1 Supplement 11.1: Measuring Temperature

This supplement expands on a topic addressed on page 201 of the textbook.

1.1 Temperature Scales

In the discussion of thermal mechanics in the textbook, I specified kelvins as the units of temperature. Why this particular unit? Although the history behind this choice is of no fundamental scientific importance, it is too interesting to leave out.

As physics blossomed in the aftermath of Newton, a wide variety of temperature-measuring techniques were proposed, and, through trial and error, the mercury thermometer (Figure 1) emerged as a practical instrument. Mercury in the thermometer’s bulb expands and contracts in direct proportion to temperature, moving the meniscus along a graduated scale. To calibrate the device, all one needs to do is to define two points on the linear relationship between temperature and the height of the mercury column.

![Figure 1: A mercury thermometer.](image)

As you might imagine, different instrument makers had different ideas for the choice these calibration points. In the early 1700s, a Dutch instrument maker, Daniel Gabriel Fahrenheit (1686-1736), used a mixture of water, ice, and
ammonium chloride to provide a repeatable low temperature that he defined as zero degrees. He took the second standard temperature as that of his wife’s armpit, designating it as 96 degrees. This, then, was the original Fahrenheit scale of temperature, as proposed in 1724. On this scale, pure water freezes at 32 degrees and boils at 205 degrees, a difference of 173 degrees. The Fahrenheit scale was subsequently revised to have 180 degrees between the freezing and boiling points of water (a rounder number, I suppose), and on this revised scale, typical human body temperature is 98.6°F. Today, the Fahrenheit scale is used officially only in the United States and Belize.

Although readily available to every scientist, human body temperature is too variable to provide a reliable standard. In 1742 the Swedish instrument maker Anders Celsius (1701-1744) proposed an alternative temperature scale, setting zero degrees to the boiling point of water and 100 degrees to water’s freezing point. Carl Linnaeus (1707-1778), the famous botanist and creator of the Linnanean system of taxonomic classification, was intrigued by Celsius’s temperature scheme, but he eschewed the inverted scale. Instead, he ordered a thermometer with 0 set to the freezing point of water and 100 at its boiling point, and this modified version of Celsius’s thermometer was an instant success. Thus, the modern centigrade scale of temperature was introduced to science by a biologist. In 1948, the centigrade scale was officially renamed the Celsius scale.

Nowadays, alcohol commonly replaces mercury in thermometers. It has a lower coefficient of expansion (and is therefore less sensitive to temperature), but it has the advantage of being relatively nontoxic.

The story of temperature doesn’t stop there, though. The kinetic theory of gases—developed long after Fahrenheit and Celsius—predicts the bulk properties of a gas (its temperature, pressure, and volume) through an understanding of the random motion of individual molecules. On the Celsius scale, if a gas behaved according to the kinetic theory, motion would cease at $-273.15\degree C$. This theoretical temperature—absolute zero—serves as an alternative zero point for a temperature scale$^1$. Accordingly, on the Kelvin scale, water freezes at 273.15 kelvins and boils at 373.15 K. The value of $k_T$ cited in the chapter (4.84 $\times$ 10$^{22}$ kelvins per joule) was chosen to make the size of a kelvin conform to the Celsius degree. Note that it is just kelvins, not degrees kelvin.

1.2 Thermometers in Biology

Let’s begin with perhaps the most elementary practical question in thermal biology: How does one measure the temperature of an organism? The standard method for measuring human body temperature—inserting a mercury or alcohol thermometer in one orifice or another and waiting a minute or two—clearly won’t work. In many cases, the organisms in question do not have orifices conveniently located for the insertion of thermometers and in some instances

$^1$In fact, gases at very low temperatures do not conform to the kinetic theory, and the deviations from theory can be explained only by quantum physics. The net result is that, even at 0 K, there must be some molecular motion.
the bulb of a thermometer is larger than the organism itself. Furthermore, mercury and alcohol thermometers are fragile and they have to be read by eye, which makes for tedious work when temperature must be recorded repeatedly.

Fortunately, there is a simple, reliable, and cheap alternative—the thermocouple. In 1821, Thomas Seebeck (1770-1831) discovered that the junction between two dissimilar metals produces a voltage proportional to absolute temperature. The precise voltage depends on the metals involved, but for any given combination, the proportionality between temperature and voltage in this thermoelectric effect allows the junction—a thermocouple—to serve as a thermometer. All one need do is weld two wires together to form a thermocouple, calibrate the device by measuring voltage at a few known temperatures, and one is set to go (Figure 2).

![Diagram of a thermocouple](image)

Figure 2: The junction between two dissimilar metals (here, copper and constantan) forms a thermocouple.

Or so it would seem. Although the basic idea of a thermocouple thermometer is incredibly simple, there is a basic complication that must be overcome. Consider, for instance, the apparatus shown in Figure 3. A copper wire is joined to a wire made of constantan (an alloy of 55% copper and 45% nickel), forming a standard type T thermocouple, and we measure the resulting temperature-dependent voltage with a voltmeter. The complication stems from the fact that wires in the voltmeter—and more importantly, their connections to the thermocouple—are themselves made of copper. That doesn’t raise any problem with the connection between the voltmeter and the copper wire of the thermocouple—because the metals are the same, no thermoelectric effect is induced—but there is a distinct problem at the connection between the voltmeter’s copper wiring and the thermocouple’s constantan component. Because these are made of dissimilar metals, their connection forms a second thermocouple, one in which the order of the metals—and hence the voltage produced by the thermoelectric effect—is reversed from that of the primary thermocouple. In short, if the primary thermocouple and the junction of the constantan wire to the voltmeter are at the same temperature, the voltage produced by one just offsets that produced by the other, and, regardless of temperature, the meter measures no voltage. A device whose output doesn’t vary with temperature
Figure 3: The problem with measuring the voltage produced by a thermocouple.

makes a lousy thermometer.

The solution to this problem is to maintain the secondary thermocouple at a constant, reference temperature. If the temperature of the primary thermocouple varies while that of the secondary, reference thermocouple does not, a net voltage is produced proportional to the difference in temperature between the two. In olden days, the reference thermocouple was held at a constant temperature by immersing it in a well-stirred ice bath at 0°C. In this configuration (Figure 4), the voltage measured by the meter is directly proportional to the Celsius temperature of the primary thermocouple.

Figure 4: A reference thermocouple solves the problem.

Straightforward as this solution is to the problem posed by the secondary thermocouple, it is inconvenient to lug an ice bath around while one makes measurements in the field. Fortunately, the wizards of the electronic industry have provided a more practical alternative. Nowadays, the reference thermocouple and ice bath can be replaced by an electronic circuit that mimics their effect. For a few hundred dollars one can purchase a battery-powered ther-
mocouple thermometer that combines an electronic ice-point reference with a sensitive voltmeter. One simply connects the thermometer to an appropriate thermocouple and one is in business.

An additional advantage of thermocouple technology is the ease with which it can interface with recording apparatus. Standard digital dataloggers take voltages as their input, making it easy to record the voltage output of a thermocouple thermometer. For an in-depth and up-to-date discussion of thermocouple thermometry, consult the *Temperature Handbook* produced by Omega Engineering, Incorporated.

Thermocouples themselves are easily made. A variety of manufacturers sell rolls of thermocouple wire composed of conjoined lengths of insulated, dissimilar metals. An appropriate length of wire is cut off, and the individual wires at one end are hooked to the electronics. The insulation is then stripped from a short length of the other end of the wires, and the exposed wires are welded or soldered together to form the thermocouple itself.

Thermocouples work just fine when wet, a huge advantage when working with living organisms. If a thermocouple breaks or corrodes, one simply cuts it off, strips a new length of wire, and makes a new one. For use in corrosive environments (e.g., seawater), the thermocouple junction can be coated with nail polish or some other coating. Thermocouple wire is available in different thicknesses, from thin 40 gauge wire handy for making tiny thermocouples to insert into small organisms, to the more robust 22 gauge wire appropriate for long-term measurements of air and water temperatures.

In summary, thermocouples provide a cost-effective and convenient way to measure temperature.

As convenient as thermocouples are, they have one important disadvantage: they are invasive. To measure the temperature of a littorine snail, for instance, the thermocouple must either be stabbed into the exposed body of the animal or inserted into a hole drilled in the shell. Allow yourself a moment to empathize with the snail in such an experiment, and you will conclude that the intrusion of a thermocouple might adversely affect the organism’s ability to behave normally. A non-intrusive method of measuring body temperature would be advantageous.

The Stefan-Boltzmann equation (see Chapter 12) provides the basic idea for a noninvasive thermometer:

\[ R_{lw} = \epsilon_{lw} \sigma_{SB} T^4. \]

Here, \( R_{lw} \) is the long-wave radiance, \( \sigma_{SB} \) is the Stefan-Boltzmann constant \( (5.67 \times 10^{-8} \text{ W} \cdot \text{m}^{-2} \cdot \text{K}^{-4}) \), \( T \) is the absolute temperature of an emitting surface, and \( \epsilon_{lw} \) is its long-wave emittance, a property of the material from which light is being emitted (again, see Chapter 12). The higher the temperature of a surface, the more infrared light it emits. Thus, if an apparatus can be built to quantify long-wave radiance, \( R_{lw} \), it can be used to measure temperature. This is the principle behind infrared thermometers.

