident radiation or the state of the matter with which the radiation interacts, the emitted terrestrial radiation is thermal radiation emitted by the heated surface and atmosphere; it has almost all of its energy at longer wavelengths than solar radiation.

Like solar radiation, terrestrial radiation consists of electromagnetic waves. But unlike solar radiation, terrestrial radiation has no visible component. The terrestrial radiative energy flux emanating from the top of Earth’s atmosphere has the bulk of its energy at wavelengths between 4 and 50 µm (Fig. 2.5). This wavelength range lies in the infrared band (Fig. 2.2). Infrared radiation is the dominant wavelength band of thermal radiation emitted by bodies with temperatures in the range found on Earth’s surface and in the atmosphere. A number of atmospheric trace constituents strongly absorb infrared radiation, for exam-
ple, water vapor, carbon dioxide, methane, and ozone. As we will discuss in more detail in section 2.4.4, absorption by such trace constituents implies that Earth’s emitted radiative energy flux in some wavelength bands is strongly reduced (e.g., in the 15 μm absorption band of carbon dioxide). By contrast, in the wavelength region between 8 and 13 μm, only ozone absorbs a substantial fraction of the terrestrial radiation in a narrow absorption band. This wavelength region is known as the atmospheric infrared window, because in it, most terrestrial radiation emitted at the surface can escape directly into space, without being absorbed in the atmosphere. However, some trace gases whose concentration is currently increasing in the atmosphere because of human activities—such as methane, nitrous oxide, and chlorofluorocarbons—have absorption bands in the atmosphere’s infrared window, making them potent greenhouse gases, as we will discuss in chapter 4.

Essentially all (99%) of the terrestrial radiative energy flux has wavelengths longer than 4 μm. Conversely, essentially all (99%) of the solar radiative energy flux has wavelengths shorter than 4 μm. Thus, for discussing Earth’s climate, we can decompose the electromagnetic spectrum into two broad and distinct bands: longwave radiation emitted by Earth, and shortwave radiation emitted by the Sun. A wavelength of 4 μm serves as a convenient separation point between longwave and shortwave radiation. The longwave band is also referred to as the thermal infrared band. For purposes of discussing Earth’s radiative energy balance, it can be taken to extend to ~50 μm. Longer wavelengths make only a small contribution to radiative processes in Earth’s atmosphere. However, they are exploited in remote sensing, for example, to infer the temperature of Earth’s atmosphere from the space-based measurements of microwave radiation (λ ≥ 1 mm) mentioned in chapter 1.

What controls the shape of the radiative energy flux spectra of the Sun and Earth? Finding the answer to that question marked the dawn of quantum mechanics at the beginning of the 20th century, which originated with a theory of how heated bodies radiate energy.

2.4 BLACKBODY RADIATION

2.4.1 Cavity Radiation

Matter at temperatures above absolute zero is in a state of perpetual microscopic motion. Molecules bend and stretch, rotate and translate, and their electrons move among different energy levels (orbitals). The molecular motions cause, for example, the jiggling Brownian motion of particles that are suspended in fluids, discovered in 1827 by the botanist Robert Brown when he observed plant pollen in water through a microscope. The mean energy of the microscopic motions increases with temperature. For gases, the mean translational kinetic energy of
molecular motion, $\langle E_k \rangle$, increases with temperature $T$ like

$$\langle E_k \rangle = \frac{3}{2} kT,$$

where $k = 1.381 \times 10^{-23} \text{ J K}^{-1}$ is the Boltzmann constant, a physical constant that forms the bridge between microscopic energies and the macroscopic temperature. This relation implies that at Earth’s mean surface temperature, $T = 289 \text{ K}$, molecules have a mean translational kinetic energy of $\langle E_k \rangle = 6 \times 10^{-21} \text{ J}$. The average molar mass of molecules in air is $m_a = 29 \text{ g mol}^{-1} = 29 \times 10^{-3} \text{ kg mol}^{-1}$ because air primarily consists of 78% molecular nitrogen ($\text{N}_2$, molar mass 28 g mol$^{-1}$) and 21% molecular oxygen ($\text{O}_2$, molar mass 32 g mol$^{-1}$), plus trace constituents. Each mole contains $N_A = 6.022 \times 10^{23}$ molecules, where $N_A$ is Avogadro’s number. So the average mass of a molecule in air is $m_a/N_A$, and the mean kinetic energy per unit mass is $N_A \langle E_k \rangle/m_a = (3/2)R_0 T/m_a$, where $R_0 = k N_A = 8.315 \text{ J K}^{-1} \text{ mol}^{-1}$ is the universal gas constant. The corresponding root-mean square velocity is $\sqrt{3R_0 T/m_a} \approx 500 \text{ m s}^{-1}$ at $T = 289 \text{ K}$. That is, we are surrounded by air molecules that zip around randomly and collide with us at an average velocity of around 500 m s$^{-1}$, or 1800 kilometers per hour. Such microscopic collisions we experience as an air temperature of $T = 289 \text{ K}$ or 16°C.

Because molecules carry electric charges and because the electrons themselves can move among orbitals, the microscopic motions imply motions of electric charges. Maxwell’s theory of electrodynamics implies in the macroscopic realm that charges that are accelerated (e.g., change direction in oscillations) radiate electromagnetic waves; the same is true in the microscopic realm. Therefore, all heated bodies radiate energy in the form of electromagnetic waves. Any body placed into a colder environment cools off by radiating electromagnetic waves. Earth’s surface cools off at night by radiating electromagnetic waves, particularly efficiently when the air is dry and the skies are clear, so that few atmospheric constituents (gases and clouds) radiate electromagnetic waves downward toward the surface.

A body can be brought into equilibrium with the electromagnetic radiation it emits by enclosing the radiation field inside a cavity that prevents it from escaping (Fig. 2.6). If the walls of the cavity have a temperature $T > 0$, they emit electromagnetic radiation. Some of the emitted radiative energy is reflected at the cavity walls, and some is absorbed. The absorbed radiation returns energy to the cavity walls. After a sufficiently long time during which the radiation bounces between the cavity walls, an equilibrium between the radiation field and the walls establishes itself. If we place a second cavity next to the first with the same temperature $T$, and let the cavities exchange radiation with each other through small holes in their walls (Fig. 2.6), the radiative energy flux going from the left to the right cavity in equilibrium must be the same as that going in the opposite direction—irrespective of the material properties of the cavities. Otherwise, there would be a spontaneous net energy transfer from one cavity
Figure 2.6: Blackbody radiation in two cavities that have the same temperature $T_1 = T_2$ and that can exchange radiation through small holes. The spectral radiative energy flux for each wavelength going from the left to the right cavity must be equal to that going in the opposite direction. Otherwise, energy could flow spontaneously from one cavity to another, possibly after insertion of a spectral filter (symbolized by the dashed line) that only lets certain wavelengths pass. This would violate the second law of thermodynamics.

to another, although they have the same temperature. This would violate the second law of thermodynamics. Moreover, the spectral radiative energy flux for each wavelength $\lambda$ going from the left to the right cavity in equilibrium must be the same as that going in the opposite direction. Otherwise, we could insert a filter between the cavities that only lets certain wavelengths $\lambda$ pass, such that there is a spontaneous net energy transfer from one cavity to another. This would again violate the second law of thermodynamics. Therefore, the spectral radiative energy flux emanating from the small holes in the cavities at each wavelength can only depend on temperature. It cannot depend on other properties of the enclosing cavities, such as their material composition. Because this thought experiment also does not depend on the orientation or shape of the cavities, the emanating radiative energy flux must be the same in all directions; that is, it must be isotropic.

The radiation emanating from the small holes in the cavities is called blackbody radiation, because the cavities act like an essentially perfect (black) absorber. Any radiative energy entering the cavity through the small hole is virtually certain to be absorbed inside if the cavity walls absorb any portion of
RADIATIVE ENERGY

it: Even if some portion is being reflected when a radiation beam is incident on the walls, multiple reflections mean that eventually the radiative energy will be absorbed, as its chances of escaping are negligible if the hole is small enough. The preceding thought experiments imply that the spectral radiative energy flux for blackbody radiation at each wavelength \( \lambda \) or frequency \( \nu \) is a universal function of temperature. Determining the form of this function was a major occupation of 19th century physics. The German physicist Max Planck succeeded at the beginning of the 20th century and established the law that now bears his name and for which he was awarded the Nobel Prize in 1918.

2.4.2 Planck’s Law

Planck’s law states that the blackbody radiative energy flux into each direction in space as a function of frequency \( \nu \) and temperature \( T \) is

\[
B_\nu(T) = \frac{2\hbar \nu^3}{c^2} \frac{1}{e^{\frac{\hbar \nu}{kT}} - 1}
\]

(2.5)

where \( B_\nu \) is known as the Planck function. Here \( h = 6.626 \times 10^{-34} \text{ m}^2 \text{ kg s}^{-1} \) is the Planck constant, a universal physical constant at the root of quantum mechanics, and \( c \) is again the speed of light and \( k \) the Boltzmann constant. More precisely, the Planck function \( B_\nu \) gives the spectral radiance of blackbody radiation. The spectral radiance is the spectral radiative energy flux that is emitted into a cone of unit solid angle aperture in a particular direction, where the solid angle is a measure of angular extent in three-dimensional space that is analogous to the planar angle in two-dimensional space (see box on p. 62). It is clear that the Planck function \( B_\nu \) does not depend on any direction measure, which is consistent with blackbody radiation being isotropic. The spectral radiance is sometimes also called the radiant intensity or simply the intensity; we will drop the “spectral” modifier where the context makes it clear that the radiance in question can depend on wavelength or frequency.

Planck obtained the function (2.5) by postulating that a blackbody consists of oscillators with frequencies \( \nu \) whose energies are quantized, so that they can only take values that are integer multiples of the quantum \( h \nu \),

\[
E_n = nh\nu, \quad n = 0, 1, 2, \ldots
\]

(2.6)

Correspondingly, as pointed out a few years after Planck by Albert Einstein, electromagnetic radiation itself is quantized: it consists of energy quanta \( h \nu \) carried by individual photons, each associated with a frequency \( \nu \). The radiation field can be viewed as a gas consisting of photons with a spectrum of frequencies; its energy increases with the number of photons in it. It is a fundamental result of equilibrium statistical mechanics that the probability that a state of energy \( E_n \) is occupied is proportional to \( e^{-E_n/(kT)} \). Using the quantization postulate for the energy and the occupation probability of energy states
Just as the planar angle measures an angular extent in two-dimensional space, the solid angle measures an angular extent in three-dimensional space. The planar angle is given in radian (rad), a measure of arc length on a unit circle: 1 rad is the angle subtended at the center of the unit circle by a unit arc length on its circumference; a full circle corresponds to $2\pi$ rad (because the circumference of a unit circle is $2\pi$). Analogously, the solid angle is given in steradian (sr), a measure of an area segment on a unit sphere: 1 sr is the solid angle subtended at the center of a unit sphere by a unit area ($d\Omega = 1$) on its surface; a full sphere corresponds to $4\pi$ sr (because the area of a unit sphere is $4\pi$).

The solid angle can be defined by two planar angles, the zenith angle $\theta$ ($0 \leq \theta \leq \pi$) and the azimuth angle $\phi$ ($0 \leq \phi \leq 2\pi$).

An increment of solid angle $d\Omega$ can be expressed through the increments of zenith angle $d\theta$ and azimuth angle $d\phi$ as $d\Omega = \sin \theta \, d\theta \, d\phi$. This is the area of the rectangle on the surface of the unit sphere that is located at zenith angle $\theta$ and is traced when the zenith angle varies by $d\theta$ and the azimuthal angle by $d\phi$. The factor $\sin \theta$ arises because the length of latitude circles (circles of constant $\theta$) and with it the basis of the rectangles shortens toward the poles.

For example, if we look up at the sky, the celestial hemisphere in our field of view subtends a solid angle of $2\pi$. The Sun and the full Moon each cover $\sim 1.1 \times 10^{-5}$ of the area of the celestial hemisphere. Thus, they each subtend a solid angle of $2\pi \times 1.1 \times 10^{-5} \approx 7 \times 10^{-5}$ sr. Although the Sun’s diameter is about 400 times that of the Moon, the Earth-Sun distance is also about 400 times the Earth-Moon distance, so that the angular extent (apparent size) of the Sun and the Moon viewed from Earth is similar. This becomes evident when the Moon blocks the entire Sun during a solar eclipse.
from statistical mechanics, one can compute both the density of the photon gas in equilibrium with a blackbody and the mean energy per photon, from which Planck’s law (2.5) follows. This was the first result of quantum mechanics. It laid the foundation for a fundamental revision of our view of matter and energy and for technological breakthroughs such as the transistor and light-emitting diodes (LEDs).

The Planck function $B_\nu$ can also be expressed as a function $B_\lambda$ of wavelength $\lambda$. The radiance $B_\lambda$ integrated over the frequency interval $[\nu, \nu + d\nu]$ must be equal to the radiance $B_\lambda$ integrated over the corresponding wavelength interval $[\lambda, \lambda - d\lambda]$, so that energy is conserved in the transformation from $B_\nu$ to $B_\lambda$. Therefore, the transformation must obey $B_\nu d\nu = -B_\lambda d\lambda$, or $B_\lambda = -B_\nu d\nu/d\lambda = B_\nu c/\lambda^2$, given that $\nu = c/\lambda$. Hence, the Planck function as a function of wavelength $\lambda$ is

$$B_\lambda(T) = \frac{2h c^2}{\lambda^5} \frac{1}{e^{h c / \lambda k T} - 1}.$$  

In what follows, we will primarily view the Planck function and derived quantities as functions of wavelength, consistent with the way in which we have displayed spectra (e.g., Figs. 2.3 and 2.5); however, all results can also be expressed as functions of frequency, using the preceding transformation rule.

The Planck function’s peak shifts toward shorter wavelengths as the temperature of the blackbody increases (Fig. 2.7). To determine the wavelength of peak radiance, one can solve $\partial B_\lambda / \partial \lambda = 0$ for $\lambda = \lambda_{\text{max}}$ (and convince oneself that $B_\lambda$ indeed has a maximum at $\lambda_{\text{max}}$). The result is known as Wien’s displacement law

$$\lambda_{\text{max}} = \frac{b}{T}, \quad b \approx 0.201 \frac{hc}{k} = 2.9 \times 10^{-3} \text{ m K}. \quad (2.8)$$

It is named after the German physicist Wilhelm Wien, who derived it semi-empirically in 1893, before Planck formulated his law. Wien’s displacement law states that the wavelength of peak radiance $B_\lambda$ is inversely proportional to the temperature: the hotter a blackbody, the shorter the wavelength of peak radiance $B_\lambda$. An iron heated in fire begins to glow red, the longest wavelength of visible light, at around 750 K. As it is heated further, its glow shifts to yellow, a shorter wavelength (Fig. 2.2). Blackbodies at temperatures typically found in Earth’s atmosphere have peak radiance in the infrared band. For example, for $T = 255$ K, Wien’s displacement law gives peak radiance at $\lambda_{\text{max}} = 11$ $\mu$m—a wavelength in the middle-infrared band. By contrast, a blackbody at the temperature $T = 5775$ K typical of the Sun’s photosphere has peak radiance at $\lambda_{\text{max}} = 0.5$ $\mu$m—the wavelength of green light (see the peaks in Fig. 2.7). Note that the frequency $\nu_{\text{max}}$ at which the radiance $B_\nu$ viewed as a function of frequency peaks is not simply related to the wavelength $\lambda_{\text{max}}$ at which the radiance $B_\lambda$ viewed as a function of wavelength peaks. That is, $\nu_{\text{max}} \neq c/\lambda_{\text{max}}$ because of the requirement of energy conservation in the transformation from $B_\nu$ to $B_\lambda$. For example, the frequency of peak radiance $B_\nu$ for $T = 5775$ K is
Figure 2.7: Normalized Planck functions for the effective temperatures of the Sun (5775 K) and Earth (255 K). The radiances are normalized so that their maximum values are equal to 1. Unnormalized, the maximum value of the Planck function for \( T = 5775 \) K is a factor \( 6 \times 10^6 \) greater than that for \( T = 255 \) K. In reality, the two curves do not intersect. A wavelength of 4 \( \mu m \) serves as a convenient separation point between solar (shortwave) and terrestrial (longwave) spectra. The solar Planck function (as a function of wavelength) peaks in the visible band; the terrestrial Planck function peaks in the mid-infrared band.

Like the actual solar and terrestrial spectra (Figs. 2.3 and 2.5), the Planck functions for \( T = 5775 \) K (Sun) and \( T = 255 \) K (Earth) normalized by their peak values are well separated, with a wavelength \( \lambda \approx 4 \mu m \) as a convenient separation point between solar (shortwave) and terrestrial (longwave) radiation (Fig. 2.7). However, it is important to keep in mind that the maximum value of the Planck function for \( T = 5775 \) K is a factor \( 6 \times 10^6 \) greater than that for \( T = 255 \) K. In reality, the two Planck functions do not intersect: Because the Planck function for each wavelength \( \lambda \) is a monotonically increasing function of temperature, the Planck function for any temperature \( T_1 \) is always greater than the Planck function for a lower temperature \( T_2 < T_1 \).
2.4.3 Emissivity, Absorptivity, and Kirchhoff’s Law

Not all bodies emit like blackbodies. For a body at temperature $T$ that emits a radiance $I_\lambda$, we can introduce the emissivity

$$e_\lambda = \frac{I_\lambda}{B_\lambda(T)},$$

(2.9)

which measures how close the emitted radiance $I_\lambda$ is to the blackbody radiance $B_\lambda(T)$. A blackbody has emissivity $e_\lambda \equiv 1$ at all wavelengths $\lambda$. A body that has an emissivity $e_\lambda < 1$ that does not depend on wavelength in a given band is called grey in that wavelength band. In general, however, because the emitted radiance $I_\lambda$ can be a function of wavelength, direction of emission, temperature, etc., so can the emissivity $e_\lambda$.

A complementary quantity is the absorptivity $a_\lambda$, which measures the fraction of the radiance $I_\lambda$ incident on a body that is absorbed: if the radiance $I_\lambda$ is incident, a body with absorptivity $a_\lambda$ absorbs the portion $a_\lambda I_\lambda$. Energy conservation demands $a_\lambda \leq 1$. We have already reasoned that a blackbody absorbs all radiance incident on it, so $a_\lambda \equiv 1$ at all wavelengths for a blackbody. But in general, like the emissivity, the absorptivity $a_\lambda$ can be a function of wavelength, direction of incidence, temperature, etc. For a body that is opaque and so does not transmit radiation, the fraction of incident radiance that is not absorbed must be reflected, for energy to be conserved. Therefore, $1 - a_\lambda$ gives the reflected fraction of the radiance incident on an opaque body; it is equal to its albedo.

