Mathematics and the environment

A. C. Fowler

Part 1

October 27, 2006

Contents

Preface

These lecture notes, still incomplete in parts, support the lecture course 'Mathematics and the environment' in its 2006 form. This edition is essentially the same as the 2005 version, except that I have removed the appendices. Also the material in section 4.3 has been reordered and revised. Note that these are draft chapters of a book in preparation, and thus contain more (but worthwhile!) material than is covered in the lecture course. As ever, any updated versions of the notes will be posted on the web as they become available.

I am grateful to Felix Ng, who rapidly and expertly produced some of the figures for chapters 1 and 2. Thanks also to Emanuele Schiavi, Thomas Vitolo and Dave Cocks for their vigilance in spotting errors in previous versions.

A. C. F.

October 27, 2006

Part 1

Chapter 1

Climate dynamics

The most noticeable facets of the weather are those which directly impinge on us: wind, rain, sun, snow. It is hotter at the equator than at the poles simply because the local intensity of incoming solar radiation is greater there, and this differential heating drives (or tries to), through its effect on the density of air, a poleward convective motion of the atmosphere: rising in the tropics, poleward in the upper atmosphere, down at the poles and towards the equator at the sea surface. The buoyancy-induced drift is whipped by the rapid rotation of the Earth into a predominantly *zonal flow*, from west to east in the northern hemisphere. In turn, these zonal flows are *baroclinically* unstable, and form waves (Rossby waves) whose form is indicated by the isobar patterns in weather charts.

All this frenetic activity obscures the fact that the weather is a rather small detail in the determination of the basic climate of the planet. The mean temperature of the planetary atmosphere and of the Earth's surface is determined by a balance between the radiation received by the Earth from the Sun (the incoming solar radiation), and that re-emitted into space by the Earth.

1.0.1 Radiation budget

We denote the incoming solar radiation by Q; it has a value $Q = 1370$ W m^{-2} (watts per square metre). A fraction a of this (the albedo) is reflected back into space, while the rest is absorbed by the Earth; for the Earth, $a \approx 0.3$. In physics we learn that a perfect radiative emitter (a *black body*) at absolute surface temperature T emits energy at a rate $E_b = \sigma T^4$, where σ is the Stefan-Boltzmann constant, given by $\sigma = 5.67 \times 10^{-8}$ W m^{-2} K^{-4}. If we assume that the Earth acts as a black body of radius R with *effective* (radiative) temperature T_e, and that it is in radiative *equilibrium*, then

$$4\pi R^2 \sigma T_e^4 = \pi R^2 (1-a) Q,$$

whence

$$T_e = \left[\frac{(1-a)Q}{4\sigma}\right]^{1/4}. \tag{1.1}$$

Computing this value for the Earth using the parameters above yields $T_e \approx 255$ K. A bit chilly, but not in fact all that bad!

Actually, if the effective temperature is *measured* (T_m) via the black body law from direct measurements of emitted radiation, one finds $T_m \approx 250$ K, which compares well with T_e. On the other hand, the Earth's (average) surface temperature is $T_s \approx 288$ K. The fact that $T_s > T_e$ is due to the *greenhouse* effect, to which we will return later. First we must deal in some more detail with the basic mechanisms of radiative heat transfer.

1.1 Radiative heat transfer

We are familiar with the idea of conductive heat flux, a vector with magnitude and direction, which depends on position $\mathbf{r}$. Radiant energy transfer is a more subtle concept. A point in a medium will emit radiation of different frequencies ν (or different wavelengths λ: they are conventionally related by $\lambda = c/\nu$, where c is the speed of light), and the intensity of emitted radiation will depend not only on position $\mathbf{r}$, but also on *direction*, denoted by $\mathbf{s}$, where $\mathbf{s}$ is a *unit* vector. Also, like heat flux, emitted radiation is an area-specific quantity (i. e., it denotes energy emitted per unit area of emitting surface), and because it depends on orientation, this causes also a dependence on angle between emitting surface and direction: the intensity you receive from a torch depends on whether it is shone at you or not.

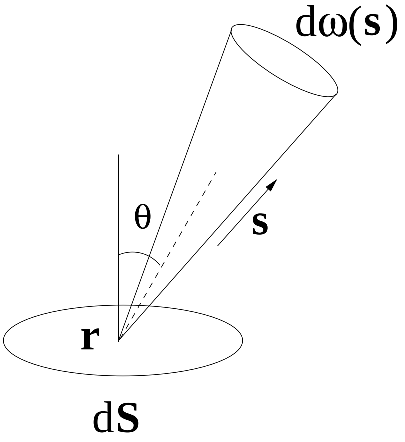

Figure 1.1: A pencil of rays emitted from a point $\mathbf{r}$ in the direction of $\mathbf{s}$.

So, the *radiation intensity* $I_\nu(\mathbf{r}, \mathbf{s})$ is defined via the relation

$$dE_\nu = I_\nu \cos\theta \, d\nu \, dS \, d\omega \, dt, \tag{1.2}$$

where dE_ν is the energy transmitted in time dt through an area dS in the frequency range $(\nu, \nu + d\nu)$ over a *pencil* of rays of *solid angle* $d\omega$ in the direction $\mathbf{s}$; see figure 2.1. θ is the angle between $\mathbf{s}$ and $d\mathbf{S}$.

The solid angle (element) $d\omega$ is the three-dimensional generalisation of the ordinary concept of angle, and is defined in an analogous way. The solid angle $d\omega$

subtended at a point O by an element of surface area $d\mathbf{S}$ located at $\mathbf{r}$ is simply

$$d\omega = \frac{\mathbf{r}\,.\,d\mathbf{S}}{r^3}. \tag{1.3}$$

The solid angle subtended at O by a surface Σ is just $\omega = \displaystyle\int_\Sigma \frac{\mathbf{r}\,.\,d\mathbf{S}}{r^3}$, and for example $\displaystyle\int_{\bigcirc} d\omega = 4\pi$, representing the solid angle over all directions from a point, and $\displaystyle\int_{\frown} d\omega = 2\pi$, representing the solid angle subtended over all upward directions.

Three processes control how the intensity of radiation varies in a medium.

- *Absorption* occurs when a ray is absorbed by a molecule, e.g. of O_2 or CO_2 in the atmosphere, or by water droplets or particles. The rate of absorption is proportional to the density of the medium ρ and the radiation intensity I_ν, and is thus given by $\rho\kappa_\nu I_\nu$, where κ_ν is the *absorption coefficient.*

- *Emission* occurs (in all directions) when molecules or particles emit radiation; this occurs at a rate proportional to the density ρ, and is thus ρj_ν, where j_ν is the *emission coefficient.*

- *Scattering* can be thought of as a combination of absorption and emission, or alternatively as a local reflection. An incident ray on a molecule or particle — a *scatterer* — is re-directed (not necessarily uniformly) by its interaction with the scatterer. The process is equivalent to instantaneous absorption and re-emission. Reflection at a surface is simply the integrated response of a distribution of scatterers. Scattering leads to an effective scattering emission coefficient $j_\nu^{(s)}$, and is discussed further below in section 1.1.6.

1.1.1 Local thermodynamic equilibrium

In order to prescribe j_ν, we will make the assumption of local thermodynamic equilibrium. More or less, this means that the medium is sufficiently dense that a local (absolute) temperature T can be defined, and *Kirchhoff's law* then defines j_ν as

$$j_\nu = \kappa_\nu B_\nu(T), \tag{1.4}$$

where $B_\nu(T)$ is the Planck function given by

$$B_\nu(T) = \frac{2h\nu^3}{c^2[e^{h\nu/kT} - 1]}, \tag{1.5}$$

where $h = 6.6 \times 10^{-34}$ J s is Planck's constant, $k = 1.38 \times 10^{-23}$ J K^{-1} is Boltzmann's constant, and $j_\nu\, d\nu$ represents the emitted energy per unit mass per unit time per unit solid angle in the frequency range $(\nu, \nu + d\nu)$.

EXERCISE. Show that $\int_{\frown}\cos\theta\, d\omega = \pi$, and deduce that $B_\nu = E_{b\nu}/\pi$, where $E_{b\nu}$ is the black body radiation emitted normally from a surface, per unit area, and (we write $E_{b\nu}$ in terms of wavelength $\lambda = c/\nu$, denoting it as $E_{b\lambda}$)

$$E_{b\nu} = E_{b\lambda} = \frac{c_1}{\lambda^5\left[\exp\left(c_2/\lambda T\right) - 1\right]}, \tag{1.6}$$

and $c_1 = 3.74 \times 10^8$ W μm^4 m^{-2}, $c_2 = 1.44 \times 10^4$ μm K (1 μm $= 10^{-6}$ m).

1.1.2 Equation of radiative heat transfer

Considering figure 2.1, the rate of change of the radiation intensity I_ν in the direction $\mathbf{s}$ is given by

$$\frac{\partial I_\nu}{\partial s} = -\rho\kappa_\nu I_\nu + \rho\kappa_\nu B_\nu, \tag{1.7}$$

and this is the equation of radiative heat transfer. Note that the meaning of $\partial I_\nu/\partial s$ in (1.7) is that it is equal to $\mathbf{s}\,.\boldsymbol{\nabla} I_\nu$, where $\boldsymbol{\nabla}$ is the gradient with respect to $\mathbf{r}$. (1.7) is easily derived from first principles, given the definition of absorption and emission coefficients.

1.1.3 Radiation budget of the Earth

We will use (1.7) below to derive a model for the vertical variation of the intensity of radiation in the Earth's atmosphere. We need to do this in order to explain the discrepancy between the effective black body temperature of the Earth (250 K) and the observed surface temperature (290 K). The discrepancy is due to the greenhouse effect of the atmosphere, which acts both as an absorber and emitter of radiation. Importantly, the absorptive capacity of the atmosphere as a function of wavelength λ is very variable. Figure 2.2 shows the variation of κ_ν (or, we might write κ_λ) as a function of λ, or more specifically, $\log_{10}\lambda$. Above it we have also the black body radiation curves (from (1.6)) for two temperatures corresponding to those of the effective Earth emission temperature, and that at the surface of the Sun. From these, we see that solar radiation is concentrated at short wavelengths, including the band of visible light ($\lambda = 0.4 - 0.7$ μm), whereas the emitted radiation is all infrared (IR). Furthermore, the absorption coefficient variation with λ is such that the atmosphere is essentially *transparent* ($\kappa \approx 0$) to solar radiation (in the absence of clouds), but (mostly) *opaque* to the emitted long-wave radiation, with the exception of an IR window between 8 and 14 μm. It is this concept of transparency to solar radiation, but the presence of only a small emission window, that leads to the analogy of a greenhouse. The incoming radiation is trapped by the atmosphere, and it is this which causes the elevated surface temperature.

The actual radiation budget of the Earth's atmosphere is shown in figure 2.3, which indicates the complexity of the transfer processes acting between the Earth's surface, the atmosphere and cloud cover, and which also shows the rôle played by *sensible* heat loss (i. e., due to convective or conductive cooling) and *latent* heat loss (due to evaporation from the oceans, for instance).

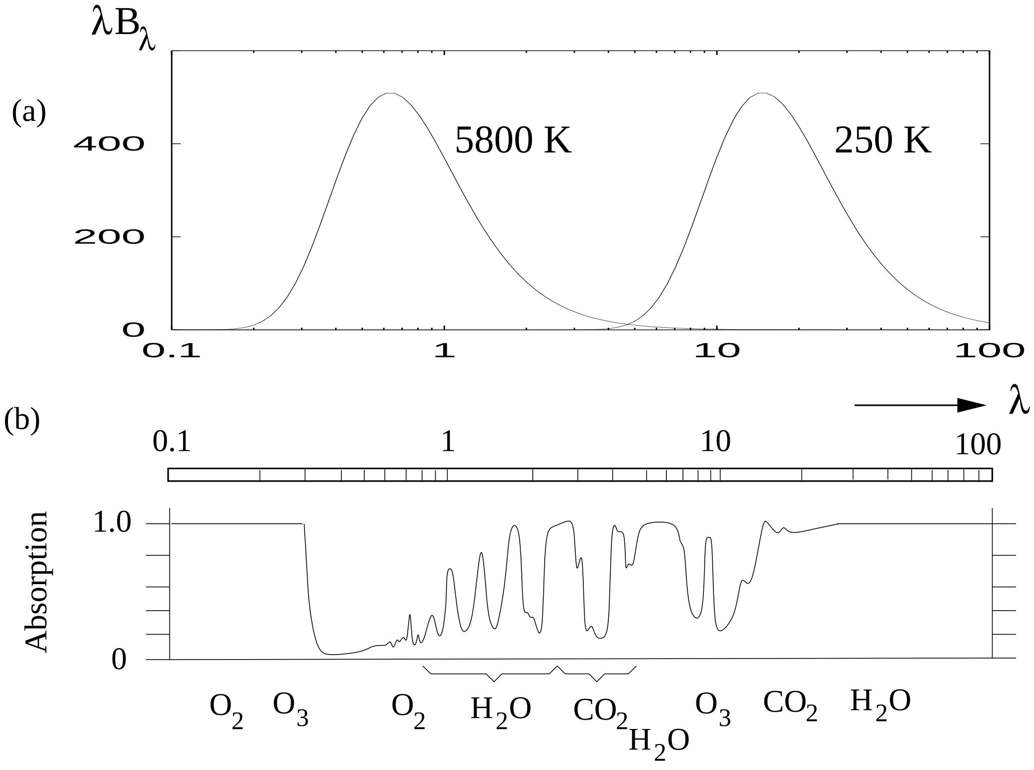

Figure 1.2: Absorption spectrum of the Earth's atmosphere. The upper graphs indicate the different wavelength dependence of the radiation emitted by the Earth and the sun. λ is measured in μm, and the solar output (from (1.6)) is scaled by 3.45×10^{-6} so that it overlays the Earth's output, if additionally λ in (1.6) is scaled by 0.043. In this case the areas under the two curves (note that $\int B_\lambda \, d\lambda = \ln 10 \int \lambda B_\lambda \, d \log_{10} \lambda$) are equal, as they should be in radiative balance. The factor 3.45×10^{-6} represents the product of $\frac{1}{4}(1 - a)$ (cf. (1.1)) with the square of the ratio of the sun's radius to the distance from the Earth to the sun. The radius of the sun is 6.96×10^8 m and the distance from the Earth to the sun is 1.5×10^{11} m, so that the value of the square ratio is about 21.53×10^{-6}. Multiplying this by the discount factor $\frac{1}{4}(1 - a)$ gives 3.45×10^{-6} if the albedo $a = 0.36$. The curves can be made to overlap for the meaqsured albedo of $a = 0.3$ by, for example, taking Earth and sun radiative temperatures to be 255 K and 5780 K, but this is largely a cosmetic exercise. The lower curve represents the absorption by atmospheric gases over a clear vertical column of atmosphere (i. e., it does not represent the absorption coefficient); we see that there is a long wave window for wavelengths between about 8 and 15 μm. This figure is redrawn from figure 2.1 of Houghton (2002).

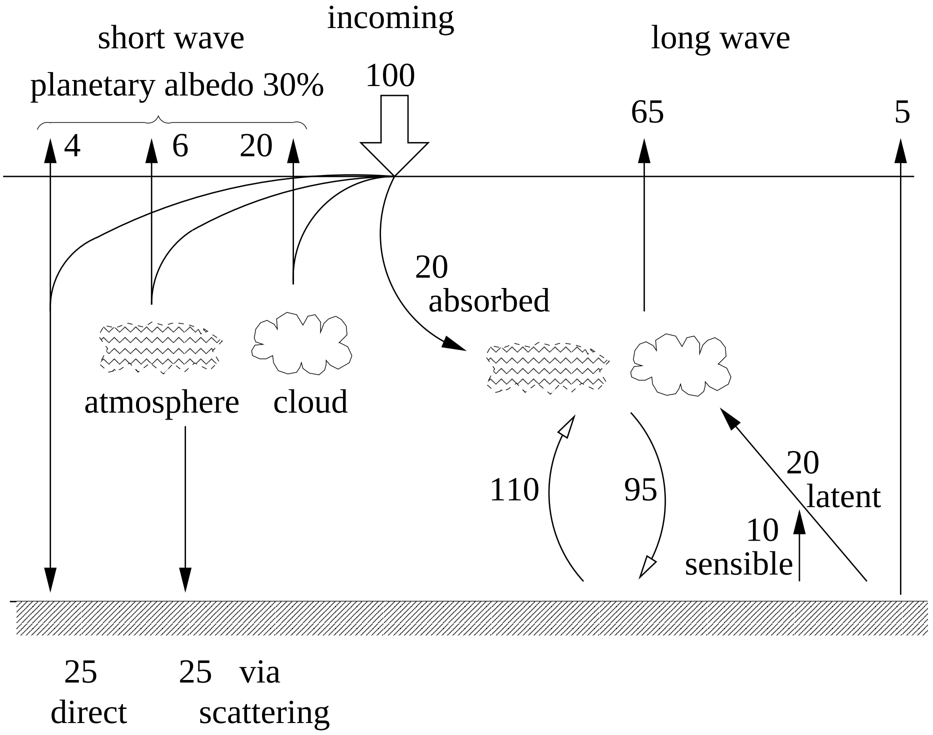

Figure 1.3: Radiation budget of the Earth. Versions of this figure, differing slightly in the numerical values, can be found in many books. See, for example, Gill (1982), figure 1.6.

As indicated in this figure, and as can be seen also from figure 2.2, one can essentially think of the short wave budget and long wave budget as separate systems. We shall be concerned here with the variation of IR radiation intensity, by solving (1.7). If κ_ν varies with ν, the problem requires computational solution. However, we can gain significant insight by introducing the idea of a *grey atmosphere*. This is one for which $\kappa_\nu = \kappa$ is independent of ν (and as mentioned, we will restrict this assumption to the long wave budget).

We then define the radiation intensity I and emission density B as

$$I = \int_0^\infty I_\nu \, d\nu, \quad B = \int_0^\infty B_\nu \, d\nu. \tag{1.8}$$

Note that, following the discussion before (1.6), we have

$$B = \frac{\sigma T^4}{\pi}, \tag{1.9}$$

where σ is the Stefan-Boltzmann constant; thus $B = E_b/\pi$. The factor of π arises because E_b represents the radiation per unit surface area emitted normally to the surface, while B represents emission per unit area per unit solid angle in any direction. It is important to understand the distinction between the two.

From (1.7), we have for a grey atmosphere

$$\frac{\partial I}{\partial s} = -\kappa\rho\,(I - B). \tag{1.10}$$

We now consider the important case of a one-dimensional atmosphere. Let z be the direction in the upward vertical, and θ the (polar) angle to the z-axis. We also define the optical depth

$$\tau = \int_z^\infty \kappa\rho\,dz, \tag{1.11}$$

and put $\mu = \cos\theta$. For a one-dimensional atmosphere, we have $I = I(\tau, \mu)$, where τ represents the vertical position, and μ represents the direction of the ray pencil in figure 2.1. Note also that $ds = dz/\mu$ (some care is needed here: z and s are independent, but this relation correctly interprets $\partial/\partial s \equiv \mathbf{s}.\boldsymbol{\nabla}_{\mathbf{r}}$ for the one-dimensional case), so that (1.10) is

$$\mu\frac{\partial I}{\partial \tau} = I - B, \tag{1.12}$$

for a *one-dimensional, grey* atmosphere.

This seems simple enough, but note that B depends on T, which is as yet unconstrained. In order to constitute B, we define the *average intensity*

$$J = \frac{1}{4\pi}\int_{\bigcirc} I\,d\omega = \tfrac{1}{2}\int_{-1}^{1} I(\tau,\mu)\,d\mu, \tag{1.13}$$

and we make the assumption of *local radiative equilibrium*[1] that $J = B$, i. e., that the total absorbed radiation at a point is equal to that determined by black body emission (note that this does not necessarily imply $I = B$ for all θ, however). The radiative intensity equation for a one-dimensional, grey atmosphere is thus

$$\mu\frac{\partial I}{\partial \tau} = I - \tfrac{1}{2}\int_{-1}^{1} I(\tau,\mu)\,d\mu, \tag{1.14}$$

and is in fact an integro-differential equation.

We require two further pieces of information to determine I completely. In view of our previous discussion, we take I as referring to long-wave radiation, and therefore it is appropriate to specify

$$I = 0 \text{ for } \mu < 0 \text{ at } \tau = 0, \tag{1.15}$$

i. e., no incoming long-wave radiation at the top of the atmosphere. Furthermore, we can see from the equation (1.14) that the net upward flux

$$\int_{\bigcirc} I\cos\theta\,d\omega = 2\pi\int_{-1}^{1} \mu I\,d\mu = \Phi \tag{1.16}$$

---

[1] This now specifically assumes that no other energy transport processes occur.

is conserved (i. e., is independent of depth). (The factor 2π is due to integration with respect to the azimuthal angle ϕ.) Since this is $2\pi\left[\int_0^1 \mu I\,d\mu - \int_{-1}^0 (-\mu I)\,d\mu\right] =$ outgoing IR radiation minus incoming IR radiation, it is in fact equal to the net emission of IR radiation. By the assumption of global radiative balance, Φ is equal to the net received short wave radiation, thus

$$\Phi = \frac{(1-a)Q}{4} = \sigma T_e^4, \tag{1.17}$$

where the factor 4 allows for the variation of received solar radiation per unit area with latitude. (Strictly, the assumption of a one-dimensional atmosphere assumes horizontal variations due to latitude are rapidly removed, e.g. by mixing, but in fact the horizontal variation is small anyway, because the atmosphere is geometrically *thin.*) In fact, even if there is global imbalance, as in climatic energy balance models (see section 2.3), we still have $\Phi = \sigma T_e^4$.

1.1.4 The Schuster-Schwarzschild approximation

The solution of (1.14) with (1.15) and (1.16) is possible but technically difficult, and is described in appendix A. A simple approximate result can be obtained by defining the outward and inward flux integrals

$$\begin{aligned} I_+ &= \int_0^1 I\,d\mu, \\ I_- &= \int_{-1}^0 I\,d\mu, \end{aligned} \tag{1.18}$$

and then approximating $\int_0^1 \mu I\,d\mu \approx \frac{1}{2}I_+$, $\int_{-1}^0 \mu I\,d\mu = -\frac{1}{2}I_-$, based on the idea that $\int_0^1 \mu\,d\mu = \frac{1}{2}$. This causes (1.14) to be replaced by

$$\begin{aligned} I_+' &= I_+ - I_-, \\ I_-' &= I_+ - I_- \end{aligned} \tag{1.19}$$

so that $I_+ - I_- = \Phi/\pi$ is the conservation law (1.16), and thus (with $I_- = 0$ at $\tau = 0$)

$$\begin{aligned} I_- &= \Phi\tau/\pi, \\ I_+ &= \frac{\Phi}{\pi}(1+\tau). \end{aligned} \tag{1.20}$$

It follows that the average intensity

$$J = \tfrac{1}{2}(I_+ + I_-) = \frac{\Phi}{2\pi}(1+2\tau) = B, \tag{1.21}$$

and using (1.9) and (1.17), we thus find the atmospheric temperature T in terms of the emission temperature T_e:

$$T = T_e \left[\frac{(1+2\tau)}{2} \right]^{1/4}. \tag{1.22}$$

The surface temperature is determined by the black body emission temperature corresponding to I_+ at the surface, where $\tau = \tau_s$, that is $I_+ = B = \sigma T_s^4/\pi$, so that the ground surface temperature is

$$T_s = T_e(1+\tau_s)^{1/4}, \tag{1.23}$$

whereas the surface air temperature T_{as} is, from (1.22),

$$T_{as} = T_e \left(\tfrac{1}{2} + \tau_s \right)^{1/4}. \tag{1.24}$$

Note that there is a discontinuity in temperature at the surface, specifically

$$T_s^4 - T_{as}^4 = 0.5 T_e^4; \tag{1.25}$$

molecular heat transport (conduction) will in fact remove such a discontinuity. If we use $T_s = 290$ K and $T_e = 255$ K, then (1.23) implies that the optical depth of the Earth's atmosphere is $\tau_s = 0.67$.

1.1.5 Radiative heat flux

Although radiative heat transfer is the most important process in the atmosphere, other mechanisms of heat transport are essential to the thermal structure which is actually observed, notably conduction and convection. In order to incorporate radiative heat transfer into a more general heat transfer equation, we need to define the *radiative heat flux.* This is a vector, analogous to the conductive heat flux vector, and is defined (for a grey medium) by

$$\mathbf{q}_R = \int_{\bigcirc} I(\mathbf{r}, \mathbf{s})\, \mathbf{s}\, d\omega(\mathbf{s}). \tag{1.26}$$

Note that $\mathbf{q}_R.\mathbf{n} = \displaystyle\int_{\bigcirc} I \cos\theta \, d\omega$ (see figure 1.1) is the energy flux density through a surface element dS with normal $\mathbf{n}$. Determination of $\mathbf{q}_R$ requires the solution of the radiative heat transfer equation for I, but a simplification occurs in the optically thick limit, when $\tau \gg 1$ (i. e., $\kappa\rho$ is small). We write

$$I = B - \frac{1}{\rho\kappa} \mathbf{s}.\boldsymbol{\nabla} I, \tag{1.27}$$

and solve for I using a perturbation expansion in powers of $1/\rho\kappa$. One thus obtains

$$I = B - \frac{1}{\rho\kappa} \mathbf{s}.\boldsymbol{\nabla} B + \ldots, \tag{1.28}$$

and substitution into (1.26) leads to the expression

$$\mathbf{q}_R \approx -\frac{4\pi}{3\kappa\rho}\boldsymbol{\nabla}B = -\frac{4\sigma}{3\kappa\rho}\boldsymbol{\nabla}T^4, \tag{1.29}$$

so that for an optically dense atmosphere, the radiative heat flux is akin to a conductive heat flux, with a nonlinear temperature-dependent (radiative) conductivity. Because of its simplicity, we will often use this expression for the radiative flux despite its apparent inappropriateness for the Earth.

1.1.6 Scattering

In a scattering atmosphere, a beam of radiation is scattered as it is transmitted, as indicated in figure 1.4. At any position $\mathbf{r}$, an incident beam of frequency ν in the direction $\mathbf{s}'$ will be deflected to a new direction $\mathbf{s}$ with a probability distribution which we define to be $p_\nu(\mathbf{s},\mathbf{s}')/4\pi$; thus the integral of p_ν over all directions is one, i. e.,

$$\int_{\bigcirc} p_\nu(\mathbf{s},\mathbf{s}')\,\frac{d\omega(\mathbf{s}')}{4\pi} = 1. \tag{1.30}$$

If all the incident radiation is scattered, then we have perfect scattering: no radiation is lost. More generally, we may suppose that a fraction a_ν is scattered (and the rest is absorbed), and a_ν is called the albedo for single scattering. Thus we define $\dfrac{a_\nu p_\nu\,d\omega}{4\pi}$ to be the probability that incident radiation from the direction $\mathbf{s}'$ will be scattered in the direction $\mathbf{s}$ over a solid angle increment $d\omega$. In general, p_ν depends on frequency, and we also suppose it depends only on the angle between $\mathbf{s}'$ and $\mathbf{s}$, thus $p_\nu = p_\nu(\mathbf{s}.\mathbf{s}')$.

Integrating this probablility over all directions $\mathbf{s}'$, we obtain the emission coefficient for scattering as

$$j_\nu^{(s)} = a_\nu\kappa_\nu \int_{\bigcirc} p_\nu(\mathbf{s},\mathbf{s}')I_\nu(\mathbf{r},\mathbf{s}')\,\frac{d\omega(\mathbf{s}')}{4\pi}, \tag{1.31}$$

where κ_ν is the emission coefficient. The equation of radiative transfer is modified from (1.7) to

$$\frac{\partial I_\nu}{\partial s} = -\rho\kappa_\nu\left[I_\nu - (1-a_\nu)\,B_\nu - a_\nu\int_{\bigcirc} p_\nu(\mathbf{s},\mathbf{s}')I_\nu(\mathbf{r},\mathbf{s}')\,\frac{d\omega(\mathbf{s}')}{4\pi}\right]. \tag{1.32}$$

Scattering in the atmosphere is most closely associated with Rayleigh's explanation for the blue colour of the sky. For the visible spectrum we can ignore short wave emission, $B_\nu = 0$. Rayleigh derived an expression for the scattering distribution of sunlight by air molecules. Importantly, the intensity of scattered radiation is proportional to ν^4 (or $1/\lambda^4$), and thus is much larger for high frequency, or short wavelength, radiation. In terms of the visible spectrum, this is the blue end. The wavelength of blue light is about 0.425 μm, while that of red light is 0.65 μm, so that blue light is scattered about five times more than red light. Hence the blue sky.

Rayleigh scattering applies to scattering by entities which are much smaller than the radiation wavelength, and in particular, molecules. Scattering by objects much larger than the wavelength (dust particles, water droplets, etc.) is called Mie scattering and is determined by WKB theory applied to the electromagnetic wave equation.

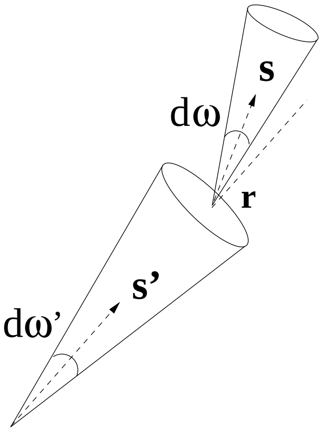

Figure 1.4: Scattering from direction $\mathbf{s}'$ to $\mathbf{s}$.

1.1.7 Troposphere and stratosphere

Thus far, we have not considered the vertical structure of the atmosphere. The principal feature of the atmosphere is that it is stratified: the density decreases, more or less exponentially, with height. This is why it becomes difficult to breathe at high altitude. The reason for this decrease is simply that the atmospheric pressure at a point depends on the weight of the overlying air, which obviously decreases with height. Since density is proportional to pressure, it also decreases with height.

To quantify this, we use the fact that for a *shallow* atmosphere (whose depth d is much less than a relevant horizontal length scale l), the pressure p is nearly hydrostatic, that is,

$$\frac{dp}{dz} = -\rho g, \tag{1.33}$$

where z is height, ρ is air density, and g is gravitational acceleration (approximately constant). If we assume (reasonably) that air behaves as a perfect gas, then

$$\rho = \frac{M_a p}{RT}, \tag{1.34}$$

where M_a is the *molecular weight*[2] of air, R is the perfect gas constant, and T is absolute temperature. For a perfect gas, the thermal expansion coefficient $-\frac{1}{\rho}\frac{\partial \rho}{\partial T}$ is simply $1/T$.

---

[2] The molecular weight is effectively the weight of a molecule of a substance. Equivalently, it is determined by the weight of a fixed number of molecules, known as a *mole*, and equal to *Avogadro's number* 6×10^{23} molecules. For air, a mixture predominantly of nitrogen (78%), oxygen (21%) and argon (0.9%), the molecular weight is given by the equivalent quantity for the mixture. It has the value $M_a = 28.8 \times 10^{-3}$ kg mole^{-1}. Useful references for such quantities and their units are Kaye and Laby (1960) and Massey (1986).

In terms of the temperature, the pressure and density are then found to be

$$p = p_0 \exp\left[-\int_0^z \frac{dz}{H}\right], \quad \rho = \rho_0 \exp\left[-\int_0^z \frac{dz}{H}\right], \tag{1.35}$$

where the *scale height* is $H = RT/M_a g$, having a value in the range 6–8 km. The temperature varies by less than a factor of two over most of the atmosphere, and an exponential relation between pressure or density and height is a good approximation.

We mentioned earlier, in deriving (1.14), that we assumed local radiative equilibrium, that is to say that radiative transport dominates the other transport mechanisms of convection and heat conduction. As we discuss further below, this is a reasonable assumption if the atmospheric density is small. As a consequence of the decrease in density with height, the atmosphere can therefore be divided into two layers. The lower layer is the *troposphere*, of depth about 10 km, and is where convective heat transport is dominant, and the temperature is *adiabatic*, and decreases with height: this is described in section 1.2 below. The troposphere is separated from the *stratosphere* above it by the *tropopause*; atmospheric motion is less relevant in the stratosphere, and the temperature is essentially governed by radiative equilibrium. In fact the adiabatic decrease in temperature in the troposphere stops around the tropopause, and the temperature increases again in the stratosphere to about 270 K at 50 km height (the *stratopause*), before decreasing again (in the *mesosphere*) and then finally rising at large distances (in the *thermosphere*, > 80 km).

The temperature structure of the atmosphere can thus be represented as in figure 1.5: the convection in the troposphere mixes the otherwise radiative temperature field to produce the adiabatic gradient which is observed.

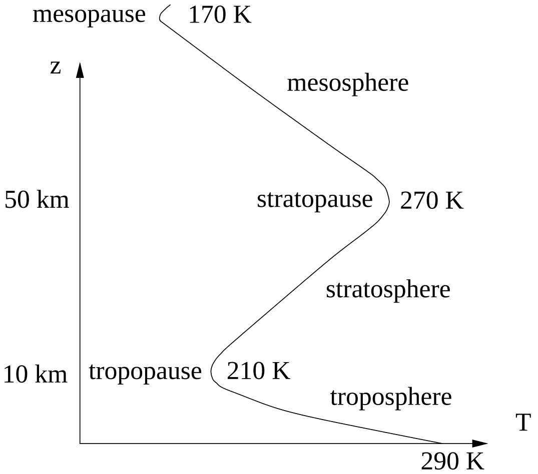

Figure 1.5: Atmospheric temperature profile. Below the tropopause, convection stirs the temperature field into an adiabatic gradient. Above it, radiative balance is dominant.

1.1.8 The ozone layer

The elevated vertical temperature profile in the stratosphere is basically due to a radiative balance between ultraviolet absorption by ozone and long wave emission by carbon dioxide. As is indicated in figure 1.2 (and as is well known), ozone (O_3) in the stratosphere is responsible for removing ultraviolet radiation, which would otherwise be lethal to life on Earth. Ozone is produced in the stratosphere through the photodissociation of oxygen. The basic sequence of reactions describing this process is due to Sydney Chapman:

$$\begin{aligned} O_2 + h\nu &\overset{j_2}{\rightarrow} 2O, \\ O + O_2 + M &\overset{k_2}{\rightarrow} O_3 + M, \\ O_3 + h\nu &\overset{j_3}{\rightarrow} O + O_2, \\ O + O_3 &\overset{k_3}{\rightarrow} 2O_2. \end{aligned} \tag{1.36}$$

The first of these reactions represents the breakdown of oxygen by absorption of ultraviolet radiation of wavelength less than 0.24 μm ($h\nu$ is Planck's quantum of energy). The next two reactions are fast. The arbitrary air molecule M catalyses the first of these. The final reaction represents the removal of ozone. Overall, the reaction can be written as

$$3O_2 \underset{r_-}{\overset{r_+}{\rightleftharpoons}} 2O_3, \tag{1.37}$$

with the first two reactions of (1.36) providing the forward reaction, and the last two the backward reaction. If we assume (as is the case) that j_3 and k_2 are sufficiently large that

$$\varepsilon = \frac{j_3}{k_2[O_2][M]} \ll 1, \quad \delta = \left[\frac{j_2 k_3}{j_3 k_2 [M]}\right]^{1/2} \ll 1, \tag{1.38}$$

then one can show (see question 1.8) that the forward and backward rates for (1.37) are

$$r_+ = \tfrac{2}{3} j_2, \quad r_- = \frac{j_3 k_3}{k_2[O_2][M]}, \tag{1.39}$$

and the (stable) equilibrium ozone concentration is given by

$$[O_3] = \left[\frac{j_2 k_2 [M]}{j_3 k_3}\right]^{1/2} [O_2]. \tag{1.40}$$

Ozone occurs principally in the *ozone layer*, at heights between 15 and 50 km (i. e., in the stratosphere), where it attains concentrations of about 10 ppmv (parts per million by volume). It is formed here because the reactions in (1.36) require UV radiation to be absorbed, which in itself requires the presence of the absorber (ozone). So at the top of the stratosphere, where the pressure and thus also density are both small, absorption is small and little ozone is formed. Deeper in the stratosphere, density increases, which allows increased production of ozone. However, this increased ozone causes further absorption of the UV radiation, and so the source for the ozone forming reaction disappears at the base of the stratosphere.

A simple model for the formation of this structure, which is called a *Chapman layer*, assumes a constant volume concentration, or mixing ratio, for ozone. The radiative transfer equation for incoming shortwave radiation of intensity I can be written

$$\frac{\partial I}{\partial z} = \kappa \rho I; \tag{1.41}$$

there is no radiative source term, and the incoming beam is unidirectional, and here taken to be vertical (the sun is overhead). We suppose a constant pressure scale height so that

$$\rho = \rho_0 \exp(-z/H). \tag{1.42}$$

With I negative, and $I \to -I_\infty$ as $z \to \infty$, the solution to this is

$$I = -I_\infty \exp\left[-\kappa \rho_0 H e^{-z/H}\right], \tag{1.43}$$

and the consequent heating rate $Q = -\dfrac{\partial I}{\partial z}$ is given by

$$Q = \frac{\tau_0 I_\infty}{H} \exp\left[-\frac{z}{H} - \tau_0 e^{-z/H}\right], \tag{1.44}$$

where

$$\tau_0 = \kappa \rho_0 H \tag{1.45}$$

is a measure of the opacity of the stratospheric ozone layer.

If τ_0 is sufficiently high, the heating rate exhibits an internal maximum, as seen in figure 1.6. This is the distinguishing feature of the Chapman layer. Since Q is also volumetric absorption rate of radiation, it indicates maximal production of ozone in the stratosphere, as is found to be the case. This structure additionally explains why the temperature rises with height through the stratosphere, because of the increased heating rate.

In the stratosphere almost all the short wave absorption is due to ozone. There is very little water vapour. The resultant heating is almost exactly balanced by long wave radiation, mostly from carbon dioxide, the remnant being from ozone again. While the resulting radiation balance controls the temperature, there is very little radiant energy lost. As can be seen from figure 1.2, the UV tail is taken off by ozone, but the visible and infra-red spectrum passes through the stratosphere unscathed.

In the troposphere, the water vapour concentration is much higher than ozone, which is virtually absent, and also of carbon dioxide. Although discussions of global warming are fixated by the greenhouse gases — carbon dioxide, methane, and so on, it needs to be borne in mind that water vapour is also a greenhouse gas, and is in fact the most important one. Adding to that the dominating influence of clouds and their somewhat mysterious influence on climate, one sees that an understanding of moisture is of principal concern in determining radiative processes in the troposphere.

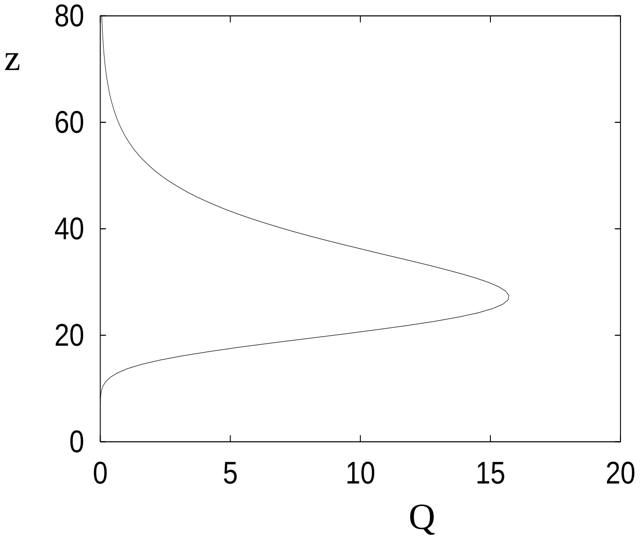

Figure 1.6: Variation of heating rate Q given by (1.44) with $I_\infty = 342$ W m^{-2}, $H = 8$ km, and $\tau_0 = 30$, this somewhat arbitrary value being chosen to show a maximum heating rate at 30 km altitude. Units of z are km, and of Q W m^{-3}. The choice of $I_\infty = 342$ W m^{-2} refers to all incoming short wave radiation, whereas it is only a small fraction of this in the ultraviolet range which is absorbed in the stratosphere.

1.2 Convection

We have seen that for a purely radiative atmosphere, a discontinuity in temperature occurs at the Earth's surface. Such a discontinuity does not occur in reality, because of molecular conduction. In fact, atmospheric motion causes heat transport in the troposphere to be more importantly due to convection rather than conduction — the transport of heat is primarily due to the motion of the atmosphere itself. The temperature of the atmosphere is described by the heat equation

$$\rho c_p \frac{dT}{dt} - \beta T \frac{dp}{dt} = k\nabla^2 T - \boldsymbol{\nabla}.\mathbf{q}_R, \tag{1.46}$$

where the terms represent respectively advection of heat, adiabatic (compression) heating, thermal conduction and radiative heat transfer. The time derivative d/dt is a *material derivative*, and represents the rate of change of a property following a fluid element. Thus, $dT/dt = 0$ means the temperature of a fluid element is conserved as it moves. It is related to the ordinary partial derivative by the relation

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \mathbf{u}.\boldsymbol{\nabla}, \tag{1.47}$$

where $\mathbf{u}$ is the fluid velocity. c_p is the specific heat, β is the thermal expansion coefficient, $\beta = -\rho^{-1}\partial\rho/\partial T$, k is the thermal conductivity, and $\mathbf{q}_R$ was defined above in (1.25).