Simple as this idea is in principle, making it work in practice is tricky. The primary practical problem is that any apparatus one builds to measure the
infrared light emitted by an object in the environment is itself likely to be at near environmental temperature, and it is therefore prone to emit a confounding amount of longwave radiation. Early models of infrared thermometers solved this problem by cooling their sensors with liquid nitrogen, but the apparatus involved was too bulky and expensive for practical field use. In recent decades, however, the electronic wizards have again come to the rescue, and, by a series of machinations I will not attempt to explain, they have produced instruments capable of accurately measuring the amount of infrared light emitted by surfaces. The less expensive of these devices measure emittance from one small area. More expensive devices—infrared cameras—simultaneously measure temperature at thousands of points in the infrared image of a surface.

Infrared thermometers and cameras are an amazing feat of technology, but they, too, have several practical disadvantages. Perhaps the most significant of these is intrinsic to the Stefan-Boltzmann equation. $R_{lw}$, the infrared radiance, is what is measured by an infrared thermometer, and it is indeed proportional to $T^4$. But this proportionality is set by $\epsilon_{lw}$, the long-wave emissivity. So, unless one has an accurate measure of $\epsilon_{lw}$, one cannot accurately measure $T$ from a simple measurement of $R_{lw}$. In other words, you can’t simply run out to the field with an infrared thermometer and immediately start taking accurate temperature measurements; you first have to measure the long-wave emissivity of the object in question. In practice, one begins by measuring the temperature of the object (using a thermocouple, for instance), and then, by comparison to the measured radiance, one calculates $\epsilon_{lw}$. Top-end infrared thermometers and cameras can be calibrated so that the temperature they read takes this measured emissivity into account. The need for an initial, separate measure of body temperature takes some of the shine off the concept of infrared thermometry. Once emissivity has been measured, however, subsequent IR measurements can be made without recalibration.

Although it can be advantageous, the nonintrusive nature of infrared thermometers has a downside as well. By their very nature, infrared thermometers measure the temperature of an object’s surface. If surface temperature differs from a body’s core temperature, an IR thermometer might not measure the temperature we desire. This effect will be most apparent for objects that have low thermal conductivity: a dry frond of intertidal seaweed, for instance, the porous shell of an acorn barnacle, or the trunk of a tree.

Temperature measured by an infrared thermometer can also be affected by an object’s surroundings. If long-wave radiation from the surroundings is reflected from the object, the additional irradiance observed by the thermometer is interpreted as an increase in temperature. Measured temperature can also be affected by the distance between thermometer and object. Water vapor in air is a potent absorber of infrared light, so some of the light emitted by an object is absorbed as it passes through humid air on its way to a thermometer, reducing the object’s apparent temperature. Of course, water vapor is also a potent emitter of infrared light, so light from water vapor is added to that of the object along the path from object to thermometer. The net effect is that the apparent temperature of an object gets closer and closer to air temperature the father the
infrared thermometer is from the object being measured. The effect is negligible over distances of 1-2 m, but is evident over larger distances. For example, one might desire to record a large-scale temperature map of a landscape. It would be easy to use an infrared camera to take a picture of the terrain from a distance, but the effect of the intervening air can skew the resulting temperature measurements. The absorption of infrared light by water also means that infrared thermometers cannot be used to measure the temperature of submerged objects. And lastly, infrared thermometers (cameras in particular) can be quite expensive.

In summary, thermocouples provide a cost-effective, reliable, but somewhat invasive method for measuring body temperatures. Infrared thermometers offer an nonintrusive alternative, but they are expensive and must be carefully calibrated for each organism and field situation.

There are many other ways in which temperature can be measured (thermistors, resistive thermal devices (RTDs), integrated circuits, etc.), but none have proven as useful as thermocouples, and we will not delve into their details.

2 Supplement 11.2: Solar Heat Influx

In this supplement, I delve into the practical matters involved in measuring (or predicting) solar heat influx.

Flux of heat from the sun, $H_{sw}$, is governed by three parameters: short-wave absorptivity, ($\alpha_{sw}$), surface area ($A_p$ and $A_{sky}$), and solar irradiance ($I_{sw,dir}$ and $I_{sw,dif}$):

$$H_{sw} = \alpha_{sw} [A_p I_{sw,dir} (t) + A_{sky} I_{sw,dif} (t)].$$

Two methods are available for estimating $H_{sw}$. In the first, one directly measures absorptivity and area for the object of interest and irradiance at the pertinent time and location. These can then be plugged into Eq. 2. This is the most accurate way of estimating solar heat influx, but its use is limited to times and places where recording instruments can be set up and maintained. Many questions in thermal biology require us to think beyond this constraint. For instance, investigation of the evolution of thermal tolerance requires lengthy time series of body temperatures, which in turn require us to predict the minute-to-minute variation in $H_{sw}$ for centuries at a time. To do so, however, we need to be able to predict the daily and annual variations of solar irradiance anywhere and for any time. IN this discussion of $H_{sw}$, we begin with empirical measurement and work our way up to these more theoretical applications.

2.1 Pyranometers

Solar irradiance is quantified using a pyranometer, a device that measures heat influx to a black surface exposed to sunlight (Figure 5). In olden days, sunlight
was allowed to fall on a bimetallic strip that had been painted black to ensure that all the incident light was absorbed. As the strip was heated, one metal expanded faster than the other, and the strip bent, thereby deflecting a pen that recorded irradiance. Modern day pyranometers also measure the intensity of sunlight by measuring the temperature of a black surface, but they do so by some means (such as a thermocouple or integrated circuit) that produces a voltage (rather than a deflection) proportional to irradiance. Typically, a pyranometer gathers light from a hemispherical view of the environment.

Both diffuse and direct light can impinge upon a pyranometer, and sunlight commonly arrives in both forms. The instrument conflates these two sources of heat, recording their combined effect. For purposes of heat-budget models, however, it is often necessary to separate the effects of diffuse and direct sunlight (e.g., they are applied to different areas, see Eq. 2), and to that end, let’s work our way through the process required to back-calculate each from pyranometer records.

When the sky is covered with a uniform layer of clouds, no direct sunlight penetrates to the ground (as evidenced by the lack of shadows), and all incident sunlight is diffuse,

\[
H_{sw} = \alpha_{sw} A_{sky} I_{sw, dif} \quad \text{(cloud-covered sky)}
\]  

(3)

By contrast, on a clear day light reaching a pyranometer has both direct and diffuse components. Direct irradiance is due to light traveling uninterrupted through the atmosphere, while clear-sky diffuse light is due to scattering of sunlight by molecules and small particles in the atmosphere. On a clear day, we tend to overlook diffuse irradiance, concentrating instead on direct sunlight. But it is diffuse sunlight that makes the daytime sky luminous. Without this scattering, the sky would be as dark during the days as it is at night. Short wavelengths are scattered more effectively than long wavelengths; this is why the sky is blue.

Let’s examine direct irradiance first. The measured irradiance of sunlight impinging directly on a planar surface depends on the angle of incidence at
which the light hits. Consider an analogy. When a flashlight shines directly down on a surface, light impinges on the surface over a relatively small area and irradiance is therefore high. By contrast, when the beam strikes the surface at an oblique angle, the same amount of light subtends a larger area, so irradiance is reduced. A consideration of the geometry of the situation (Figure 6) tells us

\[
I_{\text{meas,dir}} = I_{\text{sw,dir}} \sin \theta. \quad (4)
\]

Typically, when measuring sunlight, pyranometers are carefully oriented so that their measuring surface is horizontal, in which case \( \theta \) is the solar altitude angle, the angle the sun makes with the horizon (Figure 7).

Our next task is to account for clear-sky diffuse irradiance. To do so, we investigate the possible routes by which light arrives at Earth’s surface. The intensity of direct sunlight above Earth’s atmosphere, \( I_c \), is approximately 1366 W m\(^{-2}\) (the solar constant), but this intensity is reduced by absorption and scattering as light advances through the atmosphere. The direct intensity remaining when sunlight reaches the surface depends on the distance light has travelled through the atmosphere: the longer this path length, \( \ell \), the greater the absorption and scattering, and the less direct light remaining.

Now, the shortest path light can take from outer space to Earth’s surface is the vertical thickness of the atmosphere, \( z_{\text{min}} \) (Figure 8), and the fraction of sunlight remaining after passing through \( z_{\text{min}} \) is the transmittance, \( \tau \). Thus, when the sun is overhead, direct irradiance at the surface is

\[
I_{\text{sw,dir}} = I_c \tau. \quad (5)
\]

Transmittance would be 1 for a perfectly nonabsorbing atmosphere and 0 for a completely opaque atmosphere. In reality, for an exceptionally clear sky,
Figure 7: Irradiance measured by a horizontal pyranometer depends on solar altitude.

\[ \tau \approx 0.75. \] For a severely polluted sky, \( \tau \) can be as low as 0.45 (Liu & Jordan 1960).\(^2\)

Figure 8: The path length of sunlight through the atmosphere depends on solar altitude angle.