A thought experiment shows that the emissivity $e_\lambda$ and absorptivity $a_\lambda$ of a body in thermodynamic equilibrium are equal. Suppose a body at temperature $T$ is placed inside one of the blackbody cavities we considered before (Fig. 2.6), with cavity walls that have the same temperature $T$ as the added body inside. The cavity with the added body inside is still a blackbody cavity, with the same temperature, so the radiation field inside is still the blackbody radiation field $B_\lambda$. If the added body has an absorptivity $a_\lambda$ and emissivity $e_\lambda$, it absorbs $a_\lambda B_\lambda$ of the incident blackbody radiance and emits $e_\lambda B_\lambda$. Such absorption and emission cannot alter the blackbody radiation field. Therefore, the absorbed and emitted radiances, and with them the absorptivity and emissivity, must be equal,

$$a_\lambda = e_\lambda.$$

This is known as Kirchhoff’s law, after the German physicist Gustav Kirchhoff, who formulated and proved the law between 1859 and 1862. (He also taught the young Max Planck about thermal radiation at Berlin University in the 1870s, kindling Planck’s interest in this area of physics, which culminated in the formulation of Planck’s law.) A more elaborate version of this thought experiment also shows that the absorptivity and emissivity must have the same directional dependence. Kirchhoff’s law states that in thermodynamic equilibrium, a good absorber is a good emitter, and a poor absorber is a poor emitter.
Kirchhoff’s law is being exploited, for example, in low-emissivity (“low-e”) coatings on window glazings. Although standard glass absorbs little visible light, it has a high absorptivity and emissivity of around 0.9 for thermal infrared radiation (radiant heat). Thus, window glass absorbs radiant heat effectively, preferentially from the warmer side where more is available: from the inside in winter, and from the outside in summer. As a result, the glass heats up and then emits radiant heat to both sides. This leads to radiant heat transfer from the warmer to the colder side even if the window is insulated against thermal conduction, for example, by a gap between the window panes in double glazing. It results in radiant heat loss in winter and radiant heat gain in summer. A glass with a low-e coating can have much lower emissivity for thermal infrared radiation (down to around 0.02). It reflects most radiant heat and thus heats up much less by absorption of radiant heat. This reduces both the radiant heat loss in winter and the radiant heat gain in summer.

A corollary to Kirchhoff’s law is that the emissivity $\varepsilon$ cannot exceed 1 because energy conservation precludes an absorptivity $\alpha$ that exceeds 1. That is, the blackbody radianc $B_\lambda(T)$ is the maximum radianc a heated body at temperature $T$ can emit in thermodynamic equilibrium. Larger radiances are possible out of thermodynamic equilibrium, for example, in photoluminescence phenomena. But they are unimportant for climate. Thermodynamic equilibrium and thus Kirchhoff’s law hold locally at Earth’s surface and in the atmosphere below 60–70 km altitude. Local thermodynamic equilibrium means that sufficient interactions (collisions) among molecules occur on the macroscopic timescales of interest, so that macroscopic but still small assemblages of matter with micrometer to millimeter cross sections can be taken to be in thermodynamic equilibrium. In that case, it is meaningful to assign a temperature to the matter assemblages and view the temperature as a macroscopic measure of mean microscopic energy.

Droplets and ice crystals in clouds are very good absorbers of thermal infrared radiation (for reasons we will discuss in chapter 4), so sufficiently thick clouds absorb essentially all thermal infrared radiation incident on them. Therefore, they emit thermal infrared radiation essentially like blackbodies. This is so although they are white: while they absorb thermal infrared radiation effectively, they absorb visible light only weakly. Instead they scatter visible light, leading to their white appearance. Similarly, water, snow, and ice at Earth’s surface have a thermal infrared emissivity close to 1 (around 0.98). Land surfaces typically also have thermal infrared emissivities $\geq 0.95$, except very dry desert surfaces, whose emissivities can be $\sim 0.7$. Therefore, most land surfaces emit thermal radiation essentially like blackbodies; however, their shortwave absorption properties vary widely.

### 2.4.4 Brightness Temperature

At any wavelength $\lambda$, the Planck function $B_\lambda(T)$ is a monotonically increasing function of temperature $T$. That is, the Planck functions $B_\lambda(T)$ for different
temperatures $T$ do not intersect but lie on top of each another. This makes it possible to characterize any radiance $I_\lambda$ at a specific wavelength $\lambda$ uniquely by a brightness temperature $T_B$, which is the temperature of the blackbody that would produce the same radiance at the wavelength $\lambda$,

$$I_\lambda = B_\lambda(T_B). \quad (2.10)$$

That is, one determines an effective temperature of the emitting source by inverting the Planck function, assuming that the source has emissivity $\varepsilon_\lambda = 1$. This is especially useful in the thermal infrared band, because water, snow, ice and most land surfaces have emissivities close to 1, as do sufficiently thick clouds. Therefore, these bodies emit thermal infrared radiation essentially like blackbodies, so that their brightness temperature in the thermal infrared is close to their physical temperature.

For example, Fig. 2.8 shows the brightness temperature of the thermal infrared radiation emanating from the top of Earth’s atmosphere on October 26, 2012. The lowest brightness temperatures (~220 K) indicate emission from deep clouds, particularly in the ITCZ and in the extratropical storm tracks. These brightness temperatures roughly correspond to the physical temperatures of
the cloud tops, from where the thermal infrared radiation that escapes to space emanates. The highest brightness temperatures ($\sim 300$ K) occur over the warm, dry, and relatively cloud-free subtropical deserts. The vortex with low brightness temperature over the Atlantic off the coast of North America is Hurricane Sandy, which later transitioned into an extratropical storm dubbed “Super-storm Sandy.” In the two days before the image was taken, Sandy had crossed Jamaica and Cuba, wreaking havoc and causing dozens of deaths across the Caribbean Islands. Three days later, it made landfall on the shores of the northeastern United States, where it went down as one of the most destructive storms in history.

2.4.5 Stefan-Boltzmann Law

From the Planck function $B_\lambda$, we can calculate the total radiative energy a blackbody emits per unit time and per unit surface area. If a surface element of a blackbody subtends a solid angle $d\Omega$ to an observer that is normal to the surface element, it subtends a foreshortened solid angle $\cos \theta \, d\Omega$ to an observer at the same distance but at a zenith angle $\theta$ relative to the surface normal (Fig. 2.9). Therefore, the spectral radiative energy flux that a surface element emits into the hemisphere normal to it is

$$\int_\Omega B_\lambda(T) \cos \theta \, d\Omega = \pi B_\lambda(T).$$

Because the Planck function is isotropic, it can be pulled out of the integral on the left-hand side; the factor $\pi$ on the right-hand side arises from integrating $\cos \theta$ over all solid angles $d\Omega = \sin \theta \, d\theta \, d\phi$ in the hemisphere (see box on p. 62). To get the total radiative energy flux emerging from the surface element,
RADIATIVE ENERGY

it remains to integrate over wavelength,

\[ B(T) = \pi \int_0^\infty B_\lambda(T) d\lambda. \]

This integral is more complicated, but it can be looked up in integral tables. The result is the Stefan-Boltzmann law

\[ B(T) = \sigma T^4, \]

where the Stefan-Boltzmann constant

\[ \sigma = \frac{2\pi^5 k^4}{15c^2 h^3} = 5.670 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4} \]

is a function of the constants appearing in the Planck function. The Stefan-Boltzmann law states that the radiative energy flux emitted per unit surface area of a blackbody increases with the fourth power of temperature.

For bodies that are not black, we can again introduce an emissivity \( \epsilon \) such that the emitted radiative energy flux \( F \) is expressed as a fraction of the blackbody flux,

\[ F = \epsilon \sigma T^4. \]

The emissivity \( \epsilon \) is a spectrally integrated emissivity, obtained as the integral of the wavelength-dependent emissivities \( \epsilon_\lambda \) weighted by the Planck function \( B_\lambda \),

\[ \epsilon = \int_\Omega \int_0^\infty \epsilon_\lambda B_\lambda \cos \theta \, d\lambda \, d\Omega \]

The emissivity \( \epsilon \) can depend on temperature and other variables. Clearly, for blackbodies, \( \epsilon = 1 \).

Blackbody radiation and the physical laws governing it give us the tools for understanding the zeroth-order controls on the planetary energy balance and temperatures.

2.5 EFFECTIVE TEMPERATURE

The effective temperature of an astronomical object is its brightness temperature as a whole, integrated over all wavelengths. That is, for an object that emits a radiative energy flux \( F \), its effective temperature \( T_e \) is defined so that \( F = \sigma T_e^4 \), or

\[ T_e = \left( \frac{F}{\sigma} \right)^{1/4}. \]
This is the temperature of the blackbody that would produce the same radiative energy flux $F$.

For example, according to Eq. (2.1), the radiative energy flux $F = L/\left(4\pi r^2\right)$ emanates from the top of the Sun. With the numerical values for the solar luminosity $L$ and radius $r$ in Table 2.1, the Sun’s effective temperature follows as

$$T_e = \left(\frac{L}{4\pi\sigma r^2}\right)^{1/4} \approx 5775 \text{ K.} \quad (2.15)$$

This is a physical temperature in the Sun’s photosphere, the layer in the Sun’s outer shell from which most of the radiation that escapes to space emanates. The photosphere behaves approximately like a blackbody. The Planck function for the Sun’s effective temperature $T_e = 5775$ K is shown together with the actual solar spectral irradiance at the top of the atmosphere in Fig. 2.3. (The Planck function in the figure is integrated over a hemisphere and normalized to account for the quadratic decay of the energy flux with distance from the Sun.) It is clear that the two are close. Deviations primarily arise because the Sun’s photosphere is not isothermal, the depth from which solar radiation emanates depends on wavelength, and some gases (e.g., atomic hydrogen and molecular oxygen) absorb solar radiation in the Sun’s chromosphere, the layer just outside the photosphere.

Similarly, we can calculate the effective temperature of Earth from the solar radiative energy flux and Earth’s Bond albedo. The mean solar radiative flux incident at the top of Earth’s atmosphere is equal to the total solar irradiance $S_0 = 1362$ W m$^{-2}$. Earth intercepts this flux with a cross-sectional area equal to the area of its shadow disk in the plane perpendicular to the solar beam. Approximating Earth as a sphere with radius $r_e$, where $r_e = 6371$ km is Earth’s mean radius, the area of its shadow disk is $\pi r_e^2$ (Fig. 2.10). (Because Earth’s atmosphere is thin, $O(10 \text{ km})$, we do not need to distinguish between the radius of the solid Earth and that including the atmosphere.) Recalling that Earth’s Bond albedo is $\alpha = 29\%$, we conclude that the total solar power Earth absorbs is $S_0(1 - \alpha) \times \pi r_e^2$. This solar power is received over the entire surface of Earth, that is, over an area $4\pi r_e^2$ in the approximation of Earth as a sphere. Therefore, the mean solar radiative energy flux absorbed at the surface is

$$F = \frac{S_0(1 - \alpha)\pi r_e^2}{4\pi r_e^2} = \frac{S_0(1 - \alpha)}{4} \approx 240 \text{ W m}^{-2}.$$ 

In a steady state, Earth emits the same mean radiative energy flux as longwave radiation (Fig. 2.10). Actually, Earth’s emitted radiative energy flux is very slightly larger than the absorbed solar radiative flux. Like on many planets, an internal energy flux emanates from Earth’s interior: the geothermal heat flux. It results from radiogenic heating when long-lived radioactive isotopes in Earth’s mantle and crust decay, and from the cooling of Earth’s interior over billions of years by the release of primordial heat left over from Earth’s
Figure 2.10: Earth intercepts the solar (shortwave) radiative energy flux with a cross-sectional area equal to its shadow area in the plane perpendicular to the solar beam. It emits longwave radiation to space over its entire surface area.

formation. The geothermal heat flux drives geodynamic processes such as plate tectonics. It is conducted through Earth’s crust, slightly warms the surface and atmosphere, and is ultimately radiated away to space. However, it only amounts to ~0.1 W m$^{-2}$. This adds only a tiny fraction (0.04%) to the radiative energy flux Earth emits to space. Its magnitude is within the measurement errors of the much larger solar and terrestrial radiative energy fluxes. As important as it is for the dynamics of Earth’s interior, the geothermal heat flux is of little consequence for the energy balance of Earth’s climate system; we will neglect it.\footnote{\textsuperscript{12}} The mean longwave radiative energy flux emitted by Earth then is also $F = S_0(1 - \alpha)/4$, and Earth’s effective temperature is

$$T_e = \left( \frac{S_0(1 - \alpha)}{4\sigma} \right)^{1/4} \approx 255 \text{ K}. \quad (2.16)$$

This is about 34 K lower than Earth’s global-mean surface temperature (289 K). It corresponds to the temperature at around 5 km altitude in the midlatitude troposphere (Fig. 1.10). This is where, as we will see in chapter 6, most of the terrestrial radiation emanating from the top of Earth’s atmosphere—the so-called outgoing longwave radiation—in fact originates.
CHAPTER 2

However, where the thermal infrared radiation that escapes to space originates depends on wavelength. For example, brightness temperatures in the atmosphere’s infrared window between 8 and 13 \( \mu m \) are closer to Earth’s surface temperature, indicating emissions originating near the surface (Fig. 2.5). By contrast, brightness temperatures in strong absorption bands, for example, of carbon dioxide around 15 \( \mu m \), are well below Earth’s effective temperature \( T_e = 255 \) K. Earth’s effective temperature is merely a bulk measure characterizing the total emitted radiative energy flux.

If Earth had no atmosphere, then its mean surface temperature would be close to the effective temperature; Earth would be a frozen snowball. This is assuming two other conditions: the atmosphere-less Earth has a Bond albedo for shortwave radiation like the current Earth (although in fact the Bond albedo is dominated by scattering in the atmosphere, principally by clouds), and its surface has an emissivity for longwave radiation close to 1. Yet Earth is a habitable planet with liquid water on its surface. The reason is that Earth has an atmosphere, and the atmosphere exerts a greenhouse effect, which warms the surface.

2.6 GREENHOUSE EFFECT: ARRHENIUS MODEL

The simplest model of the greenhouse effect goes back to the Swedish scientist and Nobel laureate Svante Arrhenius, who introduced it in an 1896 paper that marks the beginning of modern climate science.\(^{13}\) It consists of a homogeneous, isothermal atmosphere in radiative equilibrium with an underlying surface. The atmosphere is assumed to be semi-grey, that is, it is transparent to shortwave radiation but has a constant emissivity \( \epsilon \leq 1 \) for longwave radiation (it is grey to longwave radiation). The surface absorbs and reflects shortwave radiation, with a shortwave albedo \( \alpha \), and it emits longwave radiation as a blackbody (Fig. 2.11). The wavelength separation between solar and terrestrial radiation is handy here: it allows us to assume that shortwave and longwave radiation interact differently with matter. For example, we can assume that the atmosphere does not absorb shortwave radiation but absorbs and emits longwave radiation. This does not violate Kirchhoff’s law because the different absorption and emission properties hold in distinct wavelength bands.

This model may be viewed as a rough first approximation of Earth’s (or another planet’s) atmosphere in the global and annual mean. The shortwave radiative energy flux incident on Earth’s surface in the global and annual mean, or the mean insolation, is \( F = S_0 \times \pi r_e^2 / (4 \pi r_e^2) = S_0 / 4 \). A portion \( aF \) of the mean shortwave flux is reflected at the surface, and a portion \( (1 - a)F \) is absorbed. The absorbed shortwave flux heats the surface to a temperature \( T_s \), and the surface emits a longwave flux \( \sigma T_s^4 \) upward, into the above-lying hemisphere, according to the Stefan-Boltzmann law. By Kirchhoff’s law, the atmosphere with emissivity \( \epsilon \) absorbs a portion \( \epsilon \sigma T_s^4 \) of the longwave flux emanating from the
Figure 2.11: Simplest model of greenhouse effect. An isothermal atmosphere with temperature $T_a$ is in radiative equilibrium with an underlying surface with temperature $T_s$. The atmosphere is transparent to shortwave radiation but grey to longwave radiation, with a longwave emissivity $\varepsilon$. The surface absorbs and reflects shortwave radiation, with a shortwave albedo $\alpha$; it emits longwave radiation as a blackbody.

surface. The remainder, $\varepsilon \sigma T_a^4$, is transmitted through the atmosphere and contributes to the outgoing longwave radiation emanating from the top of the atmosphere. A second contribution to the outgoing longwave radiation comes from the atmosphere, which is heated to a temperature $T_a$ by the longwave radiation it absorbs from the surface. It emits longwave fluxes $\varepsilon \sigma T_a^4$ both upward and downward. (Of course, it really emits in all directions, but because the model is horizontally homogeneous, we can focus on the fluxes emitted into the upward and downward hemispheres.)

The emissivity $\varepsilon$ of the atmosphere measures its longwave opacity or optical thickness. For an emissivity $\varepsilon \to 1$, the atmosphere becomes black or opaque for longwave radiation; the atmosphere is said to be optically thick in the longwave band. For $\varepsilon \to 0$, the atmosphere becomes transparent to longwave radiation; it is said to be optically thin in the longwave band. It will be instructive to consider these two limits of the model separately.
2.6.1 Energy Balance

In a steady state, energy balance at the surface demands that the absorbed shortwave flux and the downward longwave flux received from the atmosphere are balanced by the upward longwave flux that the surface emits:

\[
\frac{(1 - \alpha)S_0}{4} + \varepsilon \sigma T_s^4 = \sigma T_s^4. \tag{2.17}
\]

Similarly, energy balance in the atmosphere demands that the portion of the surface longwave flux that the atmosphere absorbs is balanced by the upward and downward longwave fluxes that the atmosphere emits:

\[
\varepsilon \sigma T_s^4 = 2 \varepsilon \sigma T_a^4. \tag{2.18}
\]

The sum of the balances at the surface and in the atmosphere gives the energy balance at the top of the atmosphere:

\[
\frac{(1 - \alpha)S_0}{4} = (1 - \varepsilon)\sigma T_s^4 + \varepsilon \sigma T_a^4. \tag{2.19}
\]

The absorbed shortwave flux (left-hand side) balances the outgoing longwave flux (right-hand side), which consists of the longwave flux originating at the surface and transmitted through the atmosphere (first term), plus the longwave flux originating in the atmosphere (second term). The total outgoing longwave flux only depends on the total solar irradiance and the surface albedo; it does not depend on the atmosphere’s longwave emissivity. However, the partitioning between longwave fluxes originating at the surface and those originating in the atmosphere changes with longwave emissivity. For an optically thick atmosphere (\( \varepsilon \to 1 \)), an observer in outer space would only detect the longwave flux originating in the atmosphere but not that from the surface. The longwave emission of the surface is fully absorbed in the atmosphere, and the longwave flux emitted by the atmosphere alone balances the absorbed shortwave flux. For an optically thin atmosphere (\( \varepsilon \to 0 \)), an observer in outer space would only detect the longwave flux originating at the surface; the longwave flux emitted by the surface alone balances the absorbed shortwave flux. For a grey atmosphere with intermediate longwave emissivity, outgoing longwave radiation in this simple model consists of a mixture of radiative fluxes originating at the surface and in the atmosphere, as is the case for Earth (Fig. 2.5).

2.6.2 Temperatures

We can use the definition (2.16) of the effective temperature \( T_e \) as a function of total solar irradiance \( S_0 \) and albedo \( \alpha \) to substitute for the insolation term and write the energy balance equations (2.17)–(2.19) solely in terms of the temperatures \( T_s, T_a, \) and \( T_e \). Solving for the atmosphere and surface temperatures in
terms of the effective temperature gives

\[ T_a = \left( \frac{1}{2 - \varepsilon} \right)^{1/4} T_e, \]  
\[ T_s = \left( \frac{2}{2 - \varepsilon} \right)^{1/4} T_e, \]  

from which it follows that

\[ T_s = 2^{1/4} T_a. \]  

In this simple model, then, the surface temperature \( T_s \) is always a factor \( 2^{1/4} = 1.19 \) greater than the atmosphere temperature \( T_a \). It is a factor \( 2/(2 - \varepsilon) \)^{1/4} \( \geq 1 \) greater than the effective temperature \( T_e \). For a given effective temperature \( T_e \), the surface temperature \( T_s \) increases with emissivity \( \varepsilon \), until the limit of a black atmosphere (\( \varepsilon = 1 \)) is reached. The warming of the surface above the effective temperature is referred to as the greenhouse effect of the atmosphere. Although the greenhouse effect warms the surface, it does not change the outgoing longwave radiation, and so it does not change the energy balance at the top of the atmosphere.