If we use the optically dense approximation (1.28), then

$$\mathbf{q}_R \approx -k_R \boldsymbol{\nabla} T, \tag{1.48}$$

where

$$k_R = \frac{16\sigma T^3}{3\kappa\rho}. \tag{1.49}$$

We will use (1.48) as a pedagogic tool rather than as an accurate model. To estimate k_R, we use values $\kappa\rho d = 0.7$, $d = 10$ km; then we find $k_R \sim 10^5$ W m^{-1} K^{-1}. This compares to a molecular thermal conductivity of order 10^{-2} W m^{-1} K^{-1}, which is therefore negligible. In fact, atmospheric flows are turbulent, and a better measure of the effective heat conduction is the eddy thermal conductivity, of order $\rho c_p U d$ times a small dimensionless drag coefficient, where d is depth scale and U is wind speed scale. Eddy conductivity is found to be comparable to the nominal radiative value deduced above.

A measure of the importance of the advective terms is provided by the *Péclet number*, which represents the size of the ratio $(\rho c_p dT/dt)/\boldsymbol{\nabla}.\mathbf{q}_R$, and is given by

$$Pe = \frac{\rho c_p U d^2}{k_R l}, \tag{1.50}$$

where U and d are velocity and depth scales as mentioned above, and l is a relevant horizontal length scale. Using values $\rho \sim 1$ kg m^{-3}, $c_p \sim 10^3$ J kg^{-1} K^{-1}, $U \sim 20$ m s^{-1}, $d \sim 10$ km, $l \sim 10^3$ km (representing the length scale of planetary waves in the atmosphere), we find $Pe \sim 20$, so that in fact atmospheric motion plays a significant rôle in the redistribution of heat. Since Pe is large, we can obtain an approximation to the vertical thermal structure of the atmosphere by neglecting the radiative and conductive transport terms altogether. This leads to the *adiabatic lapse rate*, which is determined by putting the left hand side of (1.46) to zero, and thus

$$\frac{dT}{dp} = \frac{\beta T}{\rho c_p}. \tag{1.51}$$

To obtain the variation of T with height z, we use the fact that the pressure is hydrostatic, given by (1.33), and assume a perfect gas law (1.34); then we find the (dry) adiabatic lapse rate

$$\frac{dT}{dz} = -\Gamma_d = -\frac{g}{c_p}, \tag{1.52}$$

having a value of about 10 K km^{-1}. In practice, the observed temperature gradient is nearer 6 K km^{-1}, a value which is due to the presence of water vapour in the atmosphere, the effect of which is considered below.

One of the basic reasons for the presence of convection in the troposphere is the presence of an unstable thermal gradient. The higher temperature at the ground causes the air there to be lighter; convection occurs as the warm air starts to rise, and it is the resultant overturning which causes the mixing which creates the adiabatic gradient. On a larger scale, the unstable thermal gradient which drives large scale atmospheric motion is due to the energy imbalance between the equator and the poles.

Perturbations to the adiabatic gradient occur; for example, temperature inversions can occur under clear skies at night when IR radiation from the Earth is larger. The resultant temperature structure is convectively stable (the inversion is cold and therefore heavy), and its removal by solar irradiation can be hampered by the presence of smog caused by airborne dust particles. Moreover, the cool inversion causes fog (condensed water vapour), and the condensation is also facilitated by airborne pollutant particles, which act as nucleation sites. Hence the infamous smogs in London in the 1950s, and the consequent wide-spread ban of open coal fires in cities.

While temperature inversions are convectively stable and thus persistent, superadiabatic temperatures are convectively unstable, and cannot be maintained.

1.2.1 The wet adiabat

For a parcel of air of density ρ_a containing water vapour of density ρ_v, the *mixing ratio* is defined as

$$m = \frac{\rho_v}{\rho_a}. \tag{1.53}$$

A typical value in the troposphere is $m \approx 0.02$, so that we can practically take the density of moist air as constant. As m increases, the air can become *saturated* and thus the water vapour will condense. This happens when the *partial pressure* p_v reaches the *saturation vapour pressure* p_{sv}, which depends on temperature via the Clausius-Clapeyron equation

$$\frac{dp_{sv}}{dT} = \frac{\rho_v L}{T}, \tag{1.54}$$

where L is the latent heat and T in (1.54) is the saturation value T_{sat}. Figure 2.7 shows the phase diagram for water, delineating the curves in (T, p) space at which freezing, condensation and sublimation occur. (1.54) describes the water/vapour curve in this figure. The ratio p_v/p_{sv} (normally measured as a percentage) is called the *relative humidity.* It is an anthropocentric measure of discomfort, since when the (relative) humidity is high, very little exertion will cause sweat to condense.

Let us now suppose that the atmosphere is (just) saturated. The existence of clouds actually negates this proposition, but not too badly, in the sense that we suppose rainfall removes condensed water droplets. As a moist parcel of air moves about, the increment of heat content per unit volume due to changes in T, p and ρ_v is then $\rho_a c_p dT - dp + L d\rho_v$ (using $\beta T = 1$), and thus (1.46) is modified to

$$\rho_a c_p \frac{dT}{dt} - \frac{dp}{dt} + \rho_a L \frac{dm}{dt} = k\nabla^2 T - \boldsymbol{\nabla} . \mathbf{q}_R. \tag{1.55}$$

Using the definition of m in (1.53), and the perfect gas laws

$$p = \frac{\rho_a RT}{M_a}, \quad p_{sv} = \frac{\rho_v RT}{M_v}, \tag{1.56}$$

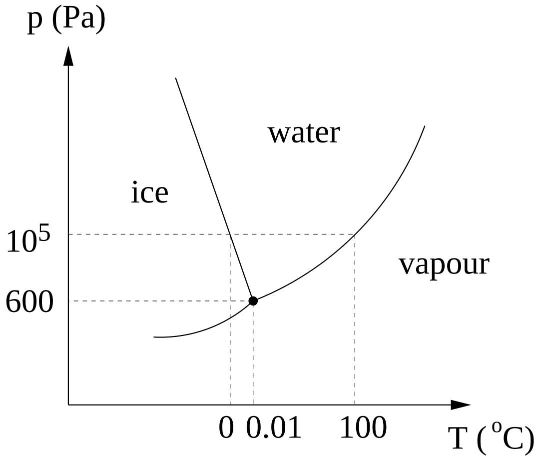

Figure 1.7: Phase diagram for water substance (not to scale).

where M_v is the molecular weight of water vapour, we find that the temperature gradient is given approximately (by ignoring the right hand side of (1.55)) by

$$\frac{dT}{dz}=-\Gamma_w=-\Gamma_d\left[\frac{1+\dfrac{\rho_v L}{p}}{1+\dfrac{\rho_v L}{p}\left(\dfrac{M_v}{M_a}\dfrac{L}{c_p T}\right)}\right], \tag{1.57}$$

which is the wet adiabat. Using values $\rho_v=0.01$ kg m^{-3}, $M_v=18\times10^{-3}$ kg mole^{-1}, $L=2.5\times10^6$ J kg^{-1}, we find a typical value $\Gamma_w\sim6$ K km^{-1}, close to that which is observed in practice.

1.3 Energy balance models

Although convective transport is the dominant mechanism of energy transfer *within* the atmosphere, the rôle of radiative transport is fundamental to the determination of the average temperature. Moreover, this is equally true if we do *not* assume radiation balance, and this allows us to study long term variations in climate which are of relevance to the evolution of paleoclimatic temperatures, quaternary ice age climates, and more recently, the effect of CO_2 levels on global temperature. All of these phenomena can be roughly understood on the basis of *energy-balance models.*

Since most of the mass of the atmosphere is contained in the troposphere, we define the mean temperature $\bar{T}$ of the atmosphere to be the vertically averaged temperature of the troposphere. Suppose the temperature is adiabatic, with constant lapse rate Γ, and of depth d. The surface temperature T is thus

$$T=\bar{T}+\tfrac{1}{2}\Gamma d. \tag{1.58}$$

In a purely radiative atmosphere, we found earlier that the greenhouse effect causes the surface temperature to be warmer than the planetary long wave emission temperature (cf. (1.23)). Let us define a greenhouse factor

$$\gamma = \left(\frac{T_e}{T} \right)^4, \tag{1.59}$$

where for (1.23), this would be

$$\gamma = \frac{1}{1+\tau_s}, \tag{1.60}$$

and this enables us to write the emitted long wave radiation in terms of the mean surface temperature, and a quantity γ which depends on atmospheric radiative properties.

We can still define a greenhouse factor by (1.59) for a radiative-convective atmosphere, but consultation of figure 1.3 shows that its theoretical determination in terms of atmospheric properties is likely to be non-trivial. Nevertheless, we shall suppose γ can be defined; for the Earth $\gamma \approx 0.61$ at present (based on $T_e = 255$ K, $T = 288$ K).

The incoming solar radiation per unit area is $(1-a)Q$ (a is the albedo, the fraction of short wave radiation which is reflected back to space), while the emitted IR radiation per unit area is σT_e^4 (units are W m^{-2}). It follows that the net received radiation over the planetary surface is $\pi R^2(1-a)Q - 4\pi R^2 \sigma T_e^4$, with units of W, and we can equate this to the rate of change of the atmospheric heat content[3], $4\pi R^2 d \rho_a c_p \dfrac{d\bar{T}}{dt}$, where d is the depth of the troposphere; c_p is the specific heat, and R is the planetary radius. Since $\rho_a R^2 d$ has units kg, c_p has units J kg^{-1} K^{-1}, and dT/dt has units K s^{-1}, this also has units W, and thus (adopting (1.59))

$$\rho_a c_p d \frac{dT}{dt} = \tfrac{1}{4}(1-a)Q - \sigma\gamma T^4, \tag{1.61}$$

in view of (1.58), since we take $\frac{1}{2}\Gamma d$ to be constant.

For constant Q, (1.61) is a simple first order differential equation with stable positive steady state, the radiative equilibrium state

$$T = T_0 = \left[\frac{(1-a)Q}{4\sigma\gamma} \right]^{1/4}. \tag{1.62}$$

The response time for small deviations from T_0 is then determined by the linearised equation, where we put $T = T_0 + \theta$, whence

$$\frac{\rho_a c_p d T_0}{(1-a)Q} \dot{\theta} \approx -\theta, \tag{1.63}$$

and the response time is

$$t_R \sim \frac{\rho_a c_p T_0 d}{(1-a)Q}. \tag{1.64}$$

---

[3] This is something of a simplification. Net addition of radiant energy to the atmosphere can cause changes in sensible heat (via temperature), latent heat (via moisture) or gravitational potential energy (via thermal expansion); we thus implicitly neglect the latter two; see also the next footnote.

With a density $\rho_a = 1$ kg m^{-3}, $c_p = 10^3$ J kg^{-1} K^{-1}, $T_0 = 288$ K, $d = 10^4$ m, $a = 0.3$, $Q = 1370$ W m^{-2}, we have $t_R \sim 35$ days, so that climatic response is relatively rapid.

1.3.1 Zonally averaged energy balance models

Energy balance models are obviously crude, but attractive nonetheless because they portray the essential truth about atmospheric energy balance. One of the more obvious features of the planetary climate is the temperature difference between equator and poles, due to the latitudinal variation of received solar variation. Indeed, it is this imbalance which drives the atmospheric weather systems. A simple modification to the 'zero-dimensional' energy balance model (1.61) is to allow a latitudinal variation in temperature. We denote latitude (angle north of the equator) by λ and we define

$$\xi = \sin\lambda, \tag{1.65}$$

thus $-1 < \xi < 1$, and $\xi = 0$ at the equator, $\xi = 1$ at the north pole. We suppose $T(\xi, t)$ is the *zonally-averaged* (i. e., integrated over longitude) temperature, and we pose the zonally-averaged energy balance equation

$$C\frac{\partial T}{\partial t} = D\frac{\partial}{\partial \xi}\left[(1-\xi^2)\frac{\partial T}{\partial \xi}\right] + \tfrac{1}{4}Q(1-a)S(\xi) - I(T). \tag{1.66}$$

In this equation, C is a heat capacity coefficient. For a dry atmosphere[4], (1.61) indicates $C = \rho_a c_p d$. D is a diffusion coefficient, scaled with $1/R^2$, and thus having (like C) units of W m^{-2} K^{-1}; it represents the poleward transport of energy through the eddy diffusive effect of large weather systems in mid-latitudes, which will be discussed further in the following chapter. $I(T)$ represents the outgoing long wave radiation, supposed to depend only on mean surface temperature. Finally, $S(x)$ represents the latitudinal variation of received solar radiation, normalised so that $\displaystyle\int_0^1 S(\xi)\,d\xi = 1$. If the albedo a is constant, then we regain (1.61) by integrating from $\xi = -1$ to $\xi = 1$, assuming T is regular at the poles. If $S \equiv 1$, then $T = T(t)$, and we also regain the earlier model.

In the formulation of (1.66), we again interpret T as the mean surface temperature, in view of (1.58), and this is what is conventionally done, though without explicit mention. It is also conventional, in view of the limited range of T, to take a linear dependence of I on T, thus

$$I = A + BT, \tag{1.67}$$

with values of A and B from measurements. Typical such values are $A = 200$ W m^{-2} and $B = 2$ W m^{-2} K^{-2}.

The resulting linear equation for T can then be solved as a Fourier-Legendre expansion if the albedo is known. For example, let us suppose that a as well as S

---

[4] For a moist, saturated atmosphere, we may take the moisture mixing ratio m to be a function of T, and in this case the latent heat $\rho_a L m$ (L being latent heat) simply modifies the heat capacity coefficient. Question 1.11 shows how to calculate $m(T)$.

is an even function of ξ (thus exhibiting north-south symmetry). It is convenient to write the equation (1.66) in terms of I, thus

$$C^*\frac{\partial I}{\partial t} = D^*\frac{\partial}{\partial \xi}\left[(1-\xi^2)\frac{\partial I}{\partial \xi}\right] + \tfrac{1}{4}Q(1-a)S(\xi) - I, \tag{1.68}$$

where $D^* = D/B$, $C^* = C/B$. We solve this in the steady state by writing

$$I = \sum_{n \text{ even}} i_n P_n(\xi), \tag{1.69}$$

where P_n is the n-th Legendre polynomial (and is an even function of ξ for n even). If we expand

$$\tfrac{1}{4}Q(1-a)S(\xi) = \sum_{n \text{ even}} q_n P_n(\xi), \tag{1.70}$$

where

$$q_n = \frac{(2n+1)Q}{4}\int_0^1 (1-a)S(\xi)P_n(\xi)\,d\xi, \tag{1.71}$$

then the coefficients i_n are given by

$$i_n = \frac{q_n}{1+dn(n+1)}. \tag{1.72}$$

For example, if we take a to be constant and the realistic approximation $S = 1 - \alpha P_2(\xi)$, $\alpha \approx 0.48$, then

$$I = \tfrac{1}{4}Q(1-a)\left[1 - \frac{\alpha P_2(\xi)}{1+6d}\right], \tag{1.73}$$

A better approximation uses $a = a_0 + a_2P_2(\xi)$, where $a = 0.68$ and $a_2 = -0.2$; this represents to some extent the higher albedo (due to ice cover) in the polar regions. The resultant two term approximation for the temperature, $T = (i_0 - A + i_2P_2(\xi))/B$, then yields a good approximation to the observed mean surface temperature if we take $D = 0.65$ W m^{-2} K^{-1}.

1.3.2 Carbon dioxide and global warming

If we are interested in the gradual evolution of climate over long time scales, then in practice we can neglect the time derivative term in (1.61), and suppose that T is in a quasi-equilibrium state. Figure 1.8 shows the rising concentration of CO_2 in the atmosphere over the last two hundred years. Essentially, the secular rise is due to the increased industrial output since the industrial revolution.

Although CO_2 is only present in small quantities, it is an important absorber for the long-wave emitted IR radiation. The effect of increasing its concentration is to increase the optical density, and thus to decrease γ. Let us suppose then that the change in CO_2 leads to a change in the greenhouse coefficient γ given by

$$\gamma = \gamma_0 - \tilde{\gamma}; \tag{1.74}$$

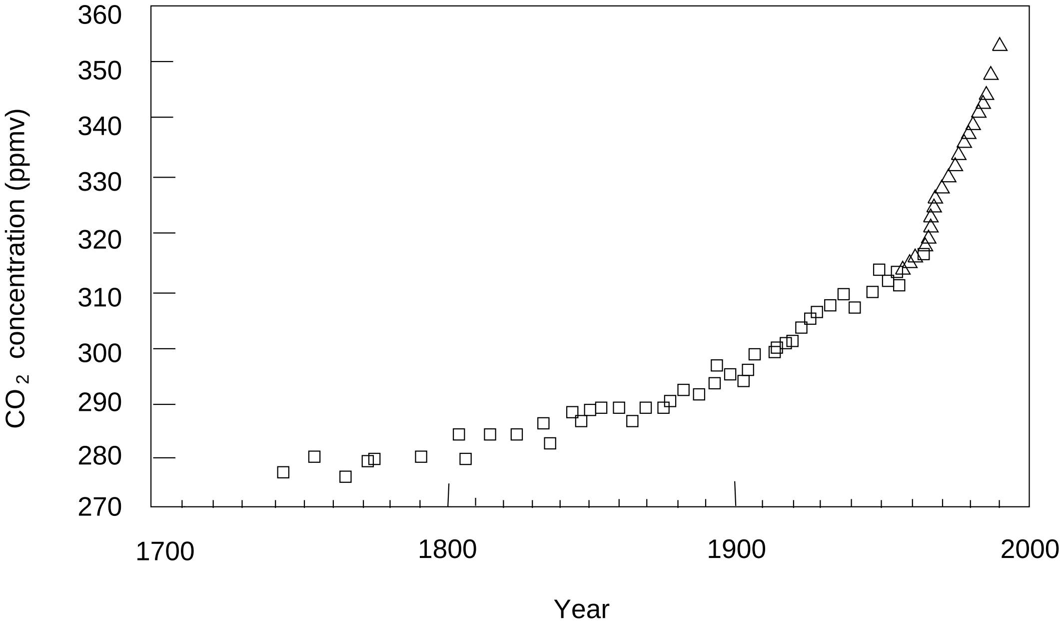

Figure 1.8: Rise in atmospheric concentration of CO_2 since 1750. The squares indicate measurements from Antarctic ice cores, and the triangles represent direct measurements from Mauna Loa observatory in Hawaii.

γ_0 is the pre-industrial reference state, and $\tilde{\gamma}$ represents the (positive) secular change due to CO_2. With $\tilde{\gamma} \ll 1$, we thus have the quasi-equilibrium given by (1.62), which leads to

$$T \approx T_0 + \frac{\tilde{\gamma} T_0}{4\gamma_0}. \tag{1.75}$$

Of course the difficulty lies in evaluating an effective dependence of $\tilde{\gamma}$ on CO_2 levels, and in reality, the problem is made more difficult by the non-greyness of the atmosphere. To understand this, let us consider the long wave thermal emission as a function of wavelength. This is shown in figure 1.9, together with black body irradiance curves at various temperatures. The emission curve divides quite neatly into a number of distinct wavelength intervals, in each of which the emission quite closely follows the black body radiation corresponding to distinct temperatures. We see a window between 10 and 13 μ, where there is little absorption, and the effective emission temperature is that at ground level. At higher wavelength, (14–16 μ), there is a CO_2 absorption band, and the radiation appears to emanate from the lower stratosphere.

In order to understand how this can be, we revisit the concept of the Chapman layer discussed above in section 1.1.8. We write the radiation intensity equation (1.7) for a one-dimensional atmosphere in the form

$$\mu \frac{\partial I_\nu}{\partial z} = -\kappa_\nu \rho_0 e^{-z/H} \left[I_\nu - B_\nu \right], \tag{1.76}$$

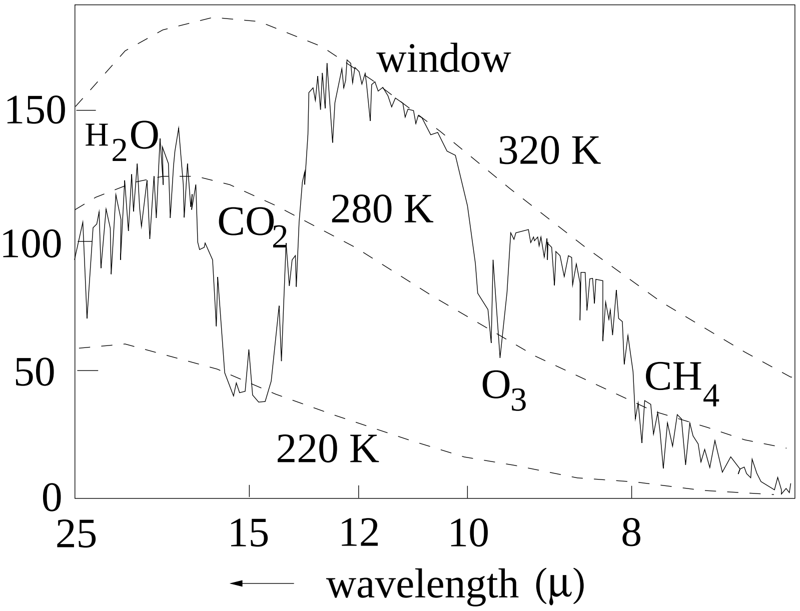

Figure 1.9: Vertical thermal emission from the Earth measured over the Sahara. The horizontal axis is linear in wavenumber, hence the irregular intervals for the wavelength in microns. The units of radiation are mW m^{-2} sr^{-1} $(\mathrm{cm}^{-1})^{-1}$, the last two indicating inverse steradian (the unit of solid angle) and wavenumber. The dashed lines are the black body radiation curves at the indicated temperatures. Redrawn from figure 12.7 of Houghton (2002).

where H is the scale height, taken as constant. For local thermodynamic equilibrium, $B_\nu(T)$ is an increasing function of temperature given by (1.5). When $\mu = 1$, the solution for upwards travelling radiation is

$$I_\nu = I_\nu^0 \exp\left[-\tau_\nu\left(1 - e^{-\zeta}\right)\right] + \tau_\nu \exp\left[\tau_\nu e^{-\zeta}\right] \int_0^\zeta B_\nu(T) \exp\left[-\zeta - \tau_\nu e^{-\zeta}\right] \, d\zeta, \quad (1.77)$$

where

$$\zeta = z/H, \quad \tau_\nu = \kappa_\nu \rho_0 H. \quad (1.78)$$

When τ_ν is small, as for the window between 10 and 13 μ, then

$$I_\nu \approx I_\nu^0. \quad (1.79)$$

When $\tau_\nu > 1$ the kernel of the integrand has an internal maximum at $\zeta = \ln \tau_\nu$, and by putting

$$\zeta = \ln \tau_\nu + Z, \quad (1.80)$$

we have if τ_ν is large the approximation

$$I_\nu \approx \exp\left[e^{-Z}\right] \int_{-\ln \tau}^{Z} B_\nu(T) \exp\left[-Z' - e^{-Z'}\right] \, dZ'. \quad (1.81)$$

The kernel $\exp\left[-Z' - e^{-Z'}\right]$ of the integrand is a peaked function with a maximum at $Z' = 0$. It thus filters out the values of B in the vicinity of $\zeta = \ln \tau_\nu$. If we idealise the kernel as a delta function centred on $\zeta = \ln \tau_\nu$, then we have

$$I_\nu|_{Z\to\infty} \approx B_\nu(T)|_{\zeta=\ln\tau_\nu}\,, \tag{1.82}$$

and it is in this sense that the thermal emission picks out black body radiation at the level corresponding to the opacity at that frequency.

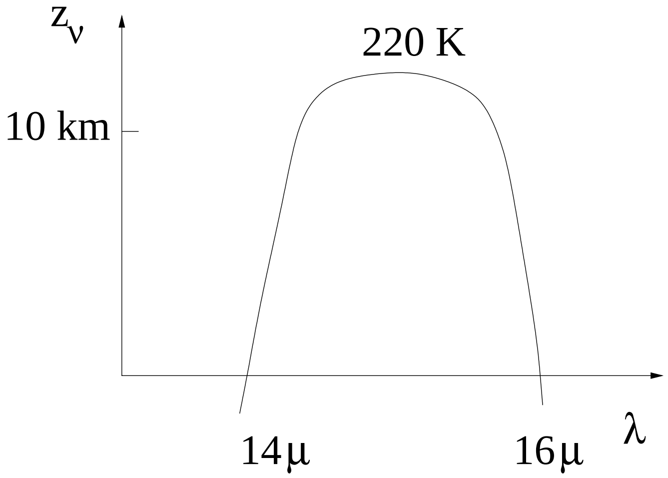

Figure 1.10: Schematic variation of the effective emission height with wavelength in the CO_2 absorption band.

We denote the effective emission altitude for a particular frequency as z_ν, thus

$$z_\nu = H \ln\left[\kappa_\nu \rho_0 H\right]. \tag{1.83}$$

Inspection of figure 1.9 then suggests that the variation of z_ν with frequency (or wavelength) in the $15\,\mu$ CO_2 absorption band is as indicated in figure 1.10, this variation being due to the variation of absorption coefficient with ν.

We can now infer the effect of increasing CO_2 density. Increasing ρ_0 has the effect of shifting the emission altitude upwards. In the stratosphere, this increases the temperature and therefore also the emission rate. Because of this, the stratosphere will cool under increased CO_2. On the other hand, the upwards shift of emission height at the fringes of the absorption band causes a cooling in the adiabatic troposphere and thus decreased emission. It is this shift of the emission height which is the cause of tropospheric heating under raised CO_2 levels.

Estimates of the consequential effect of increasing CO_2 levels is rendered uncertain because of various feedback effects which will occur in association. In particular, water vapour is also a major greenhouse gas (as can be seen from figure 1.9), and increased temperature causes increased evaporation and thus enhances the greenhouse effect. Perhaps more importantly, change of cloud cover can have a strong effect on temperature, because of its multiple influences: short wave albedo, as well as long wave absorption and emission (see figure 1.3). It is because of the uncertainty in parameterising cloud formation and structure that there is so much uncertainty associated with forecasts of global warming. Current estimates suggest that doubling CO_2 leads to a global increase of surface temperature of 2–4 K. It is politically attractive to associate the recent dry summers in England to the effects of CO_2, but although this may indeed be the cause, nevertheless the natural variability of climate on short time scales does not allow us to make this deduction with any real justification.

1.3.3 The runaway greenhouse effect

If the blanketing effect of the greenhouse gases is the cause of the Earth's relatively temperate climate, what of Venus? Its surface temperature has been measured to be in the region of 700 K, despite (see question 1.1) an effective emission temperature of 230 K. That the discrepancy is due to the greenhouse effect is not in itself surprising; the atmosphere is mostly CO_2 and deep clouds of sulphuric acid completely cover the planet. What is less obvious is why the Venusian atmosphere should have evolved in this way, since in other respects, Venus and Earth are quite similar planets.

A possible explanation of this can be framed in terms of the simple energy balance model proposed above, together with a consideration of the evolution of the amount of water vapour in the atmosphere. Initially, primitive terrestrial planets have no atmosphere (and no oceans or land ice). The heat generated by planetary accretion and by radioactive heat release is, however, enormous, and causes a huge amount of volcanism. In the eruption of magma, dissolved gases including H_2O and CO_2 are exsolved (for example, by pressure release, in much the same way bubbles form when a champagne bottle is opened). On the Earth, the increasing atmospheric density causes a slow rise in the temperature, while simultaneously the increasing partial pressure p_v of water vapour brings the atmosphere closer to saturation. On the Earth, it is supposed (see figure 2.9) that p_v reaches the saturation vapour pressure p_{sv} when $p_v > 600$ Pa (the triple point pressure). Clouds form of water droplets, and the ensuing rain forms the oceans and rivers. Most of the CO_2 is then removed from the atmosphere to form carbonate rocks.

On Venus, on the other hand, the slightly higher received solar radiation causes the (T, p_v) path which is traced to be higher. As p_v increases, so does T, and we suppose (see figure 2.9) that saturation never occurs. The water vapour continues to increase, leading to ever higher greenhouse temperatures.

A subsidiary question is then, what happens to the H_2O on Venus? The atmosphere is essentially devoid of H_2O. Here the idea is that UV radiation in the upper atmosphere dissociates the hydrogen from oxygen, the hydrogen then escapes to space, while the oxygen is used up in oxidising reactions with surface rocks.

The mechanism above is attractively simple, and can be understood using the concept of radiation balance in a grey atmosphere. The equilibrium temperature

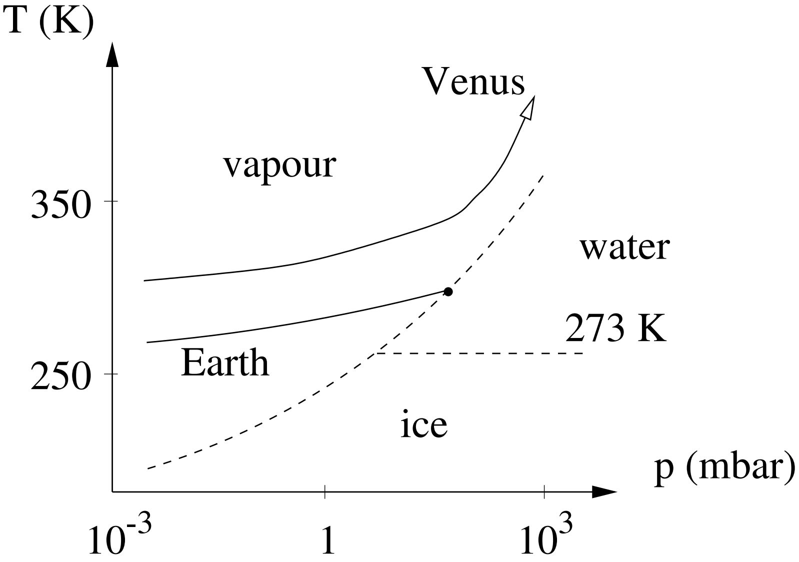

Figure 1.11: A schematic representation of the evolution of temperatures on Venus and Earth. As the atmospheric water vapour increases on Earth, condensation occurs, leading to clouds, rainfall and ocean formation. On Venus this does not occur, and the water vapour is ultimately lost through dissociation, hydrogen escape and surface oxidation reactions.

from (1.62) is

$$T = \left[\frac{(1-a)Q}{4\sigma\gamma}\right]^{1/4}, \tag{1.84}$$

and we can expect both the albedo a and greyness γ to depend on the density of water vapour ρ_v. For simplicity, take $a = 0$ and for γ we use a formula suggested by (1.60) in the optically dense limit ($\tau_s \gg 1$):

$$\gamma = \frac{1}{\tau_s}. \tag{1.85}$$

Taking $\tau_s = \kappa\rho_v d$, we then have

$$T \approx \left[\frac{Q\kappa d}{4\sigma}\rho_v\right]^{1/4}, \tag{1.86}$$

and using the perfect gas law (1.56) in the form $p_v = \rho_v RT/M_v$ gives

$$T \approx \left[\frac{Q\kappa d M_v}{4\sigma R}p_v\right]^{1/5}, \tag{1.87}$$

and we write this in the form

$$T = \left[\frac{Q\kappa d M_v p_{sv}^0}{4\sigma R}\right]^{1/5} e^{\xi}, \tag{1.88}$$

where

$$\xi = \tfrac{1}{5}\ln(p_v/p_{sv}^0), \tag{1.89}$$

and p_{sv}^0 is a reference value of the saturation vapour pressure, which we will take to be the triple point pressure, 6 mbar. On the other hand, the saturation temperature is determined by solving the Clausius-Clapeyron equation (1.54). The exact solution of this is

$$p_{sv} = p_{sv}^0 \exp\left[a\left\{1 - \frac{T_{\rm sat}^0}{T_{\rm sat}}\right\}\right], \tag{1.90}$$

where $T_{\rm sat}$ is the saturation temperature, $a = M_v L/RT_{\rm sat}^0$, and for $T_{\rm sat} - T_{\rm sat}^0 \ll T_{\rm sat}^0$, this is

$$T_{\rm sat} \approx T_{\rm sat}^0[1 + \nu\xi], \tag{1.91}$$

where

$$\nu = \frac{5}{a} = \frac{5RT_{\rm sat}^0}{M_v L}, \tag{1.92}$$

with approximate value $\nu \approx 1/4$. $T_{\rm sat}^0$ is the saturation temperature at the triple point, $T_{\rm sat}^0 \approx 273$ K. If we write $T_{\rm sat} = T_{\rm sat}^0\theta_{\rm sat}$, $T = T_{\rm sat}^0\theta$, then the planetary and saturation temperature curves are given respectively by

$$\begin{aligned} \theta &= re^{\xi}, \\ \theta_{\rm sat} &= 1 + \nu\xi, \end{aligned} \tag{1.93}$$

where

$$r = \frac{1}{T_{\rm sat}^0}\left[\frac{Q\kappa d M_v p_{sv}^0}{4\sigma R}\right]^{1/5}. \tag{1.94}$$

The definition of r here should not be taken too seriously, as we implicitly assumed that absorption was entirely due to water vapour. However, the intersection of the curves in (1.93) makes the point that the runaway effect can be expected if r is large enough, specifically if

$$r > r_c = \nu \exp\left(\frac{1-\nu}{\nu}\right) \approx 5 \tag{1.95}$$

for $\nu = 1/4$, and this corresponds to a sufficiently large value of Q. Hence, the distinction between Earth and Venus, for which the value of Q is twice that of Earth. The situation is illustrated in figure 1.11.

1.4 Ice ages

Most people are probably aware that we live in glacial times. During the last two million years, a series of ice ages has occurred, during which large ice sheets have grown, principally on the northern hemisphere land masses. The Laurentide ice sheet grows to cover North America down to the latitude of New York, while the Fennoscandian ice sheet grows in Scandinavia, reaching into the lowlands of Germany, and possibly connecting across the north sea to a British ice sheet which covers much of Britain

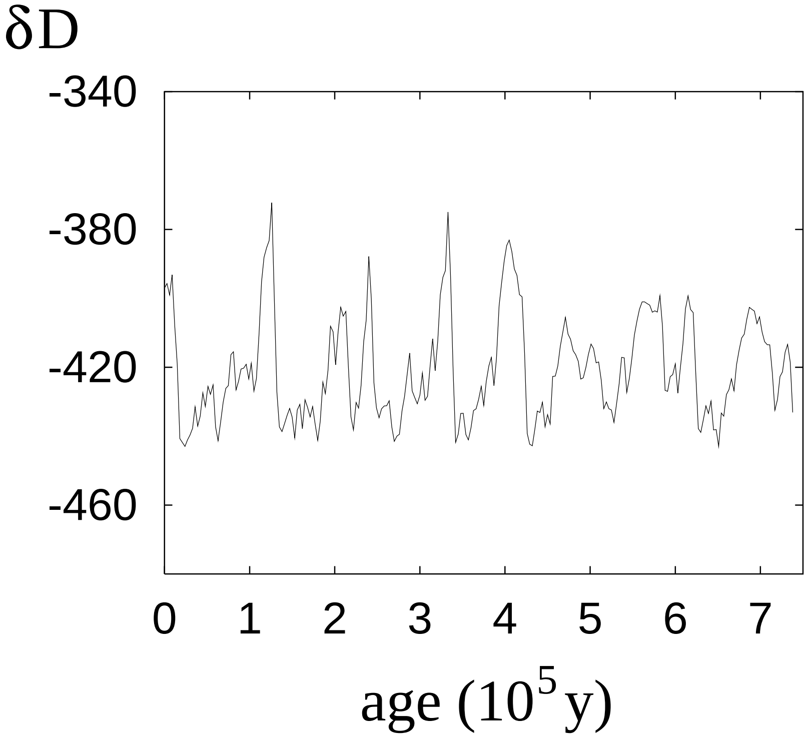

Figure 1.12: A proxy measurement from deuterium isotope data of the climate of the last 740,000 years. The measurements come from an Antarctic ice core (see the EPICA team members' paper, Agustin *et al.* 2004), and were provided by Eric Wolff. Each data point is the measurement of deuterium isotope ratios in a column of ice representing 3,000 years accumulation (i. e., the data represents 3,000 year averages). Time moves from right to left along the abscissa, and the deuterium isotope ratio is a proxy measurement of prevailing climatic temperature.

and Ireland down to Kerry in the west and Norfolk in the east. The global ice volume which grows in these ice ages is sufficient to lower sea level by some 120 metres, thus exposing vast areas of continental shelf.

These Pleistocene ice ages occur with some regularity, with a period of 100,000 years (although prior to the last half million years, a periodicity of 40,000 years appears more appropriate). The great ice sheets grow slowly over some 90,000 years, and there is then a fairly sudden deglaciation. This is illustrated in figure 1.12, which shows a proxy measurement of temperature over the last 740,000 years, obtained from an Antarctic ice core. Five sharp rises in temperature can be seen separating the last four ice ages, which show a characteristic slow decline in temperature. As we shall see, the mechanism which causes this sequence of pseudo-periodic oscillations in the climate is not very well understood.

The present glacial climate is thought to have been a result of a gradual cooling initiated by the collision of India with Asia starting some 40 million years ago, and causing the rise of the Himalayas. Although these mountains affect weather systems directly, their effect on climate may be due to the increasing precipitation and thus weathering which they induce, which leads to a removal of carbon from the atmosphere and a consequent cooling of the atmosphere. It is certainly the case that CO_2 has faithfully followed climatic temperatures through the recent ice ages, and it is difficult not to suppose that it has been a major causative factor in their explanation.

If we go further back in time, we encounter much warmer climates. The time of the dinosaurs, extending back to the Triassic, some 200 million years ago, saw a very warm climate and some very large creatures. There were no ice sheets: the Antarctic ice sheet only began to grow some tens of millions of years ago after the India–Asia collision.

Further back, however, we find evidence of major glaciated periods of Earth history, in the Carboniferous period some 300 million years ago. The glacial deposits which indicate this are located in India, South Africa, Australia and South America. But at the time of the glaciation, these continental masses were all sutured together in the great palaeo-continent of Gondwanaland, and they resided at the south pole. The break up of Gondwanaland to form the continents as we now see them only began some 200 million years ago, and is more or less coincident with the global rise in temperature and the flourishing of the dinosauria.

Even earlier in time, we have evidence of further massive glaciation on the supercontinent of Rhodinia during Proterozoic times, some 600 million years ago. The fact that these glaciations occur at then equatorial positions has led to the challenging concept of the ‘snowball Earth’, the idea that the whole planet was glaciated. Like most outrageous ideas, this is both enticing and controversial; we shall say more about it in section 1.5 below.

1.4.1 Ice-albedo feedback

The simplest type of model to explain why ice ages may occur in a sequential fashion is the energy balance model of section 1.3. On its own, it predicts a stable climatic response to solar radiative input, but when the feedback effect of ice is included, this alters dramatically. Although simple in concept, the energy balance model provides the platform for more recent models of ‘intermediate complexity’.

The mechanism of the ice-albedo feedback is this. In winter, Antarctica is surrounded by sea ice, and the Arctic ocean is permanently covered by sea ice. Land ice is also present on the Earth near the poles, or in mountainous regions. The presence of ice has a dramatic effect on the surface albedo. While the reflectivity of oceans or forest is typically 0.1, that of sea ice or snow is in the range 0.6–0.8. From figure 1.3, we see that of 50% of incoming solar radiation received at the surface, 4% is reflected from the surface: a surface albedo of 0.08. However, if the planet were covered in ice, the surface albedo might be 0.7, so that 35% would be reflected. Overall, the planetary albedo would be doubled (from 0.3 to 0.6) from this effect alone.

It is thus of interest to examine the effect on the energy balance equation of including this effect of ice and thus temperature on albedo, since the occurrence of precipitation as snow or rain is essentially related to the atmospheric temperature.

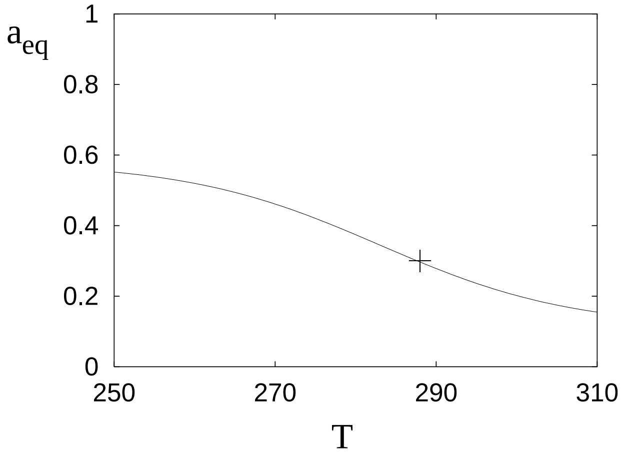

Figure 1.13: A representation of the possible variation of equilibrium surface albedo $a_{\rm eq}(T)$ due to variations in ice cover due to climatic temperature. The function plotted is $a_{\rm eq}(T)$ given by (1.98), with $a_1 = 0.58$, $a_2 = 0.47$, $T^* = 283$ K, $\Delta T = 24$ K,which tends to 0.58 for small T, and equals 0.3 at about $T = 288$ K (the point marked +).