If sunlight arrives at any angle other than vertical, its path length through the atmosphere is increased. Consideration of the geometry in Figure 8 shows

---

that the actual path length is\(^3\)

\[ \ell \approx \frac{z_{\text{min}}}{\sin \theta} \]  

(6)

Here \( \theta \) is again the sun’s altitude angle, its angle above the local horizon. The smaller \( \theta \) is—that is, the closer the sun is to the horizon—the longer sunlight’s path through the atmosphere.

Now, if fraction \( \tau \) of sunlight is absorbed in the first \( z_{\text{min}} \) of travel through the atmosphere, a fraction \( \tau \) of the remaining light would be absorbed if light travelled through another \( z_{\text{min}} \) of atmosphere. In that case, irradiance left after travelling \( 2z_{\text{min}} \) would be a fraction \( \tau \times \tau = \tau^2 \) of its value above the atmosphere. In general,

\[ I_{\text{sw,dir}} = I_c \tau^{k_Z}, \]  

(7)

where

\[ k_Z = \frac{\ell}{z_{\text{min}}} = \frac{1}{\sin \theta}. \]  

(8)

Combining Eqs. 7 and 8, we see that

\[ I_{\text{sw,dir}} = I_c \tau^{\frac{1}{\sin \theta}}. \]  

(9)

This function is shown as the solid line in Figure 9 for \( \tau = 0.75 \). For a horizontal pyranometer, measured irradiance due to direct irradiance is

\[ I_{\text{sw,meas}} = I_c \tau^{\frac{1}{\sin \theta}} \sin \theta, \]  

(10)

a function shown by the dashed line in Figure 9.

In summary, if we know the atmosphere’s transmittance (\( \tau \)) and the angle of the sun relative to the horizon (\( \theta \)), we can calculate the clear-sky direct irradiance at sea level. Note that Equations 9 and 10 are accurate only for \( \theta \) greater than approximately 20°. When the sun is closer to the horizon, refraction by the atmosphere plays a substantial role in the sun’s apparent position in the sky, making it appear higher in the sky than it actually is, and its path length relatively less. For these small angles, irradiance is greater than that predicted by this simple approach.

The intensity of clear-sky diffuse irradiance is less easily calculated. On the one hand, if less direct irradiance reaches Earth’s surface—that is, if \( \tau \) is large—it is in part because more light has been scattered. And the more light scattered, the greater the diffuse irradiance. On the other hand, the reduction in light reaching Earth’s surface is also due to absorption, suggesting that the larger \( \tau \) is, the less light there is available to scatter. It turns out that absorption dominates this interaction, and the larger \( \tau \) is, the smaller the diffuse irradiance. From personal experience, we might also predict a relationship between solar altitude angle and clear-sky diffuse irradiance: if you look at the sky directly overhead, it

\(^3\)The equation given here is an approximation based on the assumption that the Earth is flat. A more precise expression takes the curvature of the globe into account, and an even more exact expression corrects for refraction.
is dim at dawn and gets brighter as the sun rises in the sky. We would therefore expect diffuse irradiance to increase with increasing solar altitude. Liu & Jordan (1960) present an equation that quantifies these expectations:

\[ I_{sw,dif} = 0.3(1 - \frac{1}{\pi} \tau \sin \theta) I_c \sin \theta \]  

(clear sky) \hspace{1cm} (11)

This relationship is shown as a function of \( \tau \) in Figure 10.

Let’s now return to our pyranometer. On a clear day, the total irradiance impinging on the instrument is the sum of direct and diffuse irradiances,

\[ I_{sw,total} = I_{sw,dir} \sin \theta + 0.3(1 - \frac{1}{\pi} \tau \sin \theta) I_c \sin \theta \] \hspace{1cm} (12)

Solving for \( I_{sw,dir} \), we see that

\[ I_{sw,dir} = \frac{I_{sw,total}}{\sin \theta} - 0.3(1 - \frac{1}{\pi} \tau \sin \theta) I_c \] \hspace{1cm} (13)

When transmittance is high—that is, when the sky is very clear—the diffuse-irradiance term is small relative to \( I_{sw,total} \), and

\[ I_{sw,dir} \approx \frac{I_{sw,total}}{\sin \theta} \] \hspace{1cm} (14)

If the sky is turbid, the clear sky diffuse-irradiance term needs to be taken into account.

Note that Eqs. 11 to 13 apply only to clear skies. When scattered clouds are present, calculating direct and diffuse irradiances becomes more difficult.
2.2 Calculating Solar Altitude Angle and Location

It is clear from the discussion so far that, to interpret the output from a pyranometer, we need to know the sun’s altitude angle. We could measure it directly—the old fashioned way—measuring the length of the shadow of a vertical rod (a gnomon) of known height—but, once you know the trick, it is simpler to calculate it than to measure it. The ability to locate the sun in the sky is also critical for theoretical prediction (as opposed to empirical measurement) of irradiance, for specifying the projected area of organisms, and for generalized implementation of heat-budget models.

So, how does one predict the sun’s location? We tackle this problem in two steps. First, we turn to celestial mechanics, which will allow us to track the sun’s motion from the viewpoint of the center of the Earth. We then translate this geocentric information into a more familiar coordinate system, in which we express the sun’s motion relative to axes oriented north-south, east-west, and up-down. Knowing the sun’s position in this local coordinate frame, we can calculate its altitude angle and other factors of interest.

2.2.1 The Angle of the Sun Relative to the Center of the Earth

Earth rotates from west to east about its axis. As a result, to an observer on Earth’s surface, the sun appears to move from east to west, reaching its highest point in the sky (its culmination) near noon each day. If that were all there were to the sun’s motion, it would be a simple task to predict the angle of incoming sunlight as a function of time. However, because Earth orbits the
sun, and because the planet’s axis is tilted relative to the plane of its orbit, predicting the sun’s movement across the sky is less than straightforward. For example, at any given time of day, the sun’s location is never the same two days in a row. Indeed, over the course of a year, the noontime altitude angle varies by approximately 47°, and the time between sunrise and sunset changes drastically. Both fluctuations have substantial affect on body temperature. In an extreme example—during winter near the poles—the sun never rises and organisms freeze; in summer it never sets, and organisms may overheat. And, as we will see, culmination, which one would expect to occur exactly at 12:00 noon, can actually happen substantially before or after mid day.

The daily and annual variations in the sun’s apparent path across the sky are most simply explained when viewed from the perspective shown in Figure 11. Imagine that Earth is a transparent sphere (let’s call it the terrestrial sphere) and you are at its center. At any time, you could draw a line from your eye to the center of the sun, and this line would intersect the terrestrial sphere at a particular point. Our task for the moment is to describe how the point of intersection varies through time.

Figure 11: A coordinate systems for locating the direction to the sun.

First, we define a coordinate system with which to describe the location of points on the terrestrial sphere (Figure 11). Most convenient is a polar system analogous to that of latitude and longitude. Earth’s poles define the axis of the system, and the plane both perpendicular to this axis and half way between poles is the terrestrial equator. As with lines of constant longitude, lines that extend from north to south pole on the terrestrial sphere are called meridians.
Given this coordinate system, points on the sphere are specified by two angles. The first of these is the hour angle, $\omega$, the angle in the equatorial plane between the meridian of a location we are interested in (the local meridian, the longitude as you would read it from a map) and the meridian that contains the intersection of a ray of sunlight with the terrestrial sphere (the sun’s meridian). This angle varies with time of day, making one full cycle in exactly 24 hours. The second angle is the declination angle, $\delta$. Analogous to latitude, declination is the angle between a ray of sunlight and the terrestrial equator. For reasons we will explore in a moment, solar declination varies with the seasons, making one full cycle in a year (Figure 12).

Before continuing with our investigation of the sun’s location, let’s take a step back and briefly review how we keep track of time. As both an historical and practical matter, our concept of time is tied to how society has decided to define the day. There are two standard ways in which one can approach this measurement. One could take as one’s basic unit of time the interval from when a certain star is highest overhead until that same star is again highest in the sky; this is the sidereal day. However, society has chosen an alternative, the solar day, the time between successive culminations of the sun. To be precise, it is not the length of any particular solar day that we use—as I hinted above, that varies through the course of a year—but rather the mean length of the solar day averaged over an entire year$^4$. Thus, we measure time in constant units of mean solar days. One hour is defined as 1/24th of the mean solar day; one minute, 1/60th of an hour; one second, one 60th of a minute.