Some limiting cases are worth discussing explicitly:

**Optically thick atmosphere** (\( \varepsilon \to 1 \)). The atmosphere behaves like a blackbody, and its temperature is equal to the effective temperature, \( T_a = T_e \), because the outgoing longwave radiation consists only of the longwave flux originating in the atmosphere. The greenhouse effect raises the surface temperature by a factor \( 2^{1/4} \) above the effective temperature, \( T_s = 2^{1/4} T_e \).

**Optically thin atmosphere** (\( \varepsilon \to 0 \)). The atmosphere becomes essentially transparent to longwave radiation, and its temperature \( T_a = 2^{-1/4} T_e \) is a factor \( 2^{-1/4} = 0.84 \) smaller than the effective temperature \( T_e \). The surface temperature is equal to the effective temperature, \( T_s = T_e \), because the outgoing longwave radiation consists only of the longwave flux originating at the surface. Although the atmosphere does not contribute to the outgoing longwave radiation, the vanishingly small fraction \( \varepsilon \to 0 \) of the longwave flux from the surface that it absorbs heats the atmosphere to the temperature \( T_a = 2^{-1/4} T_e \). This is the so-called skin temperature that any optically thin (“skin”) layer near the top of an atmosphere assumes if it is only heated by longwave radiation that upwells from below.

For Earth’s effective temperature \( T_e = 255 \) K, this model with a black atmosphere (\( \varepsilon = 1 \)) gives a surface temperature \( T_s = 303 \) K—higher than Earth’s actual global-mean surface temperature of 289 K, but closer to it than the effective temperature. The skin temperature of an optically thin layer at the top of the atmosphere is always, irrespective of the temperature of the underlying atmosphere, a fixed fraction of the effective temperature, \( 2^{-1/4} T_e \). For Earth, the skin
temperature is \((255/2^{1/4}) \text{ K} = 214 \text{ K}\). This is close to the temperature in Earth’s lower stratosphere (Fig. 1.9), which is optically thin and is primarily heated by longwave radiation upwelling from below. However, temperatures increase with height in the stratosphere because higher layers of the stratosphere are additionally heated by absorption of solar radiation, primarily absorption of visible and ultraviolet light by ozone.

With a lower emissivity, the surface temperature in the model could be brought into closer agreement with Earth’s global-mean surface temperature. For example, with \(\epsilon = 0.8\) and \(T_\odot = 255 \text{ K}\), the model gives \(T_s \approx 289 \text{ K}\). But the primary reason the model with a black atmosphere (\(\epsilon = 1\)) gives a surface temperature that is too high is that it ignores the dynamical energy transport by atmospheric motions: Earth’s atmosphere and surface are not in radiative equilibrium with each other but exchange energy through, for example, sensible and latent heat fluxes. Moreover, energy is being transported not just radiatively but also dynamically within the atmosphere, for example, by convection (rapid, primarily vertical motion) and large-scale weather systems. We will further discuss dynamical processes and deepen our understanding of the greenhouse effect in chapter 6.

The importance of the greenhouse effect for Earth’s climate cannot be overstated. It makes Earth a habitable planet. It has probably moderated the climate since early in Earth’s history. For example, although the solar luminosity early in Earth’s history was 30% lower than it is today, geological evidence indicates that liquid water was present on Earth soon after Earth’s formation. This gives rise to what is known as the Faint Young Sun problem: how could a climate warm enough for liquid water have been sustained although the Sun was so faint? The answer almost certainly involves a strong atmospheric greenhouse effect early in Earth’s history, possibly paired with reduced cloud cover and a reduced Bond albedo.

FURTHER READING


G. B. Rybicki and A. P. Lightman, 2004: Radiative Processes in Astrophysics. Wiley-VCH. Discussion of the fundamental laws of radiation relevant for atmospheres, including, for example, a derivation of Planck’s law.

Notes


2. For models of luminosity variations over the lifetime of the Sun, see Schwarzschild (1958),
Gough (1981), and Bahcall et al. (2001).

3. The oblateness of the Sun as measured by the relative difference of polar and equatorial radii amounts to only \( \sim 10^{-5} \), making the Sun the most spherical object of the solar system (Kuhn et al., 2012; Gough, 2012).

4. Historically, the astronomical unit had been defined as the length of the semi-major axis of Earth’s orbit around the Sun. The astronomical unit is now defined as a fixed length, 1 au = 149,597,870,700 m. This is within a factor of \( 10^{-6} \) of the semi-major axis of the orbit.

5. The top-of-atmosphere spectrum in Fig. 2.3 is the American Society for Testing and Materials (ASTM) extraterrestrial reference spectrum, which is derived by combining data from a variety of satellites, aircraft, and solar telescopes with models of solar spectral irradiance. The surface spectrum is the ASTM standard reference spectrum of direct solar irradiance, derived for the U.S. standard atmosphere (Fig. 1.9), with solar zenith angle, water vapor path, and ozone concentrations representative of annual-mean midlatitude conditions. The data with detailed descriptions are available at rredc.nrel.gov/solar/spectra/am1.5/.

6. The albedo values in Table 2.2 are based on satellite measurements by Briegleb et al. (1986).

7. Figure 2.4 is the blue marble image from visibleearth.nasa.gov.


9. The spectrum is calculated with the MODTRAN model of infrared radiation in the atmosphere, available at forecast.uchicago.edu/modtran.html.

10. See, e.g., Feynman et al. (1963) or Rybicki and Lightman (2004) for derivations of Planck’s law.

11. The infrared brightness temperatures data in Fig. 2.8 are described in Janowiak et al. (2001).

12. Although they are unimportant for Earth’s climate, heat fluxes emanating from the planetary interior can be important for the climate of other planets. For example, heat fluxes of about 6 W m\(^{-2}\) and 2 W m\(^{-2}\) are emanating from the deep interiors of Jupiter and Saturn, respectively (Guillot, 1999; Guillot et al., 2004). They primarily arise because the planets are still contracting, and the gravitational energy that is released by the contraction emanates as an internal energy flux. These internal heat fluxes are of similar magnitude as the solar radiative energy fluxes absorbed on the planets and likely are essential for their climate dynamics, such as the structure of their observed jet streams (Schneider and Liu, 2009; Liu and Schneider, 2010).

13. See Archer and Pierrehumbert (2011) for a historical overview and reprint of Arrhenius’ paper.
Chapter Three

Earth’s Orbit and Insolation

The energy that drives Earth’s climate enters the climate system at the top of the atmosphere as solar radiation. The insolation at the top of the atmosphere—the solar radiative power that passes through a unit area tangential to the top of the atmosphere—depends on two factors: the solar radiative energy flux at the top of the atmosphere, and the elevation of the Sun in the sky. The solar radiative energy flux varies primarily with the Earth-Sun distance, and secondarily with the solar luminosity (whose variations dominate flux variations on timescales of billions of years). The elevation of the Sun in the sky varies with latitude, time of day, and time of year.

Thus, to understand insolation and its variations, we need to discuss how Earth orbits the Sun and how the Earth-Sun distance varies along the orbit, how to determine the elevation of the Sun in the sky, and how Earth’s orbit varies on long timescales, leading to long-term climate variations such as glacial-interglacial cycles.

3.1 ORBITAL GEOMETRY

Over the course of a year, the Earth-Sun distance varies by \(\pm 1.7\%\) around its mean value \(d_0 = 1\) au because Earth’s orbit deviates slightly from being circular (Figs. 3.1 and 3.2). The orbit is approximately an ellipse with the Sun at one of the two foci, an observation that we now call Kepler’s first law. It is named after the German scientist Johannes Kepler, who formulated it in 1609 after carefully analyzing the meticulous observations that the Danish astronomer Tycho Brahe had made with naked eyes in the preceding years.\(^1\) It later turned out that such an ellipse is the solution of what is now called the Kepler problem: the two-body problem of spherical masses moving under mutual gravitational attraction according to the laws Isaac Newton formulated in 1687. The elliptical orbit that solves the Kepler problem for Sun and Earth—or more precisely, for the Sun and the barycenter (center of mass) of the Earth-Moon system\(^2\)—is a sufficiently accurate approximation of Earth’s orbit for purposes of understanding climate. For example, Jupiter, the solar system’s most massive planet by far, exerts a gravitational pull on Earth that is about a factor \(10^{-4}\) smaller than that of the Sun. The factor \(10^{-4}\) arises because Jupiter is about a factor \(10^3\) less massive than the Sun and, at closest approach to Earth, is about a factor 4 farther away,
Figure 3.1: Earth’s orbit around the Sun viewed at an oblique angle from above the orbital plane. Day and night have equal length everywhere on the planet at the equinoxes in March and September. The shortest and longest day and night occur at the solstices: northern summer solstice, when the north pole faces the Sun, occurs in June; southern summer solstice, when the south pole faces the Sun, occurs in December. The tilt (obliquity) of Earth’s spin axis relative to the orbital plane gives rise to the seasons, defined as the periods between the solstices and equinoxes.

with the distance entering gravitational forces through an inverse square law. Such small perturbations to the Earth-Sun two-body problem do not modify the elliptical geometry of Earth’s orbit substantially enough to affect climate on timescales of years or less. However, on timescales of millennia, they do lead to variations of parameters characterizing the orbit, which are important for long-term climate variations, as we will discuss in section 3.5.

As an ellipse, Earth’s orbit is characterized by the length of the semi-major axis $a$ and by the eccentricity $e$, which together determine the distance $d_e e$ from the foci to the center of the ellipse. For $e = 0$, the ellipse is a circle; for $e \to 1$, the ellipse becomes infinitely elongated. The distance from the Sun at a focus to Earth on the elliptical orbit is given by the equation for an ellipse in heliocentric polar coordinates,

$$d = d_e \frac{1 - e^2}{1 + e \cos A}.$$  \hspace{1cm} (3.1)

This distance $d$ varies with the angle coordinate $A$, which is called the true anomaly. (The archaic term “anomaly” in celestial mechanics refers to angular distance measures. This contrasts with the use of “anomaly” for deviations from an average common in the climate sciences.) The true anomaly measures
Figure 3.2: Earth’s orbit around the Sun viewed from above the orbital plane. At perihelion (currently in January), Earth is closest to the Sun; at aphelion (currently in July), it is farthest away. The true anomaly $A$ is the angle subtended at the Sun by the orbital arc from perihelion to Earth’s position. The solar longitude $L_s$ is a shifted true anomaly, measured from vernal equinox instead of perihelion. The true anomaly and solar longitude are related by $A = L_s - \varpi$, where $\varpi$ is the longitude of perihelion, the angle subtended at the Sun by the orbital arc from vernal equinox to perihelion.

By definition of the true anomaly, perihelion is reached at $A = 0^\circ$, where the Earth-Sun distance attains its minimal value

$$d_p = d_e(1 - e).$$

Conversely, the point on the orbit when Earth is farthest from the Sun is called aphelion. Aphelion is reached at $A = 180^\circ$, where the Earth-Sun distance attains
its maximal value

\[ d_a = d_a(1 + e). \]

Earth’s orbit has a small eccentricity, currently \( e = 0.017 \). The semi-major axis \( d_0 \) had historically defined the astronomical unit and now is within a factor \( 10^{-6} \) of the current definition of the astronomical unit as a fixed length \( d_0 \). Thus, we can equate \( d_a = d_0 \) to excellent accuracy and obtain that the Earth-Sun distance varies between 0.983 au at perihelion and 1.017 au at aphelion, or by about \( \pm 1.7\% \) around its mean value of 1 au (Fig. 3.2). The apparent diameter of the Sun when viewed from Earth likewise varies by about \( \pm 1.7\% \), because the apparent diameter is inversely proportional to the distance: the Sun appears slightly larger at perihelion than at aphelion. By measuring the variation of its apparent diameter over the course of a year, the eccentricity of Earth’s orbit can be determined.

### 3.2 ORBITAL MOTION

#### 3.2.1 Kepler’s Second Law

To calculate the Earth-Sun distance \( d(t) \) at any time of year \( t \), we need to relate Earth’s true anomaly \( A(t) \) along the elliptical orbit (3.1) to the time \( t \). This is not straightforward because the true anomaly does not increase linearly with time. Instead, the true anomaly advances more rapidly when Earth is closer to the Sun, as a consequence of angular momentum conservation. If we again neglect perturbations by other planets, the only force Earth experiences along its orbit is the central gravitational pull toward the Sun. No forces act tangentially to the orbit, so relative to an axis perpendicular to the orbital plane and going through the Sun, no torques act on Earth. Therefore, the angular momentum of Earth orbiting the Sun is conserved. Because Earth’s orbital angular momentum relative to the Sun is \( h_e = d^2 \dot{A} \) per unit mass, where \( \dot{A} = dA/dt \) is the orbital angular velocity and the squared distance \( d^2 \) is the moment of inertia per unit mass, conservation of angular momentum implies that \( h_e \) is a constant of the orbital motion. This gives

\[ \dot{A} = \frac{h_e}{d^2}, \]

which is Kepler’s second law: the orbital angular velocity varies with the inverse square of the distance to the Sun. Like a pirouetting ice skater who spins more rapidly when she moves her limbs closer to her spin axis, Earth orbits the Sun more rapidly when it is closer to the Sun. The angular velocity \( \dot{A} \) is greater at perihelion than at aphelion, varying by about \( \pm 3.5\% \) between the two, because it varies like \( d^{-2} \) and the distance \( d \) varies by \( \pm 1.7\% \). As a consequence, with summer and winter defined as the seasons from the solstices (when either pole faces the Sun) to the following equinox (Fig. 3.1), northern hemisphere summer
Table 3.1: Earth's orbital parameters (January 1, 2000, 11:59 UTC)\(^5\)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semi-major axis (d_s)</td>
<td>1.000 au</td>
</tr>
<tr>
<td>Eccentricity (e)</td>
<td>0.017</td>
</tr>
<tr>
<td>Obliquity (\gamma)</td>
<td>23.44°</td>
</tr>
<tr>
<td>Longitude of perihelion (\omega)</td>
<td>282.95°</td>
</tr>
<tr>
<td>Mean anomaly (M_0)</td>
<td>357.5°</td>
</tr>
<tr>
<td>Anomalistic year (Y_a)</td>
<td>365.26 d</td>
</tr>
</tbody>
</table>

(during which aphelion occurs) currently is about 4.5 d longer than southern hemisphere summer.\(^3\) Such differences in the length of seasons can be and have been used to infer the eccentricity of Earth’s orbit.

How the orbital angular velocity \(\dot{A}\) depends on orbital parameters and on the time of year as measured by the true anomaly \(A\) becomes clearer when substituting the expression

\[
h_e = \frac{2\pi}{Y_a} (1 - e^2)^{1/2} d_s^2
\]

for the orbital angular momentum per unit mass. Here, \(Y_a\) is the orbital period known as the anomalistic year: it is the time it takes to complete one revolution from perihelion to perihelion, currently 365.26 d for Earth (Table 3.1). Substituting this expression for the orbital angular momentum \(h_e\) and the equation (3.1) of the orbital ellipse for the distance \(d\) in Kepler’s second law (3.2) leads to a differential equation for the orbital angular velocity \(\dot{A}\),

\[
\dot{A} = \frac{2\pi (1 + e \cos A)^2}{Y_a (1 - e^2)^{3/2}}. \tag{3.4}
\]

This equation shows explicitly that the rate of change of the true anomaly \(\dot{A}\) depends nonlinearly on the true anomaly \(A\) itself. Because of this nonlinearity, a closed-form solution for the true anomaly is not known, despite centuries of attempting to find one. However, the nonlinearity is only of order eccentricity, which is small for most planetary orbits, including Earth’s. Therefore, approximate solutions can be developed by expanding in power series in eccentricity.

### 3.2.2 Mean Anomaly

A first approximation of the true anomaly follows by integrating the equation (3.4) for zero eccentricity, imposing the condition that \(A = 0^\circ\) at perihelion. The
resulting zeroth-order approximation \( A \approx M + O(e) \) is called the mean anomaly

\[
M = \frac{2\pi(t - t_p)}{Y_a}.
\]  

(3.5)

The mean anomaly \( M \) increases linearly with time \( t \) since perihelion, which occurs at time \( t_p \). It is periodic with each revolution of Earth around the Sun and gives the angular distance that Earth would have moved from perihelion if it orbited the Sun along a circle (\( e = 0 \)) with constant speed and with its actual orbital period \( Y_a \).

As an alternative to measuring time relative to the time of perihelion \( t_p \), time is also often measured relative to a reference time \( t_0 \), which need not coincide with perihelion. In that case, the mean anomaly is given by

\[
M = \frac{2\pi(t - t_0)}{Y_a} + M_0,
\]  

(3.6)

where \( M_0 = 2\pi(t_0 - t_p)/Y_a \) is the mean anomaly at \( t_0 \), chosen so that \( M = 0^\circ \) at perihelion. For example, the time \( t_0 \) may define an astronomical epoch, epoch being the astronomical term for a reference moment. The time 11:59 UTC on January 1, 2000 defines the astronomical epoch called J2000 (for Julian year 2000). At this epoch, the mean anomaly was \( M_0 = 357.5^\circ \) (Table 3.1), meaning that this was \( M_0 Y_a/360^\circ \approx 363 \text{ d after or 2 d before perihelion, which occurred on January 3 both in 1999 and in 2000.} \)

The Earth-Sun distance to first order in eccentricity follows by a binomial expansion of the equation (3.1) for the orbital ellipse in eccentricity, \( d \approx d_\alpha(1 - e \cos A) + O(e^2) \), and substitution of the mean anomaly \( M \) for the true anomaly \( A \):

\[
d \approx d_\alpha (1 - e \cos M) + O(e^2). \]  

(3.7)

That is, to first order in eccentricity, the Earth-Sun distance varies sinusoidally with the time of year as measured by the mean anomaly \( M \).

### 3.2.3 True Anomaly

The next approximation of the true anomaly can be obtained by substituting the mean anomaly for the true anomaly on the right-hand side of the equation (3.4) for the orbital angular velocity. Integrating while retaining only terms of first order in eccentricity yields \( A \approx M + 2\pi \sin M + O(e^2) \). Higher-order expansions lead to a Fourier sine series in \( M \), which at third order in eccentricity is

\[
A \approx M + \left( \frac{2\pi - e^3}{4} \right) \sin M + \frac{5}{4} e^2 \sin 2M + \frac{13}{12} e^3 \sin 3M + O(e^4). \]  

(3.8)

This series satisfies \( M = A = 0^\circ \) at perihelion as required by the definition of the true anomaly, and it also satisfies \( M = A = 180^\circ \) at aphelion, as required by
the symmetry of the ellipse (Fig. 3.2).

Substituting the approximate true anomaly \(3.8\) in the equation \(3.1\) for the orbital ellipse gives the Earth-Sun distance to terms of order \(e^4\). For Earth’s current eccentricity of \(e = 0.017\), this is accurate to within a factor of about \(10^{-7}\) —more than sufficient for practical purposes in climate dynamics.

### 3.2.4 Solar Longitude

To calculate insolation at the top of the atmosphere, the position of Earth on the orbit is usually expressed in terms of a shifted true anomaly called the solar longitude,

\[
L_s = A + \omega,
\]

which gives the angular distance Earth has moved along its orbit since vernal equinox: the solar longitude \(L_s\) is the angle subtended at the Sun between Earth’s position and the position at vernal equinox (Fig. 3.2). It is obtained from the true anomaly by adding the longitude of perihelion \(\omega\) (pronounced “curly pi”), which is the angular distance along the orbit from vernal equinox to perihelion.