We write (1.61) in the form

$$c\dot{T} = R_i - R_o, \tag{1.96}$$

where $c = \rho_a c_p d$ is the specific heat capacity of the atmosphere, and

$$R_i = \tfrac{1}{4}(1-a)Q, \quad R_o = \sigma\gamma T^4, \tag{1.97}$$

are respectively the incoming short wave radiation and the emitted IR radiation.

The effect of decreasing temperature on the albedo is to increase the extent of land and sea ice, so that a will increase. It is convenient to define a family of equilibrium albedo functions

$$a_{\rm eq}(T) = a_1 - \tfrac{1}{2}a_2\left[1 + \tanh\left(\frac{T - T^*}{\Delta T}\right)\right], \tag{1.98}$$

an example of which is shown in figure 1.13. The epithet ‘equilibrium’ refers to the assumption that the land ice cover is in dynamical equilibrium with the ground surface temperature: more on this below. The effect of the albedo variation on the emitted radiation is shown in figure 1.14: R_o is an increasing function of T, but the sigmoidal nature of R_i can lead, for a range of Q and suitable choices of the albedo function, to the existence of multiple steady states. If this is the case, then the equilibrium response diagram for steady states T in terms of Q is as shown in figure 1.15.

The parameters used in figure 1.14 are chosen to illustrate the multiple intersection of R_i with R_o, but do not correspond to modern climate (for which $T = 288$ K and $R_i = R_o = 235$ W m^{-2}). The reason is that with more appropriate parameters, such as those used in figure 1.15, the two curves become very close, and the range of Q over which multiplicity occurs is very small. Insofar as these paramerisations apply

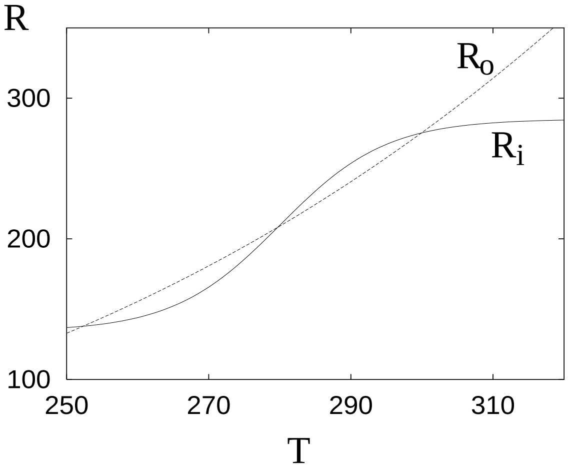

Figure 1.14: Variation of R_i and R_o with temperature T. Parameters used are $a_1 = 0.6$, $a_2 = 0.45$, $\gamma = 0.6$, $T^* = 280$, $\Delta T = 15$.

to the Earth, it does suggest that the current climate is close to a switching point, as seen in figure 1.15, corroborating this explanation for ice age formation.

It is easy to see from figure 1.14 that the upper and lower branches in figure 1.15 are stable, while the middle branch is unstable; this follows because the stability of the equilibria of (1.94) is determined by the slope of $R_i - R_o$ there: if $R_i' < R_o'$, then the equilbrium is stable, and vice versa. Thus if Q varies slowly backwards and forwards beyond Q_- and Q_+, then the temperature will vary up the lower and down the upper branch, with sudden jumps at Q_+ and Q_-. This oscillatory response exhibits *hysteresis*, it is *irreversible*, and it forms the basis for the Milankovitch theory of the ice ages, since the lower branch is associated with widespread glaciation.

1.4.2 The Milankovitch theory

The solar radiation received seasonally by the Earth is not in fact constant. Due to variations in the Earth's orbit, the value of Q varies by about $\pm$ 5% either side of its mean. Nor are these variations periodic. Because the solar system has planets (and moons) other than the Earth, and because also the planets do not act exactly as point masses, the orbit of the Earth is not precisely a Keplerian ellipse. The Earth's axis of rotation precesses, its angle of tilt (from the plane of the ecliptic) oscillates, and the eccentricity of the orbit itself oscillates. All of these astronomical features cause the value of Q to oscillate quasi-periodically, when considered for a particular latitude and a particular season. The reason for focussing on a particular season is because of the seasonal imbalance in snowfall, whence it might be supposed that, for example, it is the summer insolation received at (say) 65°N which is important, since this is likely to control the inception of northern latitude ice sheets via year round retention of snow cover and the consequent operation of the ice-albedo feedback. The importance of a

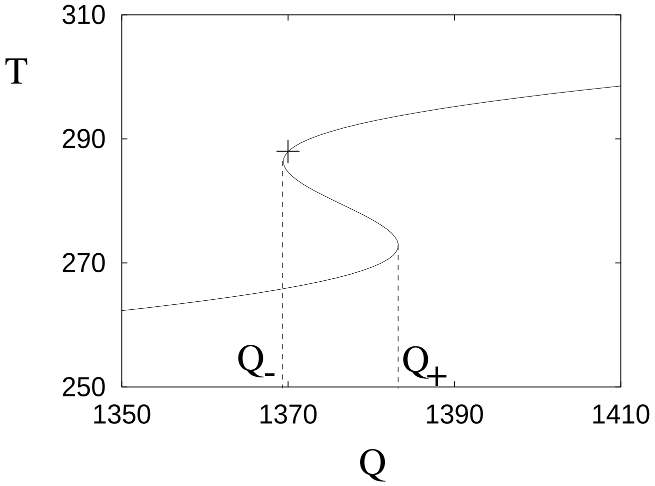

Figure 1.15: Multivalued response curve for T in terms of Q. Parameters used are $a_1 = 0.58$, $a_2 = 0.47$, $\gamma = 0.6175$, $T^* = 283$, $\Delta T = 24$. Also shown is the point (+) corresponding to current climate ($Q = 1370$ W m^{-2}, $T = 288$ K).

particular latitude is due to the fact that the seasonal insolation curves are different at different latitudes, as indeed found by Milankovitch — see figure 1.16. The major periodicities in the signals consist of one of 41,000 years due to oscillations in the tilt axis, and periods of 23,000 and 19,000 years in the precessional variation of the rotation axis. The third component, eccentricity, causes a variation over a period of 100,000 years, though its amplitude is much smaller.

The test of the Milankovitch theory, that variations in climate (and thus ice ages) are associated with the variation in Q, can then be made by computing the Fourier power spectrum of a record of past climatic temperature. Oxygen isotope ratios in deep-sea sediment cores (or in ice cores) provide a proxy measurement of temperature, and figure 1.12 showed just such a record. When a spectral analysis of records of this type is made, it is indeed found that the principal frequencies are (in order of decreasing amplitude) 100 ka (100,000 years), 41, 23 and 19 ka. This seems to serve as dramatic confirmation of the Milankovitch theory. In our simple energy balance model, the concept is enunciated by the hysteretic oscillations exhibited by the system as Q varies.

1.4.3 Nonlinear oscillations

There is currently a consensus that the Milankovitch solar variation indeed acts as pacemaker for the Quaternary ice ages, but it is as well to point out that there is an essential problem with the Milankovitch theory, even if the basis of the concept is valid. The spectral insolation frequencies do match those of the proxy climate record, with one essential discrepancy: the largest climatic signal is the 100 ka period. Ice ages essentially last 90 ka, with an interval of 10 ka between (and since the last ice

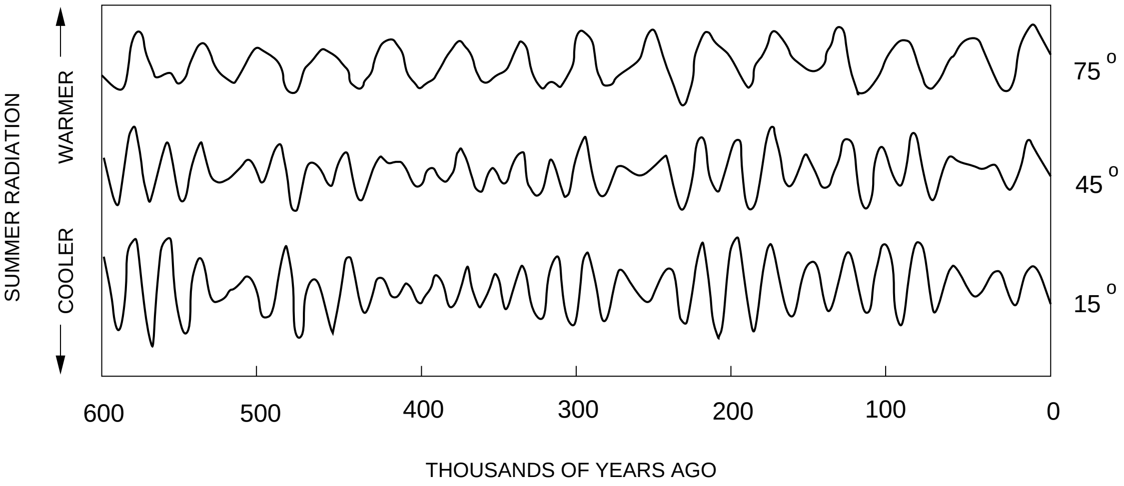

Figure 1.16: Milankovitch radiation curves for 15° N, 45° N, and 75° N. The lower two curves indicate the $\sim$ 22,000 year precession cycle, while the upper one shows more clearly the 41,000 year tilt cycle.

age terminated about 10 ka ago, as the Scottish ice sheet withdrew from the lochs in the Highlands, and the North American Laurentide ice sheet shrank from the Great Lakes, we might be on the verge of starting the next). But the 100 ka astronomical signal is very weak, and it is unrealistic to imagine that the forcing can directly drive the strong response which is observed. What may happen is that the weak 100 ka forcing *resonates* with the climatic system, suggesting that the climate is essentially a (nonlinear) oscillator, with a natural period close to 100 ka, which is tuned by the astronomical forcing. A mathematical paradigm would be the forced Van der Pol oscillator

$$\ddot{x} + \varepsilon(x^2 - 1)\dot{x} + \omega^2 x = f(t), \tag{1.99}$$

where $f(t)$ would represent the astronomical forcing. If ε is small, the oscillator has a natural frequency close to $2\pi/\omega$, and if forced by a frequency close to this, tuning can occur. If the oscillator is non-linear, other exotic effects can occur: subharmonics, chaos; no doubt these effects are present in the forced climate system too.

The simplest kind of model which can behave in an oscillatory manner is the energy balance model (1.96) subjected to an oscillating radiation input which can drive the climate back and forth between the cold and warm branches, presumably representing the glacial and interglacial periods. Two questions then arise; where does the 100,000 year time scale come from, and why is the climatic evolution through an ice age (slow development, rapid termination) so nonlinear?

There are three principal components of the climate system which change over very different time scales. These are the atmosphere, the oceans and the ice sheets. The time scale of response of ice sheets is the longest of these, and is measured by l/u, where l is a horizontal length scale and u is a horizontal velocity scale. Estimates of $l \sim 1000$ km and $u \sim 100$ m y^{-1} suggest a time scale of 10^4 y, which is within range of the value we seek. Since ice sheet extent is directly associated with albedo. it suggests that a first realistic modification of the energy balance model is to allow the ice sheets, and therefore also the albedo, to respond to a changing temperature over the slow ice sheet time scale t_i. A simple model to do this is to write

$$\begin{aligned} c\dot{T} &= \tfrac{1}{4}Q(1-a) - \sigma\gamma T^4, \\ t_i\dot{a} &= a_{\rm eq}(T) - a, \end{aligned} \tag{1.100}$$

and $a_{\rm eq}(T)$ is the equilibrium albedo represented in figure 1.13. Since the thermal response time scale is so rapid (months), we may take the temperature to be the equilibrium temperature,

$$T = T(a,Q) = \left[\frac{Q(1-a)}{4\sigma\gamma}\right]^{1/4}. \tag{1.101}$$

The energy balance model thus reduces to the first order albedo evolution equation

$$t_i\dot{a} = I(a,Q) - a, \tag{1.102}$$

where

$$I(a,Q) = a_{\rm eq}[T(a,Q)]. \tag{1.103}$$

As Q varies backwards and forwards about the critical switching values in figure 1.15, the ice extent (as indicated by a) changes on the slow time scale t_i, aiming to follow the hysteretically switching equlibria. Oscillatory inputs Q do indeed cause oscillations; if t_i is sufficiently small, these are large scale, going from warm branch to cold branch and back, whereas at larger t_i, two different oscillatory climates are possible, a cold one and a warm one (see question 1.14). None of these solutions bears much resemblance to real Quaternary ice ages, for which a more sophisticated physical model is necessary.

1.4.4 Heinrich events

The study of climate is going through some exciting times. The pulse of the ice ages can be seen in figure 1.12, but the signal appears noisy, with numerous irregular jumps. Twenty or thirty years ago, one might have been happy to ascribe these to the influence of different spectral components of the Milankovitch radiation curves on a nonlinear climatic oscillator, together with a vague reference to 'noisy' data. Increasingly in such circumstances, however, one can adopt a different view: what you see is what you get. In other words, sharp fluctuations in apparently noisy data are actually signals of real events. Put another way, noise simply refers to the parts of the signal one does not understand.

It has become clear that there are significant climate components which cause short term variations, and that these events are written in the data which is exhumed from ocean sediment cores and ice cores. Perhaps the most dramatic of these are *Heinrich events.* Sediment cores retrieved from the ocean floor of the North Atlantic reveal, among the common ocean sediments and muds, a series of layers (seven in all have been identified) in which there is a high proportion of lithic fragments. These fragments represent ice rafted debris, and are composed of carbonate mudstones, whose origin has been identified as Hudson Bay. The spacing between the layers is such that the periods between the Heinrich events is 5,000–10,000 years.

What the Heinrich events are telling us is that every 10,000 years or so (more or less periodically) during the last ice age, there were episodes of dramatically increased iceberg production, and that the ice in these icebergs originated from the Hudson Bay underlying the central part of the Laurentide ice sheet. Ice from this region drained through an ice stream some 200 km wide which flowed along the Hudson Strait and into the Labrador sea west of Greenland.

The generally accepted cause of these events is also the most obvious, but equally the most exciting. The time scale of 10,000 years is that associated with the growth of ice sheets (for example, by accumulation of 0.2 m y^{-1} and depth of 2000 m), and so the suggestion is that Heinrich events occur through a periodic surging of the ice in the Hudson Strait, which then draws down the Hudson Bay ice dome. This would sound like a capricious explanation, were it not for the fact that many glaciers are known to surge in a similar fashion; we shall discuss the mechanism for surging in chapter 4.

Another feature of Heinrich events is that they appear to be followed by sudden dramatic warmings of the Earth's climate, which occur several hundred years after the Heinrich event. Dating of these is difficult, because dating of ice cores and also of sediment cores relies on an assumption of accumulation or sedimentation rates, so precise association of timings in different such cores is risky.

How would Heinrich events affect climate? There are two obvious ways. A sudden change in an ice sheet elevation might be expected to alter storm tracks and precipitation patterns. Perhaps more importantly, the blanketing of the North Atlantic with icebergs is likely to affect oceanic circulation. Just like the atmosphere, the ocean circulation is driven by horizontal buoyancy induced by the difference between equatorial and polar heating rates. This large scale flow is called the global thermohaline circulation, and its presence in the North Atlantic is the cause of the anomalously temperate climate of Northern Europe, because of the poleward energy flux. If this circulation is disrupted, there is liable to be an immediate effect on climate.

If the North Atlantic is covered by ice, one immediate effect is a surface cooling, because of the increased albedo. This is liable to cause an increase in the thermohaline circulation, but would not cause atmospheric warming until the sea ice melted. On the other hand, the melting itself releases fresh water, which is buoyant in a saline ocean, suggesting a shutdown of ocean circulation. As we discuss further below, this can lead, following a delay, to a massive restart of ocean circulation and thus sudden warming.

1.4.5 Dansgaard–Oeschger events

There are other rapid changes in the climate which are seen during the last ice age. Figure 1.17 shows a segment of oxygen isotope measurements (a proxy for surface temperature) from the GRIP ice core on Greenland. Time marches from right to left on this diagram. There are numerous sudden rises in temperature that can be seen, followed by a more gentle sinking of temperature. These sharp rises are called *Dansgaard-Oeschger events.* Between 30,000 and 45,000 years B. P., for example there are seven of these events, thus, like Heinrich events, they occur at regular intervals, with a typical repetition period being thought to be about 1,500 years.

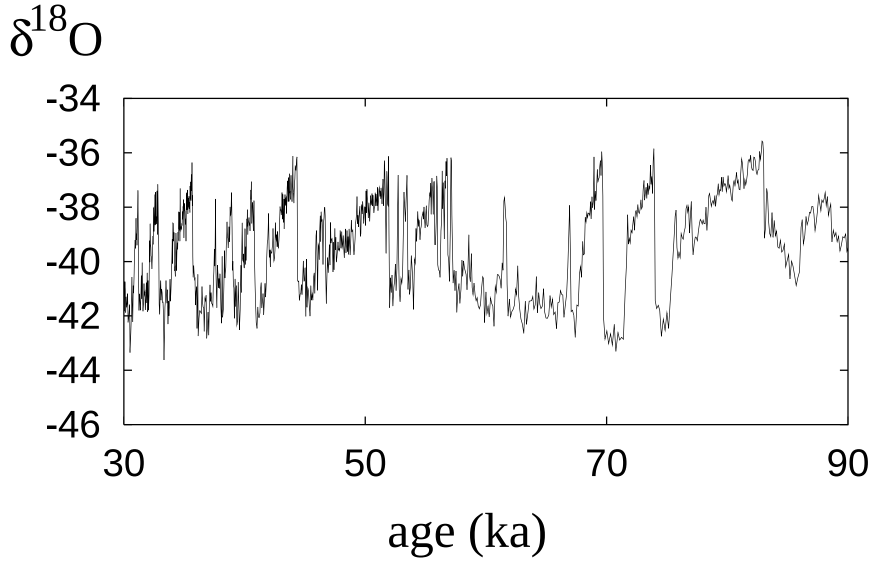

Figure 1.17: Oxygen isotope ratio (δ^{18}O) measurements from the GRIP ice core on Greenland, as a function of age in ka (1 ka = 1000 years). This is a proxy for surface temperature (with four units corresponding to about 10 K on the vertical axis). The data represents averages from segments of 55 cm length, and the age scale is determined from a model of ice burial rate. Near the surface, the separate measurements are yearly, but the compression of ice with burial causes the ice segments to encompass longer and longer time periods. At the age of 90,000 years ago (at a depth of some 2685 m), each segment is a time average of some 120 years. Thus the data is increasingly sparsely resolved further into the past.

Let us examine one of these events in greater detail, that of the D–O event between 44,000 and 45,000 years B. P. In the GRIP core, this ice lies between 2,316 and 2,330 metres depth. A higher resolution data set is that of Sigfus Johnsen, and is shown in figure 1.18. This shows that the climatic temperature changes abruptly, over a time scale of about a century. Other such inspections show that the transitions can be even shorter.

What is the cause of these warming events? Why are they so rapid, and why do they have a regular period of some 1,500 years? The idea here is that the climate in the northern hemisphere is essentially controlled by the oceanic conveyor circulation, and so the change in climate occurs because of a sudden disruption to this. Model studies have shown that an injection of a massive pulse of fresh water into the North Atlantic can cause just such a disruption.

The mechanism is, however, counter-intuitive. A warm climate is associated with

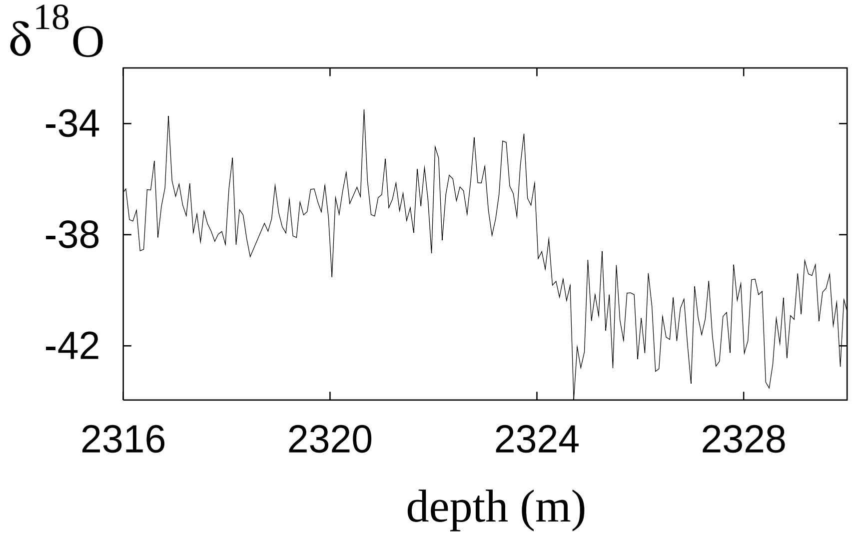

Figure 1.18: GRIP core data between 2,316 m and 2,330 m. The sharp jump near 2,324 m occurs over a range of about 1.3 metres, corresponding to a time interval of some 90 years.

a vigorous circulation, and a cold one with the circulation off, but a freshwater pulse has the initial effect, being buoyant, of switching the (relatively weak) circulation off. In an already cool climate, this has no dramatic effect. However, the effect is temporary, and a situation with no circulation is unstable. Thus when convection begins again, it does so dramatically, with deep water formation occurring further north (as it does in interglacial times), causing a sudden shift to a warmer climate. The same model studies have shown that even larger meltwater pulses, such as would be produced by melting of massive iceberg production, can produce the warming following Heinrich events.

If freshwater pulses are the cause of the sudden climate shifts, what is their origin? For Heinrich events, the ice rafted debris gives the clue; for Dansgaard–Oeschger events, there is apparently no such clue. However, it is pertinent to note that these events are associated with the presence of large ice sheets. Since we seek an origin of melt water, the most obvious (and really, the only) candidate for the source of the pulses is that they come from meltwater from the ice, and one way in which meltwater drainage is known to occur episodically is in the large subglacial floods known as jökulhlaups. As with surges, these are well documented from beneath glaciers, but not (yet) from beneath modern day ice sheets. Nevertheless, as a hypothesis it seems sensible to suggest that Dansgaard–Oeschger events arise as a consequence of semi-regular Laurentide jökulhlaups which occur with a rough periodicity. It remains to be seen whether such floods are dynamically possible, and whether they could produce the necessary fresh water at the required frequency to do the job.

1.4.6 The 8,200 year cooling event

One climatic event which is thought to have been caused by a sub-glacial flood is a sudden cooling event dated to 8,200 years B. P. This is shown in context in figure 1.19. which also shows the termination of the ice age after the Younger Dryas readvance of the ice sheets between 13,000 and 11,600 years B. P. Two Heinrich events precede the two warmings at about 15,000 and 12,000 years B. P. Following the Younger Dryas, there is a gradual return to an interglacial climate by 9,500 years B. P., and the onset of the current (Holocene) period. A cursory glance might suggest that the 8,200 year dip is just a noisy outlier, but this is not the case. It represents a genuine climatic cooling of some 4 °K.

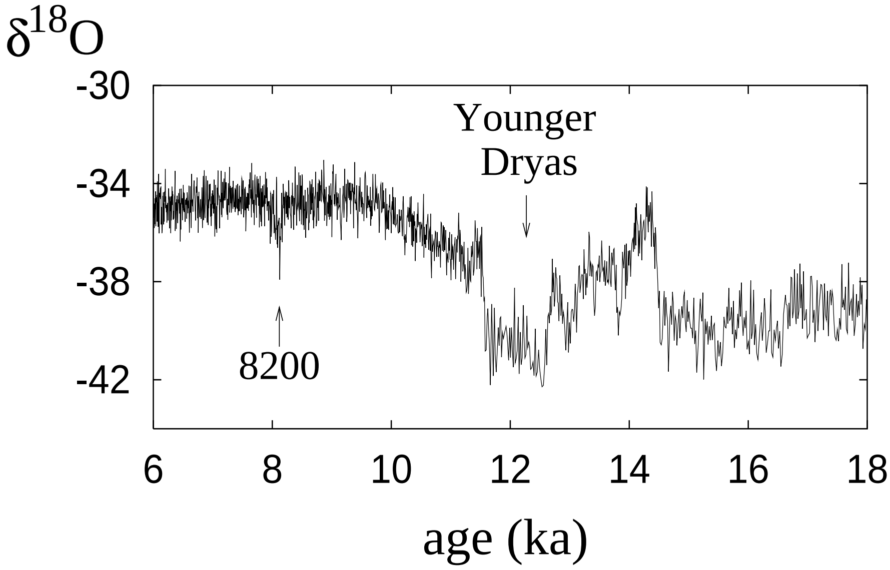

Figure 1.19: Oxygen isotope data from the GRIP core at the transition to the Holocene interglacial.

Figure 1.19 shows a high resolution record of this event. Inspection of the coarse (55 cm samples) GRIP data shows that the 8,200 year event is actually (on the age scale used) at 8,126 years B. P., and occurs in a single 55 cm segment at a depth of 1334 metres; blink, and you miss it. At this depth, six metres of ice (1,331–1,337) is considered to represent 65 years of accumulation. Plotting the data using Johnsen's higher resolution data set over a more restricted range, we can see (figure 1.20) various features.

One is that the event occupies three data points from an ice depth between 1,334 and 1,335 metres depth. These three samples are 27.5 cm long, as are the two immediately above and below. This suggests that the 'duration' of the event is between 82 and 137 cm, which corresponds to a period of between 9 and 15 years. This is incredibly fast.

The other thing to notice from figure 1.20 is that there are a good number of other large spikes and oscillations. Since, more or less, each data point represents a

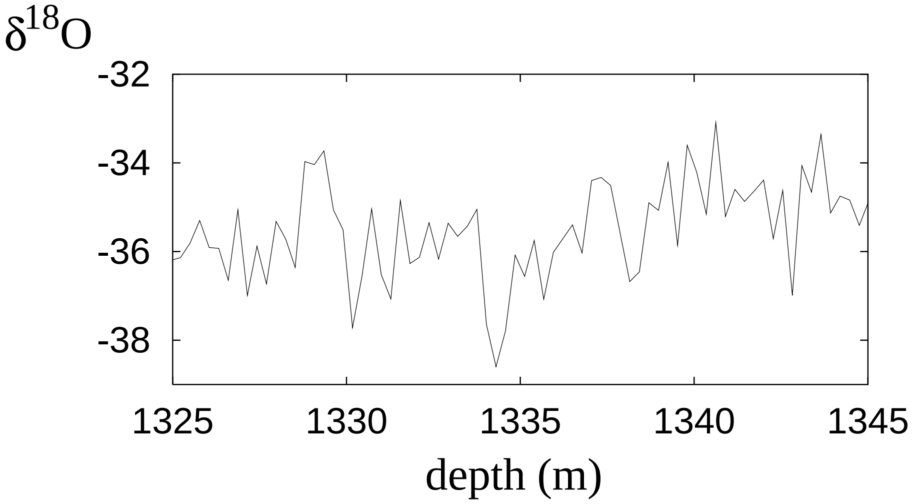

Figure 1.20: High resolution data set from 20 metres of ice near the 8,200 year event.

three year average, these are not seasonal fluctuations. Do they represent real events, or simply the natural fluctuation of the climate from year to year? If one looks at a slightly larger slice of the time series, from 1305 m to 1365 m, it is apparent (see figure 1.21) that these short term fluctuations sit on top of a broader cooling trend from about 1340 to 1315 m, with rapid decrease in the first 5 m ($\sim$ 54 years), and slower recovery over the following 20 m ($\sim$ 217 years). It is perhaps easier to imagine that this slower average trend represents the underlying event.

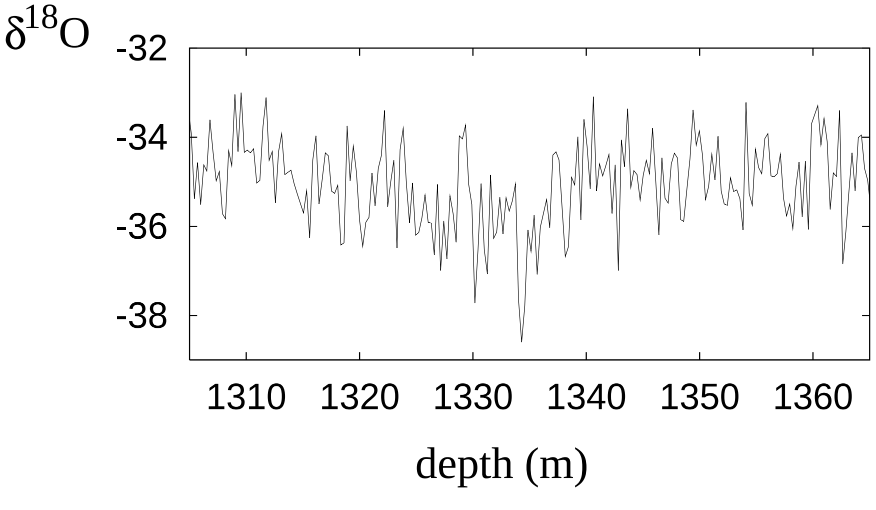

Figure 1.21: The cooling trend of the 8,200 year event.

The explanation which is currently thought to apply to this event is that it is caused by a sub-glacial jökulhlaup which drains the massive proglacial Lake Agassiz into the Hudson Strait, whence it pours into the Labrador Sea and the North Atlantic.

As the remnant of the Laurentide ice sheet dwindles, it builds up a massive proglacial lake on its southern margin. The topography is such that this lake is prevented from outflow to the south, and at some point it drains catastrophically, either over or more probably under the ice sheet to the north. The resulting fresh water efflux to the North Atlantic causes the cooling event.

One might wonder, if glacial meltwater pulses produce a warming, why would an interglacial one produce a cooling? The putative answer to this lies in our idea of what a meltwater pulse will actually do. In an interglacial climate, the ocean circulation is strong, and meltwater dims it temporarily: a cooling. In a glacial climate, the circulation is weaker, and deep water formation occurs further south, say near Iceland, than it does currently. Then a meltwater pulse may shut down the circulation entirely, which would indeed cause further cooling, but the resultant overshoot when circulation resumes causes the warming. Since Dansgaard–Oeschger events occur at the end of cooling cycles, the initial cooling is swamped by the trend. It is interesting to note that the D–O warming events in figure 1.19 are initiated at 14,500 B. P. and 11,600 B. P., the interval between these being 2,900 years. The interval between the Younger Dryas and the 8,200 event is about 3,500 years. If the D–O events are due to subglacial floods, then possibly the 8,200 event is simply the last of these. It is then tempting to look further on for similar, smaller events. There is one at 5,930 B. P., for example, and another at 5,770 B. P.; these are about another 2,400 years further on. It is a natural consequence of the hypothesis that jökulhlaups occurred from below the Laurentide ice sheet to suppose that they will occur also from beneath Greenland and Antarctica, and that this may continue to the present day. It has been suggested, for instance, that the cool period in Europe between 1550 A. D. and 1900 was due to a similar upset of the oceanic circulation.

1.4.7 North Atlantic salt oscillator

Deeply embroiled in this whole saga of Quaternary climate and the ice ages is the rôle of the North Atlantic ocean circulation. For the descriptions we have given of Heinrich events and Dansgaard–Oeschger events to work, the ocean needs to be able to circulate in different ways. That this is indeed the case has been found in a number of model studies, and the resultant flip-flop circulation is sometimes known as the ‘bipolar seesaw’. In its original form, the idea is due to Henry Stommel, and can be described with a simple ‘box’ model, as illustrated in figure 1.22.

In this model, we parameterise the thermohaline circulation in the North Atlantic by considering it to be partitioned between two compartments, an equatorial and a polar one. We label the temperature T, salinity (mass fraction of salt) S, density ρ and volume V of each box by a suffix ‘e’ or ‘p’, and we write conservation laws of mass, energy, solute and state for each box. Transports in and out of each box are considered to be a freshwater flux F_p to the polar cell, an evaporative flux F_e from the equatorial cell, and a convective flux q due to buoyancy difference from equatorial to polar cell. (The reverse flux is then $q + F_p$ in order to allow conservation of the sizes of both cells.)

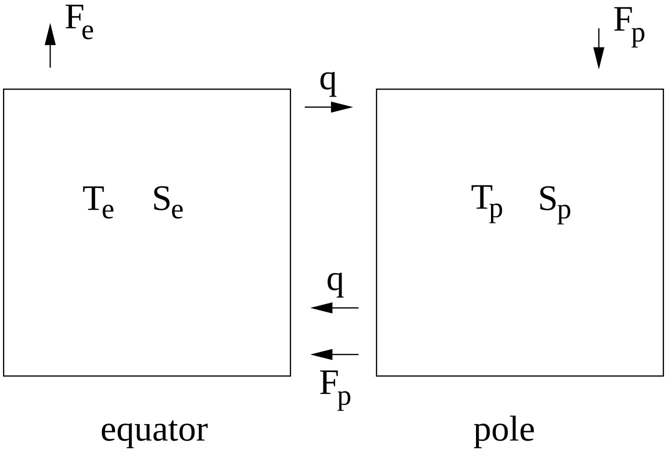

Figure 1.22: Stommel's box model of the North Atlantic circulation.

Suitable equations to describe the convective flow are then

$$\begin{aligned}
\frac{d}{dt}\left(\rho_e c V_e T_e\right) &= A_e H_e + \rho_p c T_p \left(F_p + |q|\right) - \rho_e c T_e |q|, \\
\frac{d}{dt}\left(\rho_p c V_p T_p\right) &= A_p H_p - \rho_p c T_p \left(F_p + |q|\right) + \rho_e c T_e |q|, \\
\frac{d}{dt}\left(\rho_e V_e S_e\right) &= -\rho_e |q| S_e + \rho_p S_p \left(F_p + |q|\right), \\
\frac{d}{dt}\left(\rho_p V_p S_p\right) &= \rho_e |q| S_e - \rho_p S_p \left(F_p + |q|\right), \\
\frac{d}{dt}\left[\rho_e V_e \left(1 - S_e\right)\right] &= \rho_p^0 F_p - \rho_e F_e, \\
\frac{d}{dt}\left[\rho_p V_p \left(1 - S_p\right)\right] &= 0, \\
\rho_e &= \rho_0 \left[1 - \alpha\left(T_e - T_0\right) + \beta\left(S_e - S_0\right)\right], \\
\rho_p &= \rho_0 \left[1 - \alpha\left(T_p - T_0\right) + \beta\left(S_p - S_0\right)\right], \\
q &= k\left(\rho_e - \rho_p\right). \qquad (1.104)
\end{aligned}$$

The terms in these equations are fairly self-explanatory. α and β are coefficients of thermal and saline expansion, ρ_p^0 is the freshwater density, A_e and A_p are the equatorial amd polar ocean surface areas, and for simplicity we take $A_p = A_e = A$. The heating terms H represent the heat flux to the deep ocean from the surface waters. The sea surface temperature is determined by a radiative balance, which determines equatorial and polar surface temperatures T_e^0 and T_p^0, say. We then suppose that heat transfer to the deep ocean can be parameterised by a suitable heat transfer coefficient h_T, thus we put

$$H_e = h_T(T_e^0 - T_e), \quad H_p = h_T(T_p^0 - T_p). \qquad (1.105)$$

If we add the two energy equations, we have

$$\frac{d}{dt}\left(\rho_e c V_e T_e + \rho_p c V_p T_p\right) = h_T A (T_e^0 + T_p^0 - T_e - T_p). \tag{1.106}$$

We use the facts that $\rho_{e,p} \approx \rho_0$ and we will suppose that V_e and V_p, both approximately constant, are also approximately equal, $V_{e,p} \approx V_0$. It then follows from (1.106) that, after an initial transient,

$$T_e + T_p \approx T_e^0 + T_p^0, \tag{1.107}$$

and we suppose this generally to be the case. We define T_0 by

$$\rho_e V_e T_e + \rho_p V_p T_p = 2\rho_0 V_0 T_0, \tag{1.108}$$

where

$$T_0 = \tfrac{1}{2}\left(T_e^0 + T_p^0\right), \tag{1.109}$$

and we then define the temperature excess T via

$$T_e = T_0 + T, \quad T_p = T_0 - T. \tag{1.110}$$

If we now make the assumptions that $\rho_e \approx \rho_0$, $V_e \approx V_0$ in $(1.104)_1$, then we have the approximate equation for T:

$$\rho_0 c V_0 \dot{T} = \tfrac{1}{2}\left[h_T A\left(T_e^0 - T_p^0\right) + \rho_0 c F_p\left(T_e^0 + T_p^0\right)\right] - (h_T A + 2\rho_0 c|q|)T. \tag{1.111}$$

In a similar way, we have

$$\rho_e V_e S_e + \rho_p V_p S_p = 2\rho_0 V_0 S_0, \tag{1.112}$$

where S_0 is constant, and we define

$$S_e = S_0 + S, \quad S_p = S_0 - S. \tag{1.113}$$

With the same Boussinesq type assumption, that $\rho_e \approx \rho_0$ and $V_e \approx V_0$, we obtain

$$V_0 \dot{S} = F_p S_0 - (F_p + 2|q|)S. \tag{1.114}$$

The equations (1.111) and (1.114) are essentially Stommel's box model. Their validity relies on the use of the other equations to show that it is indeed realistic to take ρ and V as constant, though these assumptions appear fairly reasonable ones. Note that with the definitions of the variables, we have

$$q = 2k\rho_0(-\alpha T + \beta S). \tag{1.115}$$

To parameterise the heat transfer coefficient h_T, we use the ideas of Reynolds averaging for turbulent flow (see appendix B). This suggests choosing

$$h_T = \frac{\varepsilon_T \rho_0 c q_0}{A}, \tag{1.116}$$

Parameter	Value
ρ_0	10^3 kg m^{-3}
c	4.2×10^3 J kg^{-1} K^{-1}
V_0	1.6×10^{17} m^3
A	0.4×10^{14} m^2
ε_T	0.01
T_e^0	300 K
T_p^0	270 K
T_0	285 K
F_p	10^5 m^3 s^{-1}
q_0	1.6×10^7 m^3 s^{-1}
α	1.8×10^{-4} K^{-1}
β	0.8
S_0	0.035

Table 1.1: Typical parameter values.

where the number ε_T is typically chosen to be in the range 0.001—0.01, and q_0 is a scale for q.

We non-dimensionalise the box model by writing

$$T = \Delta T\,\theta, \quad S = \Delta S\, s, \quad t \sim t_0, \quad q \sim q_0, \tag{1.117}$$

where we choose

$$\Delta T = \tfrac{1}{4}\left[\varepsilon_T(T_e^0 - T_p^0) + \frac{F_p}{q_0}\left(T_e^0 + T_p^0\right)\right], \quad \Delta S = \frac{F_p S_0}{2q_0}, \quad t_0 = \frac{V_0}{2q_0}. \tag{1.118}$$

Using the values in table 1.1, we find $\Delta T \approx 1$ K, $\Delta S \approx 1.1 \times 10^{-4}$, $t_0 \approx 150$ y. We use a value of q_0 as observed, rather than k, which we would in any case choose in order that q was the right size, some 16 Sv (Sverdrups: 1 Sv = 10^6 m^3 s^{-1}).

The observed surface temperature variation is of order 30 K, and the observed surface salinity variation is of order 30×10^{-4}. However, these values represent the concentrative effect of surface evaporation and heating; at depth (as is more relevant) the variations are much smaller, of order 2 K for temperature and 4×10^{-4} for salinity at 1000 m depth. The time scale is comparable to the time scales over which Dansgaard–Oeschger events occur. These features suggest that this simple model has the ring of truth.

We can write the model in dimensionless form as

$$\begin{aligned}\dot{\theta} &= 1 - (\mu + |q|)\theta,\\ \dot{s} &= 1 - (\varepsilon + |q|)s,\\ q &= \kappa(-\theta + Rs),\end{aligned} \tag{1.119}$$

where the parameters are given by

$$\begin{aligned} \varepsilon &= \frac{F_p}{2q_0}, \quad \mu = \frac{2\Delta T}{T_e^0 - T_p^0}, \\ R &= \frac{\alpha \Delta T}{\beta \Delta S}, \quad \kappa = \frac{2k\rho_0 \alpha \Delta T}{q_0}. \end{aligned} \tag{1.120}$$

Typical values of these are, from table 1.1,

$$\varepsilon \sim 0.003, \quad \mu \sim 0.07, \quad R \sim 2, \tag{1.121}$$

and we can assume without loss of generality that $\kappa = 1$, which fixes the value of q_0 (given k). Both ε and μ are small, and we will take advantage of this below.