$^4$To be even more precise, it is the mean length of the solar day in the year 1900.
With time and a coordinate system now in hand, let’s explore the hour angle, and its relationship to time of day. The hour angle is 0 when the sun culminates at our location of interest, that is, when the sun is "on the meridian." As we will see, this is close to—but seldom exactly equal to—noon as measured by local clocks. The sun’s meridian moves westward as Earth rotates to the east such that $\omega$ increases by $2\pi/24$ radians (15°) every hour. Expressed mathematically,

$$\omega = \frac{2\pi \text{ hours since local culmination}}{24} = \frac{2\pi \text{ clock time} - 12 \text{ h}}{24}. \quad (15)$$

Now for declination. Declination angle is a result of the tilt of Earth’s axis relative to the plane of the ecliptic, the plane of Earth’s orbit around the sun. If Earth’s spin axis were perpendicular to the plane of the ecliptic, the ecliptic and the terrestrial equator would coincide. If that were the case, a line from the sun to the center of the Earth would always intersect the equator, and $\delta$ would always be 0. However, Earth’s axis is currently tilted 0.4093 radians (23.44°) from being perpendicular to the plane of the ecliptic (Figure 13). As a result, when the north pole is tilted directly toward the sun, a line from the sun to Earth’s center intersects the terrestrial sphere north of the equator, at a declination angle of $+0.4093$. This happens on the northern hemisphere’s summer solstice, currently June 20 or 21. When the north pole is tilted directly away from the sun—the winter solstice—a line from sun to Earth intersects the sphere at $-0.4093(-23.44°)$. Winter solstice occurs on December 21 or 22.

Declination angle varies sinusoidally between these extremes (Figure 12); it is 0 at the vernal and autumnal equinoxes, March 20 and September 23, respectively. We can express the variation of $\delta$ mathematically:

$$\delta \approx \arcsin\left\{0.4093\cos\left[0.0172(N - 173)\right]\right\}. \quad (16)$$

Here $N$ is time within a year, expressed in units of days. For instance, $N$ varies from 0 to 1 on January 1. $N$ goes from 364 to 365 on December 31, except for leap years, when it varies between 365 and 366. $\delta$ is maximal when $N = 173$ (June 21 or 22) and minimal when $N = 355$ or 356 (December 21 or 22).

In short, if we know how many hours it has been since culmination at our location of interest and what day it is in the year ($N$), we can calculate both hour angle and solar declination, and thereby pinpoint the location of the sun from the perspective of Earth’s center.

So far, so good. Complications appear, however, when we attempt to put this scheme into practice using clocks. Clocks as we know them measure out time at a constant rate based on the mean solar day. The problem is that the sun progresses across the sky at a rate that varies through the year. Reconciling

---

5The variation in date is due to the inclusion of an extra day in leap years. In short, because it takes earth 365.242 days to orbit the sun, a year of 365 days accrues an "error" of approximately 1/4 day every year. This shift can push the date of the solstice from June 21 to June 22. The error is corrected (to a first approximation) every four years by the insertion of February 29 into the calendar.

6$N$ is sometimes mistakenly referred to as the "Julian day." Julian days are numbered from January 1, 4713 BC, not from the beginning of each year.
Figure 13: The angle of Earth’s tilt relative to the sun varies with the seasons.

the mismatch between constant clock time and variable solar time requires an extensive explanation; please bear with me.

The mismatch between clock and solar time could be avoided if, every day, each locality set its clocks to noon when the sun was on the meridian. Indeed, this was the convention until about 1880. However, with the advent of trains—and the rapid long-distance travel they engendered—this sort of continuously variable local time became problematic. We are all familiar with the irksome necessity of resetting one’s watch after a long east-west plane flight; imagine having to reset your watch every time your train stopped at a new town, and you will get the idea. To avoid this problem, the globe is divided into 24 nominal time zones, each 15° of longitude wide. Time—to be specific, local standard clock time, \( t_{\text{lsct}} \)—is held constant everywhere within a zone, set so that all clocks in a given time zone read noon when the sun is on the meridian at the zone’s central longitude. Nominal time zones are based on the standard system of longitude. For example, England’s zone extends from 7.5° east to 7.5° west longitude, 15° centered on Greenwich and the prime meridian. The time zone in which California lies, centered on 120° W, extends from 112.5° W to 127.5° W.

Note the adjective "nominal" in this description of time zones. The boundaries between actual time zones have been drawn to avoid major cities and to suit political whim; as a result, they can deviate substantially from the simple system described above\(^7\). It is always best to consult a map to figure out how a particular location fits into the standard scheme of nominal time zones. Note also that when we talk about local clock time, we mean local standard time. If the location in question is currently operating on daylight savings time, you

\(^7\)For example, Newfoundland sets its clocks half an hour early relative to Canada’s other maritime provinces.
first need to subtract an hour to adjust to standard time:

$$t_{\text{lsct}} = t_{\text{lc}} - D.$$  

Here $t_{\text{lc}}$ is local clock time measured in hours. $D = 1$ if daylight saving time is in effect; $D = 0$ otherwise. For example, if your watch reads 3:12 pm (3.2 hours after noon, 15.2 hours after midnight) and it is daylight saving time, local standard clock time, $t_{\text{lsct}}$, is 14.2 hours.

There are two details left to which we need to attend. Local clock time as we have described it is set according to the position of the sun at the central longitude of a time zone on a particular day and proceeds at a constant rate thereafter. This allows us to specify the position of the sun at clock noon on that particular day and at that particular location. If, however, we need to specify the sun’s position at another time in the year or at another location within the time zone, we need to adjust for any shift in longitude away from the central longitude, and we need to account for the fact that the length of a solar day varies through the year.

First, the matter of longitude. For localities east of a time zone’s central longitude, the sun culminates earlier than it does at the central longitude. For localities west of the central longitude, the sun culminates later. To adjust for these deviations, we require a longitude correction, $t_{\text{long}}$. For a given local longitude $L_{\text{local}}$

$$t_{\text{long}} = \frac{L_{\text{tz}} - L_{\text{local}}}{15}. \quad (18)$$

Here $L_{\text{tz}}$ is the central longitude of the time zone and $t_{\text{long}}$ is measured in hours. Thus, actual time as measured by the sun, $t_{\text{solar}}$, is

$$t_{\text{solar}} \approx t_{\text{lsct}} + t_{\text{long}}, \quad (19)$$

$$\approx t_{\text{lc}} - D + t_{\text{long}}. \quad (20)$$

Within a nominal time zone, local solar time can be anywhere from half an hour ahead to half an hour behind local standard time, depending on location.

And finally, we need to correct for annual variation in the length of the solar day. Recall that the solar day is defined as the average time from when the sun is on the meridian until it is again on the meridian, by definition exactly 24 hours. However, in reality, the length of a solar day varies by more than 30 minutes over the course of a year, a fluctuation due both to the manner in which Earth orbits the sun and to the tilt of Earth’s axis.

Earth’s orbit around the sun is elliptical. As a consequence of this eccentricity, Earth moves fastest when it is nearest the sun (perihelion) and slowest when the planet is farthest from the sun (aphelion). In the current orientation of Earth’s orbit, the planet is at perihelion on January 3 and aphelion on July 3. As Earth speeds up from aphelion to perihelion, it must rotate farther from west to east to bring the sun to culmination. In effect, due to Earth’s motion around the sun, the noontime sun appears to shift to the east relative to the stars (Figure 14), and it is this apparent eastward shift that requires the extra
rotation. For example, if, just before dawn, you could somehow shove the sun eastward (that is, back below the horizon), it would subsequently rise later than expected. This effect is greatest half way between aphelion and perhelion, on April 3 (year day 93), when culmination due to this effect is delayed relative to clock time by 7.66 minutes (0.1277 hours). Conversely, as the planet slows down from perhelion to aphelion, it must rotate less than its average amount to bring the sun to culmination, and the time of culmination happens before clock noon, an effect maximal on October 5 (year day 278) when culmination due to orbital eccentricity precedes noon by 7.66 minutes. At aphelion and perhelion, Earth’s orbital speed is neither increasing nor decreasing, so, as far as the effects of orbit eccentricity are concerned, culmination occurs at its average time, clock noon. The shift in time of culmination due to eccentricity is

\[ t_{ecc} = 0.1237 \cos \left( \frac{2\pi}{365.242} (N + 89.289) \right). \]

where \( N \) is again time of year, measured in days from midnight on January 1. This relationship is graphed in Figure 15.

Time of culmination also varies due to the tilt of Earth’s axis, its obliquity. To see how this works, let’s return to how the sun’s declination angle changes through the year (Figure 16). Imagine that this sinusoidal fluctuation is a road in the heavens along which the sun appears to move against a background of stationary stars. In that sense, motion of the sun along this graph is analogous to the apparent motion we would record if, through the course of a year, we plotted the noontime position of the sun relative to the stars. The general motion is west to east, as we have noted, but the location of the sun also moves north and south due to the change in declination. Any change in the direction of the road from directly west-to-east reduces the sun’s eastward speed. (Imagine yourself traveling generally eastward on a road that meanders north and south; your eastward progress is slower than it would be if the road ran straight east.) This effect is quantified in Figure 17. The sun’s eastward speed is maximal near where the slope of the declination graph is zero (the solstices, where the
"road" runs straight east), and minimal near where the absolute value of the
slope is greatest (the equinoxes, where the meander is maximal). For example,
the sun's apparent position moves most rapidly eastward in May and November,
and slowest in February and August. (Maximum and minimum speeds don't
occur exactly at the solstices and equinoxes because the speed of the sun along
the road varies through the year.) Recall that the rate at which the sun moves
eastward relative to the stars determines how far Earth must rotate each day to
catch up. As a result, the length of the solar day varies as the eastward speed
of the sun varies, making two cycles of fluctuation per year (Figure 17).

\[
t_{\text{obl}} = 0.1654 \cos \left[ \frac{4 \pi}{182.621} (N - 126.024) \right],
\]

a sinusoidal function with an amplitude of approximately 9.92 minutes (0.1654
hours) and a period of half a year.