The solar longitude defines the astronomical seasons (Figs. 3.1 and 3.2). By definition, \(L_s = 0^\circ\) at vernal equinox, which marks the beginning of northern spring and southern fall. Northern summer solstice, when the north pole faces the Sun, occurs at \(L_s = 90^\circ\) and marks the beginning of northern summer and southern winter. Autumnal equinox occurs at \(L_s = 180^\circ\) and marks the beginning of northern fall and southern spring. Finally, northern winter solstice or southern summer solstice, when the south pole faces the Sun, occurs at \(L_s = 270^\circ\) and marks the beginning of northern winter and southern summer. At the equinoxes, the subsolar point (the point on Earth’s surface where the Sun is directly overhead) crosses the equator, moving northward at vernal equinox and southward at autumnal equinox. The solar terminator (the line separating the dark night side from the illuminated day side of Earth) follows a meridian at the equinoxes and so is perpendicular to the equator, making night and day equally long everywhere on Earth. Thus, the amounts of solar energy incident on the northern and southern hemispheres at the top of the atmosphere are equal at the equinoxes. At the solstices, by contrast, the subsolar point reaches its maximum poleward excursion: its maximum northward excursion occurs at northern summer solstice at the Tropic of Cancer \(23.4^\circ\text{N}\), and its maximum southward excursion occurs at northern winter or southern summer solstice at the Tropic of Capricorn \(23.4^\circ\text{S}\). Thus, the amounts of solar energy incident on each hemisphere at the top of the atmosphere differ the most at the solstices.

### 3.2.5 Aligning Calendar and Seasons

To measure the time of year in a way that aligns with astronomical seasons, one could use the solar longitude \(L_s\). Using the solar longitude as a measure of the
time of year is indeed common for other planetary bodies, such as Mars, with \( L_s = 0^\circ \) marking vernal equinox and hence the beginning of northern spring. For Earth, we commonly use a calendar to identify a time of year. The internationally most widely used calendar is the Gregorian calendar, introduced in 1582 by Pope Gregory XIII. It was introduced to keep vernal equinox near March 21. The Gregorian calendar fixed a drift of the vernal equinox by about 3 days over 400 years in the until then common Julian calendar, which had been introduced by Julius Caesar in 46 BC. The Julian calendar assumed a year that was a few minutes too long. The Gregorian calendar refined the Julian system of leap years to account for the non-integer number of days in a year and fix the secular drift of vernal equinox across the calendar.

For periods deep in Earth’s geological past, we could extrapolate the Gregorian calendar system into the past. However, over timescales of tens to hundreds of thousand years, the Gregorian calendar does not prevent a secular drift of vernal equinox across the calendar. For example, around 50,000 BC, vernal equinox occurred in December [check] according to the Gregorian calendar. To use a calendar that resembles the modern calendar, with the beginning of northern spring around March 21, it is conventional for times deep in the geological past to align the calendar with astronomical seasons, by requiring vernal equinox \( (L_s = 0^\circ) \) to occur on March 21.

Aligning vernal equinox with March 21 suggests using the time \( t_{v,0} \) of vernal equinox in a specific year as the reference time to define the mean anomaly,

\[
M = \frac{2\pi(t - t_{v,0})}{\gamma_a} + M_{v,0}, \tag{3.10}
\]

where \( M_{v,0} \) is the mean anomaly at vernal equinox in the reference year. The time \( t_{v,0} \) is then simply defined as March 21 in the reference year. Because, as we will discuss in section 3.5, the longitude of perihelion advances (“precesses”) slowly in the direction of Earth’s orbital motion, completing one revolution over around 21 kyr, the mean anomaly at vernal equinox retreats slowly from year to year, by about 360°/21000 yr \( \approx 0.017^\circ \) per year. Consequently, the time it takes to complete one revolution from equinox to equinox—the period known as the tropical year—is slightly \( (1/21000 \text{ yr} \approx 35 \text{ min}) \) shorter than the anomalistic year \( \gamma_a \), the time from perihelion to perihelion.

The mean anomaly \( M_{v,0} \) at vernal equinox in the reference year can be calculated from the longitude of perihelion \( \omega \) in the reference year through an inversion of the relation \( (3.8) \) between the true anomaly and the mean anomaly. To third order in eccentricity, this gives\(^10\)

\[
M_{v,0} = -\omega + \left( e + \frac{e^3}{4} \right) (1 + \beta) \sin \omega - \frac{1}{2} e^2 \left( \frac{1}{2} + \beta \right) \sin 2\omega
+ \frac{1}{4} e^3 \left( \frac{1}{3} + \beta \right) \sin 3\omega + O(e^4), \tag{3.11}
\]
Figure 3.3: Solar radiative energy flux $S$ at the top of the atmosphere as a function of time of year. The $\pm 3.5\%$ variations between its maximum at perihelion (around January 3) and its minimum at aphelion (around July 4) are clearly evident. The dotted line shows the total solar irradiance $S_0 = 1362 \text{ W m}^{-2}$.

where $\beta = (1 - e^2)^{1/2}$. For example, for Earth’s longitude of perihelion $\omega \approx 282.95^\circ$ and eccentricity $e = 0.017$ in the year 2000 (Table 3.1), this gives a mean anomaly at vernal equinox of $M_{v,0} = -284.8^\circ$, or, by periodicity, $M_{v,0} = 75.2^\circ$. It implies perihelion (January 3, 2000 and January 4, 2001) occurred $Y_0 \times M_{v,0}/360^\circ \approx 289$ d after, or 76 d before, vernal equinox (March 21, 2000). With the relations (3.1)–(3.11), we have the tools to calculate the Earth-Sun distance $d(t)$ at any time $t$, given orbital parameters such as the eccentricity and longitude of perihelion. And from the distance $d(t)$, we can obtain the solar radiative energy flux at the top of the atmosphere.

3.3 SOLAR RADIATIVE ENERGY FLUX

3.3.1 Variation Along Orbit

The solar radiative energy flux at the top of the atmosphere is given by

$$S = S_0 \left( \frac{d_0}{d} \right)^2,$$

where the Earth-Sun distance $d$ enters through the inverse square law (2.1), and the total solar irradiance $S_0 \propto L_\odot$ varies with the solar luminosity $L_\odot$ according to its definition (2.2). Figure 3.3 shows the solar radiative energy flux at the top of the atmosphere, obtained using a fixed total solar irradiance $S_0 = 1362 \text{ W m}^{-2}$ and using the preceding series expansions to third order in eccentricity to calculate the Earth-Sun distance with the orbital parameters in Table 3.1. It is manifest that the solar radiative energy flux varies sinusoidally.
between its maximum at perihelion and its minimum at aphelion. Indeed, because Earth’s eccentricity is small, the variations are dominated by the first-order term in eccentricity,

$$S \approx S_0(1 + 2e \cos M) + O(e^2),$$  \hspace{1cm} (3.13)$$

which follows from the approximation (3.7) for the Earth-Sun distance through binomial expansion of \(d^2 = d_0^2(1 + 2e \cos M) + O(e^2)\) and equating \(d_s = d_0\).

The relative error of this first-order approximation for the solar radiative energy flux is \(10^{-3}\), implying an absolute error of order \(1 \text{ W m}^{-2}\).

The solar radiative energy flux at the top of the atmosphere varies by \(\pm 2^\circ = \pm 3.5\%\) between perihelion and aphelion. Therefore, with perihelion currently occurring during southern hemisphere summer, the peak solar radiative energy flux is about 7% higher in the southern hemisphere summer than in the northern (Fig. 3.3). But southern hemisphere summer is also 4.5 d shorter than northern hemisphere summer because the orbital angular velocity \(\dot{\lambda}\) is greater at perihelion. What is the net result of these competing effects on the total solar energy incident on Earth in either summer season?

### 3.3.2 Integrated Solar Radiative Energy Flux

Both the orbital angular velocity (3.2) and the solar radiative energy flux (3.12) at the top of the atmosphere happen to vary with the Earth-Sun distance in the same way, like \(d^{-2}\). Their competing effects on the total solar energy incident on Earth, for example, integrated over a season or a year, cancel each other. To see this, consider the solar radiative energy flux integrated between the times \(t_1 = t(\lambda_{s,1})\) and \(t_2 = t(\lambda_{s,2})\), which may mark the beginning and end of a season at solar longitudes \(\lambda_{s,1}\) and \(\lambda_{s,2}\):

$$\int_{t_1}^{t_2} S \, dt = \int_{\lambda_{s,1}}^{\lambda_{s,2}} S \left(\frac{dL_s}{dt}\right)^{-1} dL_s.$$  

Neglecting the very small rate of change of the longitude of perihelion \(\dot{\omega}\) (one revolution per 21 kyr) compared with that of the true anomaly \(\dot{A}\) (one revolution per 1 yr), we can set \(dL_s/dt = \dot{A} + d\omega / dt \approx \dot{A}\) and use Kepler’s second law (3.2) to substitute \(\dot{A} = h_c / d^2\). Therefore, the factors of \(d^2\) from the solar radiative energy flux \(S \propto d^{-2}\) and from the angular velocity \(dL_c / dt \propto d^{-2}\) exactly cancel each other in the integrand. The integrand contains only constants of motion, such as the orbital angular momentum per unit mass \(h_c\) and the total solar irradiance \(S_0\), whose variation over seasons or a year is so small that it can be ignored (e.g., the sunspot cycle leads to \(S_0\) variations of only \(\pm 0.06\%\) over its 11-year period). Hence, the integration can easily be carried out. If we use the expression (3.3) for the orbital angular momentum per unit mass \(h_c\) in terms of orbital parameters and equate \(d_s = d_0\), we obtain for the total solar energy
incident between $\lambda_{s,1}$ and $\lambda_{s,2}$

$$\int_{\lambda_1}^{\lambda_2} S \, dt = \frac{S_0}{(1 - e^2)^{1/2}} \frac{\lambda_{s,2} - \lambda_{s,1}}{2\pi} \times Y_a.$$  \hfill (3.14)

The English scientist Sir John Herschel expressed this result in 1832 as follows: “The amount of heat received by the Earth from the Sun while describing any part of its orbit is proportional to the angle described round the Sun’s center.”

Thus, because each astronomical season spans the same angular distance in solar longitude ($\lambda_{s,2} - \lambda_{s,1} = 90^\circ$), the total solar energy incident at the top of Earth’s atmosphere is the same integrated over each season, and the same holds for the integral over a year. The total solar energy received only depends on the total solar irradiance, orbital eccentricity, and length of year. But it does not depend on when perihelion occurs (up to ignoring the $O(10^{-4})$ error resulting from neglecting the rate of change of the longitude of perihelion relative to the orbital angular velocity). This important result arises because of a cosmic coincidence: the same factor $\frac{3}{2}$ appears both in the inverse square law for the dependence of the solar radiative energy flux on distance from the Sun (conservation of energy along solar beams), and in the moment of inertia of Earth’s rotation around the Sun (conservation of angular momentum along Earth’s orbit). The cancelation of the two $\frac{3}{2}$ factors has important consequences for how ice ages can arise, as we will discuss in section 3.6.

### 3.4 INSOLATION

The variation of the solar radiative energy flux at the top of the atmosphere by a few percent over the course of the year is important, especially for long-term climate changes. However, what dominates the diurnal and annual cycles of the insolation incident on a plane tangential to the top of the atmosphere is not the relatively weak annual variation of the Earth-Sun distance. It is the large variation of the incidence angle of the solar beam on the tangential plane, or the elevation of the Sun in the sky and the associated duration of daylight at any given location.

The solar elevation angle $\theta'$ measures the elevation of the Sun above the horizon. Its complement, the solar zenith angle $\theta = 90^\circ - \theta'$, is the angle the solar beam makes with the direction of normal incidence at the top of the atmosphere (Fig. 3.4). The zenith angle or elevation angle control over how large an area on the tangential plane at the top of the atmosphere a unit cross-sectional area of the solar beam is distributed. The radiative power that passes through a unit area perpendicular to the solar beam is distributed in the tangential plane over an area that is a factor $\cos^{-1} \theta$ greater, so that the power per unit area incident on the tangential plane is reduced by a factor $\cos \theta$. Thus,
the insolation at the top of the atmosphere takes the form

\[
F = S \cos \theta.
\]  

(3.15)

At sunrise and sunset, the elevation angle is \( \theta' = 0^\circ \), the zenith angle is \( \theta = 90^\circ \), and \( \cos \theta = 0 \) (up to refraction of the solar beam in the atmosphere, which leads to sunrise and sunset only approximately corresponding to a solar elevation \( \theta' = 0^\circ \)). At night, the zenith angle is \( \theta > 90^\circ \): the Sun is below the horizon, and we set \( \cos \theta = 0 \), so that the insolation \( F \) vanishes. The elevation and zenith angles depend on the latitude, time of year, and time of day. At the equinoxes, the subsolar point, where the Sun is in zenith (directly overhead), lies on the equator: the elevation angle on the equator at solar noon is \( \theta' = 90^\circ \), the zenith angle is \( \theta = 0^\circ \), and \( \cos \theta = 1 \). At the solstices, because of the tilt (obliquity) of Earth’s spin axis relative to the orbital plane, the subsolar point lies in the summer hemisphere, so that the elevation angle at solar noon is greater in the summer hemisphere than in the winter hemisphere, and the zenith angle is correspondingly smaller. The resulting insolation variations are what gives rise to the seasons (Fig. 3.1).

To use the expression (3.15) for the insolation, we need to determine how
Figure 3.5: The declination angle $\delta$ is the latitude of the subsolar point, where the Sun is in zenith. For an obliquity $\gamma$, the declination angle varies between $\delta = +\gamma$ at northern summer solstice and $\delta = -\gamma$ at northern winter solstice. The declination angle $\delta$ is zero at the equinoxes.

the zenith angle $\theta$ depends on time and location.

3.4.1 Solar Zenith Angle

To calculate how the solar zenith angle varies with time of day, time of year, and location on Earth’s surface, it is helpful to introduce two additional angles, which determine the latitude and longitude of the subsolar point:

Declination angle. The declination angle $\delta$ gives the latitude of the subsolar point. That is, it gives the latitude at which a straight line from the Sun to Earth’s center intersects Earth’s surface (Fig. 3.5). The declination angle varies with season. Its annual extremes determine where the Tropic of Cancer ($23.4^\circ$N) and Tropic of Capricorn ($23.4^\circ$S) are located: the extremes are equal to the obliquity $\gamma$ of Earth’s spin axis, which is the angle between the spin axis and the normal to the orbital plane; currently, Earth’s obliquity is $\gamma = 23.4^\circ$. The declination angle varies between $\delta = +\gamma$ at northern summer solstice (solar longitude $L_s = 90^\circ$), $\delta = 0^\circ$ at the equinoxes ($L_s = 0^\circ$ and $180^\circ$), and $\delta = -\gamma$ at northern winter solstice.
Figure 3.6: The hour angle $\eta$ is the longitude difference between the subsolar point and the location of interest. It is an angular measure of local solar time: it is zero at solar noon and completes one revolution per solar day.

$$(L_\odot = 270^\circ); \text{ see Fig. 3.1. Generally, the declination angle satisfies}$$

$$\sin \delta = \sin \gamma \sin L_\odot.$$  \hspace{1cm} (3.16)

This expression arises because the declination angle $\delta$ is related to the dot product $\hat{\Omega} \cdot \hat{d} = \cos(90^\circ - \delta) = \sin \delta$ of a unit vector $\hat{\Omega}$ in the direction of Earth’s spin axis and a unit vector $\hat{d}$ from Earth’s center to the Sun (Fig. 3.5). In an orthonormal coordinate system whose $x$ axis points from Earth’s position at vernal equinox toward the Sun and whose $z$ axis is normal to the orbital plane, the unit vector in the direction of Earth’s spin axis is $\hat{\Omega} = (0, \sin \gamma, \cos \gamma)$, and the unit vector from Earth’s center to the Sun is $\hat{d} = (\cos L_\odot, \sin L_\odot, 0)$; the expression (3.16) follows by taking the dot product of the two.

**Hour angle.** The hour angle $\eta$ is the longitude difference between the subsolar point and the location of interest (Fig. 3.6). The hour angle measures local solar time (hence the name). It increases from $\eta = 0$ at solar noon and
completes one revolution per solar day,
\[ \eta = \frac{2\pi}{T_d} (t - t_s), \]  
\[ (3.17) \]

where \( t - t_s \) is the time that has passed since local solar noon at time \( t_s \). The length of a solar day\(^\text{13} \) is the time from solar noon to solar noon on successive days: \( T_d \approx 86400 \text{ s} = 1 \text{ d} \).

In terms of the declination angle \( \delta \), hour angle \( \eta \), and latitude \( \phi \), the solar zenith angle \( \phi \) satisfies
\[ \cos \phi = \sin \phi \sin \delta + \cos \phi \cos \delta \cos \eta. \]  
\[ (3.18) \]

This expression for the zenith angle takes the solar beam incident anywhere on Earth to be parallel to the straight line from the Sun to Earth’s center—an adequate approximation because Earth’s radius is negligible relative to the Earth-Sun distance. The expression (3.18) can be derived from the relation between the zenith angle \( \phi \) and the dot product \( \hat{n} \cdot \hat{d} = \cos \theta \) of a unit vector \( \hat{n} \) normal to Earth’s surface and a unit vector \( \hat{d} \) from Earth’s center to the Sun (Fig. 3.5). In a rotating orthonormal coordinate system whose \( x \) axis lies in the equatorial plane and points from Earth’s center to the subsolar meridian and whose \( z \) axis points in the direction of Earth’s spin axis, the normal unit vector is \( \hat{n} = (\cos \phi \cos \eta, \cos \phi \sin \eta, \sin \phi) \), and the unit vector from Earth’s center to the Sun is \( \hat{d} = (\cos \delta, 0, \sin \delta) \); the expression (3.18) follows by taking the dot product of the two.

The relations (3.16)–(3.18) for the declination angle, hour angle, and zenith angle make clear, for example, that for a planet with obliquity \( \gamma = 0^\circ \), the declination angle \( \delta \) would be zero irrespective of the time of year (\( L_s \)). The zenith angle \( \theta \) would vary only with latitude \( \phi \) and with the hour angle \( \eta \). On such a planet, insolation would vary with latitude and time of day, but it would vary with time of year only to the extent that the distance to the Sun varies: That is, annual insolation variations would be only of order eccentricity \( e \); seasonality—the contrast between summer and winter—would be weak. Conversely, for a planet with obliquity \( \gamma = 90^\circ \), the declination angle would be equal to the solar longitude, \( \delta = L_s \). On such a planet, even the equator (\( \phi = 0^\circ \)) would experience extreme annual insolation variations: the zenith angle on the equator at local solar noon (\( \eta = 0^\circ \)) would vary with the solar longitude, \( \theta = L_s \), so that it goes from \( 0^\circ \) (Sun directly overhead) at the equinoxes to \( \pm 90^\circ \) (Sun at horizon) at the solstices. Uranus orbits the Sun lying on its side and has an obliquity \( \gamma = 98^\circ \) that is close to this limiting case of extreme annual insolation variations. Earth with its intermediate obliquity has intermediate annual insolation variations.
3.4.2 Daily and Annual Insolation

Because the zenith angle at sunrise and sunset is \( \theta = 90^\circ \) so that \( \cos \theta = 0 \), the relation (3.18) can be solved for the hour angle \( \eta = \eta_d \) at sunrise and sunset, leading to

\[
\cos \eta_d = -\tan \phi \tan \delta. \tag{3.19}
\]

On any latitude circle, the sunrise/sunset hour angle \( \eta_d \) is the difference between the longitude of solar noon and the longitudes of the solar terminator, where sunrise and sunset are in progress. Because of the relation (3.17) between the hour angle and the time of day, the hour angle \( \eta_d \) corresponds to half the period of daylight: The full period of daylight is \( \frac{T_d}{2} \), where \( T_d = 24 \) h. Days and nights at the equator have the same length year-round.