It is straightforward to analyse (1.119) in the phase plane. Figure 1.23 shows the steady states of q as a function of R when $\mu = 0.07$, $\varepsilon = 0.003$. Neglecting ε and taking μ to be small, we deduce that the steady states are given by

$$\begin{aligned} q &\approx (R-1)^{1/2}, \quad R > 1, \\ q &\approx -(1-R)^{1/2}, \quad R < 1, \end{aligned} \tag{1.122}$$

if $q = O(1)$, and

$$q \approx \pm \frac{\mu R}{1-R}, \quad R < 1, \tag{1.123}$$

if $q = O(\mu)$. As we might expect, the upper and lower branches are stable, and the middle one is unstable. The upper branch corresponds to present climate, with a northwards circulation at the surface. The stable lower branch corresponds to a reversed haline circulation (saline buoyancy dominates thermal buoyancy because $R < 1$).

Stommel's box model is not an oscillator as such, but it does point out the possibility of multiple convective states of the North Atlantic, and this feature has been found to be robust in other models. What appears to distinguish more realistic models from the Stommel box model is that they allow North Atlantic deep water formation to occur at different latitudes. Thus rather than simply switching from a northerly flow to a southerly one, adjustments can occur between strong northerly flows with deep water formation in the Norwegian sea, and weaker flows with deep water formation further south. It seems that these switches are instrumental in causing the rapid climatic changes during ice ages.

1.5 Snowball Earth

The story of climate on the Earth becomes perhaps more fascinating if we shift our gaze from the relatively recent geologic past to that of more ancient times. There is evidence of glaciation on Earth throughout geologic time, and on all continents. Since the continents move, through the process of plate tectonics, on time scales of hundreds of millions of years, and since their positions and configuration are instrumental in

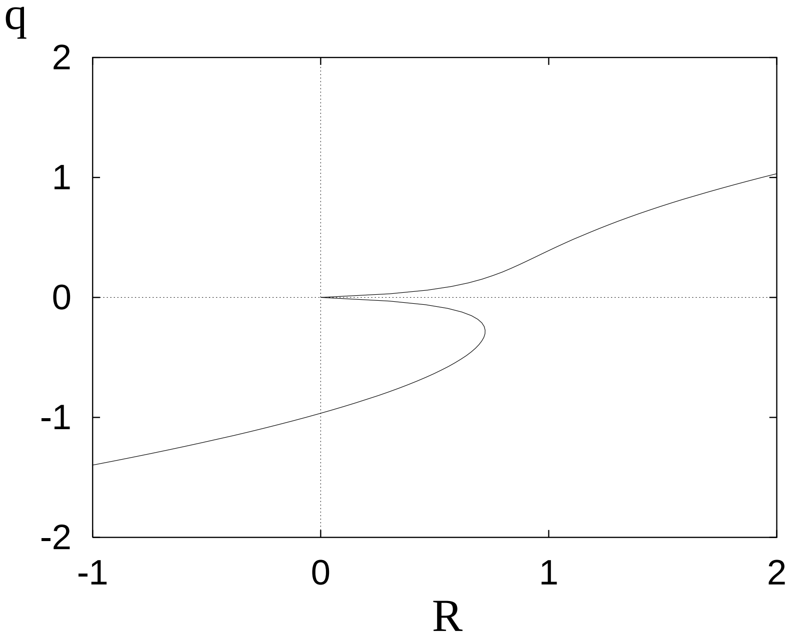

Figure 1.23: Steady states of (1.122) as a function of R.

determining ocean circulations and carbon budget (as described below), it seems that plate tectonics is implicated in the long term control of climate.

Recently, one of these periods of glaciation has been at the centre of a scientific controversy concerning what has been picturesquely termed the 'snowball Earth'. In the Neoproterozoic era, some 600 to 500 million years ago, there was a sequence of glacial episodes. At that time, the Earth's land masses were assembled into a supercontinent called Rodinia which broke up in a similar way to that in which Gondwanaland fragmented some 200 million years ago. The glaciation of Rodinia would not in itself be surprising, except for the fact that it seems that the supercontinent was located near the equator. It is not impossible for glaciers to exist in equatorial regions at high altitudes (there is an ice cap today on Mount Kilimanjaro in Tanzania), but the suggestion for the Neoproterozoic is that there were widespread ice sheets, and that in fact the land masses were covered with ice. If we suppose also that the oceans were largely ice covered, we see how the concept of a snowball Earth arises.

Although the concept of an ice-covered Earth is entirely consistent with a simple energy balance model, it is less easy to explain in detail. At that period, the Sun was 6% fainter than today. Model simulations appear able to produce equatorial glaciation providing there is very little CO_2 in the atmosphere, but it is not obvious how to produce such low levels. Nor is it easy to see how to terminate a snowball glaciation.

An interesting idea to explain this latter conundrum is the widespread occurrence of cap carbonate rocks overlying the glacigenic tillites formed from the subglacial basal sediments. The idea is that with widespread glaciation and very low temperatures, there would be no water vapour in the atmosphere. Subglacial volcanic eruptions would continue to produce CO_2 however, and with no clouds or water vapour to dissolve it (and rain it out), it would simply build up in the atmosphere. Eventually, the consequent greenhouse effect would cause a rise in temperature, followed by massive deglaciation, moistening of the atmosphere, and thus widespread acid rain. The resulting weathering processes produce the cap carbonate rocks which are seen overlying the glacial tillites.

If this end part of the story is enticing, it is not easy to initiate an equatorial glaciation. One possible way is to allow increased weathering of an equatorial supercontinent (because of tropical climate) which causes reduction of atmospheric CO_2; this then causes the cooling which initiates the glaciation. Once under way, the ice-albedo feedback effect leads to the snowball. Evidently, the whole account relies strongly on the interaction of the carbon cycle with climate. This idea is attractive, because it is widely thought that the onset of the current ice age climate originated with the collision of India and Asia some forty million years ago. The resulting (and continuing) uplift of the Himalayas resulted in massively increased weathering rates, and therefore reduction of atmospheric CO_2 and consequent planetary cooling. It is thought that the growth of the Antarctic Ice Sheet some 15 million years ago is a consequence of this cooling.

1.5.1 The carbon cycle

Just as living organisms have a variety of cycles (sleep-wake cycle, menstrual cycle, cell renewal and so on), so the Earth has a number of cycles. Water, rock, topography all go through cycles, which we will describe later in this book. There is also a carbon cycle, which we now describe, which is central to plant and animal life, and is also central to the long term control of the Earth's temperature. We have only to look at what has happened on Mars and Venus to see how delicate the control of climate is.

Carbon dioxide is produced as a by-product of volcanism. When mantle rocks melt, some CO_2 is dissolved in the melt, and depressurisation of the ascending magma causes exsolution. This eruptive production adds about 3×10^{11} kg y^{-1} to the atmosphere. On the Earth, water in the atmosphere dissolves the CO_2, forming a weak carbonic acid, and thus when rain falls, it slowly dissolves the silicate rocks of the continental crust. This process is called weathering. One typical reaction describing this dissolution is

$$\mathrm{CaSiO_3 + 2CO_2 + H_2O \longrightarrow Ca^{2+} + 2HCO_3^- + SiO_2} : \tag{1.124}$$

water dissolves calcium silicate (wollastonite) in the presence of carbon dioxide to form calcium ions, bicarbonate ions and silica. A similar reaction produces magnesium ions.

The ionic species thus produced run off in streams and rivers to the oceans, where the further reaction

$$\mathrm{Ca^{2+} + 2HCO_3^- \longrightarrow CaCO_3 + CO_2 + H_2O} \tag{1.125}$$

creates carbonate sediments. These sink to the ocean floor where they are eventually subducted back into the Earth's mantle. Overall, the pair of reactions (1.124) and

(1.125) can be summed to represent

$$\mathrm{CaSiO_3} + \mathrm{CO_2} \longrightarrow \mathrm{CaCO_3} + \mathrm{SiO_2}. \tag{1.126}$$

A very simple model to describe the evolution of the atmospheric CO_2 concentration is then

$$\dot{m}_{\mathrm{CO_2}} = -A_L W + v_{\mathrm{CO_2}}, \tag{1.127}$$

where $m_{\mathrm{CO_2}}$ is the mass of CO_2 in the atmosphere, A_L is the available land surface for weathering, W is the rate of weathering, and $v_{\mathrm{CO_2}}$ is the eruptive production rate of CO_2. It is common practice in discussing CO_2 levels to measure the amount of CO_2 as a pressure, i. e., in bars. The conversion is done by defining

$$p_{\mathrm{CO_2}} = \frac{m_{\mathrm{CO_2}} g}{A}, \tag{1.128}$$

where g is gravity and A is total planetary surface area. The current atmospheric carbon mass is 750 Gt (gigatonnes, 10^{12} kg). Multiplying by the ratio 44/12 of the molecular weights of carbon dioxide and carbon yields the current value of $m_{\mathrm{CO_2}} \approx 2.75 \times 10^{15}$ kg. Using $g = 9.81$ m s^{-2} and $A = 5.1 \times 10^{14}$ m^2, this converts to a value of $p_{\mathrm{CO_2}} = p_0 = 0.53 \times 10^{-3}$ bars, or 53 Pa. .

Weathering rate

In general we may suppose that $W = W(p_{\mathrm{CO_2}}, T, r)$, where T is temperature and r is runoff rate of water to the oceans. This dependence encapsulates the reaction rate of (1.126), and the rate of product removal by runoff. Weathering rates have been measured and range from 0.25×10^{-3} kg m^{-2} y^{-1} in arid regions to 16×10^{-3} kg m^{-2} y^{-1} in the tropics. If we suppose that (1.127) applies in equilibrium, then the consequent current average value would be $W_0 \approx 2.2 \times 10^{-3}$ kg m^{-2} y^{-1}, which appears reasonable. This uses values of $A_L = 1.5 \times 10^{14}$ m^2 and volcanic production rate $v_{\mathrm{CO_2}} = 3.3 \times 10^{11}$ kg y^{-1}.[5]

One relation which has been used to represent weathering data is

$$W = W_0 \left(\frac{p_{\mathrm{CO_2}}}{p_0}\right)^{\mu} \exp\left[\frac{T - T_0}{\Delta T_c}\right], \tag{1.129}$$

where $\mu = 0.3$, and the subscript zero represents present day values: thus $T_0 \approx 288K$, as well as the values of p_0 and W_0 given above. The current value of the Earth's runoff is $r_0 \approx 4 \times 10^{13}$ m^3 y^{-1}, and in general runoff will depend on temperature (by equating runoff to precipitation to evaporation). This dependence is subsumed into the exponential in (1.129). In general, $\partial W/\partial p_{\mathrm{CO_2}} > 0$, so that with constant production rate, CO_2 will reach a stable steady state. An inference would be that dramatic variations of climate and CO_2 levels in the past have been due to varying degrees of volcanism or precipitation on altered continental configurations, associated with long time scale plate tectonic processes.

---

[5] The current *net* annual addition of CO_2 to the atmosphere because of fossil fuel consumption and deforestation is 3.5 Gt carbon, or 1.3×10^{13} kg y^{-1}; this is forty times larger than the volcanic production rate. (The actual rate of addition is more than twice as large again, but is compensated by net absorption by the oceans and in photosynthesis.)

Energy balance

In seeking to describe how climate may depend on the carbon cycle, we use an energy balance model. Thus, we combine the ice sheet/energy balance model (1.100) with (1.127), to find the coupled system for T, a and p_{CO_2}:

$$\begin{aligned} c\dot{T} &= \tfrac{1}{4}Q(1-a) - \sigma\gamma T^4, \\ t_i\dot{a} &= a_{\mathrm{eq}}(T) - a, \\ \frac{A}{g}\dot{p}_{\mathrm{CO}_2} &= -A_L W + v_{\mathrm{CO}_2}. \end{aligned} \tag{1.130}$$

We take $a_{\mathrm{eq}}(T)$ to be given by (1.98), and W to be given by (1.129).

We model the climatic effect of the greenhouse gases CO_2 and $\mathrm{H}_2\mathrm{O}$ by supposing that γ depends on p_{CO_2}:

$$\gamma = \gamma_0 - \gamma_1 p_{\mathrm{CO}_2}; \tag{1.131}$$

the value $\gamma_0 < 1$ represents the $\mathrm{H}_2\mathrm{O}$ dependence, while the small corrective coefficient γ_1 represents the CO_2 dependence.

We have already seen that the response time of T is rapid, about a month, whereas the time scale for albedo adjustment is slower, with the time scale of growth of continental ice sheets being of order 10^4 years. An estimate for the time scale of adjustment of the atmospheric CO_2, based on this model, is

$$t_c = \frac{Ap_0}{gv_{\mathrm{CO}_2}}. \tag{1.132}$$

Using values $A = 5.1 \times 10^{14}$ y, $p_0 = 53$ Pa, $g = 9.81$ m s^{-2} and $v_{\mathrm{CO}_2} = 3.3 \times 10^{11}$ kg y^{-1}, this is $t_c \sim 0.8 \times 10^4$ y, comparable to the ice sheet growth time.

Although (1.130) is a third order system, it is clear that T relaxes rapidly to a well-defined 'slow manifold'

$$T \approx T(a, p_{\mathrm{CO}_2}) = \frac{Q(1-a)}{4\sigma\gamma(p_{\mathrm{CO}_2})}, \tag{1.133}$$

on which the dynamics is governed by the slower a and p equations. The nullclines in the (a,p) phase plane are shown in figure 1.24. The a nullcline is multivalued for the same reason that figure 1.15 indicates multiplicity, since both graphs are described by the same equations, the only difference being that p_{CO_2} (and thus γ) is used rather than Q. The horizontal axis of figure 1.15 could equally be taken to be Q/γ and thus (for fixed Q) p_{CO_2}.

The analysis of this model is indicated in question 1.16. The solutions depend on the two critical dimensionless parameters

$$w = \frac{A_L W_0}{v_{\mathrm{CO}_2}}, \quad \delta = \frac{v_{\mathrm{CO}_2} g t_i}{Ap_0}, \tag{1.134}$$

which are measures of weathering rate and volcanic production. These can vary depending on current tectonic style. The three indicated intersection points in figure

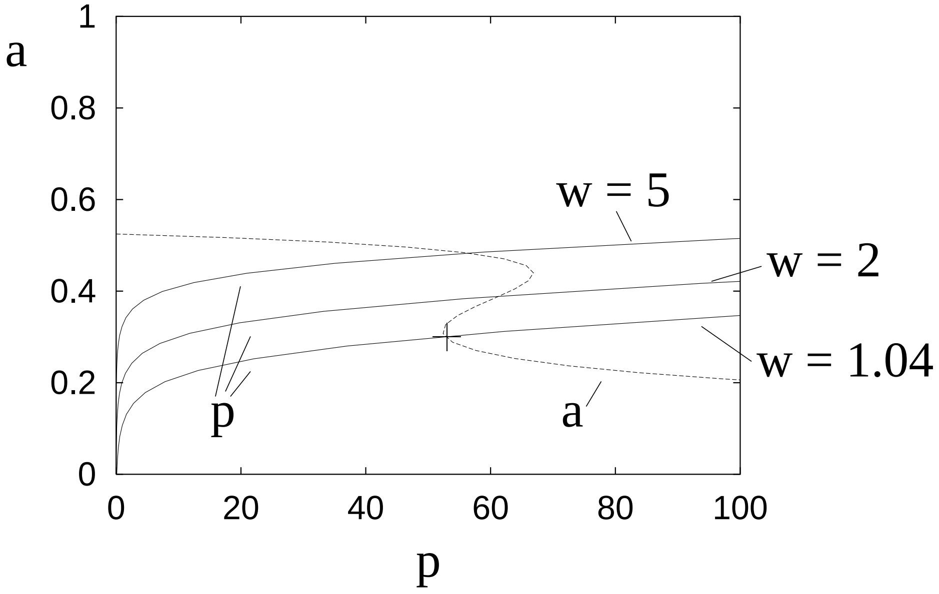

Figure 1.24: a and p nullclines for (1.130) assuming that T has rapidly equilibrated to $T(a, p_{\mathrm{CO}_2})$. The three curves occur for the values of the weathering coefficient $w = 1.04$, 2 and 5, corresponding to present climate (the point marked +), oscillatory ice ages and snowball Earth. The parameters for a_{eq} are the same as those in figure 1.13, and other values used are $p_0 = 53$ Pa, $\mu = 0.3$, $\Delta T_c = 13$ K, $T_0 = 288$ K, $\sigma = 5.67 \times 10^{-8}$ W m^{-2} K^{-4}, $Q = 1370$ W m^{-2}, $\gamma_0 = 0.64$, $\gamma_1 = 4.25 \times 10^{-4}$ Pa^{-1}.

1.24 correspond to steady states at low (current), intermediate and high weathering rates (relative to volcanic output). The solution on the upper branch indicates a snowball at enhanced weathering rates. Upper and lower branch solutions are stable, but the intermediate solution is oscillatorily unstable if δ is sufficiently small. If δ is very small, then the motion becomes relaxational. Figure 1.25 shows an oscillatory solution illustrating this discussion.

The corresponding time series is shown in figure 1.26. It does not look much like the sawtooth oscillation of the Pleistocene ice ages, and the period is too long, some half million years. No doubt one can find something more persuasive by fiddling with parameters, but it may not be worth the effort, given the enormous simplicity of the model. The main use of the model is to illustrate the point that the carbon cycle contains a feedback effect which is capable of generating self-sustaining oscillations.

1.6 Notes and references

A good, recent book which addresses most of the issues of concern in this chapter is the book by Ruddiman (2001), which provides an expert's view. The book is aimed at undergraduates, and is very accessible. It is only marred by an addiction to design

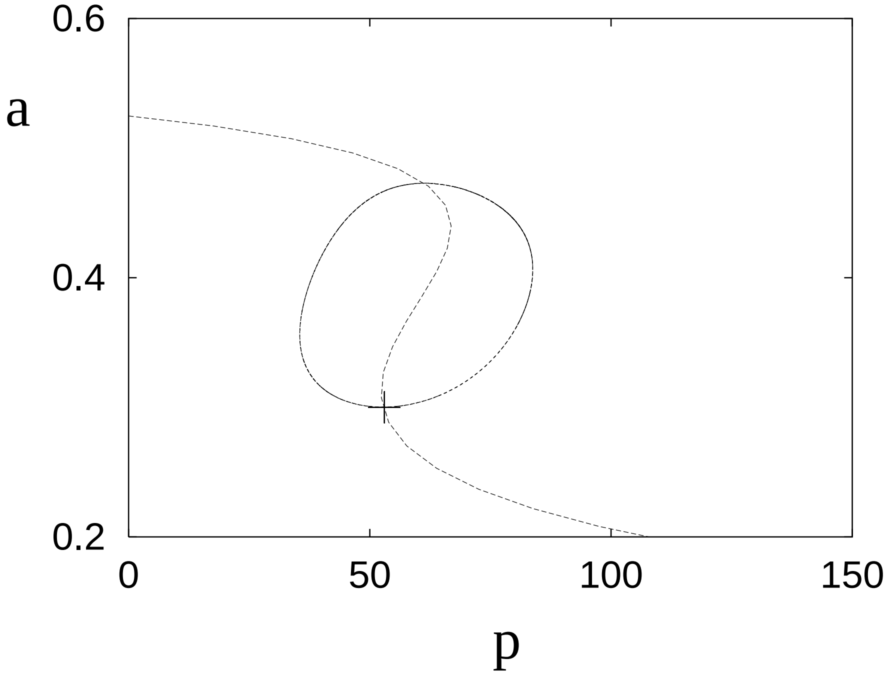

Figure 1.25: Limit cycle oscillation which passes through current climate (+) ($T = 288$ K, $a = 0.3$, $p_{\mathrm{CO}_2} = 53$ Pa). Also shown is the a nullcline of (1.130). The parameters are those of figure 1.24, with $t_i = 10^4$ y, $A = 5.1 \times 10^{14}$ m^2, $A_L = 1.5 \times 10^{14}$ m^2, $g = 9.81$ m s^{-2}. The temperature is taken to be the quasi-equilibrium value of (1.133), and the weathering and eruption rates are taken to be $W_0 = 0.26 \times 10^{-3}$ kg m^{-2} y^{-1}, $v_{\mathrm{CO}_2} = 0.2 \times 10^{11}$ kg y^{-1}. With these values, the parameters in (1.134) are $w = 1.95$ and $\delta = 0.0726$.

and graphics, which makes the book expensive and rather over the top — it is a book where there is a production team. Despite this, it is very up to date and informative.

1.6.1 Radiative heat transfer

The classic treatise on radiative heat transfer is the book by Chandrasekhar (1960), although it is dated and not so easy to follow. A more recent book aimed at engineers is that by Sparrow and Cess (1978). Most books on atmospheric physics will have some material on radiative heat transfer, for example those by Houghton (2002) and Andrews (2000). Other books are more specialised, such as those by Liou (2002) and Thomas and Stamnes (1999), but are not necessarily any easier to follow.

Rayleigh scattering is described by Strutt (1871), J. W. Strutt being Lord Rayleigh's given name.

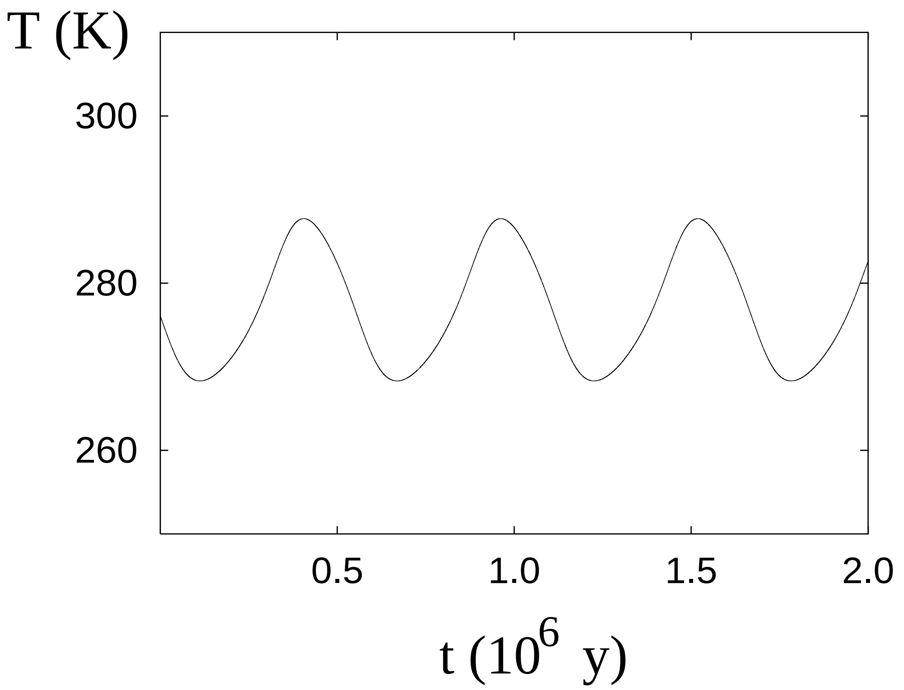

Figure 1.26: Time series of temperature for the periodic oscillation of figure 1.25.

1.6.2 The ozone layer

The description of the ozone layer dynamics essentially follows Chapman (1930). An elegant exposition is in the book by Andrews (2000). Reality is of course more complicated than the version presented here, and many more reactions can be included, in particular involving catalytic cycles, in which various chemical species catalyse the conversion of ozone to oxygen.

Chlorine species created by man-made chlorofluorocarbons have been implicated in the destruction of stratospheric ozone in the Antarctic, with the formation of the well known ‘ozone hole’.

1.6.3 Energy balance models

The original energy balance models are due to Budyko (1969) and Sellers (1969). They differ essentially only in the choice of parameterisation of emitted long wave radiation, and consider only the global balance of energy. North (1975a) allows latitude dependent albedo, and additionally allows for a parameterisation of poleward heat transport by oceans and atmosphere through a diffusive term, as in (1.66). North (1975b) added the time derivative. These meridionally averaged energy balance models do a rather good job of simulating the mean latitude dependent temperature profile, and have formed the basis for the atmospheric component of the more recent models of ‘intermediate complexity’. A later review is given by North *et al.* (1983).

1.6.4 The greenhouse effect

The first person who is generally credited with discussing the greenhouse effect is Arrhenius (1896), but Arrhenius himself refers to an earlier discussion by Fourier in 1827, where he refers to the atmosphere acting like the glass of a hothouse. Arrhenius's assessments of the effect of CO_2 are rather more severe than today's considered opinion. For a more recent discussion, see Houghton *et al.* (1996).

1.6.5 Ice ages

The data shown in figures 1.17–1.21 are taken from the GRIP (Greenland ice core project) ice core, drilled through the central part of the Greenland ice sheet project. This data is provided by the National Snow and Ice Data Centre of the University of Colorado at Boulder, Colorado, and the World Data Centre–A for Paleoclimatology at the National Geophysical Data Centre in Boulder, Colorado. This and other such data is publicly available at
`http://www.ngdc.noaa.gov/paleo/icecore/greenland/summit/index.html`
and has been reported in a number of publications, for example Johnsen *et al.* (1992), who, in particular, describe Dansgaard–Oeschger events. The higher resolution data sets in figure 1.20 and 1.21 were provided by Sigfus Johnsen, through the agency of Eric Wolff.

Abrupt climate change is documented by Severinghaus and Brook (1999) and Taylor *et al.* (1997). Taylor *et al.* (1993) find evidence of rapid ice age climate change in measurements of dust content in ice cores.

The cooling event at 8,200 years b. p. is described by Alley *et al.* (1997); Leuenberger *et al.* (1999) calibrate the temperature scale indicated by oxygen isotope variation by studying nitrogen isotope variations, suggesting that the cooling at 8,200 b. p. was of the order of 7°K; see also Lang *et al.* (1999).

1.6.6 Heinrich events

Heinrich events were first described in North Atlantic deep sea sediment cores by Heinrich (1988). MacAyeal (1993) introduced his 'binge-purge' model to explain them as a consequence of ice sheet oscillations induced by thermal instability, but assumed that a melting base would automatically cause large ice velocities. Fowler and Schiavi (1998) proposed a more physically realistic model which introduced the concept of hydraulic runaway, and Calov *et al.* (2002) showed that large scale ice sheets could oscillate somewhat as these earlier studies suggested, using a climate model of 'intermediate complexity' (essentially resolved oceans and ice sheets, and an averaged energy balance model of atmospheric fluxes).

1.6.7 Dansgaard-Oeschger events

Rahmstorf (2002) gives a nice review of the interplay of oceans and ice sheets in causing climatic oscillations during the last ice age. Ganopolski and Rahmstorf (2001) show how fluctuating freshwater delivery to the North Atlantic can cause abrupt alterations in circulation.

1.6.8 Oceans and climate

Stommel (1961) introduced the idea of different possible North Atlantic circulations. His model is not too realistic, but nevertheless simple and compelling. Rahmstorf (1995) uses a model of intermediate complexity to examine multiple circulation patterns in the North Atlantic. Depending on the freshwater flux to the North Atlantic, he finds hysteretic switches between different possible flows. Stocker and Johnsen (2003) is a more recent addition to the subject. Ganopolski and Rahmstorf (2001) provide a convincing picture of how switches of ocean circulation can cause rapid climate change. Their intermediate complexity model indicates hysteretic switches in ocean circulation due to changes in freshwater flux to the North Atlantic of unknown origin; we have suggested that the origin could be periodic subglacial floods. Broecker *et al.* (1990) and Manabe and Stouffer (1995) provide a similar thesis.

A great advocate of the 1000–2000 year rhythm in climate has been Gerard Bond; for example Bond *et al.* (1999) describe this rhythm, and also suggest that it has continued beyond the end of the ice age (into the Holocene), its most recent manifestation being the little ice age of 1500–1900. See also Bond *et al.* (1997).

1.6.9 Snowball Earth

The idea of a snowball Earth is discussed by Hoffman *et al.* (1998), for example, although the idea of ancient glaciations had been extant for a long time before that. Various modelling efforts have been made to assess the snowball's viability, for example, see Crowley and Baum (1993), Hyde *et al.* (2000), Chandler and Sohl (2000), and Pierrehumbert (2004).

1.6.10 The carbon cycle

Our (too) simple model of the interaction of the carbon cycle with ice sheet growth and climate change is based on the discussion of Walker *et al.* (1981), although their emphasis was on the rôle of CO_2 as a buffer in stabilising climate over geological time, despite the increasing solar luminosity. These ideas are elaborated by Kasting and Ackermann (1986) and Kasting (1989), who consider the effects of very large atmospheric CO_2 concentrations in early Earth history. Kasting (1989) suggests that because of the buffering effect of CO_2, a terrestrial (i. e., with liquid water) planet could be viable out as far as the orbit of Mars. In view of the plentiful evidence of water on Mars in *its* early history, this raises the intriguing prospect of a hysteretic switch from early temperate Mars to present cold Mars.

The buffering effect of CO_2 on climate and the rôle of continental location is discussed by Marshall *et al.* (1998). Berner *et al.* (1983) and Lasaga *et al.* (1985) discuss more complicated chemical models of weathering, and their effect on atmospheric CO_2 levels.

Petit *et al.* (1999) document the close relation between CO_2 levels and atmospheric temperature over the past 400,000 years. Unlike the result in figure 1.25, which together with (1.133), indicates that temperature and CO_2 will vary independently, the data shows that there is an excellent match. The model can be made consistent with this observation if the relaxation time t_i is reduced. In fact, this is not unreasonable, since the change of albedo due to sea ice coverage will be very fast, and this will shift the effective albedo time scale downwards.

Exercises

1.1 The planetary albedos of Venus, Mars and Jupiter are 0.77, 0.15, 0.58 respectively, and their distances from the sun are 0.72, 1.52, 5.20 *astronomical units* (1 a.u. = distance from Earth to the sun). Calculate the equilibrium temperature of these planets, and compare them with the measured effective black body temperatures, $T_m = 230$ K, 220 K, 130 K. Which, if any, planets appear not to be in equilibrium; can you think why this might be so?

1.2 Use Planck's law to derive the Stefan-Boltzmann law in the form

$$\int_0^\infty E_{b\lambda}\, d\lambda = \sigma T^4,$$

where σ can be written down in terms of a definite integral. By evaluating this integral, and using the values $c_1 = 3.74 \times 10^8$ W μm^4 m^{-2}, $c_2 = 1.44 \times 10^4$ μm K, evaluate the Stefan Boltzmann constant σ.

1.3 Show that $\int_{\frown} \cos\theta\, d\omega = \pi$, where θ is the polar angle, and the integral is over the range $0 < \phi < 2\pi$, $0 < \theta < \pi/2$. In a one-dimensional atmosphere, show that the average intensity is given by $J = \frac{1}{2}\int_{-1}^{1} I(\tau,\mu')d\mu'$, and show also that if the energy flux vector is $\mathbf{q}_R = \int_{\bigcirc} I(\mathbf{r},\mathbf{s})\,\mathbf{s}\,d\omega(\mathbf{s})$, then for a grey atmosphere

$$\boldsymbol{\nabla}.\mathbf{q}_R = -4\pi\kappa\rho[J - B].$$

Deduce that in radiative equilibrium, $J = B$.

1.4 For a purely absorptive atmosphere, show, by interpreting the radiation intensity along a ray path as a probability distribution function for the photon free path length (before absorption), that the mean free path is $1/\rho\kappa_\nu$. Deduce that an optically thin layer is one for which the photon mean free path is larger than the layer thickness.

1.5 In a purely scattering atmosphere, emission occurs by the scattering of radiation in all directions. Suppose that for a beam of intensity I_ν, the loss in intensity in a distance ds due to scattering is $\kappa_\nu I_\nu\, ds$, of which a fraction $P_\nu(\mathbf{s},\mathbf{s}')\, d\omega(\mathbf{s}')/4\pi$ is along a pencil of solid angle $d\omega(\mathbf{s}')$ in the direction $\mathbf{s}'$. Explain why it is reasonable to suppose that the scattering function P_ν should depend only on

$\mathbf{s}.\mathbf{s}'$, and show that the equation of radiative transfer can be written (assuming a grey atmosphere)

$$\frac{\partial I}{\partial s}=\rho\kappa\left[-I+\frac{1}{4\pi}\int_{\bigcirc}P(\mathbf{s},\mathbf{s}')I(\mathbf{r},\mathbf{s}')\,d\omega(\mathbf{s}')\right].$$

Deduce that for isotropic scattering, where $P\equiv 1$, the radiative flux $\mathbf{q}_R$ (see question 1.3) is divergence free.

For a plane parallel atmosphere in which $I=I(\tau,\mu)$, show that

$$\mu\frac{\partial I}{\partial \tau}=I-\frac{1}{4\pi}\int_{-1}^{1}\int_{0}^{2\pi}P(\mathbf{s}.\mathbf{s}')I(\tau,\mu')\,d\mu'\,d\phi',$$

where ϕ' is the azimuthal angle associated with $\mathbf{s}'$. Use spherical polar coordinates to show that

$$\mathbf{s}.\mathbf{s}'=\mu\mu'+(1-\mu^2)^{1/2}(1-\mu'^2)^{1/2}\cos(\phi-\phi'),$$

and deduce that for Rayleigh scattering, where $P(\cos\Theta)=\frac{3}{4}(1+\cos^2\Theta)$, I satisfies

$$\mu\frac{\partial I}{\partial \tau}=I-\tfrac{3}{16}\left[3I_0-I_2-\mu^2(I_0-3I_2)\right],$$

where $I_0=\displaystyle\int_{-1}^{1}I\,d\mu$, $I_2=\displaystyle\int_{-1}^{1}\mu^2 I\,d\mu$.

1.6 By non-dimensionalising the radiative heat transfer equation for a grey atmosphere using a length scale d (atmospheric depth) and an appropriate radiation intensity scale, show that in the optically thick limit, the equation takes the dimensionless form

$$I=B-\varepsilon\,\mathbf{s}.\boldsymbol{\nabla}I,$$

where $\varepsilon\ll 1$ and should be specified. Find an approximate solution to this equation, and hence show that the (dimensional) radiative energy flux vector $\mathbf{q}_R$ is given approximately by

$$\mathbf{q}_R=-\frac{4\sigma}{3\kappa\rho}\boldsymbol{\nabla}T^4.$$

1.7 The equation of radiative transfer in a grey, one-dimensional atmosphere is given by

$$\mu\frac{\partial I}{\partial \tau}=I-B,$$

with $I=0$ at $\tau=0$, $\mu<0$, and $I=B_S\equiv B(\tau_S)$ at $\tau=\tau_S$, $\mu>0$. Write down the formal solution assuming B is known, and hence show that the radiative flux $q_R=2\pi\displaystyle\int_{-1}^{1}\mu I\,d\mu$ is given by

$$q_R=2\pi\left[-\int_0^{\tau}B(\tau')E_2(\tau-\tau')\,d\tau'+B_S E_3(\tau_S-\tau)+\int_{\tau}^{\tau_S}B(\tau')E_2(\tau'-\tau)\,d\tau'\right],$$

where the exponential integrals are defined by

$$E_n(z) = \int_1^\infty \frac{e^{-zt}\,dt}{t^n},$$

and $B_S \equiv B(\tau_S)$.

Show that $E_n' = -E_{n-1}$, $E_n(0) = \dfrac{1}{n-1}$, and deduce that

$$\frac{\partial q_R}{\partial \tau} = 2\pi\left[-2B + B_S E_2(\tau_S - \tau) + \int_0^{\tau_S} B(\tau')E_1(|\tau - \tau'|)\,d\tau'\right].$$

Show also that the intensity $J = \frac{1}{2}\int_{-1}^{1} I\,d\mu$ is given by

$$J = \tfrac{1}{2}\left[\int_0^{\tau_S} B(\tau')E_1(|\tau - \tau'|)\,d\tau' + B_S E_2(\tau_S - \tau)\right].$$

By integrating the expression for q_R by parts, show that

$$q_R = 2\pi\left[B_0 E_3(\tau) + \int_0^{\tau_S} B'(\tau')E_3(|\tau - \tau'|)\,d\tau'\right]. \qquad (*)$$

If τ_S is large, so that B varies slowly with τ, show that when τ is large,

$$q_R \approx \frac{4\pi}{3}B'(\tau)$$

(essentially, this uses Laplace's method for the asymptotic evaluation of integrals).

Use the integral expression $(*)$ for q_R to show that if $q_R = \pi B_0$ at $\tau = 0$, then

$$\int_0^{\tau_S} B'(\tau')E_3(\tau)\,d\tau' = 0,$$

and deduce that the temperature gradient cannot be monotonic for such an atmosphere.

1.8 Chapman's model for the production of ozone in the stratosphere is

$$\begin{aligned}
\mathrm{O_2} + h\nu &\overset{j_2}{\rightarrow} \mathrm{2O},\\
\mathrm{O} + \mathrm{O_2} + \mathrm{M} &\overset{k_2}{\rightarrow} \mathrm{O_3} + \mathrm{M},\\
\mathrm{O_3} + h\nu &\overset{j_3}{\rightarrow} \mathrm{O} + \mathrm{O_2},\\
\mathrm{O} + \mathrm{O_3} &\overset{k_3}{\rightarrow} \mathrm{2O_2}.
\end{aligned}$$

Write down the rate equations for the concentrations X, Y and Z of oxygen atoms O, oxygen O_2 and ozone O_3, and show that

$$X + 2Y + 3Z = 2[O_2],$$

where $[O_2]$ is constant.

Suppose, as is observed, that $X \ll \frac{j_3}{k_3}$ and $Y \gg \frac{j_3}{k_2[M]}$, where [M] is the concentration of M. Use these observations to scale the equations to the form

$$\varepsilon \frac{dx}{dt} = z - xy + 2\delta y - \delta xz,$$

$$\frac{dz}{dt} = xy - z - \delta xz,$$

$$y + \tfrac{1}{2}\lambda(3z + \varepsilon x) = 1,$$

where

$$\varepsilon = \frac{j_3}{k_2[O_2][M]}, \quad \delta = \left[\frac{j_2 k_3}{j_3 k_2 [M]}\right]^{1/2}, \quad \lambda = \left[\frac{j_2 k_2 [M]}{j_3 k_3}\right]^{1/2}.$$

Assuming $\varepsilon, \delta, \lambda \ll 1$, show that the model can be partially solved to produce the approximate equation

$$\frac{dz}{d\tau} = 2\left(1 - z^2\right),$$

where $t = \tau/\delta$.

Hence show that $[O_3] \to \lambda[O_2]$ on a time scale $t \sim \left(\frac{k_2[M]}{j_2 j_3 k_3}\right)^{1/2}$, and that the reaction scheme can be represented by the overall reaction

$$3O_2 \underset{r_-}{\overset{r_+}{\rightleftharpoons}} 2O_3,$$

where

$$r_+ = \tfrac{2}{3} j_2, \quad r_- = \frac{j_3 k_3}{k_2[O_2][M]}.$$

1.9 Suppose that stratospheric heating by absorption of ultraviolet radiation is given by

$$Q = -\frac{\partial I}{\partial z},$$

where

$$I = -I_\infty \exp\left[-\tau_0 e^{-z/H}\right],$$

and

$$\tau_0 = \kappa \rho_0 H.$$

Suppose also that the (upwards) long wave radiative flux is given by

$$q_R = -k_R \frac{\partial T}{\partial z},$$

where the radiative conductivity is given by

$$k_R = \frac{16\sigma T^3 e^{z/H}}{3\kappa \rho_c}.$$

Write down the energy equation describing radiant energy transport, and show that the temperature T is given by

$$T = T_0 \left[A - \phi e^{-\zeta} - \theta \exp\left(-\tau_0 e^{-\zeta}\right)\right]^{1/4},$$

where A and ϕ are constants, and

$$T_0 = \left(\frac{3I_\infty}{4\sigma}\right)^{1/4}, \quad \theta = \frac{\tau_c}{\tau_0}, \quad \zeta = \frac{z}{H}.$$

Suppose that $\phi, \theta, A \sim O(1)$, and that $\tau_0 \gg 1$. Find approximations for T for $\zeta < \ln \tau_0$ and $\zeta \sim \ln \tau_0$, and deduce that T has a maximum at $z \sim H \ln \tau_0$. How is this related to the temperature in the stratosphere?

1.10 Using values $d = 10$ km, $\kappa \rho d = 0.67$, show that a representative value of the radiative conductivity k_R defined by $\mathbf{q}_R = -k_R \boldsymbol{\nabla} T$ for an opaque atmosphere is $k_R \simeq 1.08 \times 10^5$ W m^{-1} K^{-1}. Hence show that a typical value for the effective Péclet number

$$Pe = \frac{\rho c_p U d^2}{k_R l}$$

is about 20, if $U \approx 20$ m s^{-1}, $l \approx 1000$ km. Explain the import of this in terms of the heat equation

$$\rho c_p \frac{dT}{dt} = \boldsymbol{\nabla}.[k_R \boldsymbol{\nabla} T].$$

1.11 A wet adiabat is calculated from the isentropic equation

$$\rho_a c_p \frac{dT}{dz} - \frac{dp}{dz} + \rho_a L \frac{dm}{dz} = 0,$$

where

$$m = \frac{\rho_v}{\rho_a}, \; p = \frac{\rho_a RT}{M_a}, \; p_{SV} = \frac{\rho_v RT}{M_v},$$

and

$$\frac{dp_{SV}}{dT} = \frac{\rho_v L}{T}, \; \frac{dp}{dz} = -\rho_a g.$$

Deduce that T and p_{SV} can be calculated from the equations

$$\begin{aligned}\frac{dT}{dz} &= -\Gamma_w(\rho_v, p, T),\\ \frac{dp_{SV}}{dz} &= -\frac{\rho_v L}{T}\Gamma_w,\end{aligned}$$

where $\rho_v = \rho_v(p_{SV}, T)$, and Γ_w should be determined. Using values $M_v/M_a = 0.62$, $L = 2.5 \times 10^6$ J kg^{-1}, $T = 290$ K, $c_p = 10^3$ J kg^{-1} K^{-1}, $\rho_v = 0.01$ kg m^{-3}, $p = 10^5$ Pa, $g = 10$ m s^{-2}, $\rho_a = 1$ kg m^{-3}, show that a typical value of Γ_w is 6 K km^{-1}.