The cumulative effect of the slower than average movement of the sun from
late March to mid April is just offset by the faster than average movement from
mid April through June, so that (as regards this obliquity effect) the length of
the solar day on June 21 (the summer solstice) is equal to the average. Similarly,
the faster than average movement of the sun from June 21 to August 7 is just
offset by the slower than average movement from August 7 to September 21,
and the length of the solar day on the autumnal equinox is again equal to the
average. Similar effects apply at the winter solstice and the vernal equinox.

The overall deviation of the time of solar culmination is the sum of the effects
of orbital speed and axial tilt, an expression aptly named the Equation of Time\textsuperscript{8}

\textsuperscript{8}Here, "equation" is used in the sense of "to make equal" or "to adjust." The Equation of
EOT:

\[
EOT = t_{ecc} + t_{obl};
\]

\[
= 0.1237 \cos \left[ \frac{2\pi}{365.242} (N + 89.289) \right] + 0.1654 \cos \left[ \frac{2\pi}{182.621} (N - 126.024) \right].
\]

This equation is shown in Figure 18. The actual time of solar noon can be as much as 16.40 minutes before clock noon or 14.29 minutes after.

In sum, true local solar time \( t_{solar} \) is determined by local clock time, the longitude correction, and the Equation of Time:

\[
t_{solar} = t_{lct} + D + t_{long} + EOT.
\]

From true local solar time, we can, at long last, calculate the true hour angle:

\[
\omega = 2\pi \frac{t_{lct} - D + t_{long} + EOT - 12}{24}.
\]

In an odd historical note, all the complexity we have just tackled could be avoided if society decided to use sundials rather than clocks to tell us when it is solar noon. As we will see in a moment, in the northern hemisphere the sun is highest in the sky when directly south of any location. Thus, when a shadow cast by a vertical rod extends due north, it is exactly solar noon. (In the southern hemisphere, simply replace "north" with "south" in the last sentence.) This exceedingly simple system—requiring only a compass and a

Time adjusts clock time to make it equal to solar time.
Figure 17: The effect of obliquity varies through the year.

sundial—inherently adjusts for longitude, orbital, and axial effects. It is only because we humans have tied ourselves to the steady progression of time as measured by clocks that we must deal with these complications. To paraphrase Lennox-Boyd (2006) clocks merely tell time, sun dials define it.

2.2.2 Converting From Polar to Map Coordinates

Now that we have a method for locating the sun’s position in the sky from the perspective of Earth’s center, our next task is to transfer that information into a more familiar coordinate system. Unlike astronomers, for whom the polar system of declination and hour angle is standard, biologists (indeed, people in general) traditionally use the Cartesian coordinate system introduced in Chapter 2, with mutually perpendicular axes oriented north-south, east-west, and up-down (Figure 19), what we will call the local map coordinate system. Because the up-down axis at any location points at Earth’s center, its orientation relative to Earth’s equatorial plane is different for different latitudes, necessitating some adjustment as we translate from global to local map coordinates.

We begin by returning to the geocentric perspective of the terrestrial sphere in which the position of the sun is indicated by the arrow, $S$ (Figure 20). In this example, it is afternoon for an observer at location $A$. For simplicity, let’s


10This is an approximation. The up-down axis is defined at any location as the direction indicated by a plumb line. The mass on the end of the line is attracted by gravity toward Earth’s center, but it is also affected by the centrifugal force of Earth’s rotation. As a result, only along the equator and at the poles does a plumb line point directly toward Earth’s center. Elsewhere, the line has a slight tilt toward the equator. See Denny (2008).
Figure 18: The Equation of Time is the sum of the eccentricity and obliquity factors.

measure the dimensions of the terrestrial sphere in units of its radius, in which case $S$ has length 1.

In our Earth-centered, polar coordinate system, the position of the sun in this example would be specified by the hour angle $\omega$ and the declination angle $\delta$ of $S$. Alternatively, however, we could specify the orientation of $S$ relative to a Cartesian coordinate system with its origin at Earth’s center. In this system, one axis—call it the observer axis, $O$—lies along the intersection of the equatorial plane and the plane of A’s meridian. A second axis—the east-west axis, $E$—is perpendicular to the $O$-axis, also in the equatorial plane, and the third axis—Earth’s polar axis, $P$—is perpendicular to both the $O$- and $E$-axes. The projection of arrow $S$ onto these three axes provides an alternative means to quantify the sun’s location. To be specific, the projection of $S$ on the polar axis is

$$S_P = S \sin \delta,$$

where $S_P$ is the polar direction cosine. Recalling that $S$ has length 1,

$$S_P = \sin \delta.$$  \hspace{1cm} (26)

The projection of $S$ onto the equatorial plane is

$$S_{eq} = S \cos \delta$$  \hspace{1cm} (28)

$$= \cos \delta.$$  \hspace{1cm} (29)

and the projection of $S_{eq}$ onto the $O$ axis is, in turn, the direction cosine for the
\[ S_O = S_{eq} \cos \omega, \quad (30) \]
\[ = \cos \delta \cos \omega. \quad (31) \]

Similarly, the projection of \( S_{eq} \) on the \( E \) axis is the direction cosine for the \( E \)-axis,

\[ S_E = S_{eq} \cos \omega, \quad (32) \]
\[ = \cos \delta \sin \omega. \quad (33) \]

These three direction cosines define the orientation of \( S \), and thus the location of the sun in the sky. If an observer at Earth’s center knows \( S_O, S_E, \) and \( S_P \), he or she can point arrow \( S \) at the sun.

Our next task is to transfer our newly acquired, Earth-centered Cartesian coordinates to local map coordinates at location A. This transfer involves two actions. First, we need to move the origin of the coordinate system from the center of the Earth to \( A \). Second we need to rotate the coordinate system to adjust for \( A \)’s latitude.

We accomplish the first step by ignoring it. Relative to the distance from Earth to sun (\( 1.5 \times 10^8 \) km), the distance from Earth’s center to its surface (\( 6.3 \times 10^3 \) km) is infinitesimal, so shifting the origin of our coordinate system has negligible effect on the apparent position of the sun. In other words, we simply slide the origin of the Earth-centered system to \( A \), making no adjustment to the direction of \( S \). Having done so, to transfer our Earth-centered coordinate system to the local system of location \( A \), all we need do is account for latitude.

To see how this is done, we position ourselves relative to the geocentric Cartesian system (albeit with its center now at \( A \)) so that we are looking directly west along the \( E \) axis (Figure 21). In this perspective, we are perpendicular to the plane defined by the \( O \)- and \( P \)-axes, and these axes are aligned as shown. From this point of view, how does the local map coordinate system relate to the \( O \)- and \( P \)-axes? The up-down axis (the \( UD \)-axis) is easily specified. Because it points directly away from the center of the Earth, the \( UD \)-axis extends...
perpendicular to Earth’s surface at A. The angle between the \( UD \)- and \( O \)-axes is therefore determined by the latitude of A, that is, by the angle \( \phi \). Next, we know that the north-south axis lies in the plane of Earth’s surface at A, and at right angles to the \( UD \)-axis, as shown. And lastly, the east-west axis of the local coordinate system is the same as the \( E \)-axis of the geocentric system.

From Figure 21 it is clear that all we need do to transfer our geocentric coordinates to those of the local map system is to rotate the geocentric system to align with the local. In essence, this involves calculating the projections of \( S_E, S_P, \) and \( S_O \) onto the \( UD \)-, \( NS \)-, and \( EW \)-axes. The projection for the map \( EW \)- and geocentric \( E \)-axes is trivial because these axes are the same for both systems. Thus, the direction cosine along the \( EW \) axis is

\[
S_{EW} = S_E. \tag{34}
\]

The direction cosine of \( S \) along the \( NS \)-axis receives contributions from both \( S_P \) and \( S_O \) as shown in Figure 21. Using our identities from before (Eqs. 26, 31, and 32):

\[
S_{NS} = S_P \cos \phi - S_O \sin \phi, \tag{35}
\]

\[
= \sin \delta \cos \phi - \cos \delta \cos \omega \sin \phi. \tag{36}
\]

Similarly,

\[
S_{UD} = S_O \cos \phi + S_P \sin \phi, \tag{37}
\]

\[
= \cos \delta \cos \omega \cos \phi + \sin \delta \sin \phi. \tag{38}
\]
We have thus reached our goal. If we know the sun’s declination $\delta$, the latitude $\phi$ of location $A$, and the hour angle $\omega$ of the sun relative to $A$’s meridian, we can calculate $S_{EW}$, $S_{NS}$, and $S_{UD}$, the direction cosines of the sun in the local coordinate system. From these direction cosines, one can locate the sun.