At the equinoxes (\( \tan \delta = 0 \)), we have \( \eta_d = 90^\circ \) at all latitudes, so days and nights have the same length at all latitudes. At high latitudes, daylight stretches throughout an entire solar day when \( \tan \phi \tan \delta \geq 1 \), and it is entirely absent when \( \tan \phi \tan \delta \leq -1 \). The first case occurs during spring and summer at latitudes for which \( |\phi| \geq 90^\circ - |\delta| \) (because \( (\tan \delta)^{-1} = \tan(\pi/2 - \delta) \)). Such latitudes experience polar day: the Sun does not set below the horizon during a solar day, and we fix \( \eta_d = 180^\circ \). The second case occurs during fall and winter at latitudes for which \( |\phi| \geq 90^\circ - |\delta| \). Such latitudes experience polar night: the Sun does not rise above the horizon during a solar day, and we fix \( \eta_d = 0^\circ \). Because the extrema of the declination angle correspond to the obliquity, \( \delta = \pm \gamma \), polar day and polar night only occur poleward of the latitude circles with \( \phi = \pm(90^\circ - \gamma) \). These latitude circles are called the polar circles: the Arctic circle (\( 66.6^\circ \text{N} \)) in the northern hemisphere, and the Antarctic circle (\( 66.6^\circ \text{S} \)) in the southern hemisphere. At the polar circles, the Sun does not set for one solar day at summer solstice, and it does not rise for one solar day at winter solstice. Poleward thereof, polar day and polar night extend over longer periods. At the poles themselves, half a year of continuous polar day switches to half a year of continuous polar night and back at the equinoxes.

The diurnally averaged insolation can be computed by averaging the cosine of the zenith angle (3.18) over a solar day, setting it to zero when there is no daylight. Because the length of a solar day for Earth is short compared with the length of a year, the declination angle can be taken to be constant over a solar day. The average can then easily be calculated, giving

\[
\overline{\cos \theta} = \frac{1}{\pi} \left( \eta_d \sin \phi \sin \delta + \cos \phi \cos \delta \sin \eta_d \right). \tag{3.20}
\]

Here, the overbar denotes the mean over a solar day, \( \overline{()} = (2\pi)^{-1} \int_{-\pi}^\pi () \, d\eta \), and the sunrise/sunset hour angle \( \eta_d \) is understood to be set to \( \pi \) (i.e., \( 180^\circ \)) during polar day and to \( 0 \) during polar night. This expression captures the competing effects of the solar elevation and of the period of daylight on how insolation at
Given the orbital parameters (Table 3.1), the diurnally averaged insolation at any time \( t \) can be computed as follows:

1. Calculate the mean anomaly \( M(t) \) at time \( t \) either from Eq. (3.6) or (3.10)
2. Calculate the true anomaly \( A(t) \) given the mean anomaly \( M(t) \) (Eq. 3.8)
3. Calculate the Earth-Sun distance \( d(t) \) given the true anomaly \( A(t) \) (Eq. 3.1)
4. Get the solar longitude \( L_s(t) = A(t) + \omega \)
5. Calculate the declination angle \( \delta(t) \) given the obliquity \( \gamma \) and solar longitude \( L_s(t) \) (Eq. 3.16)
6. Calculate the sunrise/sunset hour angle \( \eta_d(t) \), setting \( \eta_d = 0 \) in polar night and \( \eta_d = \pi \) in polar day (Eq. 3.19)
7. Calculate the diurnally averaged cosine of the zenith angle \( \cos \theta \) (Eq. 3.20)
8. The diurnally averaged insolation at the top of the atmosphere then is

\[
F = S_0 \left( \frac{d(t)}{d} \right)^2 \cos \theta.
\]

the top of the atmosphere varies with latitude and time of year. For example, at the equinoxes (when \( \delta = 0^\circ \) and \( \eta_d = 90^\circ \)), the diurnally averaged cosine of the zenith angle varies only with the cosine of latitude, \( \cos \theta = \pi^{-1} \cos \phi \), which controls the insolation according to the cosine law (Fig. 2.9). By contrast, around the solstices (when \( \eta_d \) is a strong function of latitude), the fact that daylight stretches throughout the solar day near the summer pole and is entirely absent near the winter pole can dominate the variation of \( \cos \theta \) with latitude if the obliquity is sufficiently large.

These competing effects and the variations of the Earth-Sun distance over the course of a year are evident in the diurnally averaged insolation at the top of the atmosphere (Fig. 3.7), calculated for Earth’s current orbital parameters as outlined in the box on p. 94. Near the solstices, insolation is maximal at the summer pole, because the polar-day effect dominates the solar-elevation effect at the poles. (It turns out the solstitial insolation maximum is at the summer pole when the obliquity exceeds 20.7°; otherwise, the insolation maximum is at an intermediate latitude between the subsolar point and the summer pole.) Latitudes outside the tropics experience one insolation maximum during the course of a year, at summer solstice; the equator experiences two insolation maxima, at the equinoxes. Insolation around southern summer solstice (December) is more intense than around northern summer solstice (June), because perihelion currently occurs close to southern summer solstice. In the annual mean, insolation at the top of the atmosphere decreases monotonically and
symmetrically away from the equator (Fig. 3.7, right panel). At the equator, annual-mean insolation is about twice as strong as at the poles.

### 3.5 MILANKOVITCH CYCLES

To determine the insolation incident at the top of Earth’s atmosphere, it is an excellent approximation to consider the Kepler problem: the Sun and Earth (or, what is almost the same, the barycenter of the Earth-Moon system) are treated as spherical masses that interact gravitationally, undisturbed by other planets. However, Earth has an equatorial bulge and thus is not spherical, and other planets and the Moon interact with Earth gravitationally, albeit much more weakly than the Sun. The resulting small perturbations to the Kepler problem, when integrated over millennia, lead to secular (long-term) variations of Earth’s orbital parameters.¹⁴

The secular variations of Earth’s orbital parameters arise for two reasons:

1. The gravitational pull of other planets (principally Jupiter) leads to corrections to the gravitational force Earth experiences, on top of the \( d^{-2} \) inverse square law for the central gravitational force between Earth and the Sun. General relativity implies minute additional corrections, which modify the Newtonian gravitational force because of the curvature of spacetime in the vicinity of the Sun. These corrections to the inverse square law lead to secular variations of the orientation and eccentricity of Earth’s orbit around the Sun.

2. The oblateness of Earth—the equatorial radius is 21 km greater than the polar radius—and the presence of the Moon lead to corrections to the approximation of Earth as a lone spherical mass orbiting the Sun. The Moon and the Sun,
and to a lesser extent other planetary bodies, exert tidal torques on Earth’s equatorial bulge. These torques pull the equatorial bulge toward alignment with the orbital plane, which leads to a secular precession of Earth’s spin axis.

In the first half of the 20th century, the Serbian applied mathematician Milutin Milankovitch set out to explain the origin of ice ages in terms of periodic variations of Earth’s orbital parameters. He carried out laborious manual calculations of how corrections to the Kepler problem produce climatically important variations of three orbital parameters: variations of the eccentricity $e$ of the orbit, variations of the obliquity $\gamma$ of Earth’s spin axis, and precession of the longitude of perihelion $\omega$. These long-term orbital variations are now known as Milankovitch cycles.

### 3.5.1 Eccentricity Variations

The eccentricity $e$ of Earth’s orbital ellipse varies as a result of the mutual gravitational interactions among the planets. Over the past 50 Myr, the eccentricity has varied between a minimum of 0 and a maximum of 0.06 (Fig. 3.8). Eccentricity variations have several periodicities, the 5 most important being 405 kyr, 95 kyr, 124 kyr, 99 kyr, and 131 kyr (in order of decreasing amplitude). The last 4 periods have an amplitude-weighted mean value of 110 kyr and are often referred to as the “100-kyr cycle.” The eccentricity variations manifest themselves as variations of the length of the semi-minor axis. The semi-major axis of Earth’s orbit stays constant as the eccentricity varies. Because Earth’s orbital period, the length of year, only depends on the semi-major axis (Kepler’s third law), it too stays constant.

Since eccentricity variations do not affect the solar zenith angle, they only modulate the amplitude with which the solar radiative energy flux at the top of the atmosphere oscillates between perihelion and aphelion (Fig. 3.3). Thus, eccentricity variations by themselves lead to globally uniform insolation modulations over the annual cycle. Eccentricity variations also modulate the solar radiative energy flux integrated between any two fixed solar longitudes, for example, averaged over any season or over a year. Eccentricity modulates the averaged solar radiative energy flux (3.14) by a factor $(1 - e^2)^{-1/2} \approx 1 + e^2/2 + O(e^4)$. For orbital eccentricity values between 0 and 0.06, this gives small modulations of order $10^{-3}$.

### 3.5.2 Obliquity Variations

The obliquity $\gamma$ of Earth’s spin axis relative to the orbital plane varies slowly because the inclination of Earth’s orbital plane varies relative to a fixed reference plane. Over at least the past 50 Myr, the rocking motion of the orbital plane has produced obliquity variations within a narrow range between 22.1° and 24.5°, with a mean period of 41 kyr (Fig. 3.9). The obliquity variations are stabilized
EARTH'S ORBIT AND INSOLATION

Figure 3.8: Variations of the eccentricity $e$ of Earth's orbit. Eccentricity varies between 0 and 0.06, with dominant periods around 110 kyr and 405 kyr. Currently, Earth is at a relatively low point ($e = 0.017$) in its eccentricity cycle.

by the presence of the Moon. Without the Moon, Earth's obliquity would have exhibited larger variations, resulting in a less stable climate. Earth's obliquity ($\gamma = 23.4^\circ$) is currently declining. This means that the Tropics of Cancer and Capricorn, where the Sun is in zenith at the equinoxes, are moving toward the equator, at a rate of 14 m per year. Conversely, the polar circles are slowly moving toward the poles, decreasing the area that experiences polar night and day.

The obliquity of Earth's axis is responsible for the seasons. With zero obliquity, the poles would receive little light, and the only seasons Earth would experience are the weak insolation modulations owing to the eccentricity of
Obliquity varies between 22.1° and 24.5°, with a mean period of 41 kyr. Its current value, \( \gamma = 23.4° \), is close to the mean obliquity over the past several million years.

The larger the obliquity, the larger the summer-winter insolation contrast—the seasonality. Obliquity has a weak effect on insolation in the tropics, where seasonality is weak. However, it has a stronger effect in middle and high latitudes, where higher obliquities mean stronger seasonality. For example, if obliquity is increased from its present value \( \gamma = 23.4° \) to the maximum value \( \gamma = 24.5° \) it attained over the past several million years, summers in middle and higher latitudes are brighter and hence usually warmer, and winters (outside polar night) are dimmer and hence usually colder (Fig. 3.10a). Thus, the summer-winter insolation contrast in both hemispheres is enhanced, as is the annual-mean insolation. Because low-latitude changes are weaker, the equator-to-pole insolation contrast in summer is weakened, with smaller changes in the equator-to-pole insolation contrast in winter because there are no changes in polar night.

For a planet on a low-eccentricity orbit, such as Earth’s, annual-mean insolation at the poles increases relative to that at the equator with increasing obliquity. Once the obliquity is greater than the critical value \( |\gamma| \approx 54° \), annual-mean insolation at the poles exceeds that at the equator.
such high obliquity, one would expect annual-mean temperatures at the poles to be greater than those near the equator, unless a meridionally varying albedo distribution compensates the pole-equator insolation contrast. Indeed, Uranus has an obliquity of $\gamma = 98^\circ$. Uranus effectively orbits the Sun lying on its side, implying that its summer pole directly faces the Sun, with relatively intense insolation over the entire summer hemisphere, and very little insolation over the entire winter hemisphere. Annual-mean insolation on Uranus is greatest at the poles and has a minimum at the equator—a situation very different from that on Earth. However, relatively little is known about the temperature distribution and seasonal cycle on Uranus.

The seasonality changes owing to obliquity variations are symmetric between the hemispheres. Therefore, they can be expected to drive hemispherically symmetric climate changes. However, while obliquity variations affect how insolation is distributed across latitudes, they do not affect how much insolation Earth receives as a whole. That only depends on the Earth-Sun distance, which is unaffected by obliquity variations.
3.5.3 Precession

The longitude of perihelion \( \omega \) precesses (moves slowly in the direction of the orbital motion) as a result of the superposition of two motions:

**Apsidal precession** The major axis of Earth’s orbit, which connects aphelion and perihelion, rotates in the orbital plane, pivoting around the Sun in the same direction as Earth’s orbital motion (Fig. 3.11). This is called apsidal precession because aphelion and perihelion rotate relative to the fixed stars, and both are collectively referred to as apsides. The period of Earth’s apsidal precession is 112 kyr. At a constant orientation of Earth’s spin axis relative to the fixed stars, apsidal precession implies a precession of the longitude of perihelion \( \omega \) with the same period.

**Axial precession** Earth’s spin axis rotates slowly around an axis normal to the orbital plane, tracing out a pair of cones whose tips connect at Earth’s center. This motion is caused by the tidal torques on Earth’s equatorial
bulge and is called axial precession. It is similar to the precession of a spinning top. Earth’s axial precession period is 26 kyr.

Apsidal and axial precession together produce a precession of perihelion relative to the equinoxes, or, in other words, a precession of the longitude of perihelion $\omega$, with a mean period of 21 kyr (Fig. 3.11).

The longitude of perihelion affects the top-of-atmosphere solar radiative energy flux (3.12) but not the solar zenith angle. The solar radiative energy flux at the top of the atmosphere is stronger at perihelion than at aphelion, making the season in which perihelion occurs brighter and thus usually warmer. Currently, perihelion occurs during southern summer (Fig. 3.2). With the current eccentricity of 1.7%, this makes the solar radiative energy flux at the top of the atmosphere around southern midsummer about 7% stronger than it is around southern midwinter, thus enhancing seasonality in the southern hemisphere. Half a precession cycle ago, at the beginning of the Holocene 11 kyr BP, the longitude of perihelion was turned $180^\circ$. Perihelion occurred during northern summer, enhancing the solar radiative energy flux around northern midsummer relative to that around northern midwinter (Fig. 3.10b). Therefore, seasonality in the northern hemisphere was enhanced relative to today, whereas seasonality in the southern hemisphere was reduced. That is, precession-induced changes in seasonality are antisymmetric between the hemispheres.

However, Earth also moves faster around the Sun near perihelion than near aphelion. As we have seen in section 3.3.2, the more intense solar radiative energy flux near perihelion and Earth’s faster orbital motion around the Sun just cancel in their effect on the solar radiative energy flux (3.14) integrated between any two fixed solar longitudes, for example, integrated over any season or over a year. Therefore, precession of the longitude of perihelion $\omega$ has no effect on the seasonal-mean or annual-mean insolation at any location (Fig. 3.10b, right panel). It only affects how insolation is distributed across the season.

The degree to which precession modulates seasonal insolation depends on eccentricity $e$: The longitude of perihelion is irrelevant for a circular orbit ($e = 0$); the larger the eccentricity $e$ of the orbit, the larger the precession modulation of the seasonal insolation. The longitude of perihelion $\omega$ enters in the calculation of the mean anomaly $M_\omega$ at vernal equinox (Eq. 3.11) in the combination $e \sin \omega$ (and powers thereof), a quantity often called the precession index. The period of variation of the precession index depends both on precession and eccentricity variations. The precession index $e \sin \omega$ exhibits a mixture of periodicities between 19 and 24 kyr, a period range called the precession band.

### 3.5.4 Historical Insolation Variations

Insolation variations over Earth’s history are a superposition of the effects of precession, obliquity variations, and eccentricity variations. Given the observed state of the solar system now, variations of the orbital parameters can be calculated backward and forward in time. Because the solar system exhibits
 chaotic behavior—that is, sensitive dependence of the time evolution on initial conditions—the precision of the calculations deteriorates at long prediction horizons. Currently, the horizon to which orbital parameters can be calculated into the past or future is about 50 Myr. It is primarily limited by uncertainties in measurements, for example, of the small oblateness of the Sun. This means that associating climate changes deeper in the geological past with changes in orbital parameters is currently impossible. But it is possible to associate climate changes in the past several million years with changes in orbital parameters.

For example, during the Holocene Thermal Maximum around 8 kyr BP, temperatures were ~0.2 K higher than at the beginning of the 20th century (Fig. 1.5). Because of growing-season biases in the proxies underlying the temperature reconstruction, this may primarily reflect summer temperatures (chapter 1.1.3). Obliquity and eccentricity 8 kyr BP were slightly higher, and perihelion occurred during northern hemisphere summer. As a result, the insolation difference between the Holocene Thermal Maximum and now is a superposition of precession and obliquity changes like those in Fig. 3.10: Northern hemisphere summer was brighter, and seasonality in both hemispheres was enhanced, with more summer insolation and more annual-mean insolation than today in middle and higher latitudes of both hemispheres (Fig. 3.12). It is a plausible interpretation of the temperature record that the gradual cooling from the Holocene Thermal Maximum to the nadir of the Little Ice Age (Fig. 1.5) reflects primarily a mid- and high-latitude summer cooling in response to weakening summer insolation (Fig. 3.12).
3.6 ICE AGES AND ORBITALLY PACED CLIMATE VARIATIONS

Milankovitch’s intent in calculating secular insolation variations was to solve the mystery of the ice ages. In 1840, the Swiss geologist Louis Agassiz had proposed that geological features such as moraines and erratic boulders—out-of-place rocks that apparently had been transported over large distances—were remnants of vast ice sheets that once covered much of Europe, Asia, and North America. Cold glacial periods alternated with warmer interglacials. Hypotheses of why this may have occurred followed swiftly. Already two years after Agassiz’ publication, the French mathematician Joseph Adhémar suggested that ice ages were controlled by variations in Earth’s orbital parameters, and specifically the variations in the lengths of summer and winter associated with precession. With the work of the Scottish scientist James Croll in the 1860s and 1870s, the notion that ice ages are controlled by orbital variations began to be more widely discussed. Croll likewise related glacial-interglacial cycles to precession. He postulated that glacials begin when eccentricity is high and aphelion occurs during winter, so that winter is particularly cold. However, although linking ice ages to orbital variations was intriguing, the ideas of Adhémar and Croll were not entirely consistent with the accumulating geological data and were largely forgotten by the late 1800s.

In 1912, Milankovitch set out to calculate orbital insolation variations much more precisely than had been done before and relate them to climate variations—a project that was to occupy him the following three decades and culminated in the publication of his influential *Canon of Earth’s Insolation and its Application to the Ice Age Problem* (Milankovitch, 1941). Turning Croll’s argument around by 180°, Milankovitch argued that glacials begin when insolation is weak at high northern latitudes during summer. He postulated that when northern summer insolation is weak, accumulation of snow and ice in winter would exceed ice loss during the summer melt season, leading to gradual growth of northern-hemisphere ice sheets. He calculated midsummer insolation at 65°N, just south of the Arctic circle, and demonstrated how precession, obliquity variations, and eccentricity variations affect it (Fig. 3.13). At this high latitude, midsummer insolation varies by ~25% from peak to trough—a substantial change that could plausibly alter climate. But accurately dated geological data with which his hypothesis could be tested were not available during Milankovitch’s lifetime.

Spectral analyses of modern temperature reconstructions for the past 3 Myr (Fig. 1.5c) and the past 66 Myr (Fig. 1.8) reveal that orbital variations indeed leave a discernible imprint in climate records (Fig. 3.14). There is an unambiguous and strong spectral peak at the 41-kyr period of obliquity variations, both in the most recent 3 Myr and in the past 66 Myr (Fig. 3.14a and b). There also is a hint of a weak peak near the precession band, but its significance is unclear. The appearance of the 41-kyr peak in the spectrum is unsurprising given the clear 41-kyr period of ice ages before 1 Myr BP (Fig. 1.5c). But interestingly, although ice ages over the past 1 Myr seem to have had periodicities around 100 kyr, there is no discernible spectral peak around 100 kyr for the past
Figure 3.13: Insolation at 65°N at northern summer solstice. Milankovitch argued that glaciations begin when this insolation is weak. (The insolation variations here are based on modern calculations.)