By assuming that $T \approx$ constant near the ground (why?), derive a differential equation for p_{SV} as a function of p in terms of two dimensionless coefficients

$$a = \frac{M_v L}{RT}, \ \beta = \frac{M_v}{M_a}\frac{L}{c_p T},$$

and estimate their values (you will need also the values $M_v = 18 \times 10^{-3}$ kg mole^{-1}, $R = 8.3$ J mole^{-1} K^{-1}). Hence show that the *molar humidity* $h = p_{SV}/p$ decreases approximately linearly with altitude.

1.12 Show that the solution of the Clausius-Clapeyron equation for saturation vapour pressure p_{SV} as a function of temperature T is

$$p_{SV} = p_{SV}^0 \exp\left[a\left\{1 - \frac{T_0}{T}\right\}\right],$$

where for water vapour, we may take $T_0 = 273$ K at $p_{SV}^0 = 6$ mbar ($= 600$ Pa), the *triple point*, and $a = M_v L/RT_0$. Show that if T is close to T_0, then

$$p_{SV} \approx p_{SV}^0 \exp\left[a\left(\frac{T - T_0}{T_0}\right)\right].$$

If the long wave radiation from a planet is $\sigma\gamma T^4$, where T is the mean surface temperature, if the solar flux is Q (and planetary albedo is zero), and the greyness factor is taken to be given by

$$\gamma^{-1/4} = 1 + b(p_v/p_{SV}^0)^c,$$

where p_v is the H_2O vapour pressure, show that the occurrence of a runaway greenhouse effect is controlled by the intersection of the two curves

$$\theta = 1 + \lambda\xi, \ \theta = \rho(1 + be^{c\xi}),$$

where $\lambda = 1/a$, $\rho = (Q/4\sigma T_0^4)^{1/4}$. Show that runaway occurs if $\rho > \rho_c$, where

$$\rho_c + \delta = 1 + \delta \ln[\delta/b\rho_c]$$

with $\delta = \lambda/c$. Show that this determines a unique value of ρ_c, and that if δ is small,

$$\rho_c \approx 1 + \delta \ln(\delta/b) - \delta.$$

Estimate values of ρ and λ appropriate to the present Earth, and comment on the implications of these values for climatic evolution if we choose $b = 0.06$, $c = 1/4$. What are the implications for Venus, if the solar flux is twice as great? What if solar radiation were 30% lower when the planetary atmospheres were being formed?

1.13 For the energy balance model

$$c\dot{T} = R_i - R_o,$$

where $R_i = Q(1-a)$, $R_o = \sigma\gamma T^4$, and $a = a_+$ for $T < T_i$, $a = a_-$ for $T > T_w$ $(> T_i)$, $a_+ > a_-$, with $a(T)$ linear between these two ranges, show that multiple steady states are possible in a range $Q_- < Q < Q_+$, where

$$Q_- = \frac{\sigma\gamma T_w^4}{1-a_-}, \quad Q_+ = \frac{\sigma\gamma T_i^4}{1-a_+},$$

providing

$$\frac{T_w - T_i}{T_w} < \frac{a_+ - a_-}{4(1-a_-)},$$

and in this case prove that the upper and lower branches are stable, but the intermediate one is unstable.

By normalising Q and T with respect to present day values Q_0, T_0 satisfying $Q_0(1-a_-) = \sigma\gamma T_0^4$, show that the corresponding dimensionless solar fluxes and mean atmospheric temperatures q and θ satisfy

$$\begin{aligned} q_- &= \theta_w^4, \\ q_+ &= \theta_i^4 \left(\frac{1-a_-}{1-a_+}\right), \end{aligned}$$

providing

$$\frac{\theta_w - \theta_i}{\theta_w} < \frac{a_+ - a_-}{4(1-a_-)}.$$

If $\theta_w = 1$ (we are starting an ice age *now*) show that if $\theta_i = 1 - \delta$, $a_+ = a_- + \nu$, where $\delta, \nu \ll 1$, then regular ice ages will occur providing

$$\delta < \frac{\nu}{4(1-a_-)},$$

and providing the solar flux q oscillates beyond the limits $q_+ \approx 1 + \nu/(1-a_-) - 4\delta$ and $q_- = 1$.

1.14 Suppose that the planetary albedo a is given by the ordinary differential equation

$$t_i \dot{a} = I(a, Q) - a,$$

where

$$I(a, Q) = a_{\mathrm{eq}}[T(a, Q)],$$

$$a_{\mathrm{eq}}(T) = a_1 - \tfrac{1}{2} a_2 \left[1 + \tanh\left(\frac{T - T^*}{\Delta T}\right)\right],$$

$$T(a, Q) = \left[\frac{Q(1-a)}{4\sigma\gamma}\right]^{1/4}.$$

Determine the graphical dependence of I as a function of a, and how this varies with Q, and hence describe the form of oscillations if Q is periodic, and t_i is sufficiently small.

For large t_i, show that the equation can be written in the dimensionless form

$$\dot{a} = \varepsilon[I\{a, Q(t)\} - a],$$

where $\varepsilon \ll 1$. The *method of averaging* implies that a varies slowly, and thus can be written approximately as the series

$$a \sim A_0(\tau) + \varepsilon A_1(t, \tau) + \dots,$$

where $\tau = \varepsilon t$, and

$$\dot{A} = \overline{I(A_0, Q)} - A_0,$$

where $\overline{I(A_0, Q)}$ denotes the time average of I over a period of Q. Deduce that for a range of values of $\bar{Q}$, two periodic solutions can exist, and comment on their climatic interpretation.

Give explicit approximate solutions for the cases $\varepsilon \ll 1$ and $\varepsilon \gg 1$ when ΔT is very small.

1.15 Ocean temperature θ and salinity s are described by Stommel's box model

$$\dot{\theta} = 1 - (\mu + |\theta - Rs|)\theta,$$

$$\dot{s} = 1 - |\theta - Rs|)s,$$

where μ and R are positive. By analysing the equations in the phase plane, show that up to three steady states can exist, and assess their stability.

By drawing the phase portrait, discuss the nature of the solutions when there is one steady state, and when there are three.

1.16 The temperature T, $\mathrm{CO_2}$ pressure p, and planetary albedo a satisfy the ordinary differential equations

$$\begin{aligned} c\dot{T} &= \tfrac{1}{4}Q(1-a) - \sigma\gamma T^4, \\ t_i\dot{a} &= a_{\rm eq}(T) - a, \\ \frac{A}{g}\dot{p} &= -A_L W + v, \end{aligned}$$

where

$$a_{\rm eq}(T) = a_1 - \tfrac{1}{2}a_2\left[1 + \tanh\left(\frac{T-T^*}{\Delta T}\right)\right],$$

where $a_1 = 0.58$, $a_2 = 0.47$, $T^* = 283$ K, $\Delta T = 24$ K,

$$W = W_0\left(\frac{p}{p_0}\right)^{\mu}\exp\left[\frac{T-T_0}{\Delta T_c}\right],$$

and

$$\gamma(p) = \gamma_0 - \gamma_1 p.$$

Show how to non-dimensionalise the system to the dimensionless form

$$\begin{aligned} \varepsilon\dot{\theta} &= 1 - a - (1-a_0)\left(1 + \tfrac{1}{4}\nu\theta\right)^4(1-\nu\lambda p), \\ \dot{a} &= B(\theta) - a, \\ \dot{p} &= \delta\left[1 - wp^{\mu}e^{\theta}\right], \end{aligned}$$

and show that

$$\delta = \frac{vgt_i}{Ap_0}, \quad w = \frac{A_L W_0}{v}, \quad \varepsilon = \frac{4c\Delta T_c}{t_i Q}, \quad \nu = \frac{4\Delta T_c}{T_0}, \quad \lambda = \frac{\gamma_1 p_0}{\nu\gamma_0}.$$

What is the function $B(\theta)$?

Using the values $v = 3\times10^{11}$ kg y^{-1}, $g = 9.8$ m s^{-2}, $t_i = 10^4$ y, $A = 5\times10^{14}$ m^2, $p_0 = 50$ Pa, $A_L = 1.5\times10^{14}$ m^2, $W_0 = 2\times10^{-3}$ kg m^{-2} y^{-1}, $c = 10^7$ J m^{-2} K^{-1}, $Q = 1370$ W m^{-2}, $\Delta T_c = 13$ K, $T_0 = 285$ K, $\gamma_0 = 0.64$, $\gamma_1 = 4.25\times10^{-4}$ Pa^{-1}, show that

$$\delta \approx 1.2, \quad w \approx 1, \quad \varepsilon \approx 1.2\times10^{-6}, \quad \nu \approx 0.18, \quad \lambda \approx 0.3,$$

and find the value of a_0, assuming $\sigma = 5.67\times10^{-8}$ W m^{-2} K^{-4}.

Hence show that θ rapidly approaches a quasi-steady state given by

$$\theta \approx \Theta(a,p) = \alpha(a_0 - a) + \lambda p,$$

where

$$\alpha = \frac{1}{\nu(1-a_0)}.$$

In the phase plane of a and p satisfying

$$\begin{aligned} \dot{a} &= B(\Theta) - a, \\ \dot{p} &= \delta\left[1 - wp^{\mu}e^{\Theta}\right], \end{aligned} \qquad *$$

show that the p nullcline is a monotonically increasing function $a_p(p)$ of p, and that the a nullcline is a monotonically decreasing function $a_a(p)$ of p, providing $-B'(\theta) < \nu(1 - a_0)$ for all θ. Show conversely that if there is a range of θ for which $-B'(\theta) > \nu(1 - a_0)$, then the a nullcline is multivalued.

Suppose that the a nullcline is indeed multivalued, and that there is always a unique steady state. Show that at low, intermediate and high values of w, this equilibrium can lie on the lower, intermediate or upper branch of the a nullcline.

By consideration from the phase plane of the signs of the partial derivatives of the right hand sides of $(*)$ (and without detailed calculation), show that when they exist, the upper and lower branch steady states are stable, but that the intermediate steady state will be oscillatorily unstable if δ is small enough.

How would you expect the solutions to behave if $\delta \ll 1$?

Chapter 2

River flow

Much of the environment consists of fluids, and much of this book is therefore concerned with fluid mechanics. Oceans and atmosphere consist of fluids in large scale motion, and even later, when we deal with more esoteric subjects: the flow of glaciers, convection in the Earth's mantle, it is within the context of fluid mechanics that we formulate relevant models. This chapter concerns one of the most obvious common examples of a fluid in motion, that of the mechanics of rivers.

Fluid mechanics in the environment is, however, altogether different to the subject we study in an undergraduate course on viscous flow, and the principal reason for this is that for most of the common environmental fluid flows with which we are familiar, the flow is *turbulent*. (Where it is not, for example in glacier flow, other physical complications obtrude.) As a consequence, the models which we use to describe the flow are different to (and in fact, simpler than) the Navier-Stokes equations.

2.1 The hydrological cycle

Rainwater which falls in a catchment area of a particular river basin makes its way back to the ocean (or sometimes to an inland lake) by seepage into the ground, and then through groundwater flow to outlet streams and rivers. In severe storm conditions, or where the soil is relatively impermeable, the rainfall intensity may exceed the soil infiltration capacity, and then direct runoff to discharge streams can occur as overland flow. Depending on local topography, soil cover, vegetation, one or other transport process may be the norm. Overland flow can also occur if the soil becomes saturated. The hydrological cycle is completed when the water, now back in the ocean, is evaporated by solar radiation, forming atmospheric clouds which are the instrument of precipitation.

River flow itself occurs on river beds that are typically quasi-one-dimensional, sinuous channels with variable and rough cross-section. Moreover, if the channel discharge is Q ($\mathrm{m}^3\ \mathrm{s}^{-1}$), and the wetted perimeter length of the cross section is l (m), then an appropriate Reynolds number for the flow is

$$Re = Q/\nu l, \tag{2.1}$$

where $\nu = \mu/\rho$ is the kinematic viscosity (and μ is the dynamic viscosity). If $l = 10$ m, $\nu = 10^{-6}$ m^2 s^{-1}, $Q = 1$ m^3 s^{-1} (a small value), then $Re \sim 10^5$. Inevitably, river flow is turbulent for all but the smallest rivulets. Thus, to model river flow, and to explain the response of river discharge to storm conditions, as measured on flood hydrographs, for instance, one must model a flow which is essentially turbulent, and which exists in a rough, irregular channel. The classical way in which this is done is by applying a time average to the Navier-Stokes equations

$$\begin{aligned}\frac{\partial u_i}{\partial x_i} &= 0,\\ \rho\frac{\partial u_i}{\partial t} + \rho\frac{\partial}{\partial x_j}(u_i u_j) &= -\frac{\partial p}{\partial x_i} + \mu\nabla^2 u_i,\end{aligned} \tag{2.2}$$

where suffixes i represent the components, and the summation convention is used (i. e., summation over repeated suffixes is implied). If we denote time averages by an overbar, and fluctuations by a prime, thus

$$u_i = \bar{u}_i + u_i', \tag{2.3}$$

then averaging of (2.2) yields

$$\begin{aligned}\frac{\partial \bar{u}_i}{\partial x_i} &= 0,\\ \rho\frac{\partial}{\partial x_j}(\bar{u}_i \bar{u}_j) + \frac{\partial}{\partial x_j}(\rho\overline{u_i' u_j'}) &= -\frac{\partial \bar{p}}{\partial x_i} + \mu\nabla^2 \bar{u}_i.\end{aligned} \tag{2.4}$$

The second of these can be written in the form

$$(\bar{\mathbf{u}}.\boldsymbol{\nabla})\bar{\mathbf{u}} = -\boldsymbol{\nabla}p + \boldsymbol{\nabla}.\{\boldsymbol{\tau} + \boldsymbol{\tau}^T\}, \tag{2.5}$$

where $\tau_{ij} = \mu\,\dot{\varepsilon}_{ij}$ is the ordinary molecular stress, and

$$\tau_{ij}^T = -\rho\overline{u_i' u_j'} \tag{2.6}$$

is called the Reynolds stress. The essential problem in describing fully turbulent flows is to close the averaged model by prescribing the Reynolds stress.

For a flow $\mathbf{u} = (u, v, w)$ which is locally unidirectional *on average*, such as that in a river, we may take $\bar{\mathbf{u}} = (\bar{u}, 0, 0)$, and then the x component of the momentum equation becomes

$$\rho\frac{\partial}{\partial z}(\overline{u'w'}) = -\frac{\partial \bar{p}}{\partial x} + \mu\frac{\partial^2 \bar{u}}{\partial z^2}, \tag{2.7}$$

because the time averages of the transverse Reynolds stresses are zero. Integration over the depth shows that the resistance to motion is provided by the wall stress τ, and this is

$$\tau = \mu\frac{\partial \bar{u}}{\partial z} + \{-\rho\overline{u'w'}\}, \tag{2.8}$$

evaluated at the wetted perimeter of the flow. Strictly, the Reynolds stress vanishes at the boundary (because the fluid velocity is zero there), and the molecular stress changes rapidly to compensate, in a very thin laminar wall layer. It is common to evaluate (2.8) just outside this layer, close to but not at the boundary, where the molecular stress is negligible and the Reynolds stress is parameterised in some way. A common choice is to use a friction factor, thus

$$\tau = f\rho\bar{u}^2, \tag{2.9}$$

where the dimensionless number f (called the friction factor) is found to depend rather weakly on the Reynolds number. A crude but effective assumption is simply that f is constant, with a typical value for f of 0.01. Further discussion of turbulent flows, and how to model them, is given in appendix A.

2.2 Chezy's and Manning's laws

Our starting point is that the flow is essentially one-dimensional: or at least, we focus on this aspect of it. As well as the cross sectional area (of the *flow*) A and discharge Q, we introduce a longitudinal, curvilinear distance coordinate s, and we assume that the river axis changes direction slowly with s. Then conservation of mass is, in its simplest form,

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial s} = 0. \tag{2.10}$$

This ignores, for the moment, source terms due to infiltration seepage and overland flow from the catchment.

(2.10) must be supplemented by an equation for Q as a function of A, and this arises through consideration of momentum conservation. There are three levels at which one may do this: by exact specification, as in the Navier-Stokes momentum equation; by ignoring inertia and averaging, as in Darcy's law; and most simply, by ignoring inertia and applying a force balance using a semi-empirical friction factor. We begin by opting for this last choice, which should apply for sufficiently 'slow' (in some sense) flow. Later we will consider more complicated models.

We have already defined the Reynolds number Re in terms of Q and A, or equivalently a mean velocity $u = Q/A$ and a channel depth $d \sim A^{1/2}$. 'Slow' here means a small *Froude number*, defined by

$$Fr = \frac{u}{(gd)^{1/2}} = \frac{Q}{g^{1/2}A^{5/4}}. \tag{2.11}$$

If $Fr < 1$, the flow is *tranquil*; if $Fr > 1$, it is *rapid*. Gravity is of relevance, since the flow is ultimately due to gravity.

Now let l be the perimeter of a cross section, and let τ be the mean shear stress exerted at the bed (longitudinally) by the flow. If the downstream slope is α, then a force balance gives

$$l\tau = \rho g\, A \sin\alpha, \tag{2.12}$$

where ρ is density. For turbulent flow, the shear stress is given by the friction law

$$\tau = f\rho u^2, \tag{2.13}$$

where the friction factor f may depend on the Reynolds number. Since

$$u = Q/A, \tag{2.14}$$

and defining the hydraulic radius

$$R = A/l, \tag{2.15}$$

we derive the relations

$$u = (g/f)^{1/2} R^{1/2} S^{1/2}, \tag{2.16}$$

where

$$S = \sin\alpha, \tag{2.17}$$

and

$$Q = \left(\frac{g}{fl}\right)^{1/2} A^{3/2} S^{1/2}. \tag{2.18}$$

For wide, shallow rivers, l is essentially the width. For a more circular cross-section, then $l \sim A^{1/2}$, and

$$Q = (g/f)^{1/2} A^{5/4} S^{1/2}. \tag{2.19}$$

The relation (2.16) is the Chezy velocity formula, and $C = (g/f)^{1/2}$ is the Chezy roughness coefficient. Notice that the Froude number, in terms of the hydraulic radius, is

$$Fr = \frac{u}{(gR)^{1/2}} = (S/f)^{1/2}, \tag{2.20}$$

and tranquillity (at least in uniform flow) is basically due to slope.

Alternative friction correlations exist. That due to Manning is an empirical formula to fit measured stream velocities, and is of the form

$$u = R^{2/3} S^{1/2}/n' \tag{2.21}$$

where Manning's roughness coefficient n' takes typical values in the range 0.01–0.1 $\mathrm{m}^{-1/3}$ s, depending on stream depth, roughness, etc. For Manning's formula, we have

$$\begin{aligned} Q &\sim A^{4/3} \quad \text{if} \quad R \sim A^{1/2}, \\ Q &\sim A^{5/3} \quad \text{if} \quad l \ \text{ is width}, \quad R = A/l \sim A. \end{aligned} \tag{2.22}$$

Thus we see that for a variety of stream types and velocity laws, we can pose a relation between discharge and area of the form

$$Q \sim A^{m+1}, \quad m > 0, \tag{2.23}$$

with typical values $m = 1/4 - 2/3$. In practice, for a given stream, one could attempt to fit a law of the form (2.23) by direct measurement.

2.3 The flood hydrograph

We can nondimensionalise the equation for A so that it becomes

$$\frac{\partial A}{\partial t} + A^m \frac{\partial A}{\partial s} = 0, \tag{2.24}$$

a first order nonlinear hyperbolic equation, also known as a kinematic wave equation, whose solution can be written down. Specifically, take initial data parameterised as

$$A = A_0(\sigma), \quad s = \sigma > 0, \quad t = 0. \tag{2.25}$$

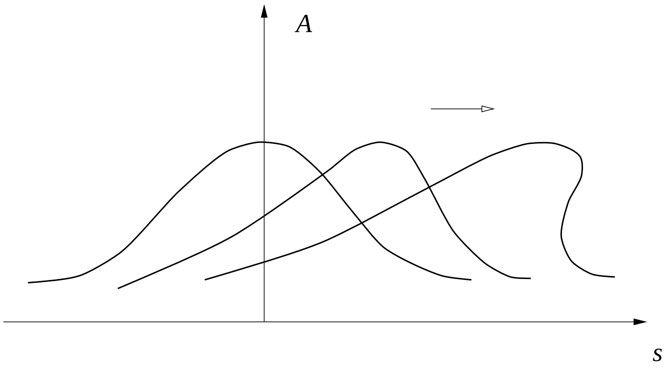

Figure 2.1: Formation of a shock wave in the solution of (2.24).

Then the characteristic equations are

$$\frac{dA}{dt} = 0,$$

$$\frac{ds}{dt} = A^m, \tag{2.26}$$

whence

$$A = A_0(\sigma), \quad s = \sigma + A_0^m t, \tag{2.27}$$

thus

$$A = A_0[s - A^m t] \tag{2.28}$$

determines A implicitly. It is a familiar fact that humped initial conditions $A_0(\sigma)$ will lead to propagation of a kinematic wave, and then to shock formation, as shown in figure 2.1, when $\partial A/\partial s$ reaches infinity. Since $\partial A/\partial s = A_s = A_0'[s - A^m t](1 - mt\,A^{m-1}A_s)$, this is on the characteristic through σ which maximises $-(A_0^m)'$, and at a time t which is the inverse of this maximum. Thereafter a shock exists at a point $s_d(t)$, and propagates at a rate given, by consideration of the integral conservation law

$$\frac{\partial}{\partial t}\int_{s_1}^{s_2} A\,ds = -[Q]_{s_1}^{s_2}, \tag{2.29}$$

by

$$\dot{s}_d = \frac{[Q]_{s_d-}^{s_d+}}{[A]_{s_d-}^{s_d+}}. \tag{2.30}$$

As an application, we consider the flood hydrograph, which measures discharge at a fixed value of s as a function of time. As an idealisation of a flood, we consider an initial condition

$$A \approx A_0\delta(s) \quad \text{at} \quad t = 0, \tag{2.31}$$

where $\delta(s)$ is the delta function, representing the overland flow of a short period of localised rainfall. Since $A = f(s - A^m t)$, it follows that $A \approx 0$ except where $s = A^m t$. The humped initial condition causes a shock to form at $s_d(t)$, with $s_d(0) = 0$, and we have

$$\begin{aligned} A &= 0, \quad s > s_d, \\ A &= (s/t)^{1/m}, \quad s < s_d, \end{aligned} \tag{2.32}$$

as shown in figure 3.2.

The shock speed is given by

$$\dot{s}_d = (Q/A)|_{s_d-} = \left.\frac{A^m}{m+1}\right|_{s_d-} = \frac{s_d}{(m+1)t}, \tag{2.33}$$

whence $s_d \propto t^{1/(m+1)}$. To calculate the coefficient of proportionality, we use conservation of mass in the form

$$\int_0^{s_d} A\,ds = A_0, \tag{2.34}$$

whence, in fact,

$$s_d = [(m+1)A_0/m]^{m/(m+1)}t^{1/(m+1)}. \tag{2.35}$$

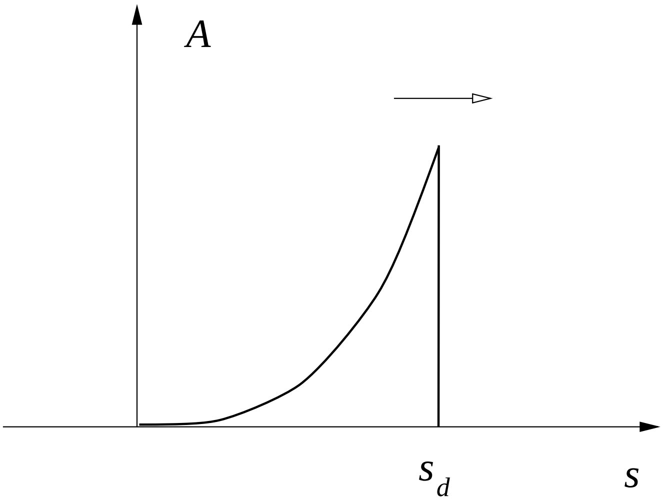

Figure 2.2: Propagation of a shock front.

Denoting $b = [(m+1)A_0/m]^{m/(m+1)}$, the flood hydrograph at a fixed station $s = s^*$ is then as follows. For $t < t^*$, where

$$t^* = (s^*/b)^{m+1}, \tag{2.36}$$

$Q = 0$. For $t > t^*$, $A = (s^*/t)^{1/m}$, and thus

$$Q = \frac{s^{*(m+1)/m}}{(m+1)} t^{-(m+1)/m}. \tag{2.37}$$

This result is illustrated in figure 2.3, together with a typical observed hydrograph. The smoothed observation can be explained by the fact that a more realistic initial condition would have delivery of the storm flow over an interval of space and time. More importantly, one can expect that a more realistic model will allow diffusive effects.

2.4 St. Venant equations

We now re-examine the momentum equation, which we previously assumed to be described by a force balance. Again consider the equations in dimensional form. Conservation of mass is written in the form

$$\frac{\partial A}{\partial t} + \frac{\partial}{\partial s}(Au) = 0, \tag{2.38}$$

and then conservation of momentum (from first principles) leads to the equation (adopting the friction law (2.13))

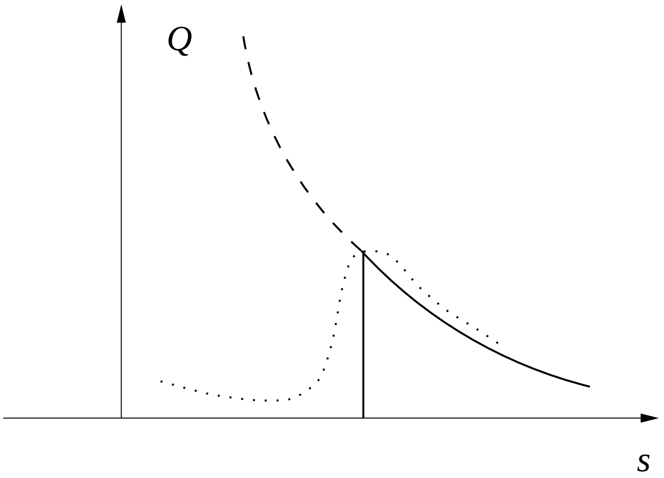

Figure 2.3: Ideal and observed hydrographs.

$$\rho\frac{\partial (Au)}{\partial t}+\rho\frac{\partial}{\partial s}(Au^2)=\rho gA\sin\alpha-\rho lfu^2-\frac{\partial}{\partial s}(A\bar{p}),\tag{2.39}$$

where $\bar{p}$ is the mean pressure. Now the pressure is approximately hydrostatic, thus $p\approx\rho gz$ where z is depth. Then $\bar{p}A\approx\int\frac{1}{2}\rho gh^2dx$ where h is total depth, x is width, and supposing $\partial h/\partial s$ is independent of x, we find[1]

$$\frac{\partial}{\partial s}(A\bar{p})=\rho gA\frac{\partial\bar{h}}{\partial s},\tag{2.40}$$

where $\bar{h}$ is the mean depth. Using (2.38), (2.39) reduces to

$$u_t+uu_s=g\sin\alpha-flu^2/A-g\frac{\partial\bar{h}}{\partial s}.\tag{2.41}$$

Equations (2.38) and (2.41) are known as the St. Venant equations.

2.4.1 Nondimensionalisation

We choose scales for $u=Q/A$, t, s, A, R (the hydraulic radius, $=A/l$), $\bar{h}$ as follows, in keeping with the assumed balances adopted earlier:

$$Au\sim Q,$$

$$g\sin\alpha\sim lfu^2/A=fu^2/R,$$

---

[1] The assumption that $\partial h/\partial s$ is constant across the stream means that along a transverse section of the river, the surface is horizontal. This is really due to the smallness of the width compared to the length. It is importantly not exactly true for meandering rivers, but is still a very good approximation.

$$
\begin{aligned}
t &\sim s/u,\\
s &\sim d/\sin\alpha,\\
\bar{h}, R &\sim d,
\end{aligned}
\tag{2.42}
$$

where we can suppose Q is a typical observed discharge, and d is a typical observed depth. Explicitly, the scales are

$$
\begin{aligned}
\left[\bar{h}\right], [R] &= d, \quad [s] = d/\sin\alpha,\\
[u] &= (gd\sin\alpha/f)^{1/2}, \quad [t] = (fd/g\sin^3\alpha)^{1/2},\\
[A] &= Q(f/gd\sin\alpha)^{1/2},
\end{aligned}
\tag{2.43}
$$

and we put $u = [u]u^*$, etc., and drop asterisks. The resulting equations are

$$
\begin{aligned}
A_t + (Au)_s &= 0,\\
F^2[u_t + uu_s] &= 1 - u^2/R - h_s,
\end{aligned}
\tag{2.44}
$$

where we would choose $h \sim R \sim A$ for a wide channel, $h \sim R \sim A^{1/2}$ for a rounded channel. In particular, for a wide channel, we have $R = h$, so that the momentum equation can be written

$$
\begin{aligned}
(wh)_t + (wuh)_s &= 0,\\
F^2(u_t + uu_s) &= 1 - u^2/h - h_s,
\end{aligned}
\tag{2.45}
$$

since $A = wh$, where w is the (dimensionless) width. The Froude number F is given by

$$
F = \frac{[u]}{(gd)^{1/2}} = (\sin\alpha/f)^{1/2}.
\tag{2.46}
$$

2.4.2 Long wave and short wave approximation

To estimate some of these scales, we take $d = 2$ m, $u = 1$ m s^{-1} and $\sin\alpha = 0.001$, typical lowland valley values. We then have the length scale $s \sim 2$ km, and the time scale $t \sim 33$ minutes, and in some sense these are the natural length and time scales for the dynamic river response. However, it is fairly clear that these scales are not appropriate either for variations over the length of a whole river, or for the shorter length and time scales appropriate to waves generated by passage of a boat, for example. Both of these situations lead to further simplifications, as detailed below.

Long wave theory

Suppose we have a river of length 100 km, and we are concerned with the passage of a flood wave along its length. It is then appropriate to rescale s and t as

$$s \sim \frac{1}{\varepsilon}, \quad t \sim \frac{1}{\varepsilon}, \tag{2.47}$$

where in this instance $\varepsilon \sim 0.02 \ll 1$. In this case the equations (2.45) become

$$\begin{aligned} h_t + (uh)_s &= 0, \\ \varepsilon F^2(u_t + uu_s) &= 1 - u^2/h - \varepsilon h_s, \end{aligned} \tag{2.48}$$

and in the limit $\varepsilon \to 0$, we regain the slowly varying flow approximation.

Short wave theory

An alternative approximation is appropriate if length scales are much shorter than 2 km. This is often the case, and particularly in dynamically generated waves, as we discuss further below. In this case, it is appropriate to rescale length and time as

$$s \sim \frac{1}{\delta}, \quad t \sim \frac{1}{\delta}, \tag{2.49}$$

where $\delta \ll 1$, and then the model equations (2.45) become

$$\begin{aligned} h_t + (uh)_s &= 0, \\ F^2(u_t + uu_s) &= \delta \left(1 - \frac{u^2}{h}\right) - h_s, \end{aligned} \tag{2.50}$$

and when δ is put to zero, we regain the shallow water equations of fluid dynamics.

2.4.3 The monoclinal flood wave

One of the suggestions made at the end of section 2.3 was that the shocks predicted by the slowly varying flood wave theory would in reality be smoothed out by some higher order physical effect. This shock structure is called the monoclinal flood wave (because it is a monotonic profile), and it can be understood in the context of the long wave St. Venant theory (2.48). The simplest version is when $F \ll 1$ as well as $\varepsilon \ll 1$, for then we can approximate the momentum equation $(2.48)_2$ by the relation

$$u \approx h^{1/2}\left(1 - \tfrac{1}{2}\varepsilon h_s \ldots\right), \tag{2.51}$$

and $2.48)_1$ becomes

$$\frac{\partial h}{\partial t} + \tfrac{3}{2}h^{1/2}h_s \approx \tfrac{1}{2}\varepsilon \frac{\partial}{\partial s}\left(h^{3/2}\frac{\partial h}{\partial s}\right). \tag{2.52}$$

This is a convective diffusion equation much like Burgers' equation, and we expect it to support a monoclinal wave which provides a shock structure joining values h_- upstream to lower values h_+ downstream. We analyse this shock structure by writing

$$s = s_f + \varepsilon X, \tag{2.53}$$

where s_f is the flood wavefront, and X is a local coordinate within the shock structure. To leading order we then obtain the equation

$$-ch_X + \left[h^{3/2}\left(1 - \tfrac{1}{2}h_X\right)\right]_X = 0, \tag{2.54}$$

where $c = \dot{s}_f$ is the wave speed. Integrating this, we obtain

$$ch = h^{3/2}\left(1 - \tfrac{1}{2}h_X\right) + K, \tag{2.55}$$

where we require

$$K = ch_- - h_-^{3/2} = ch_+ - h_+^{3/2} \tag{2.56}$$

(which gives the shock speed determined in the usual way by the jump condition $c = \left[h^{3/2}\right]_-^+ / [h]_-^+$). Hence h is given by the quadrature

$$2X = \int_h^{h_0} \frac{h^{3/2}\,dh}{[ch - h^{3/2}] - K}, \tag{2.57}$$

where the arbitrary choice of $h_0 \in (h_+, h_-)$ simply fixes the origin of X. (2.57) can be simplified to give

$$X = \int_w^{w_0} \frac{w^4\,dw}{(w - w_+)(w_- - w)(w + C)}, \tag{2.58}$$

where $w = h^{1/2}$, and

$$C = \frac{w_+ w_-}{w_+ + w_-}, \tag{2.59}$$

and $X(w)$ can of course be evaluated.

Of particular interest is the small flood limit, in which $\Delta w = w_- - w_+$ is small. In this case $C \approx \frac{1}{2}w_+$, and h can be found explicitly, as the approximation

$$h = \left[\frac{h_+^{1/2} + h_-^{1/2}e^{-X/\Delta X}}{1 + e^{-X/\Delta X}}\right]^2, \tag{2.60}$$

where

$$\Delta X = \frac{2w_+^3}{3\Delta w} = \frac{4h_+^2}{3\Delta h} \tag{2.61}$$

is the shock width. A further simplification (because $\Delta h = h_- - h_+$ is small) is

$$h = h_+ + \frac{\Delta h\, e^{-X/\Delta X}}{1 + e^{-X/\Delta X}}. \tag{2.62}$$

In dimensional terms, the shock width is of order

$$\frac{d^2}{\Delta d \sin\alpha}, \tag{2.63}$$

where d is the depth. Following a storm, if a river of depth two metres and bedslope 10^{-3} rises by a foot (thirty centimetres), the shock width is about thirteen kilometres: not very shock like! Figure 2.4 shows the form of the monoclinal flood wave (as given by (2.60)).

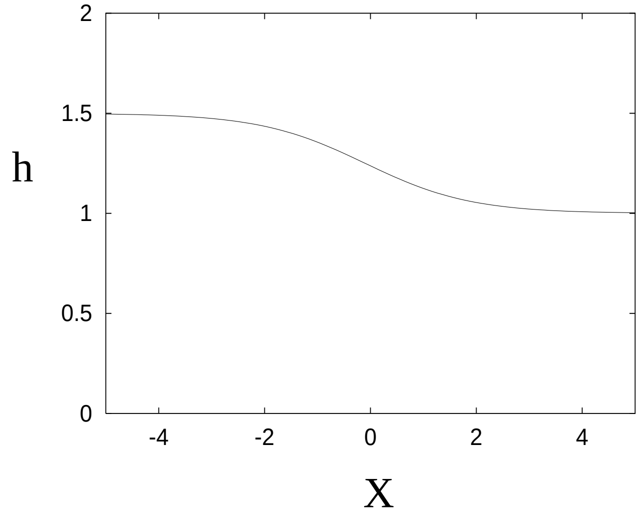

Figure 2.4: The monoclinal flood wave given by (2.60), with $h_- = 1.5$, $h_+ = 1$, $\Delta X = 1$.

2.4.4 Waves and instability

The monoclinal flood wave is one example of a river wave. More generally, we can expect disturbances to a uniformly flowing stream to cause waves to propagate, and in this section we study such waves. In particular, we will find that if the basic flow is sufficiently rapid, then disturbance waves will grow unstably. Such waves are commonly seen in fast flowing rivulets, for example on steep pavements during rainfall, and even on car windscreens.

To analyse waves on rivers, we take the basic river flow as being (locally) constant, thus in (2.45) (with $R = h$)

$$u = h = 1, \tag{2.64}$$

and we examine its stability by writing

$$u = 1 + v, \quad h = 1 + H, \tag{2.65}$$

and linearising. We obtain the linear system

$$\begin{aligned} H_t + H_s + v_s &= 0, \\ F^2(v_t + v_s) &= -2v + H - H_s, \end{aligned} \tag{2.66}$$

whence

$$F^2 \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s} \right)^2 v = -2 \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s} \right) v - v_s + v_{ss}. \tag{2.67}$$

Solutions $v = \exp[iks + \sigma t]$ exist, provided σ satisfies

$$F^2(\sigma+ik)^2+2(\sigma+ik)+ik+k^2=0, \tag{2.68}$$

or

$$\tilde{\sigma}=-i\tilde{k}-1\pm[1-i\tilde{k}-\tilde{k}^2/F^2]^{1/2}, \tag{2.69}$$

where we write

$$\sigma=\tilde{\sigma}/F^2, \quad k=\tilde{k}/F^2. \tag{2.70}$$

There are thus two wave like disturbances. The possibility of instability exists, if either value of $\tilde{\sigma}$ has positive real part. We define the positive square root in (2.69) to be that with positive real part. Specifically, we define

$$p+ikq=\left\{1-i\tilde{k}-\frac{\tilde{k}^2}{F^2}\right\}^{1/2}, \tag{2.71}$$

where we take $p>0$; thus, the real and imaginary parts of $\tilde{\sigma}$ are given by

$$\tilde{\sigma}_R=\pm p-1, \quad -\frac{\tilde{\sigma}_I}{\tilde{k}}=1\mp q, \tag{2.72}$$

and the criterion for instability is that $\tilde{\sigma}_R>0$, i. e., $p>1$. In this form, the growth rate of the wave is $\tilde{\sigma}_R/F^2$, while the wave speed is $-\tilde{\sigma}_I/\tilde{k}$. From (2.71), we find

$$q=-\frac{1}{2p}, \quad L(p)\equiv p^2-\frac{\tilde{k}^2}{4p^2}=1-\frac{\tilde{k}^2}{F^2}. \tag{2.73}$$

As illustrated in figure 2.5, $L(p)$ is a monotonically increasing function of p, and therefore the instability criterion $p>1$ is equivalent to $L(p)>L(1)$. Since p is determined by $L(p)=1-(\tilde{k}^2/F^2)$, while from (2.73), $L(1)=1-(\tilde{k}^2/4)$, we see that instability occurs if

$$F>F_c=2. \tag{2.74}$$

Thus, for tranquil flow, $F<O(1)$, the flow is stable. For rapid flow, $F>O(1)$, it can be unstable. The wave which goes unstable (when $p=1$) propagates downstream, because its wave speed is $1-q=\frac{3}{2}$, and in fact the $p>0$ wave always propagates downstream. The other wave, always stable, propagates downstream unless $1+q<0$, i. e., if and only if $p<1/2$, or equivalently,

$$F<F_-=\frac{2\tilde{k}}{(3+4\tilde{k}^2)^{1/2}}. \tag{2.75}$$

Note that F_- depends on $\tilde{k}$, and that $0<F_-<1$; we therefore have three distinct ranges for F:

$F>2$: two waves downstream, one unstable;

$1<F<2$: two waves downstream, both stable;

$F<1$: stable waves can propagate upstream and downstream.