We can immediately put our direction cosines to work. From the definition of solar altitude angle we know that $S_{UD} = S \sin \theta$, so the solar altitude angle $\theta$ is

$$\theta = \arcsin \frac{S_{UD}}{S}.$$  \hspace{1cm} (39)

But $S = 1$, so

$$\theta = \arcsin S_{UD},$$  \hspace{1cm} (40)

$$= \arcsin(\cos \delta \cos \omega \cos \phi + \sin \delta \sin \phi).$$  \hspace{1cm} (41)

Thus, from knowledge of solar declination, hour angle, and latitude, we can calculate $\theta$ for any time and location. With $\theta$ thus in hand, Eqs. 7 and 8 can be used to calculate solar irradiance at sea level. Cool.

2.2.3 Converting From Map to Organismal Coordinates

The ability to locate the position of the sun relative to local map coordinates is useful for calculating solar altitude and irradiance, but there is one final step we need to take to generalize our knowledge. As satisfying as it may be to know where the sun is in the sky, we still need to relate that position to the orientation of a specific plant or animal. For example, we noted in Chapter 11
that some littorine snails adjust the orientation of their shells so that the apex points at the sun during the hottest time of day, thereby minimizing projected area. To predict the optimal alignment of a snail for a particular day, we need to know how the shell’s orientation matches up with the path of the sun.

The procedure for transferring solar to organismal coordinates is similar to that we just used to transfer geocentric to map coordinates. We first define a Cartesian coordinate system for the organism, and then rotate that system to align with the local east-west, north-south, up-down system. An example is shown in Figure 22 using a limpet as an example. The orientation of the limpet is defined by the orientation of its anterior-posterior (AP), left-right, (LR) and dorsal-ventral (DV) axes, which, like the axes of the map coordinate system, are mutually perpendicular. We desire to know how these axes are aligned with respect to the sun’s location in our the east-west (EW), north-south (NS), and up-down (UD) map system. Fully describing the animal’s orientation requires measurement of three angles:

- $\psi$, the angle (measured in the horizontal north-south, east-west plane) between the AP axis of the animal and the NS-axis;
- $\xi$, the angle (measured in the vertical up-down, east-west plane) between the LR axis of the animal and the EW-axis;
- and $\zeta$, the angle (measured in the vertical up-down, north-south plane) between the DV axis of the animal and the UD-axis.

For those familiar with aeronautical terminology, the angles $\psi$, $\xi$, and $\zeta$ correspond to yaw, roll, and pitch, respectively. Note that in each case, the angle is positive if rotation is in a counter-clockwise direction when viewed down the axis of rotation. For example, looking down (that is, toward the down end

Figure 22: The coordinate system of an organism, here a limpet.
of the \( UD \)-axis) on a horizontal limpet with its head facing north, if the animal reorients such that its head now faces northwest, the rotation involved is counter-clockwise and the angle \( \psi \) between the animal’s \( AP \)-axis and the \( NS \)-axis is positive (Figure 23). If, instead, the animal moves to a position in which its head faces northeast, the rotation involved is in the negative sense. For the right-left axis, we look along the axis from right to left. For the anterior-posterior axis, we look from anterior to posterior.

![Figure 23: The relationship between local and geocentric coordinate systems.](image)

Shifting our understanding of the sun’s location from one coordinate system to another involves rotation about any or all of a system’s axes. In the relatively simple example we worked through in Figure 21, a shift between geocentric and local coordinate systems required rotation only about the east-west axis to account for latitude. Other shifts are less constrained. An animal can orient itself in any direction relative to the local map coordinates by rotating about all three of its body-centered axes. Fortunately, the mathematical machinations involved in the rotation of axes are straightforward, if somewhat laborious. The calculations are described most succinctly in terms of rotation matrices.

Suppose we know the direction cosines \( S_{NS} \), \( S_{EW} \), and \( S_{UD} \) of the sun’s position in the local coordinate system, and we want to know the equivalent direction cosines \( S_{AP} \), \( S_{LR} \), \( S_{DV} \) for a coordinate system centered on a limpet. Let’s first explore a simple example in which the limpet sits on a horizontal substratum. As long as the limpet is horizontal, its dorsal-ventral axis is parallel to the up-down axis, but the animal can rotate its anterior-posterior axis to point in any direction it pleases. Note that any shift in the \( AP \) axis automatically entails a shift in the \( LR \) axis because the two must remain perpendicular. The effects on direction cosines of rotating the shell by an angle \( \psi \) around the \( DV \)-
axis is succinctly described by the matrix equation:

\[
\begin{bmatrix}
S_{AP} & S_{LR} & S_{DV}
\end{bmatrix} =
\begin{bmatrix}
\cos \psi & -\sin \psi & 0 \\
\sin \psi & \cos \psi & 0 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
S_{NS} \\
S_{EW} \\
S_{UD}
\end{bmatrix}.
\] (42)

The matrix multiplication specified by this equation is interpreted as follows, where the values in a row of the 3 \times 3 rotation matrix are matched up with the values in the right-most column matrix:

\[
S_{AP} = (S_{NS} \cdot \cos \psi) + (S_{EW} \cdot -\sin \psi) + (S_{UD} \cdot 0),
\] (43)

\[
S_{LR} = (S_{NS} \cdot \sin \psi) + (S_{EW} \cdot \cos \psi) + (S_{UD} \cdot 0),
\] (44)

\[
S_{DV} = (S_{NS} \cdot 0) + (S_{EW} \cdot 0) + (S_{UD} \cdot 1).
\] (45)

Similar rotation matrices exist for rotation about the animal’s other two axes. For rotation by angle \(\xi\) about the AP-axis:

\[
\begin{bmatrix}
S_{AP} & S_{LR} & S_{DV}
\end{bmatrix} =
\begin{bmatrix}
1 & 0 & 0 \\
0 & \cos \xi & -\sin \xi \\
0 & \sin \xi & \cos \xi
\end{bmatrix}
\begin{bmatrix}
S_{NS} \\
S_{EW} \\
S_{UD}
\end{bmatrix}.
\] (46)

And for rotation by angle \(\zeta\) about the LR-axis:

\[
\begin{bmatrix}
S_{AP} & S_{LR} & S_{DV}
\end{bmatrix} =
\begin{bmatrix}
\cos \zeta & 0 & \sin \zeta \\
0 & 1 & 0 \\
-\sin \zeta & 0 & \cos \zeta
\end{bmatrix}
\begin{bmatrix}
S_{N} \\
S_{E} \\
S_{P}
\end{bmatrix}.
\] (47)

If the animal shifts its orientation about more than one axis, the individual rotations are calculated sequentially. For example, a shift around all three axes by angles \(\psi\), \(\xi\), and \(\zeta\) is calculated by carrying out three matrix multiplications, one after the other:

\[
\begin{bmatrix}
S_{ap}'' & S_{lr}'' & S_{dv}''
\end{bmatrix} =
\begin{bmatrix}
\cos \psi & -\sin \psi & 0 \\
\sin \psi & \cos \psi & 0 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
S_{NS} \\
S_{EW} \\
S_{UD}
\end{bmatrix},
\] (48)

\[
\begin{bmatrix}
S_{ap}' & S_{lr}' & S_{dv}'
\end{bmatrix} =
\begin{bmatrix}
1 & 0 & 0 \\
0 & \cos \xi & -\sin \xi \\
0 & \sin \xi & \cos \xi
\end{bmatrix}
\begin{bmatrix}
S_{ap}'' \\
S_{lr}'' \\
S_{dv}''
\end{bmatrix},
\] (49)

\[
\begin{bmatrix}
S_{ap} & S_{lr} & S_{dv}
\end{bmatrix} =
\begin{bmatrix}
\cos \zeta & 0 & \sin \zeta \\
0 & 1 & 0 \\
-\sin \zeta & 0 & \cos \zeta
\end{bmatrix}
\begin{bmatrix}
S_{ap}' \\
S_{lr}' \\
S_{dv}'
\end{bmatrix}.
\] (50)

It is this sequence of operations that leads to the final results. These calculations look more forbidding than they actually are. Given the day of the year, time of day, and latitude of a location, we can predict \(S_{NS}\), \(S_{EW}\), and \(S_{UD}\), and these values can be inserted into Equations 48 through 50.
Knowing the orientation of the organism in our local coordinate system, we can measure the angles $\xi$, $\zeta$, and $\psi$, and we can then proceed to crank through the calculations.

A brief review may be in order at this point. The sun’s position in the sky determines the intensity of direct sunlight that reaches sea level and the direction in which that light impinges on a particular organism. We have at long last constructed a mechanism through which we can predict these important values as a function of clock time.