3 Myr (the same is true for spectra for the past 1 Myr). The spectrum simply appears “red” overall, meaning that variance decreases with frequency. Red spectra commonly arise in forced-dissipative systems because higher frequencies usually are more strongly damped than lower frequencies (which is why, for example, low-frequency sound waves travel farther than higher-frequency sound waves). Clear peaks at eccentricity periods around 110 kyr and 405 kyr appear in the spectrum of temperature reconstructions over the past 66 Myr (Fig. 3.14b).

How spectral peaks at the 41-kyr period of obliquity variations arise is relatively easy to explain: low obliquity leads to cooler summers in high latitudes that promote the growth and preservation of ice sheets, which in turn feed back onto temperature, for example, through ice-albedo feedback. Feedbacks from the carbon cycle also seem to play a role in glacial-interglacial cycles, because carbon dioxide concentrations co-vary with temperature (Fig. 1.7): they amplify and globalize the temperature variations because carbon dioxide is well mixed in the atmosphere on timescales longer than years. But how spectral peaks at eccentricity periods arise is mysterious, given the weak effect (order $10^{-3}$) eccentricity variations have on insolation. If these peaks are not artifacts of the analysis, there must exist strong amplifying feedbacks in the climate system that convert weak, sinusoidal orbital insolation variations at the top of the atmosphere into strong, sawtooth-patterned climate variations on the ground.

By now, it is well established that orbital variations pace glacial-interglacial cycles. How obliquity variations can do so is relatively clear. But precisely how other orbital variations affect glacial-interglacial cycles remains unclear. The strong precession effect seen in the northern midsummer insolation (Fig. 3.13) has no effect on insolation averaged over the summer, which may be a better proxy for summer melting of ice. So it is not clear how precession would affect ice volume; indeed, precession has not left a clear spectral signature in the temperature reconstructions. Adding to the mystery, it is clear that before 1 Myr BP, glacial occurred roughly every 41 kyr, apparently paced by
Figure 3.14: Power spectral density of reconstructed temperature variations. (a) Power spectrum of the reconstructed temperature variations over the past 3 Myr in Fig. 1.5c. (b) Power spectrum of the reconstructed temperature variations over the past 66 Myr (Cenozoic era) in Fig. 1.8. The grey dotted lines and shading mark periods of orbital variations: 405, 124 and 95 kyr are periods of eccentricity variations; 41 kyr is the dominant period of obliquity variations; and the grey band at 19–24 kyr is the precession band. The power spectral densities are normalized so that the frequency integral of the power is equal to the mean square amplitude of the time series in Figs. 1.5 and 1.8.

While it is generally accepted that variations in Earth’s orbit pace ice ages and other long-term climate changes, we do not have a mechanistic understanding of how this happens. What are the crucial nonlinearities in the climate system obliquity variations (Fig. 1.5). Thereafter, the periodicity changed to irregular oscillations with periods between 80 and 120 kyr. It is now widely thought that the small insolation changes owing to eccentricity variations cannot cause glacial-interglacial cycles. But what drives the glacial-interglacial cycles of the past 1 Myr? Explaining the periodicity of glaciations over the past 1 Myr is known as the “100-kyr problem.” It has been postulated that the periods around 100 kyr arise because the deglaciations occur every four or five precession cycles, or every two or three obliquity cycles.21
that convert sinusoidal orbital insolation variations into the sawtooth-shaped climate variations of glacial-interglacial cycles, with gradual cooling followed by relatively rapid warming (Fig. 1.5)? What processes give rise to the co-variation of temperature and carbon dioxide concentrations, which amplify glacial-interglacial cycles and help making them global? What role did the long-term cooling trend over the past 50 Myr (Fig. 1.8) play in glacial-interglacial cycles, and what may it have to do with the periodicity change 1 Myr ago? These are the questions at the frontier of current knowledge.

FURTHER READING

J. Imbrie and K. P. Imbrie, 1979: Ice Ages: Solving the Mystery. Harvard UP. A fascinating historical account of how ice ages were discovered and were related to orbital variations.

R. Fitzpatrick, 2012: An Introduction to Celestial Mechanics. Cambridge UP. Introduction to orbital mechanics and perturbation theories with which orbital variations can be calculated.


Notes

1. Galileo Galilei made the first telescopic observations of the sky in the months immediately after Kepler published his Astronomia Nova in 1609, which contained the first two of Kepler’s laws.

2. The barycenter of the Earth-Moon system lies inside Earth, displaced roughly 3/4 of Earth’s radius from Earth’s center.

3. See, for example, Fitzpatrick (2012, chapter 3.13) for a calculation of the lengths of the seasons.

4. See, e.g., Fitzpatrick (2012, chap. 3.8) for a derivation of the expression (3.3) for the orbital angular momentum per unit mass.

5. Orbital parameters for Earth and other planets are available at nssdc.gsfc.nasa.gov/planetary/factsheet and ssd.jpl.nasa.gov/?horizons.

6. The epoch J2000 is defined as 12:00 TT on January 1, 2000, where TT stands for terrestrial time. Terrestrial time is an astronomical time standard that, in the year 2000, deviated from Coordinated Universal Time (UTC) by about one minute. Irregular deviations between TT and UTC arise because of irregularities in Earth’s rotation, as UTC is pegged to solar time at 0° longitude, whereas TT is defined independently of Earth’s rotation.

7. See Fitzpatrick (2012, appendix A.10) for a derivation of the series expansion (3.8) for the true anomaly.

8. Vernal equinox (from Latin ver, “spring”) refers to the northern spring equinox, and autumnal (from Latin autumnus, “fall”) to northern fall equinox.

9. The longitude of perihelion is sometimes given as a geocentric longitude, i.e., as the angular distance between the position of the Sun at perihelion and at vernal equinox as viewed from Earth. This geocentric longitude of perihelion is related to the heliocentric longitude of perihelion \( \omega \) we use here by \( \omega = 180° \).

10. See Brouwer and Clemence (1961, chap. 2.4) for a derivation of the approximation (3.11).

11. These and other times between equinox and perihelion should be interpreted as being for the barycenter of the Earth-Moon system. The times between equinox and perihelion for Earth alone differ slightly because of the Moon’s orbit around Earth.

12. See Herschel (1832).
13. The solar day is infinite for a tidally locked planet, which completes one revolution around its spin axis in one orbital period (the planetary “year”), so that the same subsolar point always faces the Sun. (Earth’s Moon is tidally locked with Earth, so that the same side always faces Earth.) For such a planet, one would define the hour angle directly through longitude differences.

14. See Murray and Dermott (1999, chapter 7) and Fitzpatrick (2012, chapter 4) for perturbation theories with which the secular variations can be calculated.

15. Milankovitch collected the results of his calculations into a masterful monograph (Milankovitch, 1941), which went to press in Belgrade just days before the German attack on Yugoslavia. The printing house was destroyed during the bombing of Belgrade on 6 April 1941, but Milankovitch’s masterwork survived, and the printing was completed later in the year. The interruption of the printing process is evident in extant original copies of his book: Only lower-quality, acidic paper was available after the war started, and this paper browned strongly over time. So the pages in later parts of his book, which were printed later, now are much more strongly browned than those in earlier parts.

16. See Berger (1976, 1978), Berger and Loutre (1991), and Laskar et al. (2004, 2011) for detailed modern calculations of the climatically important orbital parameters, on which the numerical values in what follows and Figs. 3.8–3.13 are based.

17. For the role of the Moon in stabilizing obliquity, see Laskar et al. (1993) and Lissauer et al. (2011).


19. Artifactual spectral peaks may arise from orbital tuning of age models underlying the temperature reconstructions, that is, using correlations between calculated orbital insolation variations and estimated ages of records to fine-tune the ages. This may play a role for some of the records underlying the temperature reconstruction in Fig. 1.8 (Zachos et al., 2001).

20. In 1976, the American-English team of James Hays, John Imbrie, and Nicholas Shackleton published a landmark paper, which was celebrated as providing the first clear evidence that orbital variations leave spectral signatures in climate reconstructions and pace the ice ages (Hays et al., 1976). See Maslin (2016) for an overview of the history.

21. See Raymo and Huybers (2008) and Maslin (2016) for a summary of the current state of discussions about what may control glacial-interglacial cycles.
Chapter Four

Radiation and Matter

Understanding the interactions between electromagnetic radiation and matter requires a glimpse at the molecular and quantum realm. When photons meet a molecule, they can scatter (change direction without permanently altering the molecule’s energetic state), or they can be absorbed (converted to the molecule’s energy). Absorption makes the excited molecule unstable and eventually subject to losing the same energy via an inverse process to absorption, emission.

This chapter surveys the physical processes involved in absorption, emission, and scattering that are important for understanding Earth’s atmospheric thermal structure and optical characteristics. This will lead us better to understand why trace atmospheric constituents, such as water vapor and carbon dioxide, are primarily responsible for the greenhouse effect, how absorption and emission are dependent on temperature and pressure, why the sky is blue and thin clouds are white, and how Earth’s albedo is controlled.

4.1 ABSORPTION AND EMISSION

As we saw from Kirchhoff’s law and energy conservation earlier, an object that is a good absorber of a particular wavelength is a good emitter of that wavelength. This implies that at thermal equilibrium absorption and emission are symmetric processes, or that emission is absorption with time running backwards. But how is this exactly achieved?

Absorption and emission increase and decrease energies of atoms and molecules, and can only occur when photons’ energy quanta ($h\nu$) match the energy required ($\Delta E$) for possible transitions within the atom or molecule to occur:

$$\Delta E = h\nu = \frac{hc}{\lambda}. \tag{4.1}$$

A photon can increase an atom’s energy by absorption which may bump the atom’s electrons to orbits of higher quantized energy levels, or by acceleration and increasing the atom’s kinetic energy. If several atoms form a molecule, other mechanisms that can increase the molecular energy come into play, whereby the molecule’s atoms can vibrate and rotate relative to each other. The overall
internal energy change of a molecule is thus:

$$\Delta E = \Delta E_{\text{electronic}} + \Delta E_{\text{translational}} + \Delta E_{\text{vibrational}} + \Delta E_{\text{rotational}}.$$ 

the components of which are referred to as energy transitions. We now discuss these transitions in more detail.

### 4.1.1 Electronic Transitions

Take, say, the hydrogen atom, which has a nucleus with an orbiting electron. Each orbital shell is associated with a particular energy level (or state). The shell closest to the nucleus, referred to as the ground state, contains the least energetic electrons, and the energy required to reside at each concentric shell increases with the distance from the atom's nucleus. It is possible for an electron to jump to higher-energy shells, provided that it is excited by absorption of a photon with the appropriate wavelength (and thus energy). If a photon's energy does not match any of the atom's quantum energy levels, the photon will pass by the atom without interaction. Because of these discrete shells and the ability to only absorb specific energy quanta, the hydrogen atom only absorbs light at particular wavelengths. This can be observed by a spectrometer in a lab as very narrow absorption lines at those wavelengths.

Upon absorption, the new excited state is not stable and the electron will eventually, depending on the molecule's radiative decay rate, drop down to the original energy level while emitting the same wavelength photon as the photon that was earlier absorbed. If it is emitted immediately after absorption, this is essentially scattering and is discussed in the next section.

The structure of atoms as we know it today and the consequent deduction of how quantized light interacts with atomic energy levels via electron exchanges was first postulated by Niels Bohr in 1913, which led to him being awarded a Nobel Prize a decade later. [box with bohr model](#)

Electronic transitions require most amount of energy to occur, and so they only interact with higher-frequency photons of visible and ultraviolet ($\lambda \leq 1\ \mu\text{m}$). It is these transitions that are responsible for UV light absorption by ozone in the stratosphere, or for photochemical smog above large cities immersed in high concentrations of nitrogen oxides.

### 4.1.2 Translational Transition

Translational transition occurs when the kinetic energy of an atom or molecule is altered. It relates to the temperature of the gas volume, recalling that $< u_{\text{rms}} >= (3RT_0/m_d)^{1/2}$. This transition is not quantized, and it is a major player in broadening absorption lines, as it can aid molecules in absorbing wavelengths that do not normally correspond to their quantum states. Under sufficient pressure (collisions), translational energy transitions may dominate the radiative decay and convert some of the quantized energies into heat.
CHAPTER 4

There are two types of dipole moments: electric and magnetic. If a molecule has opposite charges, $Q$, separated by a distance $l$, then the electric moment is $p_e = Ql$. The electric field in an electromagnetic wave can then exert a torque on that moment and cause the molecule to rotate. Similarly, a magnetic dipole moment of a charged loop (with the loop representing, e.g., the orbit of an electron), the product of the current through the loop and the loop’s area, $p_m = IA$, can interact with the magnetic field of an electromagnetic wave and exert torque on the molecule, analogous to a compass.

4.1.3 Vibrational Transitions

Vibrational transitions may occur because molecular covalent bonds can behave like springs, causing the bound atoms to oscillate upon excitation, whereby the atoms are repulsed when their nuclei are too close together and attracted when they are far away.

This is a quantized transition, so it requires a particular energy quanta to become excited. There are several ways (or modes) in which a molecule can become excited. These modes correspond to the different ways in which atoms of a molecule oscillate relative to each other, and will be discussed more specifically for $\text{H}_2\text{O}$ and $\text{CO}_2$ below.

These modes require that a photon exerts a torque on a molecule, which is only possible if the molecule has an electric or magnetic dipole moment. A dipole moment is effectively a lever that can interact with the electromagnetic field (4.1.3). In the presence of a moment, each mode can be excited by incident photons of different wavelengths.

Because vibrational transitions require less energy than the electronic transitions, they absorb mainly in the near infrared and thermal infrared part of the spectrum ($1 \mu m \lesssim \lambda \lesssim 20 \mu m$). Interaction with the infrared radiation makes this transition most important for Earth’s climate. In particular, the bending vibrational mode in $\text{CO}_2$ and $\text{CH}_4$ can break the molecular symmetry and leads to oscillating dipole moments, and lead to vibrational and rotational transitions.

4.1.4 Rotational Transitions

Rotational transitions are also quantized and depend on the presence of an electric or magnetic dipole moment. They require even less energy to be excited and thus interact primarily with far infrared and microwave photons ($\lambda \gtrsim 20 \mu m$).

Some diatomic homonuclear (two identical atoms) molecules have no permanent electric or magnetic dipole moment. Hence the most abundant atmospheric gas, $\text{N}_2$, has a weak role in the radiative transfer. On the other hand, $\text{O}_2$
has no electric dipole moment but its magnetic dipole gives rise to rotational transitions. It is this property of $O_2$ that allows us to retrieve microwave depth soundings of different frequencies that correspond to different $O_2$ absorption lines. From these soundings we can recover the brightness temperature, and attain faithful continuous global coverage since the beginning of the satellite era, 1979. These soundings have revolutionized weather forecasting.

### 4.1.5 Collision-Induced Temporary Transitions

Molecular collisions can induce temporary dipole moments which, in the presence of the suitable-wavelength photons, result in vibrational-rotational transitions even in diatomic gases, such as $N_2$. This induced absorption/emission is especially important in dense atmospheres. For example, Titan’s atmosphere has high enough pressure that collisions make the $N_2$-rich atmosphere very opaque to infrared radiation. Collision-induced absorption is similarly important on Jupiter and Saturn. The strong dependence on pressure means that more collisions occur near the surface, which can lead to the lower part of the atmosphere being warmer and more buoyant, thus inducing strong convection on those planets. On Earth this process is less effective due to the smaller number of collisions, though collisions themselves are important for broadening absorption line shapes, as will be discussed in the following sections.

### 4.2 EARTH’S PRIMARY ABSORBERS

#### 4.2.1 Water

Water vapour is a dominant greenhouse gas, along with $CO_2$, as they both absorb infrared radiation selectively and happen to be transparent to shortwave radiation.

$H_2O$ is a triatomic molecule with three distinct vibrational modes (Fig. 4.1). The symmetric stretching mode occurs when the atoms move away from each other. This transition corresponds to 2.7 $\mu m$ wavelength.

The bending mode is essentially corresponds to the flattening of a molecule and its energy corresponds to wavelengths of 6.3 $\mu m$, giving rise to particularly strong absorption features in this wavelength in the Earth’s outgoing longwave spectrum (see Fig. XX).

The asymmetric stretch mode with an asymmetrical deformation of the molecule is also a strong absorber of wavelengths at 2.7 $\mu m$.

#### 4.2.2 Carbon Dioxide

Another strong greenhouse gas, about equally important as water vapour, is $CO_2$. Unlike $H_2O$, $CO_2$ does not have a permanent dipole moment. A tempo-
Figure 4.1: Vibrational modes of H2O.
rare dipole moment can be triggered by bending bonds temporarily breaking the molecular symmetry and creating an electric dipole moment. Its symmetric, antisymmetric and bending vibrational modes operate at energies corresponding to wavelengths of 2.7 μm, 4.3 μm and 15 μm, respectively. The 15 μm band occupies wavelengths near the center of Earth’s infrared emission spectrum (see Fig. XX), and it is so efficient at absorbing this wavelength that virtually all of the photons of this wavelength that are detected from space originate in the upper layers of the atmosphere. The photos from lower levels of the atmosphere are absorbed.

4.2.3 Other Greenhouse Gases

There are other strong absorbers that are worth noting. For example, the CO₂ window region is the range of the spectrum where CO₂ is ineffective at absorbing wavelengths, but in this window even small amounts of other greenhouse gases, such as O₃, CH₄ and N₂O, can have a large effect on the atmosphere’s radiation (Seinfeld XX).

Similar to CO₂, CH₄ can induce temporary electric dipole moments, break symmetry and cause vibrational and rotational modes. Technically, CH₄ is a more effective absorber than CO₂, but there is less of it. That said, (for reasons we will discuss later) greenhouse gases affect the climate logarithmically with concentration, so the effect of a particular absorber should be considered with respect doubling of the absorber’s concentration. With CH₄ being less abundant, its doubling is achieved more easily, which makes it a very potent greenhouse gas.

Other notable gases are CO and NO₂, both with permanent dipole moments, but lower concentrations than CO₂. O₃ has a permanent dipole and is a strong absorber of UV radiation, as well as being a notable greenhouse gas.

There are also unnatural gases that affect the absorption properties of the atmosphere, such as the chlorofluorocarbons which, as well as being the major actor in O₃ hole depletion, are strong greenhouse gases in the atmospheric window, where the natural components of the atmosphere are largely transparent to infrared radiation.

4.3 ABSORPTION LINE SHAPES

If one computes all energetic modes of a molecule, or measures them in a pristine lab environment, even a simple triatomic molecule will yield a complicated spectrum intercepted by many infinitesimal absorption lines at frequencies of its transition modes.

However, this is not observed in the atmosphere. Instead there are wider absorption bands, centered around the thin absorption lines, which are intimately responsible for the ability of the atmosphere to capture energy sufficiently for
Figure 4.2: Stretching and bending modes of CO$_2$ that induce temporary dipole moments.
life on Earth. We already saw this spectral broadening in Earth’s outgoing longwave radiation spectrum, where absorption lines are indeed widened and smeared, forming broader absorption bands. Broadening can and does occur via three main processes as follows.

4.3.1 Natural Width

Broadening can arise from a spontaneous radiative decay, where the excited atomic/molecular states have a finite natural lifetime. Heisenberg’s principle tells us that lifetime and energy cannot be known simultaneously with absolute certainty. This stochastic component implies that absorption lines must be broadened, an effect that is measurable by Mössbauer spectroscopy. However, this effect is very small compared to the following two types of broadening so, in general, it can be neglected.