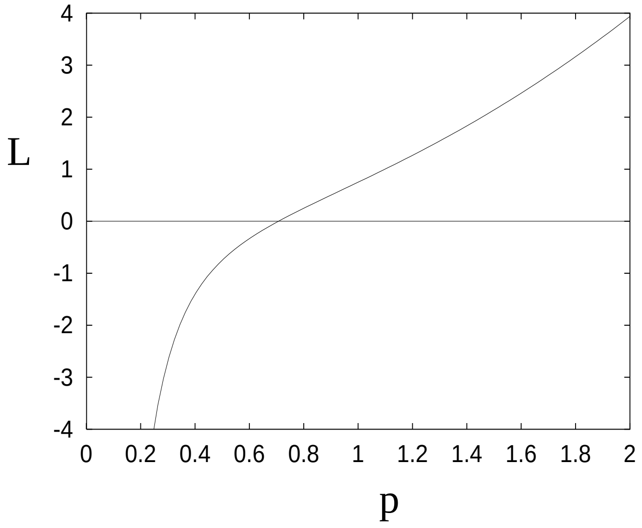

Figure 2.5: The function $L(p)$ defined by (2.73), with $\tilde{k} = 1$.

To go further than this requires a study of the nonlinear system (2.44). We see that the transition at $F = 1$ is associated with the ability of waves to propagate upstream. The transition at $F = 2$ is sometimes called Vedernikov instability, and is associated with the formation of downstream propagating *roll waves*.

2.5 Nonlinear waves

When $F > 2$, linear disturbances will grow, and nonlinear effects become important in limiting their eventual amplitude. Because of the hyperbolic form of the equations, we might then expect shocks to form. To examine this hyperbolic form, we put

$$\gamma = 1/F. \tag{2.76}$$

The equations are then

$$\begin{gathered} h_t + (hu)_s = 0, \\ u_t + uu_s + \gamma^2 h_s = \gamma^2 \left[1 - \frac{u^2}{h} \right], \end{gathered} \tag{2.77}$$

and they can be written in the form

$$\frac{\partial}{\partial t} \begin{pmatrix} h \\ u \end{pmatrix} + \begin{pmatrix} u & h \\ \gamma^2 & u \end{pmatrix} \frac{\partial}{\partial s} \begin{pmatrix} h \\ u \end{pmatrix} = \begin{pmatrix} 0 \\ \gamma^2 \left[1 - \dfrac{u^2}{h} \right] \end{pmatrix}. \tag{2.78}$$

2.5.1 Characteristics

The eigenvalues of $B = \begin{pmatrix} u & h \\ \gamma^2 & u \end{pmatrix}$ are given by

$$\lambda = u \pm \gamma h^{1/2}, \tag{2.79}$$

and the matrix P of eigenvectors and its inverse P^{-1} are given by

$$P = \begin{pmatrix} \sqrt{h} & \sqrt{h} \\ \gamma & -\gamma \end{pmatrix}, \quad P^{-1} = \frac{1}{2\gamma\sqrt{h}} \begin{pmatrix} \gamma & \sqrt{h} \\ \gamma & -\sqrt{h} \end{pmatrix}. \tag{2.80}$$

The integral

$$\int P^{-1} \begin{pmatrix} dh \\ du \end{pmatrix} = \int \begin{pmatrix} \dfrac{dh}{2\sqrt{h}} + \dfrac{du}{2\gamma} \\ \dfrac{dh}{2\sqrt{h}} - \dfrac{du}{2\gamma} \end{pmatrix} = \begin{pmatrix} \sqrt{h} + \dfrac{u}{2\gamma} \\ \sqrt{h} - \dfrac{u}{2\gamma} \end{pmatrix} \tag{2.81}$$

is well-defined, and determines the characteristic variables (the *Riemann invariants*, so called because they are constant on the characteristics in the absence of the forcing gravity and friction terms, as in shallow water theory). The equations can thus be compactly written in the characteristic form

$$\left[\frac{\partial}{\partial t} + (u \pm \gamma\sqrt{h})\frac{\partial}{\partial s}\right]\left[u \pm 2\gamma\sqrt{h}\right] = \gamma^2\left[1 - \frac{u^2}{h}\right]. \tag{2.82}$$

Nonlinear waves propagate downstream if $u/\gamma h^{1/2} > 1$, but one will propagate upstream if $u/\gamma h^{1/2} < 1$. This is consistent with the preceding linear theory (since $u/\gamma h^{1/2}$ is the local Froude number). Because the equations (2.82) are of second order, simple shock wave formation analysis is not generally possible. The equations (2.82) are very similar to those of gas dynamics, or the shallow water equations, and we expect the equations will support the existence of propagating shocks in a similar way.

2.5.2 Roll waves

There is a good deal of evidence that solutions of (2.77) do indeed form shocks, and these are called *roll waves*. They are seen in steep flows with relatively smooth beds (and thus low friction), but this combination is difficult to find in natural rivers. It is found, however, in artificial spillways, such as that shown in figure 2.6, which shows a famous photograph of roll waves propagating down a spillway in Switzerland. Roll waves can be found forming on any steep incline. Film flow down steep slopes during heavy rainfall will inevitably form a sequence of periodic waves, and these are roll waves; I used to see them frequently at my daughter's school, for example.

To describe roll waves, we seek travelling wave solutions to (2.77), in the form $h = h(\xi)$, $u = u(\xi)$, where $\xi = s - ct$ is the travelling wave coordinate, c being

Figure 2.6: Roll waves propagating down a spillway at the Grünnbach, Switzerland.

the wave speed. Substitution of these into (2.77) yields the two ordinary differential equations

$$\begin{aligned} -ch' + (uh)' &= 0, \\ -cu' + uu' &= 1 - \frac{u^2}{h} - \gamma^2 h'. \end{aligned} \tag{2.83}$$

The first equation has the integral

$$(u - c)h = -K, \tag{2.84}$$

where K is a positive constant. The reason that it must be positive is that the positive characteristics (those with speed $u + \gamma h^{1/2}$) must run into (not away from) the shock, that is,

$$u_+ + \gamma h_+^{1/2} < c < u_- + \gamma h_-^{1/2}, \tag{2.85}$$

where h_+ and h_- are the values of h immediately in front of and immediately behind the shock. Hence

$$\gamma h_+^{3/2} < K < \gamma h_-^{3/2}. \tag{2.86}$$

Substitution of (2.84) into the second equation yields a single first-order equation for u, or h. We choose to write the equation for h, thus

$$h' = \frac{h^3 - (ch - K)^2}{\gamma^2 h^3 - K^2}. \tag{2.87}$$

As indicated in figure 2.7, we aim to solve this equation in $(0, L)$, with $h = h_+$ at $\xi = 0$ and $h = h_-$ at $\xi = L$. The quantities involved in this equation and its boundary conditions are L, c, h_-, h_+ and K, and these have to be determined. Solution of the differential equation (2.87) from 0 to L yields one condition,

$$L = \int_{h_+}^{h_-} \frac{\gamma^2 h^3 - K^2}{h^3 - (ch - K)^2}\, dh, \tag{2.88}$$

which determines L in terms of the other quantities. Thus four extra conditions need to be specified to determine these.

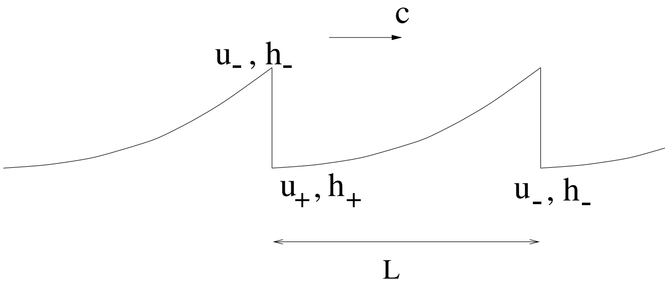

Figure 2.7: Schematic form of roll waves.

There are two jump conditions to apply across the shock. These are conservation of mass, which we omit, as it is automatically satisfied by (2.84), and conservation of momentum, which has the form

$$c = \frac{\left[hu^2 + \frac{1}{2}\gamma^2 h^2\right]_-^+}{[hu]_-^+}. \tag{2.89}$$

Simplification of this using (2.84) gives

$$\left[\tfrac{1}{2}\gamma^2 h^2 + \frac{K^2}{h}\right]_-^+ = 0. \tag{2.90}$$

Evidently, consideration of the graph of $\frac{1}{2}\gamma^2 h^2 + \dfrac{K^2}{h}$ shows that this determines h_+ in terms of h_-, for given K, see figure 2.8.

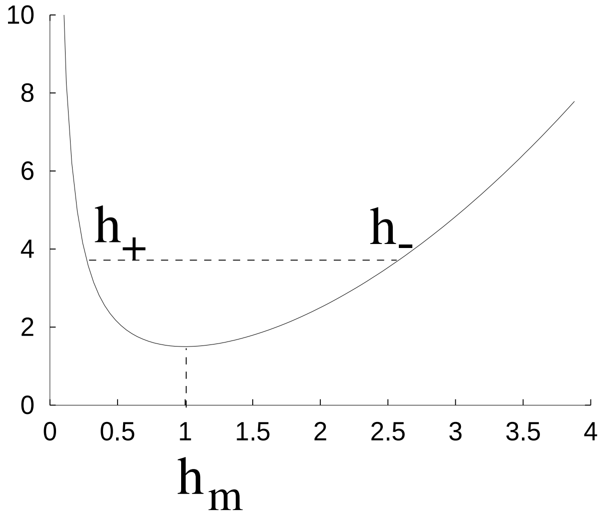

Figure 2.8: Supercritical and subcritical values of h across a shock: graph of $\frac{1}{2}\gamma^2h^2 + K^2/h$, $\gamma = K = 1$.

We denote the critical value of h at the minimum in figure 2.8 as h_m, thus

$$\gamma^2 h_m^3 = K^2; \tag{2.91}$$

clearly we must have $h_- > h_m$ and $h_+ < h_m$ (this is also implied by (2.86)), that is to say, the flow is subcritical behind the shock and supercritical in front of it. In particular, there is a value of $\xi \in (0, L)$ with $h = h_m$, and in order that the derivative in (2.87) remain finite, it is necessary that the numerator also vanish at this point. Since $K > 0$, this implies

$$ch_m - K = \frac{K}{\gamma}. \tag{2.92}$$

We have added an extra quantity h_m to the other unknowns L, h_-, h_+, K and c. To determine these six quantities, we have the four equations (2.88), (2.90), (2.91) and (2.92). This appears to imply that the roll waves described here form a two parameter family, with (for example) the wavelength and wave speed being arbitrary. This is at odds with our expectation that a sensibly described physical problem will have just the one solution. In order to understand this, we need to reconsider the hyperbolic form of the describing equations (2.77). A natural domain on which to solve these equations is the semi-infinite real axis $s > 0$, in which case appropriate boundary conditions are to prescribe h and u on $t = 0$ and $s = 0$. The initial conditions are prescribed to represent the experimental start-up, and the boundary conditions at $s = 0$ must represent the inlet conditions. The effect of the initial conditions is washed out of the system as the characteristics progress down stream, and the roll waves which are observed are determined by the boundary conditions at $s = 0$.

Of course, these inlet conditions are not generally consistent with a periodic travelling wave solution, but we would expect that prescribed values of u and h at the inlet would provide the extra two parameters to fix the solution precisely. One such parameter is easy to assess. Because mass is conserved, the mean volume flux

$$Q = \frac{1}{L}\int_0^L (ch - K)\, d\xi \tag{2.93}$$

must be equal to that at the inlet, and we can take Q in (2.93) to be prescribed, and in fact we can choose the velocity and depth scales so that $Q = 1$.

It is not as obvious how to provide the other recipe, because the mean momentum flux is not conserved downstream; its value at the inlet does not tell us its value downstream. This is because of the gravity and friction terms. However, it *is* the case that these terms must balance on average, that is to say,

$$\int_0^L (h - u^2)\, d\xi = 0; \tag{2.94}$$

this actually follows by integrating the momentum equation (written in conservation form) over a wavelength. The momentum advection and pressure gradient terms vanish because of (2.90), leaving (2.94). This appears to give a final condition to close the system: but it does not, as (2.94) actually reduces to (2.90) when the integration is carried out. An appropriate final condition will be described after first reducing the conditions above to simpler form.

We rewrite the relations (2.88), (2.90), (2.92) and (2.93) using h_m as the defining parameter, and putting

$$h_+ = h_m\phi_+, \quad h_- = h_m\phi_-; \tag{2.95}$$

then we have K and c given by

$$K = \gamma h_m^{3/2}, \quad c = h_m^{1/2}(1+\gamma), \tag{2.96}$$

and L, ϕ_+ and ϕ_- are determined, after some algebra, by

$$\begin{aligned} L &= \gamma^2 h_m \int_{\phi_+}^{\phi_-} \frac{(\phi^2+\phi+1)\, d\phi}{(\phi-\gamma)^2 - \gamma^2\phi}, \\ 1 &= \frac{\gamma^2 h_m^{5/2}}{L}\int_{\phi_+}^{\phi_-} \frac{(\phi^2+\phi+1)\{\phi+\gamma(\phi-1)\}\, d\phi}{(\phi-\gamma)^2-\gamma^2\phi}, \\ &\left[\tfrac{1}{2}\phi^2 + \frac{1}{\phi}\right]_-^+ = 0. \end{aligned} \tag{2.97}$$

The second of these can be written independently of L as

$$q = \frac{\displaystyle\int_{\phi_+}^{\phi_-} \frac{(\phi^2+\phi+1)\{\phi+\gamma(\phi-1)\}\, d\phi}{(\phi-\gamma)^2-\gamma^2\phi}}{\displaystyle\int_{\phi_+}^{\phi_-} \frac{(\phi^2+\phi+1)\, d\phi}{(\phi-\gamma)^2-\gamma^2\phi}}, \tag{2.98}$$

where

$$q = 1/h_m^{3/2}. \tag{2.99}$$

The profile of ϕ is given by the scaled version of (2.87), which is

$$\phi' = \frac{(\phi-\gamma)^2 - \gamma^2\phi}{\gamma^2 h_m(\phi^2+\phi+1)}. \tag{2.100}$$

The numerator must be positive, and since $\phi = 1$ for some ξ, a necessary condition for this to be true is that $\gamma < 1/2$. In terms of the Froude number, this is $F > 2$, which is the condition under which the roll wave instability occurs in the first place. This nicely suggests that the roll waves bifurcate as a non-uniform solution from the steady state at $F = 2$.

It is apparent from the above discussion that the crux of the determination of the roll wave parameters is the solution of $(2.97)_3$ and (2.98) for given positive q. If ϕ_+ and ϕ_- can be found for any such q, then they can be found for any h_m, after which L, K and c follow directly from (2.96) and $(2.97)_1$.

To find the solutions of $(2.97)_3$ and (2.98), we note that ϕ_+ and ϕ_- are uniquely defined in terms of the ordinate of the graph in figure 2.8; in fact, for any $\phi_+ \in (0,1)$, $(2.97)_3$ gives the explicit solution

$$\phi_- = \tfrac{1}{2}\left[-\phi_+ + \left\{\phi_+^2 + \frac{8}{\phi_+}\right\}^{1/2}\right]; \tag{2.101}$$

then (2.98) gives $q = q(\phi_+;\gamma)$. The other constants are then given explicitly by (2.96), $(2.97)_1$ and (2.98), and in particular, if we define

$$\begin{aligned} N(\phi_+) &= \int_{\phi_+}^{\phi_-} \frac{(\phi^2+\phi+1)\{\phi+\gamma(\phi-1)\}\,d\phi}{(\phi-\gamma)^2-\gamma^2\phi}, \\ D(\phi_+) &= \int_{\phi_+}^{\phi_-} \frac{(\phi^2+\phi+1)\,d\phi}{(\phi-\gamma)^2-\gamma^2\phi}, \end{aligned} \tag{2.102}$$

(thus $q = N/D$), then using

$$h_m = \left(\frac{D}{N}\right)^{2/3}, \tag{2.103}$$

we have

$$L = \frac{\gamma^2 D^{5/3}}{N^{2/3}}, \quad c = \frac{(1+\gamma)D^{1/3}}{N^{1/3}}. \tag{2.104}$$

The equations (2.101), (2.103) and (2.104) determine ϕ_-, h_m, L and c in terms of ϕ_+. The final relation which determines ϕ_+ must arise from the prescription of the boundary conditions at the inlet. In general, it is not obvious how this should be done; we will therefore restrict ourselves to the particular case where the inlet conditions are almost constant.

For the particular case of constant inlet conditions $u = h = 1$, the solution is uniform, and no waves develop. Let us then suppose that the inlet conditions at

$s = 0$ are $u = 1 + U(t)$ and $h = 1 + H(t)$, where the time average of U is zero, and also the time-averaged flux $\overline{uh} = 1$, and the perturbations H and U are very small.

In figures 2.9–2.11 we plot the wave height $h_m(\phi_- - \phi_+$, wavelength L and speed c (all dimensionless) as a function of the Froude number F.

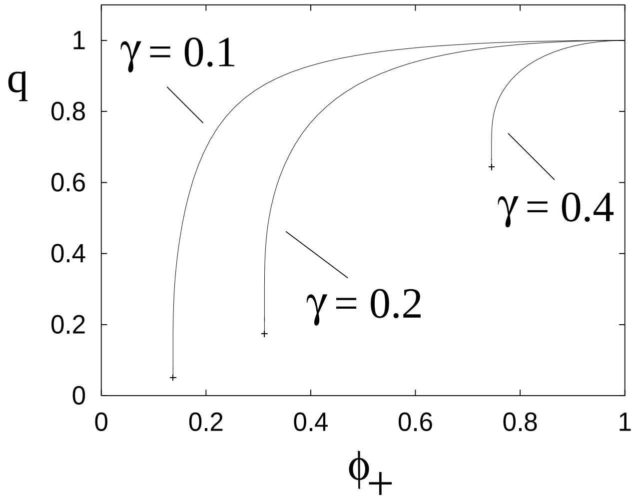

Figure 2.9: Graphs of q as a function of ϕ_+ for $\gamma = 0.4$ ($F = 2.5$), $\gamma = 0.2$ ($F = 5$), and $\gamma = 0.1$ ($F = 10$).

A feature of figure 2.9 is the termination of the curves at a finite value. The integrals which define N and D in (2.102) can be explicitly evaluated. If we define the two (positive) roots of $(\phi - \gamma)^2 - \gamma^2\phi = 0$ to be

$$\alpha_\pm = \frac{\gamma}{2}\left[2 + \gamma \pm \{\gamma^2 + 4\gamma\}^{1/2}\right], \tag{2.105}$$

thus $\alpha_+ > \alpha_- > 0$, then we restrict $\phi_+ > \alpha_+$ so that $\phi' > 0$ in (2.100). Consideration of N and D then shows that

$$D = -A\ln(\phi_+ - \alpha_+) + O(1), \quad N = -C\ln(\phi_+ - \alpha_+) + O(1) \tag{2.106}$$

as $\phi \to \alpha_+$. From this it follows that $q \to q_+$ as $\phi \to \alpha_+$, where $q_+ = C/A$, and is given explicitly by

$$q_+ = (1 + \gamma_+)\alpha_+ - \gamma. \tag{2.107}$$

These termination points are marked by a cross at the end of the curves in figure 2.9. Because $q = q_+ + O\left(\frac{1}{-\ln(\phi_+ - \alpha_+)}\right)$, the slope of the curves is infinite at these points. (This also makes it hard to draw the figures. To get within 0.02 of q_+, for example, we can expect to have to take $\phi_+ - \alpha_+ \approx \exp(-50) \approx 10^{-22}$!)

As $\phi_+ \to 1$, then also $\phi_- \to 1$, and hence both N and D are $O(1)$. Direct consideration of (2.102) shows that $q \to 1$ as $\phi_+ \to 1$. As a consequence of these

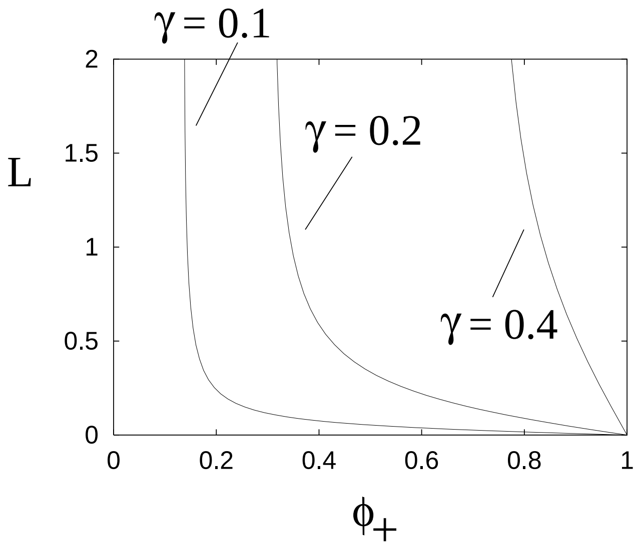

Figure 2.10: Wavelength L in terms of ϕ_+, with $Q = 1$.

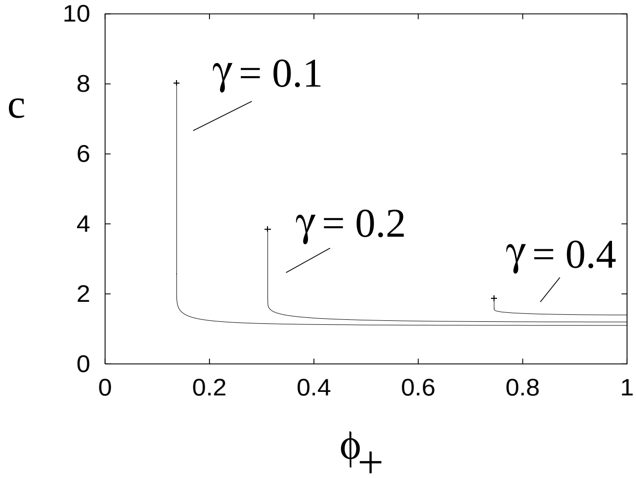

Figure 2.11: Wave speed c in terms of ϕ_+, with $Q = 1$.

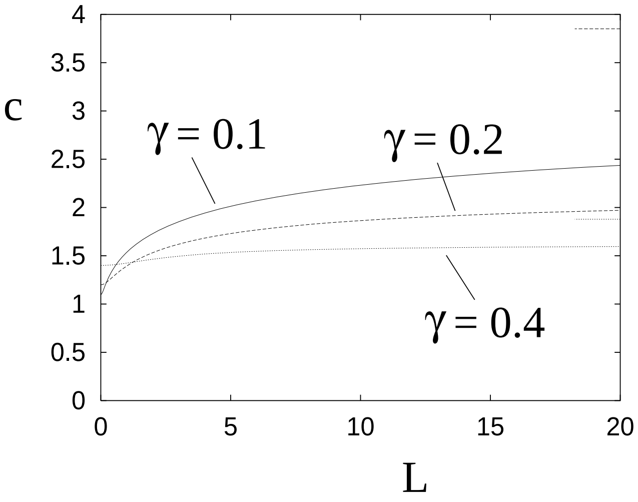

Figure 2.12: Wave speed c as a function of L, with $Q = 1$, for $\gamma = 0.1$, $\gamma = 0.2$ and $\gamma = 0.4$. The horizontal lines at the right indicate the corresponding asymptotes c_+ for $\gamma = 0.1$ and $\gamma = 0.2$. That for $\gamma = 0.4$ is not visible, since $c_+ \approx 8.03$ in that case.

limiting behaviours, $L \to 0$ and c is finite as $\phi_+ \to 1$, while $L \to \infty$ as $\phi_+ \to 1$, but c tends to a finite limit just as q does. As shown in figures 2.9–2.11, all three quantities vary monotonically between $\phi_+ = \alpha_+$ and $\phi_+ = 1$, and consequently c is a monotonically increasing function of L, which tends to a limit c_+ as $L \to \infty$, where

$$c_+ = \frac{(1+\gamma)Q^{1/3}}{q_+^{2/3}}. \tag{2.108}$$

This is shown in figure 2.12. Analysis of the limit $\phi_+ \to \alpha_+$ shows that $c = c_+ + O(1/L)$ as $L \to \infty$, but evidently the approach to the limit is extremely slow, particularly at low γ (high Froude number).

2.5.3 Tidal bores

A bore on a river is a shock-like wave which travels upstream, and it occurs because of forcing at the mouth of the river due to tidal variation in sea level. In England the best known example is the Severn bore, which occurs because of the very high tidal range in the Severn estuary. Large crowds come to view the bore, which manifests itself as a wall of water about a metre high advancing up river at a speed of some four to five metres a second. Figure 2.13 shows a photograph of the Severn bore. Bores occur on certain rivers due to a confluence of factors. The tidal range has to be very large, and this can be caused by tidal resonance in an estuary; in addition, the river must narrow dramatically upstream, so that the estuary acts like a funnel. The wave then forms because the rapidly rising water level in the estuary causes a large upstream water flux, and with a sufficiently large funnelling effect, a shock wave will

Figure 2.13: The Severn bore.

be formed. Bores occur all over the world, for example in the Amazon, the Seine, the Petitcodiac river which flows into the Bay of Fundy, and the Tsien Tang river in China. Where they occur, they are spectacular, but relatively few rivers have them, because of the severity of the necessary conditions for their formation.

Figure 2.14 shows the geometry of the Severn river and estuary. The bore forms near Sharpness, and is best viewed at various places further upstream, notably Minsterworth and Stonebench, where public access is available. Figure 2.15 shows a profile of the river during passage of a bore. There are certain features evident in this figure which are relevant when we formulate a model. The river depth at low stage is about a metre, whereas the tidal range is much greater than this. In the Severn estuary, it can be 14.5 metres, and at Sharpness, it is 9 metres in the figure. The other feature of importance is the apparent alteration in the bedslope as the estuary is approached. As an idealisation of this, figure 2.16 shows the basic geometry of a river–estuary system, which we can use to explain bore formation.

The river in figure 2.16 flows into a tidal basin, where the water level fluctuates tidally with a period of slightly more than twelve hours. Such fluctuations cause the river/estuary boundary point to migrate back and forth. In particular, approaching high tide this point moves upstream. The idea behind bore formation is that if the upstream velocity of this boundary is faster than the upstream characteristic wave speed,[2] a smooth wave cannot occur, and a shock must form, as indicated in figure 2.16.

We want to study this phenomenon in the context of the St. Venant equations

---

[2] We assume the Froude number F is less than one at low stage, which is the realistic condition; in that case, one wave travels upstream. If $F > 1$, a standing wave would form at the boundary.

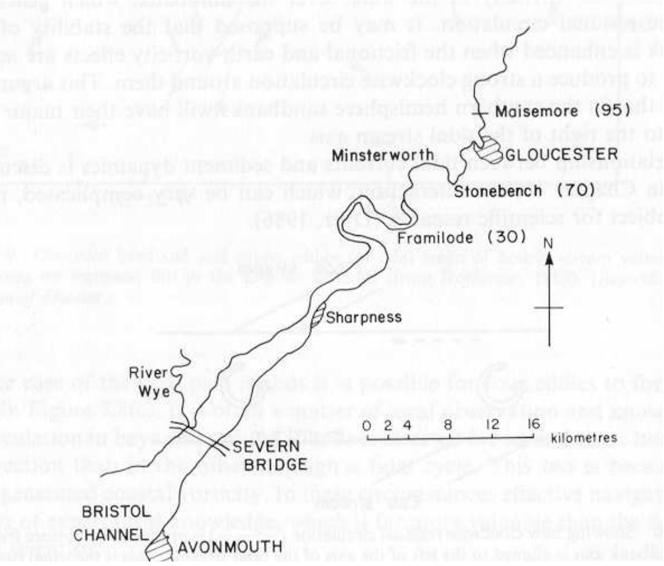

Figure 2.14: A sketch map of the river Severn.

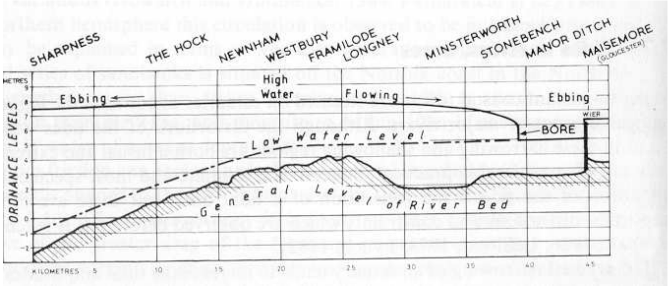

Figure 2.15: Profile of the Severn during passage of a bore.

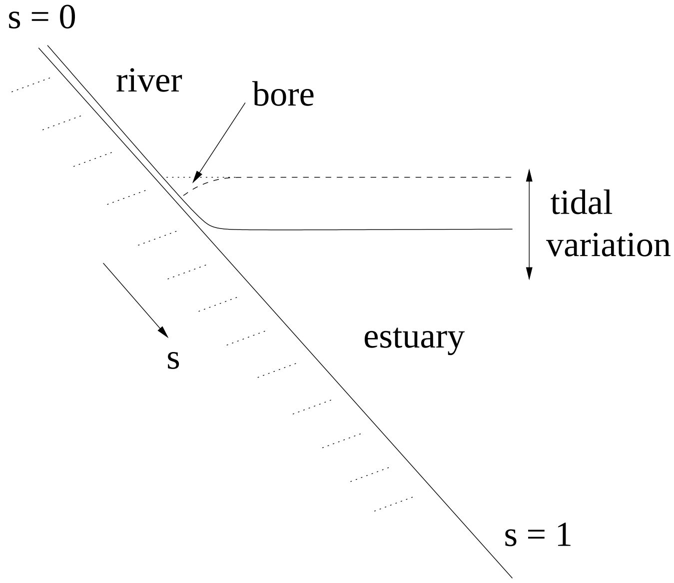

Figure 2.16: Idealised river basin geometry.

(2.44), where for a wide channel, we choose

$$R = h, \quad A = wh, \tag{2.109}$$

where w is the width, and is taken to be a prescribed function of s. The phenomenon of concern occurs over the length of the river, so that long wave theory is appropriate. From figure 2.15, a suitable length scale is of the order of 45 km, where the length scale used in writing (2.44) is $d/\sin\alpha$, and is 2 km if we take $d = 2$ m and $\sin\alpha = 10^{-3}$. If we take a typical velocity upstream as 2 m s^{-1}, then the corresponding time scale is 10^3 s, or 15 minutes, and the Froude number is about 0.3. The scale up in distance is thus of order 22, while that in time to the half-period of tidal oscillations is similar. This suggests that we rescale both time and space as

$$t \sim \frac{1}{\varepsilon}, \quad s \sim \frac{1}{\varepsilon}, \tag{2.110}$$

where a plausible value of ε may be of order 0.05. In this case (2.44) can be written in the form

$$\begin{aligned} wh_t + (wuh)_s &= 0, \\ \varepsilon F^2(u_t + uu_s) &= 1 - \frac{|u|u}{h} - \varepsilon h_s, \end{aligned} \tag{2.111}$$

or equivalently in the form

$$\varepsilon\left[\pm F\frac{\partial}{\partial t} + (\sqrt{h} \pm Fu)\frac{\partial}{\partial s}\right]\left[2\sqrt{h} \pm Fu\right] = 1 - \frac{|u|u}{h} \mp \frac{\varepsilon F w'\sqrt{h}u}{w}, \tag{2.112}$$

which shows explicitly that the characteristic wave speeds are

$$\pm\frac{\sqrt{h}}{F}+u. \tag{2.113}$$

Finally, we wish to study the situation shown in figure 2.16, where the tidal range is significantly larger than the river depth. The simplest choice is to suppose the tidal amplitude is also $O(1/\varepsilon)$, so that appropriate boundary conditions for (2.111) are

$$\begin{aligned} wuh &= 1 \quad \text{at} \quad s=0, \\ h &= \frac{H_1(t)}{\varepsilon} \quad \text{at} \quad s=1, \end{aligned} \tag{2.114}$$

representing a constant upstream volume flux, and a prescribed tidal range.

The assumption that $\varepsilon \ll 1$ allows us to solve (2.111) asymptotically. The solution has two parts, river and estuary, joined at a front which we denote by $s=s_f$. Upstream, for $s<s_f$, the flow is quasi-stationary, and we have, to leading order,

$$wuh \approx 1, \quad 1-\frac{u|u|}{h}\approx 1, \tag{2.115}$$

whence

$$u\approx w^{-1/3}, \quad h=w^{-2/3}. \tag{2.116}$$

The steady solution of (2.111) is appropriate, because the sub-characteristic wave propagates downstream, and after any initial transient, the upstream boundary condition leads to a steady flow.

Downstream, for $s>s_f$, we write

$$h=\frac{H}{\varepsilon}, \tag{2.117}$$

so that

$$\begin{aligned} wH_t+(wuH)_s &\approx 0, \\ 1-H_s &\approx 0, \end{aligned} \tag{2.118}$$

whence

$$H\approx s-1+H_1, \tag{2.119}$$

from which there follows

$$u\approx\frac{-\dot{H}_1\displaystyle\int_{s_f}^{s} w\,ds}{wH}, \tag{2.120}$$

where we choose the integration constant for matching purposes at s_f. Also to match the solution to that in $s<s_f$, we need to take

$$s_f=1-H_1. \tag{2.121}$$

Transition region

At the front, we define

$$s=s_f+\varepsilon X,\quad \dot{s}_f=c,\quad w_f=w\left[s_f(t)\right]; \tag{2.122}$$

then to leading order we have

$$\begin{aligned} -cw_fh_X+(w_fhu)_X &= 0,\\ F^2(u-c)u_X &= 1-\frac{u|u|}{h}-h_X, \end{aligned} \tag{2.123}$$

with boundary conditions

$$\begin{aligned} h\to w_f^{-2/3},\quad u\to w_f^{-1/3} \text{ as } X\to-\infty,\\ h\sim X,\quad u\sim c \text{ as } X\to\infty, \end{aligned} \tag{2.124}$$

in order to match to the upstream and downstream solutions. Only the conditions on h are necessary, those on u following automatically. A first integral of the mass conservation equation $(2.123)_1$ gives

$$(u-c)h=\kappa=\left[w_f^{-1/3}-c\right]w_f^{-2/3}, \tag{2.125}$$

and from this we find

$$h_X=\frac{h^3-|\kappa+ch|(\kappa+ch)}{h^3-\kappa^2F^2}. \tag{2.126}$$

This can be compared with (2.87). The difference in the present case is that κ and c are given, and the question is only whether a solution exists joining $h=h_-=w_f^{-2/3}$ upstream to the downstream solution $h\sim X$. Note that as $X\to-\infty$, $\kappa+ch\to w_f^{-1}$, so that $h\to w_f^{-2/3}$ can consistently be satisfied.

Let us suppose that the tide is coming in, thus $c<0$. From (2.113), we suspect there may be trouble if

$$-c>w_f^{-1/3}\left(\frac{1}{F}-1\right) \tag{2.127}$$

(assuming $F<1$). If we suppose that the opposite inequality holds, i. e., $-c<w_f^{-1/3}\left(\frac{1}{F}-1\right)$, then a little algebra shows that this is precisely the criterion that $h_-=w_f^{-2/3}>(\kappa F)^{2/3}$, i. e., the denominator of (2.126) is positive. To see that there is a solution of this problem in this case, we need to show that the numerator of (2.126) is also positive, for then h will increase indefinitely as required.

The numerator, N, is given by

$$N=\left\{h^3-w_f^{-2}\right\}-\left\{\left|w_f^{-1}+c\left(h-w_f^{-2/3}\right)\right|\left[w_f^{-1}+c\left(h-w_f^{-2/3}\right)\right]-w_f^{-2}\right\}. \tag{2.128}$$

Both expressions in curly brackets are zero when $h=h_-$ at $X=-\infty$; for h slightly greater than h_-, the left curly bracketed expression is positive, while the right curly

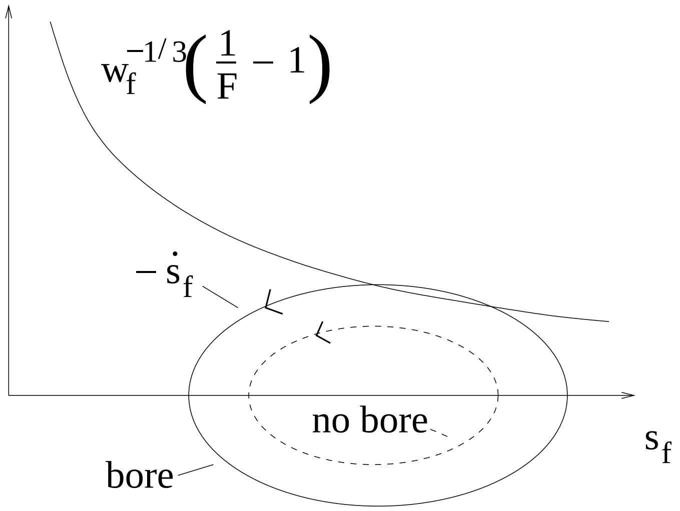

Figure 2.17: Bore formation occurs for large tides and rapidly widening rivers with reasonably sized Froude numbers.

bracketed expression decreases, since $c < 0$. The numerator is thus positive for $h - h_-$ small and positive, and remains so. From this it follows that a solution of the transition problem exists if $-c < w_f^{-1/3}\left(\dfrac{1}{F} - 1\right)$.

It remains to be shown that no solution exists if the opposite inequality, (2.127), holds. In this case the denominator is initially negative. As before, the numerator is positive if $h > h_-$, and equivalently negative if $h < h_-$, thus implying $h_X < 0$ if $h > h_-$, and $h_X > 0$ if $h < h_-$. This means solutions of (2.126) can only approach h_- as $X \to \infty$, and no transition solution exists. This suggests another form of solution, one in which a discontinuity forms at the critical condition

$$-\dot{s}_f = w(s_f)^{-1/3}\left(\frac{1}{F} - 1\right), \tag{2.129}$$

and thereafter propagates upstream as a shock front. This is the bore. Figure 2.17 shows a schematic illustration of the criterion (2.129) for bore formation.

Propagation of the bore

The outer river and estuary solutions (2.116), (2.119) and (2.120) remain valid after the formation of a shock, but the transition region is replaced by a shock at s_f, where the values of h_- and u_- (given by (2.116) with $w = w_f$) jump (up) to values h_+ and u_+, which have to be determined along with s_f. Initially h_+ and u_+ are $O(1)$, and as long as this remains true (which we suspect is all the time), we require that s_f is still given by

$$s_f = 1 - H_1 + O(\varepsilon); \tag{2.130}$$

the location of the bore is essentially determined by the tidal range. Jump conditions of mass and momentum across the developing bore then imply that the bore speed $-\dot{s}_f = v$ satisfies

$$-v = \frac{[hu]_-^+}{[h]_-^+} = \frac{\frac{1}{2}[h^2]_-^+ + F^2[hu^2]_-^+}{[hu]_-^+}, \tag{2.131}$$

and these two relations serve to determine h_+ and u_+, since $v = \dot{H}_1$.

2.6 Notes and references

Books on hydrology tend to be geographical in nature, describing the processes important in the hydrological cycle. For example, Chorley (1969) or Ward and Robinson (1990) are useful introductions. Books on hydraulics, on the other hand, concentrate on the fluid dynamics of the river flow itself. An example is the book by French (1984); an older classic is that by Ven te Chow (1959). A nice book which bridges the gap, and also includes discussion of sediment transport and channel morphology and pattern, is that by Richards (1982). A more detailed account of sediment transport is given by Allen (1985). Flood waves and roll waves have been discussed from the present perspective by Whitham (1974). The effect of tidal variations on river flow is discussed by Pugh (1987); in particular, he describes the phenomenon of the river bore.

Exercises

2.1 Find a relationship between the hydraulic radius R and the area A for triangular (notch shaped) or rectangular (canal shaped) cross sections. Hence show that Chezy's and Manning's laws both lead to a general relationship of the form

$$Q = \frac{cA^{m+1}}{m+1},$$

with $0 < m < 1$, giving explicit prescriptions for c and m. For a canal of depth h, show that the flow is turbulent if

$$h \gtrsim 10^2 \nu^{2/3} \left(\frac{f}{Sg}\right)^{1/3},$$

where ν is the kinematic viscosity, f is the friction factor, S is the slope and g is gravity. Taking $\nu = 10^{-6}$ m^2 s^{-1}, $f = 0.01$, $S = 10^{-3}$, $g = 10$ m s^{-2}, find a critical depth for turbulence. Is the Isis turbulent?

2.2 For flow in a pipe, the friction factor f in the formula $\tau = f\rho u^2$ is often taken to depend on the Reynolds number, for example, Blasius's law of friction has $f \propto Re^{-1/7}$. By taking $Re = UR/\nu$, where R is the hydraulic radius, find modifications to Chézy's law if $f \propto Re^{-\beta}$. Comment on whether you can obtain Manning's flow law this way.

2.3 The cross sectional area of a river A is assumed to satisfy the wave equation

$$\frac{\partial A}{\partial t}+cA^m\frac{\partial A}{\partial s}=0,$$

where s is distance downstream. Explain how this equation can be derived from the principle of conservation of mass. What assumptions does your derivation use?

A river admits a steady discharge $Q=Q_+$. At $t=0$, a tributary at $s=0$ is blocked, causing a sudden drop in discharge to $Q_- < Q_+$. Solve the equation for A using a characteristic diagram and show that an *expansion fan* branches from $s=0$, $t=0$. What is the hydrograph record at a downstream station $s=s_0>0$?