2.3 Measuring Projected Area

Now that we can calculate $I_{sw}$, our next task is to quantify the heat influx that ensues given the irradiance we have just calculated. Heat absorbed from direct short-wave radiation depends on projected area $A_p$, the area available to intercept sunlight. For a few simple shapes (e.g., spheres, ellipsoids, and cones) it is possible to calculate projected area. For other objects, projected area must be measured. In practice, projected area is usually determined by taking a picture of the object in question from the perspective of the sun, that is, with the camera’s shadow falling on the object. Area measured on the resulting two-dimensional photograph (adjusted for magnification) is the object’s projected area. Because the sun’s location in the sky varies through the day, $A_p$ is also likely to vary. Only spheres have the same projected area independent of the angle of illumination, so unless one is dealing with a spherical organism, one needs to take the time-dependent variation in $A_p$ into account. This can be accomplished by taking pictures of the organism from a series of angles that track the location of the sun as it moves relative to the plant or animal, motion that we now know how to predict.

2.4 Measuring Absorptivity

The actual amount of heat absorbed from sunlight depends not only on the light’s intensity and the projected area of an organism, but also on the absorptivity of the organism’s surface. In Chapter 11, we expressed these factors—intensity and absorptivity—as single, overall values, but it is often instructive to parse each factor into its wavelength-by-wavelength components. Consider, for instance, the hypothetical example shown in Figure 24, an idealized absorption spectrum for a green alga (solid line). Due to the presence of chlorophyll a in this seaweed, light in the blue and red portions of the spectrum is effectively absorbed, and light in the green range of the spectrum is reflected. It is this reflected light that gives the seaweed its green color. A contrasting hypothetical absorption spectrum is shown by the dashed line, in this case for a purplish snail. Here, light in the green and yellow wavelengths is absorbed, and light in the red and blue is reflected.

Because the area under these two spectra is the same, one might easily suppose that the overall absorptivities of the two organisms would be similar. That supposition is based, however, on the implicit assumption that the same
amount of energy is available to be absorbed at all wavelengths. In reality, that is not the case. The spectrum of solar irradiance has its peak in the green and yellow wavelengths, with much less energy in the blue and red. Thus, even though the green seaweed is effective at absorbing red and blue light, there is relatively little energy to be absorbed at these wavelengths and overall it absorbs less solar energy than does the snail.

In practice, to calculate the overall absorption coefficient $\alpha_{sw}$ for an organism, one must weight the wavelength-by-wavelength data from an absorption spectrum by the corresponding wavelength-by-wavelength fractions of overall solar irradiance:

$$\alpha_{sw} = \sum_{i=1}^{n} \alpha_{sw}(\lambda_i) \frac{I_{sw}(\lambda_i)}{I_{sw, total}}$$

Here $\alpha_{sw}(\lambda_i)$ is the short-wave absorptivity at wavelength $\lambda_i$, $I_{sw}(\lambda_i)$ is the irradiance at that wavelength, and $I_{sw, total}$ is total solar irradiance.

Note that because the efficiency of atmospheric scattering of sunlight varies as a function of wavelength, the spectrum of $I_{sw, dir}$ differs from that of $I_{sw, dif}$. ($I_{sw, dif}$ is augmented in the short wavelengths.) As a result, the absorptivity for diffuse light can differ from that for direct light, but the effect is typically small.

2.5 Solar Heat Influx: A Summary

We have covered a lot of ground in the last few pages, and a brief summary may help to fix ideas. Using a pyranometer, it is possible to directly measure the
intensity of light impinging on a horizontal surface. The utility of this information is constrained, however. To interpret pyranometer measurements in terms of heat influx to an organism, one must know the location of the sun in the sky, both to calculate $I_{sw}$ and its components ($I_{sw,dir}$, and $I_{sw,dif}$) and to measure $A_p$. To facilitate these calculations, we have developed a method for translating clock time into solar location. Thus, given a time series of pyranometer measurements, we can at last calculate $H_{sw}$, the shortwave flux of heat into an organism.

In addition to its utility in interpreting empirical pyranometer data, our ability to predict the sun’s location provides a means to calculate $H_{sw}$ at times and places for which empirical data are not available.

3 Supplement 11.3: Measuring the Heat-Transfer Coefficient

*This supplement expands on a topic addressed on page 209 of the textbook.*

Convective transfer of heat is modeled using Newton’s law of cooling:

$$H_{cv} = h_c A_{conv} (T_f - T_b).$$  

This relationship tells us that the instantaneous rate of energy lost from a heated object depends on the difference between the object’s temperature and that of the fluid. Thus, if we heat an object and measure how fast it cools, we can deduce the value of $h_c$. A little bit of math highlights what we need to measure and what to expect if we actually perform this experiment.

Much real-world complexity has been shoehorned into the heat transfer coefficient, $h_c$, which is sensitive to the shape of an organism and the ratio of inertial to viscous forces in the fluid, which in turn depends on the fluid’s speed. The interaction among these factors is such that it is difficult to predict $h_c$ accurately, which means that we have to measure it.

To do so, we jump to our calculation of thermal response time in Supplement 12.1. In that analysis, we will deal with a simple heat-budget model, $H_{cv} = H_{net}$. In the process of solving for response time, we arrive at the conclusion that

$$\Delta T = \Delta T_0 \exp \left( \frac{-h_c A_{conv}}{mC_p} t \right).$$  

where $\Delta T = (T_b - T_{air})$ and $\Delta T_0$ is $\Delta T$ at the start of the experiment, $t = 0$. We can then take the log of each side of this equation:

$$\ln \Delta T = \left( \frac{-h_c A_{conv}}{mC_p} \right) t + \ln \Delta T_0.$$  

This expression is the equation for a line. If we plot $\ln \Delta T$ as a function of time, the line has $\ln \Delta T_0$ as its $y$ intercept and $\frac{-h_c A_{conv}}{mC_p}$ as its slope (Figure 25).
In other words,

\[ \text{slope, } \vartheta = \frac{-h_c A_{\text{conv}}}{m C_p}, \quad (55) \]

\[ h_c = \frac{-m C_p \vartheta}{A_{\text{conv}}}. \quad (56) \]

This, then, is a recipe for measuring \( h_c \). In essence, we heat a model of known \( m, C_p, \) and \( A_{\text{conv}} \), record how fast it cools, and plot the data appropriately.

Of course, there are a few details to which we must attend. For instance, it is seldom practical to use the plant or animal itself to make these measurements. Due to the relatively low thermal conductivity of animals’ bodies (approximately that of water, \( 0.6 \, \text{W·m}^{-1} \cdot \text{K}^{-1} \)), an organism in the experiment described above would have a substantial gradient from its warm interior to a relatively cool exterior. Because it is the temperature at the surface that determines convective transfer, and this temperature is affected by the rate at which heat moves internally by conduction, the low thermal conductivity of water would confound the measurement of \( h_c \). This problem can be circumvented by using a model of the organism made from a material that has a thermal conductivity sufficiently large that any internal temperature gradient is negligible. The materials that best fit this bill are silver and copper, with thermal conductivities of 429 and 401 \( \text{W·m}^{-1} \cdot \text{K}^{-1} \), respectively. Copper is cheaper than silver, but not as readily molded into organism-shaped models. By contrast, because silver is used in jewelry, techniques for molding it are well established and readily available at local shops or colleges. The lost wax procedure is particularly handy. In short, one makes a rubber mold of the organism in question. The mold is then used to cast a wax replica of the plant or animal, which, in turn, is encased in plaster. Channels (called sprews) connect the wax to the outside of the plaster casing.
so that, when the plaster is heated, the wax melts and runs out of the casing, leaving an organism shaped cavity. (This is the lost wax part of the procedure.) Molten silver is poured into this cavity and allowed to cool. When the plaster is cracked open, the end result is an accurate silver model of the organism. A small hole drilled into the center of the model can accommodate a thermocouple, allowing for continuous measurement of the model’s temperature. The specific heat of pure silver is 235 J kg\(^{-1}\) K\(^{-1}\), that of stirling silver is very nearly the same, 231 J kg\(^{-1}\) K\(^{-1}\).

The convective area of the model can be estimated by carefully molding and trimming aluminum foil to cover the appropriate area of the model and comparing the weight of the foil to that of a known area of foil. The model is then weighed, heated, and inserted into air flowing at a known speed, and its temperature monitored as described above. The resulting data are used to calculate \(h_c\) for that wind speed.

There is one remaining detail to take care of. In the derivation above, we have implicitly assumed that the model’s heat can be lost only to convective transfer. This would be a good approximation if the substratum on which the model sits were a perfect insulator, but that is not possible in practice. Foam insulation, the typical choice for the substratum, is a good insulator, but not a perfect one. The effect of heat loss to the substratum can be estimated by performing the experiment as previously described, but without wind and with the heated model covered with a piece of foam insulation to minimize convective heat transfer. The heat transfer coefficient for this control situation is then subtracted from the \(h_c\) measured in the presence of convection to estimate the true \(h_c\).
4 Supplement 11.4: A Non-Traditional Mass Transfer Coefficient

This supplement expands on a topic addressed in a footnote on page 215 of the textbook.