4.3.2 Doppler Broadening

Doppler shifts are related to the velocities of randomly moving molecules (and thus temperature, recalling \(< u_{\text{rms}} >= (3RT_0/m_d)^{1/2}\)). If an emission spectrum is measured with a spectrometer, emitting molecules flying towards the spectrometer are measured to have slightly higher frequencies than those that are flying away. This describes the classical Doppler shift, with the Doppler frequency shift is given by:

\[
\nu_r = \nu \left(1 - \frac{u}{c}\right)^{1/2} \approx \nu(1 - \frac{u}{c})
\]

where \(\nu\) is the original frequency at which the light is emitted, \(\nu_r\) is the frequency measured by the receiver, and \(\frac{u}{c}\) is the ratio of the molecular speed and the speed of light. This ratio is quite small but not negligible (using our estimated value for \(< u_{\text{rms}} > \approx 500 \text{ m s}^{-1}\), this would be of order \(10^{-6}\)). The first equality is the relativistic form, whereas the second one is the classical form that assumes \(\frac{u}{c} \ll 1\). Note that the higher the temperature the larger the Doppler shift. This process is largely independent of pressure.

The distribution of molecules’ velocities is Gaussian (or the Maxwell distribution), so the absorption line broadening from the Doppler effect is also Gaussian, with a standard deviation of order \(10^{-6}\):

\[
\Phi_D(\nu) \propto \exp \left( -\frac{\delta \nu^2}{\Delta \nu_D^2} \right)
\]

where \(\delta \nu\) is the distance from line center and \(\Delta \nu_D\) the Doppler line width.
4.3.3 Pressure Broadening

Pressure broadening depends on the collision rate and thus pressure, so lines broaden with higher pressure. This broadening is largely independent of temperature.

The distribution of the absorption line width from pressure broadening is given by the Lorentzian distribution:

$$\Phi_L(\nu) \propto \frac{\Delta \nu_L}{\delta \nu^2 + \Delta \nu_L^2}$$

where $\delta \nu$ is the distance from the line center and $\Delta \nu_L$ is the pressure-broadened width.

The Lorentzian distribution is used extensively in satellite remote sensing at TOA. If the measured radiation has wavelengths close to a line center, it is more likely to originate at higher altitudes, where the atmospheric pressure is lower. Conversely, radiation measured far away from the line center is more likely emitted from lower levels of the atmosphere, where pressure is higher.

The collisional relaxation time is much shorter than the lifetime of the excited state, so this broadening dominates the natural broadening. The pressure-dependent Lorentzian broadening compared to the temperature-dependent Doppler broadening, as well as their combined effect, called Voigt broadening, is shown at different heights in Fig XX (Christian fig). The exponential decrease in pressure makes this broadening a strong function of height. In contrast, temperature changes by around 30%, so there the Doppler broadening is still important at higher altitudes. Note that the energy integrated across all wavelengths must remain the same for any amount of broadening, so as to conserve energy. Therefore a broader functional will be associated with a smaller peak.

Ground-based microwave measurements at the oxygen absorption bands have been used since the 1960s to recover temperature profiles, and were mounted onto satellites two decades later to obtain a global coverage. The advantage of this technique is that O$_2$ concentration is similar everywhere, thus partly alleviating problems from unfaithful calibration methods.

4.4 Absorption Spectrum

Although we saw the absorptive effect of Earth’s main absorbers in Fig. XX, we can study their individual effect using their clear-sky (i.e. no clouds) absorption spectra (Fig. XX). Evidently, there is a little absorption is in visible band, which is why it will be later reasonable to assume that the clear-sky atmosphere is transparent to solar radiation.

Water and CO$_2$ dominate the infrared absorption with ozone incising the their absorption window rather markedly. HDO is a water molecule whose one
hydrogen atom is replaced by deuterium, and interestingly, it has very different absorption properties compared to H$_2$O. Despite its small amounts in the atmosphere, it can be tracked by satellites as a tracer and its vertical profile is used to infer the origins of water before it mixes. It is clear that if these absorption spectra are combined, a similar pattern to Fig. XX will emerge. However, upon a closer look, both the incoming shortwave and outgoing longwave spectra include additional features besides the absorption lines or their associated broadened absorption bands. Some parts of the spectra are absorbed more gradually with respect to wavenumber, features that are referred to as continuum absorption. In the case of outgoing infrared radiation, one can observe continuum absorption around $10^{14}$ $\mu$m, likely due to more complex clustering of water molecules. Similarly, the spectrum of solar radiance at Earth’s surface is shaped by a continuum absorption, mainly due to energetic photons ionising the air (photoionisation) or breaking down molecules (photodissociation).

4.5 SCATTERING

Photons can be redirected via the process of scattering. By contrast to absorption and emission, scattering does not permanently alter the energetic state of the scattering matter, and the incident photons will not change in frequency (though their phase and polarisation may change). In effect scattering can be viewed as absorption immediately followed by re-emission, and it can be conceptually understood without explicitly considering quantum effects. Since the re-emission occurs in all directions, only some of the solar radiance reaches the surface. Thus both absorption and scattering reduce the amount of radiation reaching the surface. A useful measure of the extent to which a particle scatters incident radiation is the scattering cross section. It can be viewed as an imaginary area fraction of incident light that is scattered, rather than absorbed (and is often divided into forward scatter and back scatter). There are several types of scattering, of which two we outline here.

4.5.1 Rayleigh Scattering

Rayleigh scattering occurs when light is deflected by particles whose diameter is substantially smaller that the wavelength of the incident photon (with a diameter $2\tau \leq 1/10\lambda$). Most of this type of scattering is off atmospheric molecules, most commonly nitrogen and oxygen on Earth. The main result this theory is that the extent to which light is scattered is inversely proportional to the fourth power of the photon wavelength. This is often quantified using the Rayleigh scattering cross section

$$\sigma_s \propto \frac{\lambda^6}{\lambda^4}$$

(4.2)
where \( r \) is the particle radius and \( \lambda \) is the photon wavelength. The full relation can be derived from classical-physics considerations of how electromagnetic waves drive oscillations of charges in matter, and how these oscillating charges, in turn, induce radiative electromagnetic waves—the scattered waves (see p. 119).

This strong dependence on wavelength, is the reason why the sky is blue during the day (as the blue photos of white sunlight are scattered more readily) and why it is red at sunrise and sunset (as along the longer path through the atmosphere the remaining radiation that did not get scattered away is that of longer wavelengths). Why is the daytime sky blue rather than violet? It is true that violet wavelengths are scattered even more readily than blue wavelengths. In fact, the sky is violet blue when measured with spectroscopic analysis. The answer lies in the sensitivity of the human eye (dark background needed; see history article; sensitivity of human eye? then: is the blue color of the sky real?)

The ‘blue’ color of the sky is not specific to Earth. Unless there are other stronger scatterers, such as dust on Mars or aerosols on Titan, other planetary atmospheres would assume the same scattering properties as Earth. On Mars the atmosphere is concentrated with dust particles that are large enough to only scatter red light, whereas the higher frequencies get absorbed, which is what gives the planet its reddish appearance and its name (after the Roman god of war, because of its reddish appearance).

The \( r^6 \) dependence of Rayleigh scattering is useful in weather radar imaging. Since most cloud droplets are in the Rayleigh scattering regime (radar wavelengths are approximately 10 cm), radar reflectivity is proportional to the 6th moment of the droplet size distribution (assuming spherical droplets). One can thus infer cloud or rain droplet sizes from the return beams that were deflected by rain or cloud droplets. Rain rates can in turn be estimated from fact that heavier rain has bigger droplets.

### 4.5.2 Mie Scattering

Mie scattering is a more general solution of scattering problem, but it requires a more laborious treatment than Rayleigh scattering, namely spherical harmonic expansions of Maxwell’s equations. The Mie scattering theory is applied to photon deflection by particles that are about the same diameter as the photon’s wavelength (i.e., with a particle diameter \( 2r \approx \lambda \)). Generally this scattering has a weaker dependence on wavelength than Rayleigh scattering, with the Mie scattering cross section:

\[
\sigma_s \propto \lambda^2
\]

and it scatters preferentially in the direction of the incident beam (unlike Rayleigh scattering, which is isotropic).

This scattering is responsible for the reddish light on a hazy day above urban areas where aerosol particles act as condensation nuclei and thereby facilitate formation of small water particles. This causes the light to be diffuse.
Electromagnetic waves drive oscillations of charges in matter, whose amplitude $x$ approximately satisfies the forced harmonic oscillator equation

$$\frac{d^2}{dt^2} x + \omega_0^2 x = \frac{F}{m},$$

where $\omega_0$ is the natural oscillation frequency, $F$ is the force acting on the charges, and $m$ is their mass. The force an electric field $E$ exerts on a charge $q_e$ is $F = q_e E$. For an incident electric field $E = \hat{E} \cos(\omega t)$ with frequency $\omega$, the solution for the charge oscillations is $x = \hat{x} \cos(\omega t)$, with amplitude

$$\hat{x} = \frac{q_e \hat{E}}{m(\omega_0^2 - \omega^2)}.$$

The oscillating charges in turn radiate electromagnetic waves. These are the scattered waves, and on average they carry away the power

$$P \propto \frac{q_e^2 \omega^4 \hat{x}^2}{\epsilon_0 c^3} = \frac{q_e^4 \hat{E}^2}{m^2 \epsilon_0 c^3 (\omega^2 - \omega_0^2)^2}.$$

Because electromagnetic waves travel with the speed of light and have an energy density $\epsilon_0 \hat{E}^2$, the incident radiative energy flux is $\epsilon_0 c \hat{E}^2$. The ratio of the scattered power to the incident energy flux thus is

$$\sigma = \frac{\omega^4}{(\omega^2 - \omega_0^2)^2} \sigma_T \quad \text{with} \quad \sigma_T = \left( \frac{q_e^2}{m \epsilon_0 c^2} \right)^2.$$

This scattering cross-section $\sigma$ expresses the scattered power in terms of the area of the incident beam that would have to be intercepted if all energy that impinges on that area were scattered. If $\omega \gg \omega_0$, the scattering cross-section $\sigma = \sigma_T$ is independent of frequency. This is the case of Thompson scattering on an unbound ($\omega_0 \to 0$) charge, i.e. free electron. If $\omega_0 \gg \omega$, the natural frequency is much higher than the frequency of the incident wave. In this case, the scattering cross section $\sigma = (\omega/\omega_0)^4 \sigma_T$ is a strong function of frequency $\omega$. This is the case of Rayleigh scattering on a bound charge. The Rayleigh scattering cross-section for macroscopic scatterers such as rain drops can also be expressed as $\sigma \propto \omega^4 r^6 \propto r^6 / A^4$ because both their total charge $q_e$ and mass $m$ scale with the volume $r^3$ of the scatterer.
and reddish. In general, larger particles become less dependent on the incident wavelength.

It is Mie scattering off larger particles in clouds that is responsible for the white color of clouds, despite the little amount of water that forms them. This makes clouds optically significant and important for modifying our climate.

Note that the microwave radiation is not efficiently scattered or absorbed by clouds, so it can be used to infer temperatures by remote sensing even in cloudy regions.

### 4.6 Earth’s Albedo

#### 4.6.1 What Albedo Depends Upon

Albedo is the incident solar radiation that is reflected or scattered back to space. The relative albedos of some of the listed surface types are intuitive. For example, bright snow has a higher albedo than a dark forest canopy. But how
Figure 4.4: Scattering by optically thick clouds. Stronger scattering for lower angles of incidence because of forward scattering.

other albedos arise is less clear and deserves further comment.

The albedo of the ocean surface depends on the surface roughness (waviness), and it depends on the angle of incidence of the sunlight. For normal incidence, the albedo is lower (~3%); for oblique incidence, it is higher (up to ~40% for a typical ocean roughness). The dependence on the angle of incidence arises because water not only reflects light but also refracts it, transmitting radiative energy below the surface where much of it is eventually absorbed (if it is not scattered back out of the water, e.g., by phytoplankton in the water). At more oblique incidence, more of the incoming sunlight is reflected and less is refracted—a result from geometric optics that for smooth surfaces is encapsulated in what are known as Fresnel’s reflection formulas. Fresnel’s reflection formulas also apply locally to the face of a water wave, because the face is much larger than the wavelength of sunlight. When a ray of sunlight strikes the mean ocean surface at an oblique angle, its incidence on a steep wave face can be close to normal. Fresnel’s reflection formulas imply that more light is refracted into the wave than would be the case for oblique incidence on a flat ocean surface; the albedo is locally reduced. Thus, at oblique angles of incidence
on the mean ocean surface, waviness reduces the albedo. Conversely and for analogous reasons, at closer to normal incidence on the mean ocean surface, waviness increases the albedo. But it remains true also for a wavy ocean surface that its albedo increases the more oblique the incidence of the sunlight. This dependence on the angle of incidence is consistent with a common experience: more sunlight is reflected from a water surface at sunrise or sunset than at noon. Because the sun on average is lower in the sky in high latitudes than in low latitudes, the average angle of incidence of sunlight is more oblique in high latitudes. Consequently, the mean albedo of the ocean surface increases with latitude. Its global-mean value is about 10%.

For clouds, the albedo similarly depends on the angle of incidence of sunlight and properties such as the cloud water path, the typical size of droplets, and the shape of any ice crystals that may be present. For example, the albedo generally is higher for smaller droplets or a larger cloud water path. Each, holding the other fixed, increases the total cross-sectional area of scatterers that intercept rays of sunlight, increasing the albedo. The albedo of clouds generally is also greater for more oblique incidence of sunlight. Here the dependence on the angle of incidence can arise for two reasons. First, for optically thin (i.e., almost transparent) and plane clouds, a ray of sunlight that strikes a cloud at an oblique angle has a longer path through the cloud and therefore a greater chance of being intercepted by a cloud particle than a ray incident normal to the cloud plane. This increases the albedo for oblique incidence on thin clouds. Second, particles that are greater than the wavelength of light, such as cloud droplets or ice crystals, preferentially scatter the incident light into a narrow cone around the forward direction—the direction into which the light ray was traveling before it was scattered. This forward scattering is the reason why hair or dust specks in air glow brightly when we view them backlit by a light source, or why the glare on a dirty windshield is most intense when we drive toward a low sun. Preferential forward scattering means for plane cloud facets that a greater fraction of the incident sunlight will be scattered upward toward space when the incidence is glancing, so that the forward scattering cone contains an upward component. It also reduces the albedo of clouds compared to what it would be if the cloud particles scattered isotropically, because forward scattering directs more sunlight into the cloud rather than reflecting it back.

However, Earth’s albedo at the top of the atmosphere is not constant but varies in space and time. In the annual mean, it is lowest (15–20%) over subtropical oceans, where cloudless skies often expose open ocean water with low albedo (Fig. 4.5a, left panel). It is highest (≥70%) over the bright snow and ice of the polar regions, particularly over the elevated ice sheets of Greenland and Antarctica. Subtropical deserts such as the Sahara also stand out as regions of high albedo compared with surrounding areas. But otherwise, land–ocean contrasts and other variations of Earth’s surface properties only leave a weak signature on top-of-atmosphere albedo variations. Instead, the dominant
Figure 4.5: Earth’s annual-mean albedo at the top of atmosphere. (a) Total albedo, obtained as the ratio of total upward to downward solar radiative energy flux at the top of the atmosphere. (b) Clear-sky component of the albedo, obtained as the ratio of upward solar radiative energy flux from cloudless regions to total downward flux. All radiative energy fluxes were measured by NASA’s space-based Clouds and Earth’s Radiant Energy System (CERES) instruments. They were averaged over the 13 years from March 2000 through February 2013 prior to the computation of the albedos. The right panels show the zonal means of the left panels. The dashed red line in the upper right panel is the fit from the heuristic model in the box on p. ??.

The albedo variation is the gradual increase from ~20% in the tropics, to ~45% at 60°N/S, to ≥70% near the poles. This is clearly evident in the zonal-mean albedo (Fig. 4.5a, right panel), which increases nearly symmetrically away from the equator—despite the strongly hemispherically asymmetric distribution of continents. The zonal albedo variations (the variations along latitude belts) are comparatively weak. They correlate with measures of cloudiness, such as cloud water path (Fig. 1.21) and cloud fraction (Fig. 1.23). For example, the albedo is elevated (~30%) over the subtropical eastern Pacific with its high fraction of low clouds (cf. Fig. 1.23), and it is reduced over the subtropical western Pacific with its lower cloud water path and cloud fraction. These observations indicate that clouds exert the primary control on Earth’s albedo variations. But variations of surface properties are also important, especially regionally, for example, near
the poles.

This contention is confirmed by the clear-sky albedo: the albedo contribution from cloud-free regions of the sky (Fig. 4.5b, left panel). That is, the clear-sky albedo is computed by dividing the upward solar radiative energy flux emanating at the top of Earth’s atmosphere by the total solar radiative energy flux incident at the top of the atmosphere. The difference between the total albedo (Fig. 4.5a) and the clear-sky albedo (Fig. 4.5b) gives the contribution of cloudy regions of the sky to the total albedo. The clear-sky albedo is substantially lower than the total albedo in most regions, except near the poles and in some regions with few clouds, for example, over subtropical oceans and deserts, where the two are similar. The clear-sky albedo increases more slowly away from the equator than the total albedo. In the zonal mean, it increases from \( \sim 10\% \) in the tropics, to \( \sim 20\% \) at 60°N/S, to \( \geq 70\% \) near the poles (Fig. 4.5b, right panel). This confirms the central role of clouds in controlling Earth’s Bond albedo and albedo variations at the top of the atmosphere. In particular, most of the gradual increase of the albedo from the tropics to the polar regions must be associated with clouds.

How does the gradual increase of the top-of-atmosphere albedo with latitude arise? [it’s not the oceans—see clear-sky albedo] The relatively large cloud fraction and large surface albedo near the poles play a role. But the albedo increase is more gradual and more uniform across land–ocean boundaries than can be accounted for by such variations in cloud cover or surface albedo (compare, e.g., the albedo in Fig. 4.5a with the cloud fraction in Fig. 1.23). The dependence of the albedo of individual clouds on the angle of incidence of sunlight probably also contributes. Another important reason probably is that the overall albedo of cloud fields composed of several clouds also depends on the angle of incidence of sunlight, for different reasons than the albedo of individual clouds. If one looks at a cloud field directly from above, at a viewing angle close to the normal, holes in the cloud field are evident. Over oceans, they expose darker surfaces below the clouds. By contrast, if one looks at a cloud field with vertical protrusions at a more oblique viewing angle, looking toward the horizon, the cloud field appears more uniform, because the vertical protrusions mask holes and hide the darker surfaces below (Fig. 4.6). Instead of the surface, one sees facets of clouds tilted toward the viewing direction. A result is that when the sun is lower in the sky and sunlight strikes a cloud field more obliquely, with closer-to-normal incidence on some tilted facets, the overall albedo of the cloud field and of an underlying darker surface is higher than at more normal incidence. Because the average angle of incidence of sunlight on cloud fields in high latitudes is more oblique than at low latitudes, the mean albedo increases with latitude. The box on p. ?? presents an illustrative heuristic model of how reflection from cloud fields can lead to a latitude-dependent albedo at the top of the atmosphere.

Part of the gradual increase of Earth’s albedo with latitude may arise because cloud fields are not plane and parallel to Earth’s surface but have vertical protrusions. At oblique angles of incident of sunlight, the protrusions cast shadows on
Figure 4.6: Clouds over the Amazon Delta as viewed from the International Space Station. Openings between the clouds are clearly visible in the foreground, at a nearly normal viewing angle. Toward the horizon, at more oblique viewing angles, openings are masked by vertical cloud protrusions, and the cloud field appears more uniform. (Image credit: NASA)

openings between the clouds and reflect sunlight from their illuminated faces. This increases the overall albedo of a cloud field. Quantifying this effect has shown it to be a significant but not sufficient factor for explaining the symmetry of Earth’s albedo, and the complete picture remains to be unveiled by future research.