Later, the tributary is re-opened, causing a sudden rise from Q_- to Q_+. Draw the characteristic diagram, and show that a shock wave propagates forwards. What is its speed?

2.4 Use the method of characteristics to find the general solution of the equation describing slowly-varying flow of a river. Show also that in general shocks will form, and describe in what situations they will not. What happens in the latter case?

Either by consideration of an integral form of the conservation of mass equation, or by consideration from first principles, derive a jump condition which describes the shock speed. In terms of the local water speed, what is the speed of a shock (a) when it first forms; (b) when it advances over a dry river bed?

2.5 A river of rectangular cross section with width w carries a steady discharge Q_0 ($\mathrm{m}^3\ \mathrm{s}^{-1}$). At time $t=0$, a rainstorm causes a volume V of water to enter the river at the upstream station $s=0$. Assuming Chézy's law, find the solution for the resulting flood profile (sketch the corresponding characteristic diagram), and derive a (cubic) equation for the position of the advancing front of the flood. Without solving this equation, find an expression for the discharge Q_l at the downstream station $s=l$.

2.6 Derive the St. Venant equations from first principles, indicating what assumptions you make concerning the channel cross section. Derive a non-dimensional form of these equations assuming Manning's roughness law and a triangular cross section.

A sluice gate is opened at $s=0$ so that the discharge there increases from Q_- to Q_+. The hydrograph is measured at $s=l$. Using l as a length scale, and with a corresponding time scale $\sim l/u$, derive an approximate expression for the dimensionless discharge in terms of A, if the Froude number is small, and also $\varepsilon=[\bar{h}]/Sl\ll 1$, where $[\bar{h}]$ is the scale for the mean depth and S is the slope.

Hence show that A satisfies the approximate equation

$$\frac{\partial A}{\partial t} + \tfrac{4}{3}A^{1/3}\frac{\partial A}{\partial s} = \tfrac{1}{4}\varepsilon\frac{\partial}{\partial s}\left[A^{5/6}\frac{\partial A}{\partial s}\right].$$

What do you think the difference between the hydrographs for $\varepsilon = 0$ and $0 < \varepsilon \ll 1$ might be?

2.7 Why should the equation

$$A_t + cA^m A_s = M$$

represent a better model of slowly varying river flow than that with $M = 0$? Find the general solution of the equation, given that $A = 0$ at $s = 0$, and $A = A_0(s)$ at $t = 0$, $s > 0$. Find also the steady state solution $A_{eq}(s)$. How would you expect solutions representing disturbances to this steady profile to behave?

Suppose $A_0 = A_{eq} + \overline{A}\delta(s)$, representing an initial flood concentrated at $s = 0$; show that the resulting flood occurs in $s_- < s < s_+$, and show that the profile of A between s_- and s_+ is given implicitly by

$$A^{m+1} - (A - Mt)^{m+1} = \frac{(m+1)Ms}{c},$$

and deduce that

$$s_- = \frac{cM^m t^{m+1}}{(m+1)}.$$

What happens as $M \to 0$?

2.8 A dimensionless long wave model for slowly varying flow of a river of depth h and mean velocity u is given in the form

$$h_t + (uh)_s = M(s),$$

$$0 = 1 - \frac{u^2}{h} - \varepsilon h_s,$$

where $\varepsilon \ll 1$.

How would you physically interpret the positive source term $M(s)$?

Show that for small ε, the model can be reduced to the approximate form

$$h_t + (h^{3/2})_s = M(s) + \tfrac{1}{2}\varepsilon[h^{1/2}h_s]_s.$$

Show that if $h = 0$ at $s = 0$, then an approximate steady state solution is given by

$$h = \left\{\int_0^s M(s)\,ds\right\}^{2/3}.$$

Find this approximate solution if $M \equiv 1$, and show that it is in fact an exact solution of the differential equation. Write down the general solution of the differential equation in this case.

Explain why the condition of a horizontal water surface might be an appropriate boundary condition to apply at $s = 1$, and show that in terms of the scaled variables, this implies $h_s = 1/\varepsilon$ at $s = 1$. Show that with this added boundary condition, the approximate solution (when $M \equiv 1$) is still appropriate, except in a boundary layer near the outlet.

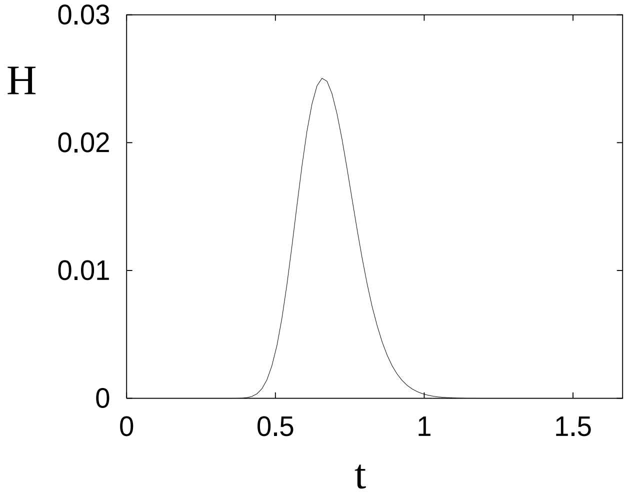

Figure 2.18: $H(s,t)$ plotted at fixed $s = 1$ as a function of t, using values $\varepsilon = 0.03$, $l = 0.005$, $\delta = 1$.

Next, suppose that $M = 0$ for large enough s, and that $\displaystyle\int_0^\infty M(s)\,ds = 1$. Write down the linear equation satisfied by small perturbations H to the steady state $h = 1$ when s is large.

By seeking solutions of the form $\exp[\sigma t + iks]$, show that small wave-like disturbances travel at speed $\frac{3}{2}$ and decay on a time scale $t \sim O(1/\varepsilon)$.

Show that if $\zeta = s - \frac{3}{2}t$, $\tau = \frac{1}{2}\varepsilon t$, then $H_\tau = H_{\zeta\zeta}$, and deduce that if $H = \delta \exp[-s^2/l^2]$ at $t = 0$, then

$$H = \delta \left(\frac{t_0}{t_0 + t} \right)^{1/2} \exp \left[\frac{-(s - \frac{3}{2}t)^2}{2\varepsilon(t_0 + t)} \right]$$

for $t > 0$, where $t_0 = \dfrac{l^2}{2\varepsilon}$. (A typical hydrograph described by this function is shown in figure 2.18. It is asymmetric, but the steep shock-like rise is limited by the linearity of the model.)

2.9 Using Chézy's law with a rectangular cross section, show how to non-dimensionalise the St. Venant equations, and show how the model depends on the Froude number, which you should define. Choose or guess suitable values for the Thames in London, the Isis/Cherwell in Oxford, an Alpine (or other) mountain stream, and determine the corresponding natural length and time scales, and the Froude number, for these flows. Show also that in the case of long wave and short wave motions, the equations effectively become those of slowly varying flow and the shallow water equations, respectively.

2.10 Derive the appropriate forms of the St. Venant equations assuming Manning's roughness law, and a triangular river cross section, and show in detail that small disturbances to the steady state can propagate up and down stream if $F < F_1$, but can only propagate downstream if $F > F_1$, and that they are unstable if $F > F_2$. What are the values of F_1 and F_2?

2.11 *The hydraulic jump*

Using the dimensionless form of the mass and momentum equations (for a canal), show that discontinuities (shocks) in the channel depth travel at a (dimensionless) speed V given by

$$V = \frac{[Au]_-^+}{[A]_-^+} = \frac{[F^2Au^2 + \frac{1}{2}A^2]_-^+}{[F^2Au]_-^+},$$

where ± refer to the values on either side of the jump, and F is the Froude number. Show that a stationary jump at $s = 0$ is possible (this can be seen when a tap is run into a basin) if $Au = Q$ in $s > 0$ and $s < 0$, and

$$\left[\frac{F^2Q^2}{A} + \frac{A^2}{2}\right]_-^+ = 0.$$

Deduce that for prescribed Q and A_-, a unique choice of $A_+ \neq A_-$ is possible. Show also that the locally defined Froude number is

$$Fr = \frac{FQ}{A^{3/2}},$$

and deduce that the hydraulic jump connects a region of *supercritical* ($Fr > 1$) flow to a *subcritical* ($Fr < 1$) one. (In practice, $A_- < A_+$ if $Q > 0$; if $A_- > A_+$, the discontinuity cannot be maintained.)

2.12 Evaluate the integrals to find explicit expressions for N and D in (2.102), and show that as $\phi_+ \to \alpha_+$,

$$D = -A\ln(\phi_+ - \alpha_+) + D_0 + o(1), \quad N = -C\ln(\phi_+ - \alpha_+) + N_0 + o(1),$$

and find explicit expressions for A, C, D_0 and N_0. Hence show that as $\phi_+ \to \alpha_+$,

$$\ln\left(\frac{1}{\phi_+ - \alpha_+}\right) \approx b(L + L^*) + O\left(\frac{1}{(L + L^*)}\right),$$

where the constant b should be determined, and deduce that

$$c \approx c_+ - \frac{k}{L+L^*} + O\left(\frac{1}{(L+L^*)^2}\right),$$

where k and L^* should be found. By evaluating k and L^* for different values of γ, show that both quantities increase rapidly as γ is reduced, and hence explain why the convergence of c to c_+ in figure 2.12 is so slow. Compare this asymptotic result with a direct numerical evaluation of $c(L)$. How good is the asymptotic result?

2.13 Why does the numerator have to be positive in (2.100)?

Mathematics and the environment

Part 2

Chapter 3

Dunes

The muddy colour of many rivers and the milky colour of glacial melt streams are due to the presence in the water of suspended sediments such as clay and silt. The ability of rivers to transport sediments in this way, and also (for larger particles) by rolling or saltation as *bedload* transport, forms an important constituent of the processes by which the Earth's topography is formed and evolved: the science of geomorphology.

Sediment transport occurs in a variety of different (and violent) natural scenarios. Powder flow avalanches, sandstorms, lahars and pyroclastic flows are all examples of violent sediment laden flows, and the kilometres long black sandur beaches of Iceland, laid down by deposition of ash-bearing floods issuing from the front of glaciers, are testimony to the ability of fluid flows to transport colossal quantities of sediment. In this chapter we will consider some of the landforms which are built through the interaction of a fluid flow with an erodible substrate; in particular we will focus on the formation of *dunes* and *anti-dunes* in rivers, and *Aeolian dunes* in deserts.

3.1 Patterns in rivers

There are two principal types of patterns which are seen in rivers. The first is a pattern of channel form, i. e. the shape taken by the channel as it winds through the landscape. This pattern is known as a meander, and an example is shown in figure 3.1.

The second type of pattern consists of variations in channel profile, and there are a number of variants which are observed. A distinction arises between profile variations transverse to the stream flow and those which are in the direction of flow. In the former category are bars; in the latter, dunes and anti-dunes. The formation of lateral bars results in a number of different types of river, in particular the braided and anastomosing river systems (described below).

All of these patterns are formed through an erosional instability of the uniform state when water of uniform depth and width flows down a straight channel. The instability mechanism is simply that the erosive power of the flowing water increases with water speed, which itself increases with water depth. Thus a locally deeper flow will scour its bed more rapidly, forming a positive feedback which generates

Figure 3.1: A meandering river. Photograph courtesy Gary Parker.

the instability. The different patterns referred to above are associated with different geometric ways in which this instability is manifested.

River meandering occurs when the instability acts on the banks. A small oscillatory perturbation to the straightness of a river causes a small secondary flow to occur transverse to the stream flow, purely for geometric reasons. This secondary flow is directed outwards (away from the centre of curvature) at the surface and inwards at the bed. As a consequence of this, and also because the stream flow is faster on the outside of a bend, there is increased erosion there, and this causes the bank to migrate away from the centre of curvature, thus causing a meander.

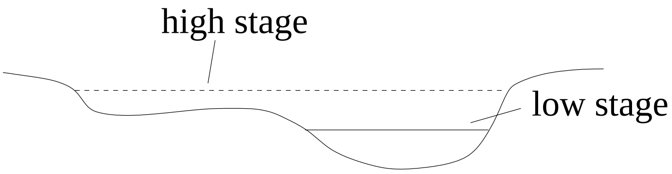

Figure 3.2: Cross section of a braided river with one lateral bar, which is exposed when the river is at low stage (i. e., the river level is low). The instability which causes the bar is operative in stormflow conditions, when the bar is submerged.

Braided rivers form because of a lateral instability which forms perturbations called bars. This is indicated schematically in figure 3.2. A deeper flow at one side of a river will cause excess erosion of the bed there, and promote the development of a lateral bar in stormflow conditions. The counteracting (and thus stabilising) tendency is for sediments to migrate down the lateral slope thus generated. Bars commonly form in gravel bed rivers, and usually interact with the meandering tendency to form alternate bars, which form on alternate sides of the channel as the flow progresses downstream. In wider channels, more than one bar may form across the channel, and the resulting patterns are called multiple row bars. In this case the stream at low stage is split up into many winding and connected braids, and the river is referred to as a braided river.

It is fairly evident that the scouring conditions which produce lateral bars and braiding only occur during bank full discharge, when the whole channel is submerged. Such erosive events are associated with major floods, and are by their nature occasional events. In between such floods, vegetation may begin to colonise the raised bars, and if there is sufficient time, the vegetative root system can stabilise the sediment against further erosion. A further stabilising effect of vegetation is that the plants themselves increase the roughness of the bed, thus diminishing the stress transmitted to the underlying sediment. If the bars become stably colonised by vegetation, then the braided channels themselves become stabilised in position, and the resulting set of channels is known as an anastomosing river system.

The final type of bedform is associated with waveforms in the direction of flow. Depending on the speed of the flow, these are called dunes or anti-dunes. At high values of the Froude number ($Fr > 1$), anti-dunes occur, and at low values ($Fr < 1$) dunes occur. A related feature is the ripple, which also occurs at low Froude number. Ripples are distinguished from dunes by their much smaller scale. Indeed, ripples and dunes often co-exist, with ripples forming on the larger dunes. The rest of this chapter focusses on models to describe the formation and evolution of dunes.

3.2 Dunes

Dunes are perhaps best known as the sand dunes of wind-blown deserts. They occur in a variety of shapes, which reflect differences in prevailing wind directions. Where wind is largely unidirectional, transverse dunes form. These are ridges which form at right angles to the prevailing wind. They have a relatively shallow upslope, a sharp crest, and a steep downslope which is at the limiting angle of friction for slip. The air flow over the dune separates at the crest, forming a separation bubble behind the dune. Transverse dunes move at speeds of metres per year in the wind direction.

Linear dunes, or seifs, form parallel to the mean prevailing wind, but are due to two different prevailing wind directions, which alternatively blow from one or other side of the dune. Such dunes propagate forward, often in a snakelike manner.

Other types of dunes are the very large star dunes (which resemble starfish), which form when winds can blow from any direction, and the crescentic barchan dunes, which occur when there is a limited supply of erodible fine sand. They take the shape of a crab-like crescent, with the arms pointing in the wind direction. Barchan dunes have been observed on Mars.

As already mentioned, dunes also occur extensively in river flow. At very low flow rates, ripples form on the bed, and as the flow rate increases, these are replaced by the longer wavelength and larger amplitude dunes. These are regular scarped features, whose steep face points downstream, and which migrate slowly downstream. They form when the Froude number $Fr < 1$ (the lower régime), and are associated with river surface perturbations which are out of phase, and of smaller amplitude. The wavelength of dunes is typically comparable to the river depth, the amplitude is somewhat smaller than the depth.

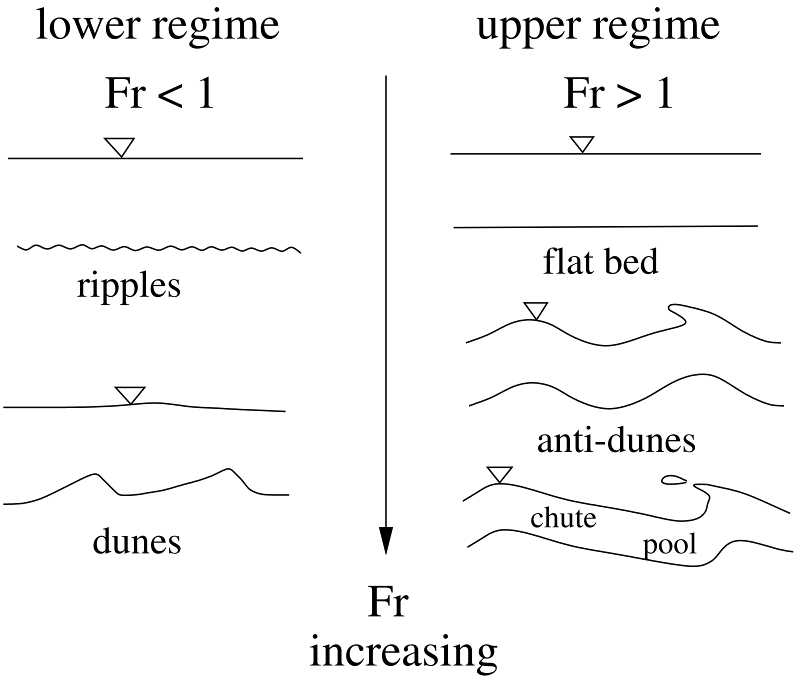

Figure 3.3: The succession of bedforms which are observed as the Froude number is increased. In the lower régime, where $Fr < 1$, we see first ripples and then the larger dune features. Surface perturbations are small. In the upper régime, $Fr > 1$; dunes disappear, giving a flat bed, and then anti-dunes are formed, in phase with surface waves. These are often transient features, occurring in flood conditions, and they are likely to be time dependent also.

When the Froude number increases further, the plane bed re-forms at $Fr \approx 1$, and then for $Fr > 1$, we obtain the upper régime, wherein anti-dunes occur. Whereas dunes are analogous to shock waves, anti-dunes are typically sinusoidal, and are in phase with the surface perturbations, which can be quite large. They may travel either upstream or (more rarely) downstream. Indeed, for the more rapid flows, backward breaking shocks occur at the surface, and chute and pool sequences form. Anti-dunes can be found on rapid outlet streams on beaches; for example I have seen them on beach streams on the West coast of Normandy, where the velocity is on the order of a metre per second, and the flow depth may be several centimetres. A common observed feature of such flows is their time dependence: anti-dunes form, then migrate upstream as they steepen, leading to hydraulic jumps and collapse of the pattern, only for it to re-form elsewhere The succession of bedforms as the Froude number increases is illustrated in figure 3.3. Anti-dunes do not form in deserts simply because the Froude number is never high enough.[1]

Dunes and anti-dunes clearly form through the erosion of the underlying bed, and thus mathematical models to explain them must couple the river flow mechanics with those of sediment transport. Sediment transport models are described below. There are two main classes of bedform models. The most simple and appealing is to combine the St. Venant equations with an equation for bedform erosion. There are two ways in which sediment transport occurs, as bedload or as suspended load. Each transport mechanism gives a different model, and we shall find that a suspended load transport model can predict the instability which forms anti-dunes, but not dunes, which indeed may occur in the absence of suspended sediment transport.[2] On the other hand, the St. Venant equations coupled with a simple model of bedload transport cannot predict instability, although such a model can explain the shape and speed of dunes.

The other class of model which has been used describes the variation of stream velocity with depth explicitly. One version employs potential theory, as is customarily done in linearised surface wave theory. At first sight, this appears implausible insofar as the flow is turbulent, and indeed the model can then only explain dunes when the bed stress is artificially phase shifted. In order to deal with this properly, it is necessary to include a more sophisticated description of turbulent flow, and this can be done using an eddy viscosity model, which is then able to explain dune formation. The issue of analysing the model beyond the linear instability régime is more difficult, and some progress in this direction is described in this chapter.

3.2.1 Sediment transport

Transport of grains of a cohesionless bed occurs as *bedload* or *in suspension.* At a given flow rate, the larger particles will roll along the bed, while the smaller ones are lifted by turbulent eddies into the flow. Clearly there is a transition between the two modes of transport: saltating grains essentially bounce along the bed.

Relations to describe sediment transport are ultimately empirical, though theory suggests the use of appropriate dimensionless groups. The basic quantity is the Shields stress (Shields 1936), defined as the dimensionless quantity

$$\tau^* = \frac{\tau}{\Delta\rho g D_s}. \tag{3.1}$$

Here τ is the basal shear stress, $\Delta\rho = \rho_s - \rho_w$ is the excess density of solid grains over water (ρ_s is the density of the solid grains, ρ_w is the density of water), g is gravity, and D_s is the grain size. In general, grain sizes are distributed, and the Shields stress depends on the particle size. The shear stress τ at the bed is usually related to the mean flow velocity u by the semi-empirical relation (2.13), i. e.,

$$\tau = f\rho_w u^2, \tag{3.2}$$

---

[1] The Froude number corresponding to a wind of 20 m s^{-1} = 45 miles per hour over a boundary layer depth of 1 km is 0.2.

[2] This also seems to be true of anti-dunes.

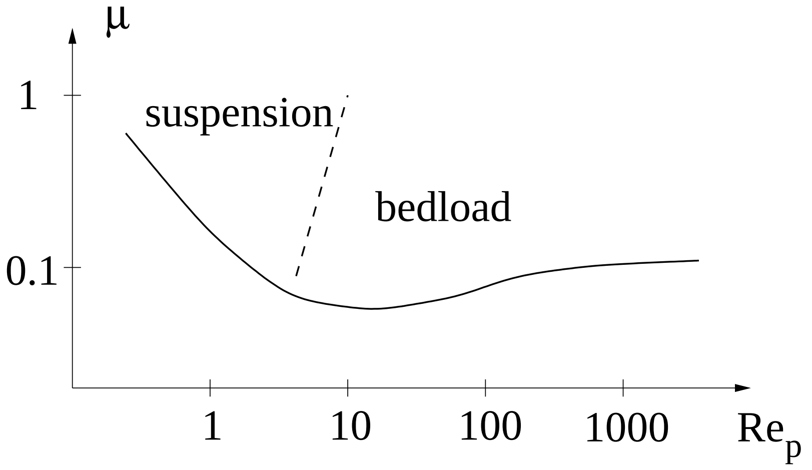

Figure 3.4: The critical Shields stress for the onset of sediment transport, weakly dependent on the particle Reynolds number $Re_p = u_* D_s/\nu$.

where f is a dimensionless friction factor, of typical value 0.01–0.1. (Larger values correspong top rougher channels.)

Shields found that sediment transport occurred if τ^* was greater than a critical value τ_c^*, which itself depends on flow rate via the particle Reynolds number

$$Re_p = \frac{u_* D_s}{\nu}; \tag{3.3}$$

(The friction velocity is defined to be

$$u_* = (\tau/\rho_w)^{1/2}.) \tag{3.4}$$

Figure 3.4 shows the variation of τ_c^* with $u_* D_s/\nu$; except at low flow rates, $\tau_c^* \approx 0.05$.

3.2.2 Bedload

Various recipes have been given for bedload transport, that due to Meyer-Peter and Müller (1948) being popular:

$$q^* = K[\tau^* - \tau_c^*]_+^{3/2}, \tag{3.5}$$

where $[x]_+ = \max(x, 0)$. Here $K = 8$, $\tau_c^* = 0.047$, and q^* is the dimensionless bedload transport rate (Einstein 1950), defined by

$$q^* = \frac{q_b}{(\Delta\rho g D_s^3/\rho_w)^{1/2}}, \tag{3.6}$$

q_b being the bedload measured as volume per unit stream width per unit time.

3.2.3 Suspended sediment

Suspended sediment transport is effected through a balance between an erosion flux v_E and a deposition flux v_D, each having units of velocity. The meaning of these is that $\rho_s v_E$ is the mass of sediment eroded from the bed per unit area per unit time, while $\rho_s v_D$ is the mass deposited per unit area per unit time.

3.2.4 Erosion

It is convenient to define a dimensionless erosion rate E via

$$v_E = v_s E, \tag{3.7}$$

where v_s is the particle settling velocity, given by Stokes's formula

$$v_s = \frac{\Delta \rho g D_s^2}{18\eta}, \tag{3.8}$$

η being the dynamic viscosity of water. Various expressions for E are due to Garcia (1989), Van Rijn (1984) and Smith and McLean (1977). They share the feature that E is a concave increasing function of basal stress. Typical is Van Rijn's relationship

$$E \propto (\tau^* - \tau_c^*)^{3/2} Re_p^{1/5}; \tag{3.9}$$

typical measured values of E are in the range $10^{-3} - 10^{-1}$.

3.2.5 Deposition

The calculation of deposition flux v_D is more complicated, as it is analogous to the calculation of basal shear stress in terms of mean velocity via an eddy viscosity model. We can write the dimensionless deposition flux D in the form

$$\rho_s v_D = v_s \bar{c} D, \tag{3.10}$$

where $\bar{c}$ is the mean column concentration of suspended sediment, measured as mass per unit volume of liquid, and D depends on a modified Rouse number $R = v_s/\varepsilon_T \bar{u}$. (Here ε_T is related to the eddy viscosity; specifically ε_T^{-1} is the Reynolds number based on the eddy viscosity, so the Rouse number is a Reynolds number based on particle fall velocity and eddy viscosity.) D increases with R, with $D(0) = 1$, and a typical form for D is

$$D = \frac{R}{1 - e^{-R}}. \tag{3.11}$$

3.3 The potential model

Kennedy (1963) introduced a model of fluid flow over an erodible bed, invoking potential flow for the fluid. We restrict our attention to two dimensional motion in the (x, z) plane: x is distance downstream, z is vertically upwards. The bed is at $z = s(x, t)$, the free water surface is at $z = \eta(x, t)$, so that the depth h is given by

$$h = \eta - s; \tag{3.12}$$

the geometry is shown in figure 3.5. In Kennedy's potential flow model, the usual equations for the fluid flow potential ϕ apply:

$$\begin{aligned} \nabla^2\phi &= 0 \text{ in } s < z < \eta, \\ \phi_z &= \eta_t + \phi_x\eta_x \quad \text{on} \quad z = \eta, \\ \phi_t + g\eta + \tfrac{1}{2}|\boldsymbol{\nabla}\phi|^2 &= \text{constant} \quad \text{on} \quad z = \eta, \\ \phi_z &= s_t + \phi_x s_x \text{ on } z = s. \end{aligned} \tag{3.13}$$

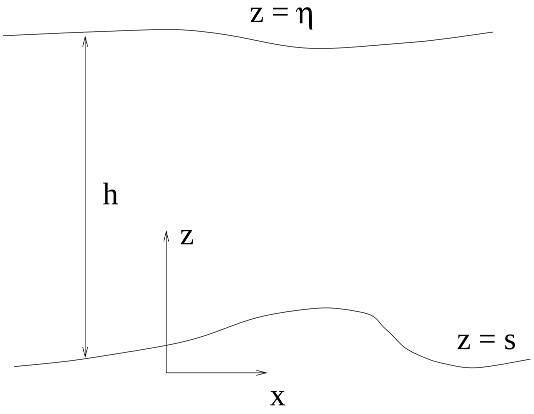

Figure 3.5: Geometry of the problem.

The extra equation required to describe the evolution of s is the *Exner equation*:

$$(1 - n)\frac{\partial s}{\partial t} + \frac{\partial q_b}{\partial x} = 0, \tag{3.14}$$

where n is the porosity of the bed; this assumes bedload transport only, and we may take (see equations (3.5) and (3.2)) $q_b = q_b(u)$, where $q_b'(u) > 0$. Implicitly, we suppose a (turbulent) boundary layer at the bed, wherein the basal stress develops through a shear layer; the basal shear stress will then depend on the outer flow velocity. We define

$$q = \frac{q_b}{1 - n} \tag{3.15}$$

and following Kennedy, we take

$$q = q\left(u|_{x-\delta}\right), \tag{3.16}$$

that is to say, u is evaluated at $x - \delta$ and $z = s$, where the phase lag δ is included to model the notion that in shear flow over a boundary, such a lag is indeed present

(Benjamin 1959). Of course (3.16) is a crude and possibly dangerous way to model this effect.

To examine the linear stability of a uniform steady state we write $s = 0$, $\eta = h$,

$$\phi = ux + \Phi, \quad q = q(u) + Q, \quad \eta = h + \zeta, \tag{3.17}$$

and then linearise the equations and boundary conditions (which are applied at the unperturbed boundaries $z = 0$ and $z = h$) to obtain

$$\begin{aligned} \nabla^2\Phi &= 0 \ \text{ in } \ 0 < z < h; \\ \Phi_z &= \zeta_t + u\zeta_x, \\ \Phi_t + g\zeta + u\Phi_x &= 0 \quad \text{on} \quad z = h; \\ \Phi_z &= s_t + us_x, \\ s_t + Q_x &= 0 \quad \text{on} \quad z = 0, \end{aligned} \tag{3.18}$$

where

$$Q = q'(u)\Phi_x|_{x-\delta, z=0} \,. \tag{3.19}$$

For a mode of wavenumber k, we put

$$(\zeta, s, Q) = (\bar{\zeta}, \bar{s}, \bar{Q}) \times e^{ikx+\sigma t}, \tag{3.20}$$

and write

$$\Phi = e^{ikx+\sigma t}[A\cosh kz + B\sinh kz], \tag{3.21}$$

so that the boundary conditions together with (3.19) become

$$\begin{gathered} k[A\sinh kh + B\cosh kh] = (\sigma + iku)\bar{\zeta}, \\ (\sigma + iku)[A\cosh kh + B\sinh kh] + g\bar{\zeta} = 0, \\ kB = (\sigma + ik)\bar{s}, \\ \sigma\bar{s} + ik\bar{Q} = 0, \\ Q = q'ike^{-ik\delta}A. \end{gathered} \tag{3.22}$$

Some straightforward algebra leads to

$$\sigma k[(\sigma + iku)^2 + gk\tanh kh] + \sigma + iku)k^2q'e^{-ik\delta}[(\sigma + iku)^2\tanh kh + gk] = 0, \tag{3.23}$$

a cubic for $\sigma(k)$.

Solution of this is facilitated by the observation that we can expect two modes to correspond to upstream and downstream water wave propagation, while the third corresponding to erosion of the bed may be much smaller, basically if q_b is sufficiently small. Specifically, let us assume (realistically) that $q \ll hu$. Then $q' \ll h$, and for small q', the roots of (3.23) are approximately the (stable) wave modes

$$\frac{\sigma}{-ik} \approx u \pm (\frac{g}{k}\tanh kh)^{1/2}, \tag{3.24}$$

and the erosive mode

$$\sigma \approx -k^2uq'[\sin k\delta + i\cos k\delta]\tanh kh \frac{\left[F^2 - \dfrac{\coth kh}{kh}\right]}{\left[F^2 - \dfrac{\tanh kh}{kh}\right]}, \tag{3.25}$$

where we define the Froude number by

$$F^2 = \frac{u^2}{gh}. \tag{3.26}$$

For the erosive mode, the growth rate is

$$\operatorname{Re}\sigma = -k^2uq'\sin k\delta\tanh kh \frac{\left[F^2 - \dfrac{\coth kh}{kh}\right]}{\left[F^2 - \dfrac{\tanh kh}{kh}\right]}, \tag{3.27}$$

and the wave speed is

$$-\frac{\operatorname{Im}\sigma}{k} = kuq'\cos k\delta\tanh kh \left[\frac{F^2 - \dfrac{\coth kh}{kh}}{F^2 - \dfrac{\tanh kh}{kh}}\right]. \tag{3.28}$$

This gives us the typical instability diagram shown in figure 3.6. For $\delta < 0$ the regions above and below the two curves are unstable, corresponding to dunes and anti-dunes. The curves are given by $F^2 = (\coth kh)/kh$ and $F^2 = (\tanh kh)/kh$, respectively.

The phase relation between surface and bed for the erosive bed is given by

$$\frac{\bar{\zeta}}{\bar{s}} \approx \frac{F^2\operatorname{sech} kh}{\left[F^2 - \dfrac{\tanh kh}{kh}\right]}, \tag{3.29}$$

and this defines wave forms below the lower curve in figure 3.6 as dunes, and those above as antidunes.

Figure 3.6 is promising, at least if $\sin k\delta < 0$, as it will predict both dunes and anti-dunes. To get the wave speed positive, we need in fact to have $\cos k\delta > 0$, thus $0 > k\delta > -\pi/2$ (we can take $-\pi < k\delta < \pi$ without loss of generality), whereas we would generally want $k\delta < -\pi/2$ for anti-dunes to migrate backwards.

There is a serious problem with this model, beyond the fact that the phase shift δ is arbitrarily included. The spatial delay is unlikely to provide a feasible model for nonlinear studies; indeed, we see that $\operatorname{Re}\sigma \sim k^2$ at large k, and in the unstable régime this is one of the hallmarks of ill-posedness.

Having said that, it will indeed turn out to be the case that a phase lead ($\delta < 0$) really is the cause of instability. A phase lead means that the stress, and thus the bedload transport, takes its maximum value on the upstream face of a bump in the

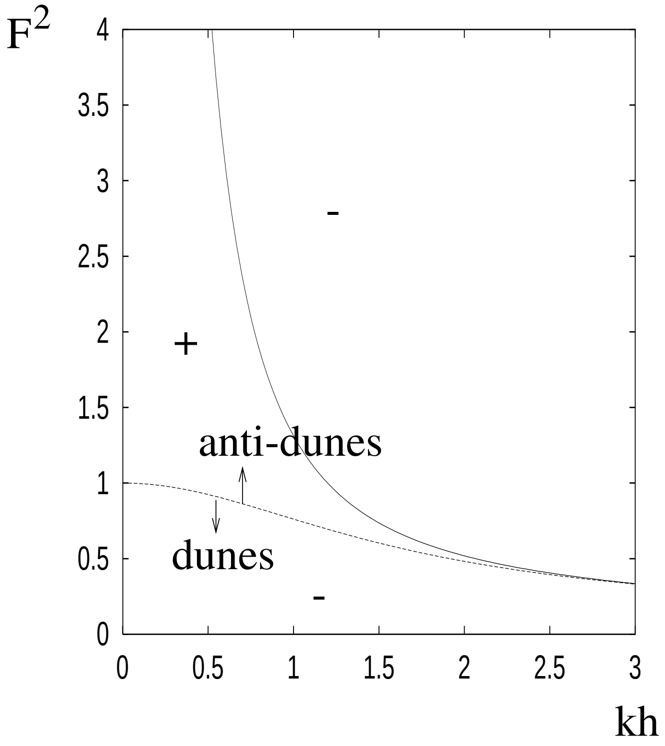

Figure 3.6: Instability diagram for the potential flow model. The regions marked with a minus sign, above the upper curve and below the lower curve, are regions of instability if $\delta < 0$, more specifically if $\sin k\delta < 0$. The marked distinction between dunes and anti-dunes is based on the surface/bed phase relation (see (3.29). Wave motion is downstream if $\cos k\delta > 0$, upstream if $\cos k\delta < 0$.

bed. A phase lead will occur because of the effect of the bump on the turbulent velocity structure above, as we discuss further below. It can also occur through an effect of bedload inertia (Parker 1975) (see also question 3.5).

The choice of wave speed in this theory is unclear, since $\cos k\delta$ can be positive or negative. The possibly more likely choice of a positive value implies positive wave speed.

3.4 St. Venant type models

Since river flow is typically modelled by the St. Venant equations, it is natural to try using such a model together with a bed erosion equation to examine the possibility of instability. This has the added advantage of being more naturally designed for nonlinear studies. A St. Venant/Exner model was studied by Reynolds (1965). The model can be written in the form

$$s_t + q_x = 0,$$

$$\begin{aligned} h_t + (uh)_x &= 0, \\ u_t + uu_x &= gS - \frac{fu^2}{h} - g\eta_x, \end{aligned} \tag{3.30}$$

where S is the downstream slope, $q = q(\tau)$ and $\tau = f\rho_w u^2$, and $\eta - s = h$. It is convenient to take advantage of the limit $q \ll hu$, just as we did before, and we do so by first non-dimensionalising the equations. We choose scales as follows:

$$s, x, h, \eta \sim h_0, \quad u \sim u_0, \quad q \sim q_0, \quad t \sim h_0^2/q_0, \tag{3.31}$$

and we choose h_0, u_0 by balancing terms as follows: $uh \sim Q_0$, $gS \sim fu^2/h$; here Q_0 is the (prescribed) volume flow per unit width. We choose q_0 as the size of the bedload transport equation in (3.5).

With these scales, the dimensionless equations corresponding to (3.30) are

$$\begin{aligned} s_t + q_x &= 0, \\ \varepsilon h_t + (uh)_x &= 0, \\ F^2(\varepsilon u_t + uu_x) &= -\eta_x + \delta(1 - \frac{u^2}{h}), \\ h &= \eta - s, \end{aligned} \tag{3.32}$$

where the parameters are

$$F = \frac{u_0}{\sqrt{gh_0}}, \quad \varepsilon = \frac{q_0}{Q_0}, \quad \delta = S. \tag{3.33}$$

If we now suppose $\varepsilon \ll 1$ and $\delta \ll 1$, both of them realistic assumptions, then we have approximately

$$\begin{aligned} uh &= 1, \\ \tfrac{1}{2}F^2u^2 + \eta &= \tfrac{1}{2}F^2 + 1, \end{aligned} \tag{3.34}$$

supposing that $u, h \to 1$ at large distances. Eliminating h and η, we have

$$s = 1 - \frac{1}{u} + \tfrac{1}{2}F^2(1 - u^2), \tag{3.35}$$

whose form is shown in figure 3.7. In particular, $s'(1) = (1 - F^2)$, so the basic state $u = 1$ corresponds to the left hand or right hand root of $s(u)$ depending on whether $F < 1$ or $F > 1$.

We also have

$$\frac{ds}{d\eta} = \frac{F^2 - h^3}{F^2}, \tag{3.36}$$

so that small perturbations to $h = 1$ are out of phase (dunes) if $F < 1$ and in phase (anti-dunes) if $F > 1$. If we take the dimensionless bedload transport as $q \approx \tau^{3/2} = u^3$ (the dimensionless basal stress having been scaled with $f\rho_w u_0^2$), so that $u = q^{1/3}$, then we see from (3.35) that $s = s(q)$, and $s(q)$ has the same shape as $s(u)$, as shown in figure 3.7.

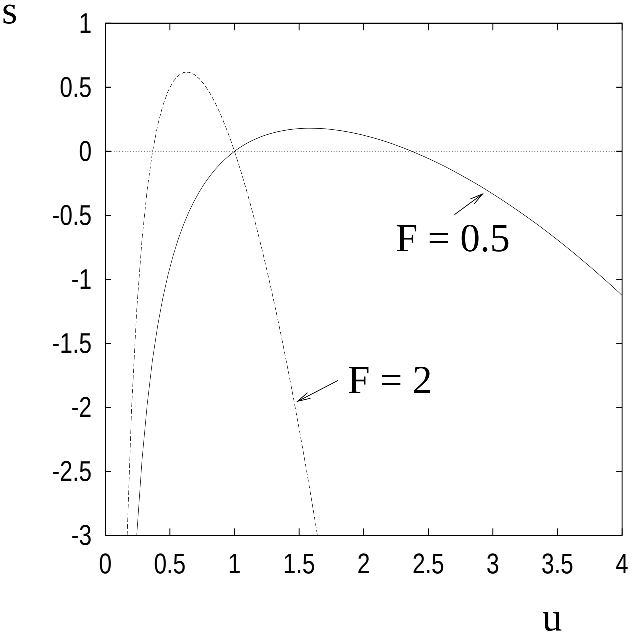

Figure 3.7: $s(u)$ as given by (3.35) for two typical cases of rapid and tranquil flow.

The whole model reduces to the single first order equation

$$s'(q)q_t + q_x = 0. \tag{3.37}$$

Disturbances to the uniform state $q = 1$ will propagate at speed $v(q) = 1/s'(q)$, where v is shown in figure 3.8. For $F < 1$, $v(1) > 0$ and $v'(1) > 0$, thus waves in q (and thus s) propagate downstream and form forward facing shocks; this is nicely consistent with dunes. For $F > 1$, $v < 0$ and $v'(q)$ is normally positive. Backward facing shocks form, these are elevations in s if $v' > 0$.