The mass transfer coefficient cited in the chapter differs from that of many texts. Traditionally, the rate of evaporative mass transport is written as

\[
\frac{dm}{dt} = h_{m,trad}A_{evap}(\rho_{v,b} - \rho_{v,air}),
\]

where \(\rho_{v,b}\) and \(\rho_{v,air}\) are the densities (kg m\(^{-3}\)) of water vapor at the organism’s surface and in the bulk air, respectively. For our purposes, it is convenient to deal with the difference in mole fraction, \(C_v\), rather than the difference in vapor density. This can be accomplished by noting that

\[
C_v = \frac{\rho_vRT}{m_{mol,air}},
\]

where \(m_{mol}\) is the molecular mass of water (0.018 kg mol\(^{-1}\)) and \(R\) is the universal gas constant (8.3143 J mol\(^{-1}\)K\(^{-1}\)). Thus,

\[
\frac{dm}{dt} = h_{m,trad}A_{evap}(C_{v,b} - C_{v,air}).
\]

We then define \(h_m\)

\[
h_m = h_{m,trad}\left(\frac{m_{mol,air}}{RT}\right).
\]

5 Supplement 11.5: Relating Mass Transfer to Heat Transfer

This supplement expands on a topic addressed on page 215 of the textbook.

Heat can flow into or out of an organism as water evaporates or condenses. The rate of this evaporative exchange is

\[
H_{evap} = \Lambda_{wat}A_{evap}h_m(C_{v,b} - C_{v,air})
\]

where \(\Lambda_{wat}\) is the latent heat of evaporation, \(A_{evap}\) is the area available for evaporation, and \(C_{v,b}\) and \(C_{v,air}\) are the mole fractions of water vapor at the body’s surface and in the surrounding air, respectively. As with convective heat transfer, much of the complexity associated with evaporative heat exchange has been squeezed into the mass-transfer coefficient, \(h_m\). Fortunately, having delved in some detail into the measurement of heat-transfer coefficients, we are in a position to estimate the mass-transfer coefficient. The jump from heat to mass is based on the notion that, at the small scale of the boundary layer adjacent to
an organism, both processes are governed by the rate of diffusion of molecules. In essence, we propose that the ratio of heat transfer to mass transfer should be proportional to the ratio of the rates at which heat and mass diffuse:

$$\frac{h_c}{h_m} \propto \frac{D_H}{D_m}$$

(62)

Here, $D_H$ is the diffusivity of heat in air and $D_m$ is the diffusivity of mass. A detailed consideration of the physics of diffusive transfer (see Incropera & DeWitt 2001)$^{11}$ suggests that

$$\frac{h_c}{h_m} \propto \left( \frac{D_H}{D_m} \right)^{2/3}$$

(63)

a proposition known as the Reynolds analogy, in honor of Osborne Reynolds (he of the Reynolds number). It would take considerable explaining to justify the exponent of 2/3 in this equation (see Bird et al. 1964)$^{12}$, and I will not attempt it here. As luck (and physics) would have it, for air, the diffusivities of heat and mass are quite similar ($21.5 \times 10^{-6}$ and $24.2 \times 10^{-6}$ m$^2$s$^{-1}$, respectively) so their ratio is close to 1, and the exact power to which that ratio is raised is therefore of little practical consequence ($(21.5/24.2)^{2/3} = 0.924$).

As diffusivities, both $D_H$ and $D_m$ have the same units (m$^2$s$^{-1}$), so the righthand side of this equation is dimensionless. In contrast, $h_c$ and $h_m$ have different units, so to make this proportionality an equality, we must incorporate a constant with the appropriate dimensions:

$$\frac{h_c}{h_m} = \frac{\rho_{\text{air}} C_{p,\text{air}} R T}{m_{\text{mol air}}} \left( \frac{D_H}{D_m} \right)^{2/3}.$$  

(64)

Here $\rho_{\text{air}}$ and $C_{p,\text{air}}$ are the mass density and specific heat of air, respectively, and $R$ and $T$ are the universal gas constant (8.3143 J·mol$^{-1}$·K$^{-1}$) and absolute temperature. $p_{\text{air}}$ is atmospheric pressure (in N·m$^{-2}$, approximately $10^5$ at sea level) and, because it is the transport of water vapor that concerns us, $m_{\text{mol}}$ is the molecular weight of water (0.018 kg·mol$^{-1}$). At 20°C and sealevel

$$\frac{\rho_{\text{air}} C_{p,\text{air}} R T}{m_{\text{mol air}}} \left( \frac{D_H}{D_m} \right)^{2/3} = 1502.$$  

(65)

the value cited in Chapter 11.

6 Supplement 11.6: Calculating Wet-Bulb Temperature

This supplement expands on a topic addressed on page 218 of the textbook.

---


Our task is to solve Eq. 11.41 in the textbook:

\[ C_{p,\text{air}} (T_{\text{air}} - T_{wb}) = \Lambda_{\text{wat}} [C_s(T_{wb}) - h_r C_s(T_{\text{air}})] \]  \hfill (66)

for the wet bulb temperature, \( T_{wb} \). The first step is to rearrange the equation and give it a new name

\[ f(T_{wb}) = C_{p,\text{air}} (T_{\text{air}} - T_{wb}) - \Lambda_{\text{wat}} [C_s(T_{wb}) - h_r C_s(T_{\text{air}})] . \]  \hfill (67)

We are searching for the value of \( T_{wb} \) at which \( f(T_{wb}) = 0 \) making Eq. 66 true. An efficient method for this search is outlined in Figure 26. Here I have supposed that \( f(T_{wb}) \) is a smooth function of \( T_{wb} \), shown by the solid curve. In a graphic sense, we are searching for location on the \( x \)-axis (the axis of \( T_{wb} \)) at which the curve of \( f \) crosses 0.

Figure 26: Newton’s method applied to the calculation of wet-bulb temperature, \( T_{wb} \).

To start the search, we make a guess. Any reasonable guess will do; let’s pick air temperature, \( T_{\text{air}} \), as shown on the graph. Evaluating \( f \) at \( T_{\text{air}} \), we see that our guess isn’t particularly good—\( f(T_{\text{air}}) \) is negative. Nonetheless, this initial estimate gives us a starting point, the dot labelled 1 on the curve. Given this point, we can use the slope of \( f \) at \( T_{\text{air}} \) to calculate a new guess for \( T_{wb} \). Knowing the slope of \( f \) at dot 1, we can extend a line with this slope from dot 1 to the \( x \)-axis. This intersection is our new guess for \( T_{wb} \). Calculating \( f \) for this new guess, tells us we are still a bit off (dot 2), but closer than we were before to the \( T_{wb} \) at which the solid curve crosses 0. We can then calculate the slope of \( f \) at dot 2, extend a new line to the \( x \)-axis, and repeat the whole procedure. In this fashion, we can efficiently home in on an accurate value for \( T_{wb} \). This elegant search tactic was first worked out by Isaac Newton, and it is therefore known as Newton’s method.

Before we can use this method to solve Eq. 66, we need to know how the saturation mole fraction of water vapor varies with temperature. Campbell &
Norman (1998) tell us that

\[ C_s(T) \approx k_1 \exp \left( \frac{k_2 T}{T + k_3} \right), \tag{68} \]

where

\[ \begin{align*}
  k_1 &= 6.03 \times 10^{-4}, \tag{69} \\
  k_2 &= 17.502, \tag{70} \\
  k_3 &= 240.975 \degree C. \tag{71}
\end{align*} \]

Here \( T \) is measured in \( \degree C \). This expression allows us to restate Eq. 66:

\[ f(T_{wb}) = C_{p,air} T_{air} - C_{p,air} T_{wb} - \Lambda_{wat} k_1 \exp \left( \frac{k_2 T_{wb}}{T_{wb} - k_3} \right) + \Lambda_{wat} k_1 \exp \left( \frac{k_2 T_{air}}{T_{air} - k_3} \right). \tag{72} \]

Taking the derivative of this expression (to calculate the slopes needed for Newton’s method) involves two generic relationships from calculus:

\[ \frac{d \exp a}{dx} = \exp a \frac{da}{dx}, \tag{73} \]

\[ \frac{d}{dx} \left( \frac{u}{v} \right) = \frac{v \frac{du}{dx} - u \frac{dv}{dx}}{v^2}. \tag{74} \]

Applying these relationships to the task at hand, we find that

\[ f'(T_{wb}) = -C_p(T_{wb}) - k_1 \exp \left( \frac{k_2 T_{wb}}{T_{wb} - k_3} \right) \left( \frac{k_2 T_{wb}}{T_{wb} + k_3} \right). \tag{75} \]

Newton’s method then specifies that, given a guess \( T_{wb,old} \) for wet-bulb temperature, a new, better guess is

\[ T_{wb,new} = T_{wb,old} - \frac{f(T_{wb,old})}{f'(T_{wb,old})}. \tag{76} \]

One can iterate this calculation as many times as necessary. Calculating \( f \) for each new guess tells you how close you are to the precise value at which \( f(T_{wb}) = 0 \).

\[ ^{13}\text{Campbell, G.S. and J.M. Norman. 1998. An Introduction to Environmental Biophysics. (2d Ed.) Springer-Verlag, N.Y.} \]