Notes

1. See Cox and Munk (1955) for an early theoretical calculation of the ocean albedo given its wave spectrum and Payne (1972) and Jin et al. (2004) for measurements.
2. See Loeb et al. (2009) and Stephens et al. (2012) for a description of the data on which Fig. 4.5 is based.
**Constants and Parameters**

**UNIVERSAL PHYSICAL CONSTANTS**

- \( c = 2.998 \times 10^8 \text{ m s}^{-1} \) : Speed of light in vacuum
- \( h = 6.626 \times 10^{-34} \text{ m}^2 \text{ kg s}^{-1} \) : Planck constant
- \( k = 1.381 \times 10^{-23} \text{ m}^2 \text{ kg s}^{-2} \text{ K}^{-1} \) : Boltzmann constant
- \( N_A = 6.022 \times 10^{23} \) : Avogadro constant
- \( R_0 = 8.315 \text{ J K}^{-1} \text{ mol}^{-1} \) : Universal gas constant
- \( \sigma = 5.670 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4} \) : Stefan-Boltzmann constant

**EARTH PARAMETERS**

- \( d_0 = 1.495978707 \times 10^{11} \text{ m} \) : Astronomical unit (mean Earth-Sun distance)
- \( d_s = 1.00 \text{ au} \) : Semi-major axis of orbit
- \( r_e = 6.371 \times 10^6 \text{ m} \) : Earth's mean radius
- \( R \) : Specific gas constant of air
- \( S_0 = 1362 \text{ W m}^{-2} \) : Total solar irradiance (solar constant)
- \( Y_a = 365.26 \text{ d} \) : Anomalistic year (period from perihelion to perihelion)
Symbols and Abbreviations

FREQUENTLY USED SYMBOLS

$\alpha, \alpha_{\lambda}$  Albedo (integrated over solar spectrum or per wavelength)
$\gamma$  Obliquity of Earth’s rotation axis
$\delta$  Solar declination angle
$\epsilon, \epsilon_{\lambda}$  Emissivity (broadband or per wavelength)
$\eta$  Hour angle
$\theta$  Solar zenith angle
$\theta'$  Solar elevation angle
$\lambda$  Wavelength, longitude
$\nu$  Frequency
$\omega$  Longitude of perihelion
$\rho$  Density (of air unless otherwise noted)
$\rho_v$  Density of water vapor
$\phi$  Latitude, azimuth angle
$\alpha_{\lambda}$  Absorptivity
$A$  True anomaly
$B_{\nu}, B_{\lambda}$  Planck function
$c_s$  Specific heat capacity
$C_d$  Drag/transfer coefficient
$c_p$  Specific heat at constant pressure
$c_v$  Specific heat at constant volume
$d$  Earth-Sun distance
$e$  Eccentricity of orbit/vapor pressure
$e^*$  Saturation vapor pressure
$E$  Eccentric anomaly
$F$  Insolation
$h$  Height scale
$h_e$  Earth’s orbital angular momentum around Sun
$\mathcal{H}$  Relative humidity
$I_v, I_{\lambda}$  Radiance
$L_s$  Solar longitude
$q$  Specific humidity
$F_L$  Latent energy flux
$F_S$  Sensible heat (dry enthalpy) flux
$p$  Pressure
Specific humidity, saturation specific humidity

Mean anomaly

Solar radiative energy flux

Time of perihelion/vernal equinox

Solar noon

Temperature

Eastward velocity

Three-dimensional velocity vector \( u = (u, v, w) \)

Northward velocity

Horizontal velocity vector \( v = (u, v) \)

Upward velocity

Vertical (height) coordinate

ABBREVIATIONS

CE Common Era (Gregorian calendar year)

BP Before present (here, before the year 2000)

DJF December, January, February

ITCZ Intertropical Convergence Zone

JJA June, July, August

NASA National Aeronautics and Space Administration

TOA Top of atmosphere

UTC Coordinated Universal Time

UNITS

Astronomical unit \((1 \text{ au} = 1.495978707 \times 10^{11} \text{ m})\)

Day \((1 \text{ d} = 86400 \text{ s})\)

Hour \((1 \text{ h} = 3600 \text{ s})\)

Joule \((1 \text{ J} = 1 \text{ kg m}^2 \text{ s}^{-2})\)

Year \((1 \text{ kyr} = 10^3 \text{ yr}, 1 \text{ Myr} = 10^6 \text{ yr})\)

Steradian

Sverdrup \((1 \text{ Sv} = 10^9 \text{ kg s}^{-1})\)

Watt \((1 \text{ W} = 1 \text{ J s}^{-1})\)
Bibliography


Kay, J. E., and A. Gettelman, 2009: Cloud influence on and response to seasonal


Masson-Delmotte, V., and Coauthors, 2010: EPICA Dome C record of glacial


## Index

- absorptivity, 65
- Adhémar, Joseph, 103
- aerosols, 6
  - color, 120
- African Humid Period, 33
- Agassiz, Louis, 103
- albedo, 54, 65
  - Bond, 54, 70
  - clear sky, 124
  - clouds, 122
  - measurement, 55
  - ocean, 121
  - relation to color, 55
  - surface types, 55
  - top of atmosphere, 123
- angular momentum, 23
  - orbital, 81, 87
- annual temperature range, see temperature anomaly
  - use in celestial mechanics, 79
  - use in climate sciences, 6
- Antarctic circle, see polar circle
- aphelion, 80
  - definition, 80
- apsis, 100
- Arctic circle, see polar circle
- arid climate, 31
- Assmann, Richard, 17
- astronomical unit, 52
- atmospheric infrared window, 72
  - austral, 48
- azimuth angle, 62
- Baroclinic eddies, 193
- barycenter, 78
- Beer’s law, 128
- Beer, August, 128
- bernoulli principle, 167
- blackbody radiation, 58
  - radiance, 61
- Boltzmann constant, 59
- boreal, 48
- Bouguer, Pierre, 128
- Brahe, Tycho, 78
- brightness temperature, 66, 69
  - thermal infrared, 67
- Brown, Robert, 58
- Brownian motion, 58
- calendar
  - Gregorian, 85
  - Julian, 83, 85
- CALIPSO, 40
- carbon dioxide
  - atmospheric concentration, 13
- cavity radiation, see blackbody radiation
- Cenozoic, 16, 105
- chaos, 102
- cloud fraction, 40, 41
- cloud water path, 38, 39
- clouds, 38
  - albedo, 122
  - cirrus, 54
  - color, 120
- CloudSat, 40
- Columbus’ voyage, 22
- Columbus, Christopher, 21
- convection, 31, 33
- cosine law, 68, 94
- Croll, James, 103
data assimilation, 2
day
  length, 28
daylight, 93
de Saussure, Horace-Bénédict, 17
deciliation angle, 90, 90
dipole moments, 110
Dust Bowl, 6
Earth orbit, 78, 79
  Earth-Sun distance, 79–86
  parameters, 79, 82
  variations, see Milankovitch cycles, 12–14, 95
Earthshine, 55
easterlies, see zonal wind
eccentricity, 79
Eemian, 12
effective temperature, 64, 69
  Earth, 57, 71
  Sun, 54, 70
electromagnetic spectrum, 52, 58
electromagnetic waves, 53, 57, 59
emissivity, 65
Eocene, 15
  Early Eocene Climatic Optimum, 15
  Middle Eocene Climatic Optimum, 15
epoch (astronomy), 83
equinox, 79
  autumnal, 84, 106
  vernal, 84, 106
European Centre for Medium-Range Weather Forecasting, 2
evaporation, 27, 27, 28
  equatorial minimum, 28
  global and annual mean, 29
  global distribution, 28
  limitation by water availability, 29
extratropics, 2
feedback
  carbon cycle, 104
ice-albedo, 55, 104
Ferrel cell, 25
Ferrel, William, 25
foraminifera, 9, 10, 13, 16
Fresnel’s reflection formulas, 121
Galilei, Galileo, 106
gas constant
  universal, 59
geothermal gradient, 47
glacial period, 11
  inception, 11, 103
  termination, 11
  glacial-interglacial cycle, see ice age
  greenhouse effect, 72
  greenhouse gas, 8
  water vapor, 38
grey body, 65
Gulf Stream, 29
Hadley cell, see Hadley circulation
Hadley circulation, 25, 31, 35
Hadley, George, 25
Hays, John, 107
heat capacity
  constant volume, 163
Held-Hou model, 189
Herschel, John, 88
Hide’s theorem, 178
Holocene, 8, 33
  Thermal Maximum, 10
horse latitudes, 20
  hour angle, 91, 91
  at sunrise and sunset, 93
humid climate, 31
hurricane
  Sandy, 68
hyperthermal, 15
ice age, 8, 11, 103
  100-kyr problem, 105
  periodicity, 14
ice core, 9, 11
ice sheet, 3, 10, 11, 13
Antarctic, 15
Fennoscandian, 11
Last Glacial Maximum, 12
Laurentide, 11
ice-albedo feedback, see feedback
Imbrie, John, 107
infrared radiation, 27
insolation, 72
calculation, 94
definition, 52
top of atmosphere, 88, 95
interglacial period, 11
International Geophysical Year, 1
Intertropical Convergence Zone, see ITCZ
irradiance, 53
definition, 52
spectral, 53
isotopes
fractionation, 11
ice cores, 11
oxygen, 8, 13
precipitation, 11
isotopologue, 11
ITCZ, 22, 25, 30, 40
connection to Hadley circulation, 25
precipitation rate, 29
jet stream, 23
Kepler
first law, 78
second law, 81, 87
third law, 96
Kepler problem, 78, 95
Kepler, Johannes, 78
Kirchhoff’s law, 65
Kuroshio Current, 28
Lambert, Johann Heinrich, 128
Lambert-Beer law, 128
lapse rate, 17, 18
Last Glacial Maximum, 11
latent heat, 27
lidar, 40
Little Ice Age, 10
local thermodynamic equilibrium, 66
longitude of perihelion, see perihelion
longwave radiation, 58
absorption by condensate, 66
low-emissivity coating, 66
luminosity, see Sun, luminosity
main-sequence star, 51
mean anomaly, 82
at vernal equinox, 85
Medieval Warm Period, 10
meridional overturning circulation, 25, 25
zonal mean, 26
meridional plane, 25
meridional wind, 20
mesosphere, 17
meter
definition, 47
microwave measurement, 2, 33, 38, 58
middle atmosphere, 17
Milankovitch cycles, 95
eccentricity, 96, 97
obliquity, 96, 98
precession, 85, 100, 100
apsidal, 100
axial, 100
precession band, 101, 105
precession index, 101
Milankovitch, Milutin, 96, 103
moist convection, 145
monsoon
Asian, 20, 23, 38
net precipitation, 31, 31
global distribution, 32
Newton, Isaac, 78
obliquity, 14, 79, 89
definition, 90
<table>
<thead>
<tr>
<th>Index</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>variations, see Milankovitch cycles</td>
<td>11</td>
</tr>
<tr>
<td>ocean currents, 44</td>
<td>11</td>
</tr>
<tr>
<td>Oligocene, 15</td>
<td>11</td>
</tr>
<tr>
<td>opacity, 73</td>
<td>11</td>
</tr>
<tr>
<td>optical thickness, 73</td>
<td>11</td>
</tr>
<tr>
<td>orbit, see Earth orbit</td>
<td>11</td>
</tr>
<tr>
<td>orbital (electron), 58</td>
<td>11</td>
</tr>
<tr>
<td>outgoing longwave radiation, 71, 73, 74</td>
<td>11</td>
</tr>
<tr>
<td>ozone, 54, 76</td>
<td>11</td>
</tr>
<tr>
<td>Pacific Decadal Oscillation, 6</td>
<td>11</td>
</tr>
<tr>
<td>Paleocene-Eocene Thermal Maximum, 15</td>
<td>11</td>
</tr>
<tr>
<td>Paleolithic Age, 14</td>
<td>11</td>
</tr>
<tr>
<td>perihelion, 80</td>
<td>11</td>
</tr>
<tr>
<td>definition, 80</td>
<td>11</td>
</tr>
<tr>
<td>longitude of, 80, 84</td>
<td>11</td>
</tr>
<tr>
<td>precession, see Milankovitch cycles</td>
<td>11</td>
</tr>
<tr>
<td>time of, 83</td>
<td>11</td>
</tr>
<tr>
<td>photon gas, 61</td>
<td>11</td>
</tr>
<tr>
<td>Planck constant, 61</td>
<td>11</td>
</tr>
<tr>
<td>Planck function, 61–69</td>
<td>11</td>
</tr>
<tr>
<td>Sun and Earth, 64</td>
<td>11</td>
</tr>
<tr>
<td>transformation from frequency to wavelength, 63</td>
<td>11</td>
</tr>
<tr>
<td>Planck’s law, 61</td>
<td>11</td>
</tr>
<tr>
<td>Planck, Max, 61</td>
<td>11</td>
</tr>
<tr>
<td>plate tectonics, 15, 71</td>
<td>11</td>
</tr>
<tr>
<td>Pleistocene, 14</td>
<td>11</td>
</tr>
<tr>
<td>Pliocene, 14</td>
<td>11</td>
</tr>
<tr>
<td>polar circle, 97</td>
<td>11</td>
</tr>
<tr>
<td>definition, 93</td>
<td>11</td>
</tr>
<tr>
<td>polar day, 93</td>
<td>11</td>
</tr>
<tr>
<td>polar night, 93</td>
<td>11</td>
</tr>
<tr>
<td>power, 51</td>
<td>11</td>
</tr>
<tr>
<td>precession, see Milankovitch cycles</td>
<td>27, 29</td>
</tr>
<tr>
<td>precipitable water</td>
<td>27, 29</td>
</tr>
<tr>
<td>definition, 33</td>
<td>27, 29</td>
</tr>
<tr>
<td>global distribution, 34</td>
<td>27, 29</td>
</tr>
<tr>
<td>precipitation, 27, 29</td>
<td>27, 29</td>
</tr>
<tr>
<td>global and annual mean, 29</td>
<td>27, 29</td>
</tr>
<tr>
<td>global distribution, 30</td>
<td>27, 29</td>
</tr>
<tr>
<td>isotopic composition, 11</td>
<td>11</td>
</tr>
<tr>
<td>net, see net precipitation</td>
<td>11</td>
</tr>
<tr>
<td>storm track, 31</td>
<td>11</td>
</tr>
<tr>
<td>storms, 34</td>
<td>11</td>
</tr>
<tr>
<td>quantization postulate, 61</td>
<td>11</td>
</tr>
<tr>
<td>radar, 23, 40</td>
<td>11</td>
</tr>
<tr>
<td>radiance, 61</td>
<td>11</td>
</tr>
<tr>
<td>radiant intensity, see radiance</td>
<td>11</td>
</tr>
<tr>
<td>radiative energy flux</td>
<td>11</td>
</tr>
<tr>
<td>blackbody, 68</td>
<td>11</td>
</tr>
<tr>
<td>solar, 52, 78</td>
<td>11</td>
</tr>
<tr>
<td>spectral, 53</td>
<td>11</td>
</tr>
<tr>
<td>radiometric dating, 10</td>
<td>11</td>
</tr>
<tr>
<td>radiosonde, 1, 38</td>
<td>11</td>
</tr>
<tr>
<td>rawinsonde, 23</td>
<td>11</td>
</tr>
<tr>
<td>Rayleigh scattering, 117</td>
<td>11</td>
</tr>
<tr>
<td>cross-section, 119</td>
<td>11</td>
</tr>
<tr>
<td>reanalysis data, 2</td>
<td>11</td>
</tr>
<tr>
<td>red giant, 51</td>
<td>11</td>
</tr>
<tr>
<td>relative humidity</td>
<td>11</td>
</tr>
<tr>
<td>definition, 35</td>
<td>11</td>
</tr>
<tr>
<td>zonal mean, 37</td>
<td>11</td>
</tr>
<tr>
<td>residence time</td>
<td>11</td>
</tr>
<tr>
<td>atmospheric water, 33</td>
<td>11</td>
</tr>
<tr>
<td>ocean water, 29</td>
<td>11</td>
</tr>
<tr>
<td>Roaring Forties, 20</td>
<td>11</td>
</tr>
<tr>
<td>salinity, 31</td>
<td>11</td>
</tr>
<tr>
<td>satellite, 1</td>
<td>11</td>
</tr>
<tr>
<td>cirrus cloud detection, 54</td>
<td>11</td>
</tr>
<tr>
<td>geostationary, 67</td>
<td>11</td>
</tr>
<tr>
<td>lidar, 40</td>
<td>11</td>
</tr>
<tr>
<td>microwave radiometer, 2, 33, 38</td>
<td>11</td>
</tr>
<tr>
<td>polar-orbiting, 2</td>
<td>11</td>
</tr>
<tr>
<td>radar, 40</td>
<td>11</td>
</tr>
<tr>
<td>seasonality, 92, 98</td>
<td>11</td>
</tr>
<tr>
<td>seasons</td>
<td>88</td>
</tr>
<tr>
<td>astronomical definition, 84</td>
<td>88</td>
</tr>
<tr>
<td>incident solar radiative energy,</td>
<td>88</td>
</tr>
<tr>
<td>length of, 82, 87</td>
<td>88</td>
</tr>
</tbody>
</table>
semi-grey atmosphere, 72
Shackleton, Nicholas, 107
shortwave radiation, 58
skin temperature, 75, 76
solar constant, see total solar irradiance
solar day
definition, 92
tidally locked planet, 107
solar longitude, 80, 84
solar radiation, 50
atmospheric absorption, 53
spectrum, 53, 53, 54
solar terminator, 84, 93
solid angle, 61, 62
solstice, 79
northern summer, 84
northern winter, 84
southern summer, 84
Somali jet, 20
South Pacific Convergence Zone, 29, 30
specific humidity, 35
definition, 35
zonal mean, 36
spectrum
red, 16, 104
speed of light, 53
Stefan-Boltzmann constant, 69
Stefan-Boltzmann law, 68
storm track, 20
clouds, 40
precipitation rate, 31
stratocumulus clouds, 140
stratopause, 17
stratosphere, 17
temperature, 76
sublimation, 27
subsolar point, 84, 89
latitude and longitude, 90
subtropics, 3
Sun, 50
chromosphere, 70
facula, 51
faint young, 76
fusion reaction, 50
luminosity, 50, 51, 70, 76, 78, 86
photosphere, 63, 70
properties, 50
radiative energy flux, 70, 86
sunspot cycle, 51, 52, 87
Sverdrup, 26
Sverdrup, Harald, 48
Teisserenc de Bort, Léon, 17
temperature
annual range, 3, 5
global- and annual-mean, 3
land-ocean seasonal contrast, 4
lapse rate, 17, 18
long-term variations, 9, 16, 8–16
power spectrum, 105
proxies, 8, 9, 13
recent variations, 6, 5–8
relation to molecular motion, 59
surface, 2, 4
vertical structure, 18, 17–20
zonal mean, 19
terrestrial radiation, 56
spectrum, 57
window region, 58
thermal infrared radiation, see long-wave radiation
thermodynamics
second law, 60
Thompson scattering, 119
tidally locked planet, 107
total solar irradiance, 53, 86, 87
definition, 52
trade winds, 20–23
transpiration, 27
Tropic of Cancer, 2, 84, 90, 97
Tropic of Capricorn, 2, 84, 90, 97
tropics, 2
tropopause, 17, 20
definition, 18
height, 18
troposphere, 17
true anomaly, 79, 80
volcanism, 15
wavenumber, 53
weathering, 15
westerlies, see zonal wind
western boundary current, 29
Wien's displacement law, 63
wind
jet stream, 23
shear, 23
surface, 20, 21

year
anomalous, 82
tropical, 85
Younger Dryas, 13

zenith angle, 62, 88, 89, 90
diurnal average, 93
zonal mean, 18
zonal wind, 20
easterlies, 20, 22, 23
westerlies, 20, 23
zonal mean, 24