Unfortunately, the hyperbolic equation does not admit instability. It is straightforward to insert a lag as before, by writing $q(x,t) = q[s(x-\delta,t)]$, or equivalently $s(x,t) = s[q(x+\delta,t)]$. Perturbation of

$$\begin{aligned} s_t + q_x &= 0, \\ q &= q[s(x-\delta,t)], \end{aligned} \tag{3.38}$$

via

$$s = \bar{s}e^{ikx+\sigma t}, \quad q = 1 + \bar{q}e^{ikx+\sigma t}, \tag{3.39}$$

leads to

$$\begin{aligned} \sigma\bar{s} + ik\bar{q} &= 0, \\ \bar{q} &= q'e^{-ik\delta}\bar{s}, \end{aligned} \tag{3.40}$$

and thus

$$\sigma = kq'[-\sin k\delta - i\cos k\delta]. \tag{3.41}$$

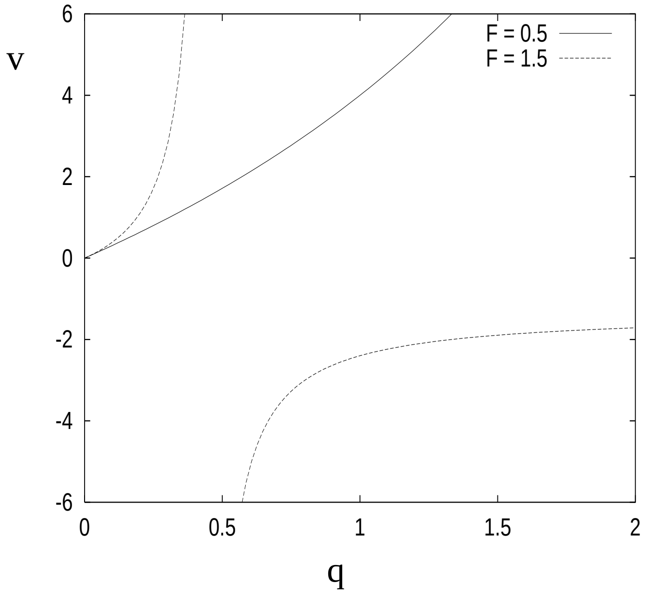

Figure 3.8: The wave speed $v(q) = 3q^{4/3}/(1 + F^2)$ for the tranquil and rapid cases $F = 0.5$ and $F = 1.5$.

This requires $\delta < 0$ for instability if $q'(s) > 0$ ($F < 1$) and $\delta > 0$ if $q'(s) < 0$ ($F > 1$). The long wavelength limit of (3.25) in which $kh \to 0$ is precisely (3.41), bearing in mind that (3.25) is dimensional and that $q' = dq/du$ there, whereas $q' = dq/ds$ in (3.41).

3.5 A suspended sediment model

The shortcoming of both the potential model and the St. Venant/Exner model is the lack of a genuine instability mechanism. We now show that the inclusion of suspended load can produce instability. Ideally, we would hope to predict anti-dunes, since dunes certainly do not require suspended sediment transport. A St. Venant model including both bedload and suspended sediment transport is

$$
\begin{aligned}
h_t + (uh)_x &= 0, \\
u_t + uu_x &= g(S - \eta_x) - \frac{fu^2}{h}, \\
\frac{\partial}{\partial t}(hc) + \frac{\partial}{\partial x}(hcu) &= \rho_s(v_E - v_D), \\
(1-n)\frac{\partial s}{\partial t} + \frac{\partial q_b}{\partial x} &= -(v_E - v_D),
\end{aligned}
\tag{3.42}
$$

where c is the column average concentration (mass per unit volume) of suspended sediment (written as $\bar{c}$ earlier). The distinction between suspended sediment transport and bedload lies in the source terms due to erosion and deposition, v_E and v_D, and it is these which will enable instability to occur. We have $\eta - s = h$, and we suppose $q_b = q_b(\tau)$, $\tau = f\rho_w u^2$, whence $q = q(u)$. Additionally (see (3.7) and (3.10)), we write

$$v_E = v_s E, \quad \rho_s v_D = v_s c D, \tag{3.43}$$

and expect that $E = E(u)$ and $D = D(u)$, with $E' > 0$, $D' < 0$; typically $E < 1$, $D > 1$.

We scale (3.42) as before in (3.31), except that we choose the time scale t_0, downstream length scale x_0, and concentration scale c_0 via

$$c_0 = \rho_s \frac{E_0}{D_0}, \quad t_0 = \frac{(1-n)h_0}{v_s E_0}, \quad x_0 = \frac{Q_0}{v_s D_0}, \tag{3.44}$$

where we write

$$E = E_0 E^*(u/u_0), \quad D = D_0 D^*(u/u_0), \tag{3.45}$$

and choose E_0 and D_0 so that E^* and D^* are $O(1)$, and so that these are consistent with typical observed suspended loads of 10 g l^{-1}. With this choice of scales, we obtain the dimensionless set of equations

$$\begin{aligned} \eta - s &= h, \\ \varepsilon h_t + (uh)_x &= 0, \\ F^2(\varepsilon u_t + uu_x) &= \delta\left(1 - \frac{u^2}{h}\right) - \eta_x, \\ h(\varepsilon c_t + uc_x) &= E^* - cD^*, \\ s_t + \beta q_x &= -(E^* - cD^*), \end{aligned} \tag{3.46}$$

where the parameters ε, F, δ and β are now given by

$$\begin{aligned} \varepsilon &= \frac{E_0}{(1-n)D_0} = \frac{c_0}{\rho_s(1-n)}, \quad \delta = \frac{u_0 S}{v_s D_0}, \\ F &= \frac{u_0}{(gh_0)^{1/2}}, \quad \beta = \frac{q_{b0} D_0}{Q_0 E_0} = \frac{\rho_s q_{b0}}{c_0 Q_0}. \end{aligned} \tag{3.47}$$

Here q_{b0} is the scale for q_b rather than $q = q_b/(1-n)$. The Froude number is the same as before, but the parameters ε and δ are different: ε is a measure of the suspended sediment density relative to the bed density, and is always small; δ is the ratio of the (small) bed slope to the ratio of settling velocity to stream velocity. For more rapidly flowing streams, we might expect $\delta \sim 1$. However, if we suppose that wavelengths of anti-dunes are comparable to the depth (so $x_0 \sim h_0$), then (3.44) implies $\delta \sim S \ll 1$. Thus $\delta \sim 1$ implies $x_0 \sim h_0/S \gg h_0$. The parameter β is a direct measure of the ratio of bedload ($\rho_s q_{b0}$) to suspended load ($c_0 Q_0$). For $\beta \gg 1$, we would revert to our preceding bedload model and its scaling, and neglect the suspended load. If we adopt the Meyer-Peter/Müller relation in (3.5) and (3.6), then (noting that $fu_0^2 = gSh_0$)

$$q_{b0} = \frac{K\rho_l}{\Delta\rho}(ghS)^{3/2}, \tag{3.48}$$

and we can write

$$\beta = \left\{ \frac{K\rho_l}{(1-n)\Delta\rho} \right\} \frac{S^{3/2}}{\varepsilon F}; \tag{3.49}$$

both small or large values are realistic.

To analyse (3.46), we ignore bedload (put $\beta = 0$) and take $\varepsilon \to 0$. Then

$$\eta = h + s, \quad uh = 1, \tag{3.50}$$

so that

$$\begin{gathered} c_x = E^*(u) - cD^*(u) = -s_t, \\ \frac{\partial}{\partial x}\left[\tfrac{1}{2}F^2u^2 + \frac{1}{u} + s\right] = \delta(1-u^3). \end{gathered} \tag{3.51}$$

If, in addition, $\delta \ll 1$, then, taking $s = 0$ when $h = 1$,

$$s = s(u) = \tfrac{1}{2}F^2(1-u^2) + 1 - \frac{1}{u}, \tag{3.52}$$

and the entire suspended load model is

$$s'(u)\frac{\partial u}{\partial t} = cD^*(u) - E^*(u) = -\frac{\partial c}{\partial x}. \tag{3.53}$$

The function $s(u)$ is the same as we derived before in (3.35) and shown in figure 3.7. We can in fact write (3.53) as a single equation for u, by eliminating c; this gives

$$c = \frac{E^*(u)}{D^*(u)} + \frac{s'(u)}{D^*(u)}\frac{\partial u}{\partial t},$$

$$s'(u)\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left[\frac{E^*(u)}{D^*(u)} + \frac{s'(u)}{D^*(u)}\frac{\partial u}{\partial t}\right] = 0, \tag{3.54}$$

and the equation for u (or the pair for u, c) is of hyperbolic type. Note that natural initial-boundary conditions for (3.53) are to prescribe u at $t = 0$, $x > 0$, and c at $x = 0$, $t > 0$.

Let us examine the stability of the steady state $u = 1$, $c = 1$. We put

$$u = 1 + Ue^{ikx+\sigma t}, \quad c = 1 + Ce^{ikx+\sigma t}, \tag{3.55}$$

and linearise, to obtain (noting $E^*(1) = D^*(1) = 1$)

$$ikC = [E^{*\prime}(1) - D^{*\prime}(1)]U - C = -\sigma s'(1), \tag{3.56}$$

and thus

$$\sigma = \left[\frac{E^{*\prime}(1) - D^{*\prime}(1)}{s'(1)}\right]\left(\frac{-k^2 - ik}{1+k^2}\right). \tag{3.57}$$

If we suppose $E^{*\prime} > 0$, $D^{*\prime} < 0$ as previously suggested, then this model implies instability ($\mathrm{Re}\,\sigma > 0$) for $s'(1) < 0$, i.e. $F > 1$, and that the wave speed is $-\mathrm{Im}\,(\sigma)/k < 0$; thus this theory predicts upstream-migrating anti-dunes.

Two features suggest that the model is not well posed if $F > 1$. The first is the instability of arbitrarily small wavelength perturbations; the second is that the unstable waves propagate upstream, although the natural boundary condition for c is prescribed at $x = 0$.

Numerical solutions of (3.53) are consistent with these observations. In solving the nonlinear model (3.53) in $0 < x < \infty$, we note that

$$\frac{d}{dt}\int_0^\infty s(u)\,dx = -[c]_0^\infty, \tag{3.58}$$

which simply represents the net erosion of the bed downwards if the sediment flux at infinity is greater than at zero. It thus makes sense to fix the initial boundary conditions so that

$$\begin{gathered} c = 1 \quad \text{on} \quad x = 0, \\ u \to 1 \quad \text{as} \quad x \to \infty,\ t = 0. \end{gathered} \tag{3.59}$$

For $F < 1$, numerical solutions are smooth and approach the stable solution $u = c = 1$. However, the solutions are numerically unstable for $F > 1$, and u rapidly blows up, causing breakdown of the solution.

Some further insight into this is gained by consideration of the solution at $x = 0$. If $c = c_0(t)$ on $x = 0$ and $u = u_0(x)$ on $t = 0$, then we can obtain u on $x = 0$ from (3.54), by solving the ordinary differential equation

$$\frac{\partial u}{\partial t} + \frac{E^*(u)}{s'(u)} = \frac{D^*(u)}{s'(u)}c_0(t) \tag{3.60}$$

with $u = u_0(0)$ at $t = 0$. If we suppose that $c = 1$ at $x = 0$, then it is easy to show that if $F < 1$ and $u(0,0) < 1/F^{2/3}$, then $u(0,t) \to 1$ as $t \to \infty$. If on the other hand, $F > 1$ and $u(0,0) < 1$, then $u(0,t) \to 1/F^{2/3}$ in finite time, and the solution breaks down as $\partial u/\partial t \to \infty$; if $u(0,0) > 1$, then $u(0,t) \to \infty$, again in finite time if, for example, $E^* \propto u^3$. More generally, breakdown of the solution when $F > 1$ occurs in one of these ways at some positive value of x. Thus this suspended sediment model shares the same weakness of the Kennedy phase shift model in not providing a well posed nonlinear model.

3.6 Eddy viscosity model

The relative failure of the models above to explain dune and anti-dune formation led Engelund (1970) and Smith (1970) to the consideration of a full fluid flow model, but rather than suppose that the flow was shear free and that viscous effects were confined to a turbulent boundary layer, as did Kennedy (1963), they considered a rotational model of turbulent shear flow incorporating an eddy viscosity, together with the Exner equation for bedload transport. This allows for a linear stability analysis of the uniform flow over a flat bed via the solution of a suitable Orr-Sommerfeld equation, and this is what Smith and Engelund did. We shall in fact proceed in somewhat more generality. As an observation, fully-formed dunes have relatively small height to length ratios, and thus the fluid flow over them can be approximately linearised. Although we use a linear approximation to derive the stress at the bed, we may retain the nonlinear Exner equation for example. In this way we may derive a nonlinear evolution equation for bed elevation.

3.6.1 Turbulent St. Venant equations

In this section, we rederive the St. Venant equations (2.38) and (2.41) of hydraulics and sediment flow, but not from first principles, but by averaging the point forms of the Navier-Stokes equations, in which we include an eddy viscosity, which is taken to be constant.

We suppose that we have two-dimensional turbulent flow down a slope of gradient S, governed by the Reynolds equations

$$\begin{aligned} u_t + uu_x + wu_z &= -\frac{1}{\rho}\frac{\partial p}{\partial x} + \nu_T \nabla^2 u + gS, \\ w_t + uw_x + ww_z &= -\frac{1}{\rho}\frac{\partial p}{\partial z} + \nu_T \nabla^2 w - g(1-S^2)^{1/2}, \\ u_x + w_z &= 0, \end{aligned} \tag{3.61}$$

where (u, w) are the velocity components and ν_T is an eddy viscosity associated with the Reynolds stress terms. In the second equation, we can take $g(1-S^2)^{1/2} \approx g$ since S is small.

The suspended sediment concentration c is described by

$$c_t + uc_x + (w - v_s)c_z = \nu_T \nabla^2 c, \tag{3.62}$$

where v_s is the particle fall velocity, assuming the turbulent sediment diffusivity equals the eddy (kinematic) viscosity.

Boundary conditions for these equations are

$$\begin{gathered} p = 0, \quad w = \eta_t + u\eta_x, \\ \rho_w \nu_T \frac{\partial u}{\partial z} \approx \sigma_{nt} = 0, \quad v_s c + \nu_T \frac{\partial c}{\partial z} = 0 \quad \text{at} \quad z = \eta; \\ \rho_s \left[(1-n)\frac{\partial s}{\partial t} + \frac{\partial q_b}{\partial x} \right] = v_s c + \nu_T \frac{\partial c}{\partial z}, \\ u = 0, \quad w = s_t + us_x \quad \text{at} \quad z = s. \end{gathered} \tag{3.63}$$

Here σ_{nt} is the shear stress, approximately equal to $\rho_w \nu_T \partial u/\partial z$ for small surface slopes.

Now let us formally suppose that bedforms and surface waves have small amplitudes (or, particularly, small slopes). We should not expect this to invalidate the derivation of a useful theory. In the spirit of lubrication theory, it then follows that $\partial/\partial z \gg \partial/\partial x$, $w \ll u$, and therefore p is approximately hydrostatic:

$$p \approx \rho_w g(\eta - z). \tag{3.64}$$

(This applies even with the viscous terms, which are of order $\nu_T\nabla^2 u \sim \nu_T u/d^2 \sim \varepsilon_T u^2/d$, and thus $\lesssim u^2/d$).

Integration of the x momentum equation from s to η then yields

$$\frac{\partial}{\partial t}\int_s^\eta u\,dz+\frac{\partial}{\partial x}\int_s^\eta u^2\,dz=gh(S-\eta_x)-\nu_T\left.\frac{\partial u}{\partial z}\right|_s+\int_s^\eta \nu_T u_{xx}\,dz. \tag{3.65}$$

Furthermore, mass conservation is represented by

$$\frac{\partial h}{\partial t}+\frac{\partial}{\partial x}\int_s^\eta u\,dz=0, \tag{3.66}$$

and vertical integration of (3.62) yields

$$\frac{\partial}{\partial t}\int_s^\eta c\,dz+\frac{\partial}{\partial x}\int_s^\eta uc\,dz=-\nu_T\left.\frac{\partial c}{\partial z}\right|_s-v_s c|_s+\int_s^\eta \nu_T c_{xx}\,dz. \tag{3.67}$$

The relation of these to the St. Venant equations is clear. Define the column averages by an overbar, thus

$$\bar{u}=\frac{1}{h}\int_s^\eta u\,dz,\quad \bar{c}=\frac{1}{h}\int_s^\eta c\,dz; \tag{3.68}$$

this leads to

$$\begin{aligned}
h_t+(\bar{u}h)_x&=0,\\
(h\bar{u})_t+(D_u h\bar{u}^2)_x&=gh(S-\eta_x)-\frac{\tau}{\rho_w}+\int_0^\eta \nu_T u_{xx}\,dz,\\
(h\bar{c})_t+(D_c h\bar{u}\bar{c})_x&=\rho_s(v_E-v_D)+\int_s^\eta \nu_T c_{xx}\,dz,
\end{aligned} \tag{3.69}$$

where, in the parlance of the literature of two-phase flow, the profile coefficients are

$$D_u=\frac{\overline{u^2}}{\bar{u}^2},\quad D_c=\frac{\overline{uc}}{\bar{u}\bar{c}}, \tag{3.70}$$

and

$$\rho_s v_E=-\nu_T\left.\frac{\partial c}{\partial z}\right|_s,\quad \rho_s v_D=v_s c|_s,\quad \tau=\rho_w\nu_T\left.\frac{\partial u}{\partial z}\right|_s. \tag{3.71}$$

The reasonable choices $D_u=D_c=1$ then lead to the St. Venant forms

$$\begin{aligned}
h_t+(\bar{u}h)_x&=0,\\
\bar{u}_t+\bar{u}\bar{u}_x&=g(S-\eta_x)-\frac{\tau}{\rho_w h}+\frac{1}{h}\int_s^\eta \nu_T u_{xx}\,dz,\\
h(\bar{c}_t+\bar{u}\bar{c}_x)&=\rho_s(v_E-v_D)+\int_s^\eta \nu_T c_{xx}\,dz,
\end{aligned} \tag{3.72}$$

together with the Exner equation, which from (3.63) and (3.71) is

$$(1-n)\frac{\partial s}{\partial t}+\frac{\partial q_b}{\partial x}=v_D-v_E. \tag{3.73}$$

It would be natural to suppose

$$\frac{1}{h}\int_s^\eta \nu_T u_{xx}\,dz\approx \nu_T\bar{u}_{xx},\quad \int_s^\eta \nu_T c_{xx}\,dz\approx h\bar{c}_{xx}, \tag{3.74}$$

if inclusion of such terms were necessary. However, these are stabilising terms. In seeking an improved expression for τ, we are motivated by the fact that with $\tau = f\rho_w\bar{u}^2$ and in the absence of suspended sediment transport, the bed is neutrally stable: even a small correction to τ may render the bed unstable.

3.6.2 Orr-Sommerfeld equation

Suppose, therefore, that we consider perturbations to a basic shear flow $u(z)$ in $s < z < \eta$ which satisfies (3.61) with ν_T taken as constant. (Later, we shall study a more realistic eddy viscosity model.) It is convenient first of all to non-dimensionalise the equations (3.61). In the basic uniform state, with $s = 0$ and $\eta = h_0$, the shear flow satisfies

$$\nu_T\frac{\partial u}{\partial z}=gS(h_0-z), \tag{3.75}$$

whence

$$u=\frac{gS}{\nu_T}\left(h_0z-\tfrac{1}{2}z^2\right), \tag{3.76}$$

and the column mean flow is

$$u_0=\frac{1}{h_0}\int_0^{h_0} u\,dz=\frac{gS}{3\nu_T}h_0^2. \tag{3.77}$$

Taking $\nu_T = \varepsilon_T u_0 h_0$, we find that

$$\tau=\rho_w\nu_T\left.\frac{\partial u}{\partial z}\right|_0=f\rho_w u_0^2, \tag{3.78}$$

where $f = 3\varepsilon_T$. This gives the relationship between the empirical f and the semi-analytic ε_T. It may in fact be more appropriate to think of ε_T as being constant, and if the bed and hence the flow is perturbed, we would only retain constant ν_T if the volume flux (more precisely, the volume flux scale) is the same; this we therefore assume.

We now non-dimensionalise the variables by writing

$$(u,w)\sim u_0,\quad (x,z)\sim h_0,\quad t\sim h_0/u_0,\quad p-\rho g(h_0-z)\sim\rho_w u_0^2. \tag{3.79}$$

The dimensionless equations are

$$u_t+uu_x+wu_z=-p_x+\frac{1}{R}\nabla^2u+\frac{S}{F^2},$$

$$w_t + uw_x + ww_z = -p_z + \frac{1}{R}\nabla^2 w,$$
$$u_x + w_z = 0, \tag{3.80}$$

and the parameters are a turbulent Reynolds number and the Froude number squared:

$$R = \frac{u_0 h_0}{\nu_T}, \quad F^2 = \frac{u_0^2}{g h_0}. \tag{3.81}$$

The dimensionless basic velocity profile is then

$$u = \frac{g S h_0^2}{\nu_T u_0}\left(z - \tfrac{1}{2}z^2\right), \tag{3.82}$$

and the dimensionless mean velocity is

$$\bar{u}^* = \frac{g S h_0^2}{3\nu_T u_0}. \tag{3.83}$$

We have $\nu_T = \varepsilon_T u_0 h_0$, ε_T constant; for consistency with our previous definition of ν_T, we choose the mean steady velocity to be u_0, i. e., we choose $\bar{u}^* = 1$ in (3.83), and thus $u_0 = (g S h_0/f)^{1/2}$. In particular, the dimensionless basic velocity profile is

$$u = U(z) = 3\left(z - \tfrac{1}{2}z^2\right). \tag{3.84}$$

We now suppose that s and η are perturbed by small amounts; we may thus linearise (3.80). We put

$$(u, w) = (U(z) + \psi_z, -\psi_x), \tag{3.85}$$

whence it follows for small ψ that ψ satisfies the steady state Orr-Sommerfeld equation

$$U\nabla^2\psi_x - U''\psi_x = R^{-1}\nabla^4\psi, \tag{3.86}$$

where we assume stationary solutions in view of the anticipated fact that s evolves on a slower time scale.

The condition of zero pressure at $z = \eta$ is linearised to be

$$\eta = 1 + F^2 p\,|_{z=1}. \tag{3.87}$$

If F^2 is small, then we may take η to be constant, and we do so as we are primarily interested in dunes. However, the dimensionless pressure p is only determined up to addition of an arbitrary constant, which implies that the value of the constant η is unconstrained. This represents the vertical translation invariance of the system. If a uniform perturbation to s is made, then the response of the (uniform) stream is to raise the surface by the same amount. We can remove the ambiguity by prescribing $\eta = 1$, with the implication that the mean value of s is required to be zero.

The other boundary conditions on $z = s$ and $z = 1$ are no slip at the base, no shear stress at the top, and the perturbed volume flux is zero. These imply

$$\psi = 0, \quad \psi_{zz} = 0 \quad \text{on} \quad z = 1,$$

$$\int_0^s U(z)\,dz + \psi = 0, \quad U + \psi_z = 0 \;\; \text{on} \;\; z = s. \tag{3.88}$$

Linearisation of this second pair about $z = 0$ gives

$$\psi = 0, \quad \psi_z = -U_0' s \;\; \text{on} \;\; z = 0, \tag{3.89}$$

where $U_0' = U'(0)$. Our aim is now to solve (3.86) with (3.88) and (3.89) to calculate the perturbed shear stress. The dimensional basal shear stress is then

$$\tau = \rho_w \varepsilon_T u_0^2 U_0' \left[1 + s\frac{U_0''}{U_0'} + \frac{1}{U_0'}\psi_{zz}|_0 \right], \tag{3.90}$$

and since $f = 3\varepsilon_T = \varepsilon_T U_0'$, $u_0 = \bar{u}$, we may write this as

$$\tau = f\rho_w \bar{u}^2 \left[1 + \frac{sU_0''}{U_0'} + \frac{1}{U_0'}\psi_{zz}|_0 \right]. \tag{3.91}$$

The problem to solve for ψ is linear and inhomogeneous, and so we suppose that

$$s = \int_{-\infty}^{\infty} \hat{s}(k) e^{ikx}\,dk, \quad \psi = \int_{-\infty}^{\infty} \hat{\psi}(k) e^{ikx}\,dk. \tag{3.92}$$

(Note $\hat{s}$ will evolve slowly in time.) For each wave number k, we obtain

$$ik \left[U(\hat{\psi}'' - k^2\hat{\psi}) - U''\hat{\psi} \right] = \frac{1}{R} \left[\hat{\psi}^{iv} - 2k^2\hat{\psi}'' + k^4\hat{\psi} \right], \tag{3.93}$$

with boundary conditions

$$\begin{gathered} \hat{\psi} = \hat{\psi}'' = 0 \;\; \text{on} \;\; z = 1, \\ \hat{\psi} = 0, \quad \hat{\psi}' = -U_0'\hat{s} \;\; \text{on} \;\; z = 0, \end{gathered} \tag{3.94}$$

and thus we finally define

$$\hat{\psi} = -U_0'\hat{s}\Psi(z,k), \tag{3.95}$$

where Ψ satisfies the canonical problem

$$\begin{gathered} ik \left[U(\Psi'' - k^2\Psi) - U''\Psi \right] = \frac{1}{R} \left[\Psi^{iv} - 2k^2\Psi'' + k^4\Psi \right], \\ \Psi = \Psi'' = 0 \;\; \text{on} \;\; z = 1, \\ \Psi = 0, \quad \Psi' = 1 \;\; \text{on} \;\; z = 0. \end{gathered} \tag{3.96}$$

In terms of Ψ, the basal (dimensional) shear stress is

$$\tau = f\rho_w \bar{u}^2 \left[1 - s - \int_{-\infty}^{\infty} e^{ikx}\hat{s}(k)\Psi''(0,k)\,dk \right]. \tag{3.97}$$

Using the convolution theorem, this is

$$\tau = f\rho_w \bar{u}^2 \left[1 - s + \int_{-\infty}^{\infty} K(x-\xi)s'(\xi)\,d\xi \right], \tag{3.98}$$

where $s' = \partial s/\partial x$, and

$$K(x) = -\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\Psi''(0,k)}{ik} e^{ikx}\,dk. \tag{3.99}$$

Depending on K, we can see how τ may depend on displaced values of s. The form of (3.98) illustrates our previous discussion of the vertical translation invariance of the system. For a possible uniform perturbation s = constant, we would obtain a modification to the basic friction law, $\tau = f\rho_w \bar{u}^2$. This is excluded by enforcing the condition that s has zero mean in x,

$$\int_{-\infty}^{\infty} s(x)\,dx = 0, \tag{3.100}$$

which corresponds to prescribing

$$\hat{s}(0) = 0. \tag{3.101}$$

Note that this excludes, for example, the flow over an isolated bump. In that case, if we have $s \to 0$ as $x \to \pm\infty$, then $\eta \to 1$ there, and there is no restraint on s. The assumption of zero mean is really designed for the case of a general, or periodic, perturbation.

To determine K, we need to know the solution of (3.96) for all k. In general, the problem requires numerical solution. However, note that $R = 1/\varepsilon_T$, and is reasonably large ($R = 30$ if $\varepsilon_T = 0.034$). This suggests that a useful means of solving (3.96) may be asymptotically, in the limit of large R. The fact that we can obtain analytic expressions for $\Psi''(0,k)$ means this is useful even when R is not dramatically large, as here.

The solution of the Orr-Sommerfeld equation at large R has a long pedigree, and it is a complicated but mathematically interesting problem. It can be shown that, for $k > 0$,

$$\Psi''(0,k) \approx -3(ikRU_0')^{1/3}\mathrm{Ai}(0) + O(1), \tag{3.102}$$

where Ai is the Airy function. For $k < 0$, $\Psi''(0,k) = \overline{\Psi''(0,-k)}$, and this leads to

$$\frac{\Psi''(0,k)}{ik} \approx \begin{cases} -ce^{-i\pi/3}|k|^{-2/3}, & k > 0 \\ -ce^{i\pi/3}|k|^{-2/3}, & k < 0, \end{cases} \tag{3.103}$$

where

$$c = 3(RU_0')^{1/3}\mathrm{Ai}(0), \tag{3.104}$$

and $c \approx 1.54R^{1/3}$ for $U_0' = 3$, as $\mathrm{Ai}(0) = 1/3^{2/3}\Gamma(2/3) \approx 0.355$. From (3.99), we find

$$K(x) = \frac{c}{\pi}\int_0^{\infty} \frac{\cos[kx - \frac{\pi}{3}]\,dk}{k^{2/3}}. \tag{3.105}$$

Evaluating the integral,[3] we obtain the simple formula

$$K(x) \;=\; \frac{\mu}{x^{1/3}}, \quad x > 0,$$

---

[3] How do we do that? The blunt approach is to consult Gradshteyn and Ryzhik (1980), where the relevant formulae are on page 420 and 421 (items 4 and 9 of section 3.761). The quicker way, using complex analysis, is to evaluate $\int_0^\infty \theta^{\nu-1}e^{i\theta}\,d\theta$ (after a simple rescaling of k, $k|x| = \theta$) by rotating the contour by $\pi/2$ and using Jordan's lemma. Thus $\int_0^\infty \theta^{\nu-1}e^{i\theta}\,d\theta = \Gamma(\nu)e^{i\pi\nu/2}$.

$$K(x) \;=\; 0, \quad x < 0, \tag{3.106}$$

where

$$\mu = \frac{3^{2/3}R^{1/3}}{\{\Gamma(\frac{2}{3})\}^2} \approx 1.98R^{1/3}. \tag{3.107}$$

For stability purposes, note that

$$K = \int_{-\infty}^{\infty} \hat{K}(k)\, e^{ikx} dk, \tag{3.108}$$

where

$$\hat{K} = -\frac{\Psi''(0,k)}{2\pi ik} = \frac{c\exp\left[-\dfrac{i\pi}{3}\operatorname{sgn} k\right]}{2\pi|k|^{2/3}}. \tag{3.109}$$

3.6.3 Orr-Sommerfeld-Exner-St. Venant model

We return to (3.72), or equivalently we can go right back to (3.32). Since the scales used there are the same as above, it is clear that the modified model[4] based on (3.98) is

$$\begin{aligned} s_t + q_x &= 0, \\ \varepsilon h_t + (uh)_x &= 0, \\ F^2(\varepsilon u_t + uu_x) &= -\eta_x + \delta\left[1 - \frac{u^2}{h}\left(1 - s + \int_{-\infty}^{\infty} K(x-\xi)\frac{\partial s}{\partial \xi}(\xi,t)\,d\xi\right)\right], \\ h &= \eta - s. \end{aligned} \tag{3.110}$$

If $\delta \ll 1$, then, since $K \sim 2R^{1/3} \approx 6$ for $R = 30$, it is still reasonable to neglect the term in K in $(3.110)_3$. Put ε and δ to zero; then, as before,

$$\begin{aligned} uh &= 1, \\ \tfrac{1}{2}F^2u^2 + \eta &= \tfrac{1}{2}F^2 + 1, \\ h &= \eta - s, \end{aligned} \tag{3.111}$$

and $s = s_0(u) = \frac{1}{2}F^2(1-u^2) + 1 - \dfrac{1}{u}$, as in (3.52). We suppose $q = q(\tau)$, where the dimensionless basal stress (τ scaled with $f\rho_w u_0^2$) is

$$\tau = u^2\left[1 - s + \int_{-\infty}^{\infty} K(x-\xi)\frac{\partial s}{\partial \xi}(\xi,t)d\xi\right]. \tag{3.112}$$

---

[4] There is a subtle point here concerning the modified stress, since in (3.110) we allow the basic (dimensional) bed stress $f\rho_w\bar{u}^2$ to vary with mean velocity, whereas in the derivation of (3.103), we specifically kept the mean flow constant. The two differing views can be reconciled if one views (3.103) as providing a local correction to the stress on short length scales, whereas the variation of u in (3.110) is supposed, formally, to be over a longer length scale.

The Exner equation is thus

$$\frac{\partial s}{\partial t} + q'(\tau)\frac{\partial \tau}{\partial x} = 0. \tag{3.113}$$

For small F, we have

$$u \approx \frac{1}{1-s}, \tag{3.114}$$

so that

$$\tau \approx \frac{1}{1-s} + \frac{1}{(1-s)^2}\int_{-\infty}^{\infty} K(x-\xi)\frac{\partial s}{\partial \xi}(\xi, t)\, d\xi. \tag{3.115}$$

We linearise by writing $s = \hat{s}\, e^{ikx+\sigma t}$, $\tau = 1 + \hat{\tau} e^{ikx+\sigma t}$, so that

$$\sigma\hat{s} + ikq'(1)\hat{\tau} = 0,$$

$$\hat{\tau} = \hat{s} + \hat{K}ik\hat{s}, \tag{3.116}$$

and thus

$$\sigma = q'(1)[k^2\hat{K} - ik]. \tag{3.117}$$

When $\operatorname{Re}\hat{K} > 0$, as for (3.109), the steady state is unstable, with $\operatorname{Re}\sigma \sim k^{4/3}$ as $k \to \infty$. Specifically, the growth rate is (taking k positive)

$$\operatorname{Re}\sigma = \frac{q'(1)c}{4\pi}k^{4/3}, \tag{3.118}$$

while the wave speed is

$$-\frac{\operatorname{Im}\sigma}{k} = q'(1)\left[1 + \frac{\sqrt{3}ck^{1/3}}{4\pi}\right]; \tag{3.119}$$

Thus waves move downstream.

3.6.4 Well-posedness

The effect of (3.115) is to cause increased τ where s_x is positive, on the upstream slopes of bumps. Since u is in phase with s, this implies τ leads u (i.e., τ is a maximum before s is); it is this phase lead which causes instability. However, the unbounded growth rate at large wave numbers is a sign of ill-posedness. Without some stabilising mechanism, arbitrarily small disturbances can grow arbitrarily rapidly. In reality, another effect of bed slope is important, and that is the fact that sediment wants to roll downslope: in describing the Meyer-Peter/Müller result, no attention was paid to the variations of bed slope itself.

For a particle of diameter D_s at the bed, the streamflow exerts a force of approximately τD_s^2 on it, and it is this force which causes motion. On a slope, there is an additonal force due to gravity, approximately $-\Delta\rho g D_s^3 s_x$. Thus the net stress causing motion is actually

$$\tau - \Delta\rho g D_s s_x. \tag{3.120}$$

In dimensionless terms, we therefore modify the bedload transport formula by writing

$$q = q[\tau_e], \quad \tau_e = \tau - \beta s_x, , \tag{3.121}$$

where

$$\beta = \frac{\Delta\rho D_s}{\rho_w h_0 S}. \tag{3.122}$$

Typical values are $\Delta\rho/\rho_w \approx 2$, $D_s \sim 10^{-3}$ m, $h_0 \sim 2$ m, $S \sim 10^{-3}$, whence $\beta \sim 1$. Larger particles or steeper slopes cause larger values, but generally we will suppose that $\beta \sim O(1)$.

The effect of this is to replace the definition of τ in (3.115) by

$$\tau_e = \frac{1}{1-s} + \frac{\mu}{(1-s)^2}\int_{-\infty}^{\infty} K(x-\xi)\frac{\partial s}{\partial \xi}(\xi,t)\,d\xi - \beta s_x, \tag{3.123}$$

(together with (3.113)) and in the stability analysis, $\hat{\tau} = \hat{s}[1 + ik\hat{K} - ik\beta]$, whence

$$\sigma = q'(1)\left[-ik\left\{1 + \frac{\sqrt{3}ck^{1/3}}{4\pi}\right\} + \frac{c}{4\pi}k^{4/3} - \beta k^2\right]. \tag{3.124}$$

This exhibits the classical behaviour of a well-posed model. The system is stable at high wavenumber, and the maximum growth rate is at $k = (c/\pi\beta)^{3/2}$. This would be the expected preferred wavenumber of the instability.

One can try and push the study of this instability slightly further, into the non-linear régime. (3.123) is valid if $s \ll 1$. If we expand beyond the linear term, we have

$$\tau_e \approx 1 + s + s^2 + \mu\int_0^{\infty} |\xi|^{-1/3}\frac{\partial s(x-\xi,t)}{\partial x}\,d\xi - \beta s_x, \tag{3.125}$$

while also we can take q' as constant, if $\tau \approx 1$. Then if we write

$$x - q'(1)t = X \tag{3.126}$$

(so X is the spatial coordinate relative to the moving dune frame), and rescale the variables as

$$X = \left(\frac{\beta}{\mu}\right)^{3/2}\xi, \quad s = \frac{\beta}{2}\left(\frac{\mu}{\beta}\right)^{3/2}\phi, \quad t = \frac{\beta^2}{q'(1)\mu^3}\tau, \tag{3.127}$$

we derive an appealing canonical form of the Exner equation:

$$\phi_\tau + \frac{\partial}{\partial\xi}\left[\tfrac{1}{2}\phi^2 + \int_0^{\infty} |\eta|^{-1/3}\frac{\partial\phi(\xi-\eta,\tau)}{\partial\xi}\,d\eta - \phi_\xi\right] = 0. \tag{3.128}$$

Consulting the sizes of the various corrective terms in (3.125), we find that they are small if $\dfrac{\mu^3}{2\beta} \ll 1$. Since $\mu \approx 6$, this requires that β should be large. This can be achieved for large sediment size and/or steep slopes, but is likely to be uncommon. Thus (3.128) serves more as a pedagogical example than as a practical predictive tool. Even were it accurate, its application is limited by the inevitable development of separated flow in the lee of the developing dune.

The equation (3.128) is very similar to the Kuramoto-Sivashinsky equation, where the destabilising integral term is replaced by $\phi_{\xi\xi}$, and the stabilising diffusive term is replaced by $\phi_{4\xi}$. Numerical solutions obtained using a pseudo-spectral numerical method show that random initial data evolves into travelling waves with steep fronts. Eventually the waves merge, until one wave occupies the computational domain. An example of this is shown in figure 3.9.

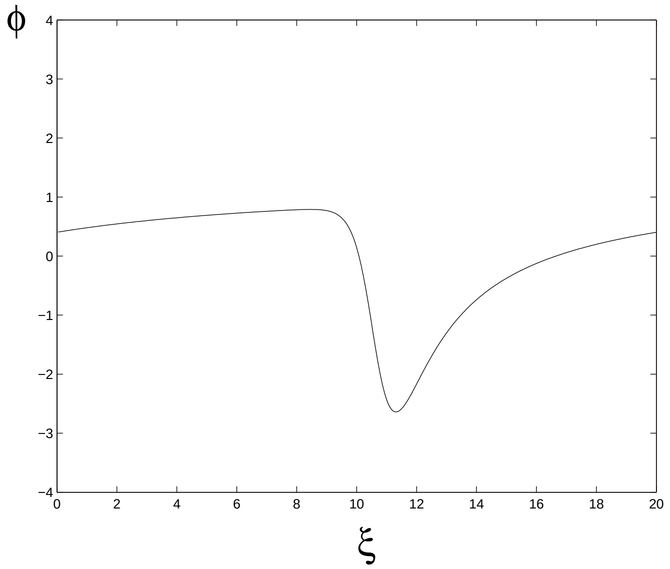

Figure 3.9: The eventual travelling dune which arises from solving (3.128) on a domain of length 20.

This wave has a steep front and a more gentle rear slope. It is easy to understand how this can arise from (3.128). When the length scale is large, both the integral term and the diffusive term are small. Thus the waves generated by the instability form shocks, and these are smoothed by the diffusive term. At large time, the integral term is irrelevant, and the equation is essentially Burgers' equation. The wave speed of a shock front is thus the mean of the values of ϕ either side of the front, and so larger shocks overtake smaller ones. What eventually limits the wave growth is the domain size, and so this model lacks any mechanism to determine the length scale of fully evolved dunes on an infinite domain.

3.7 Mixing-length model for aeolian dunes

Measurements of turbulent fluid flow in pipes, as well as air flow in the atmosphere (and also in wind tunnels), show that the assumption of constant eddy viscosity is not a good one, and the basic shear velocity profile is not as simple as assumed in the preceding section. In actual fact, the concept of eddy viscosity introduced by Prandtl was based on the idea of momentum transport by eddies of different sizes, with the transport rate (eddy viscosity) being proportional to eddy size. Evidently, this must go to zero at a solid boundary, and the simplest description of this is Prandtl's mixing length theory. In this section, we generalise the previous approach a little to allow for such a spatially varying eddy viscosity, and we specifically consider the case of aeolian dunes, in which a kilometre deep turbulent boundary layer flow is driven by an atmospheric shear flow.

3.7.1 Mixing-length theory

The various forms of sand dunes in deserts were discussed earlier; the variety of shapes can be ascribed to varying wind directions, a feature generally absent in rivers. Another difference from the modelling point of view is that the fluid atmosphere is about ten kilometres in depth, and the flow in this is essentially unaffected by the underlying surface, except in the atmospheric boundary layer, of depth about a kilometre, wherein most of the turbulent mixing takes place. Within this boundary layer, there is a region adjoining the surface in which the velocity profile is approximately logarithmic, and this region spans a range of height from about forty metres above the surface to the 'roughness height' of just a few centimetres or millimetres above the surface.

Consider the case of a uni-directional mean shear flow $u(z)$ past a rough surface $z = 0$, where z measures distance away from the surface. If the shear stress is constant, equal to τ, then we define the friction velocity u_* by

$$u_* = (\tau/\rho)^{1/2}, \tag{3.129}$$

where ρ is density. Observations support the existence near the surface of a logarithmic velocity profile of the form

$$u = \frac{u_*}{\kappa} \ln \left(\frac{z}{z_0} \right), \tag{3.130}$$

where the Von Kármán constant $\kappa \approx 0.4$, and z_0 is known as the roughness length: it represents the effect of surface roughness in bringing the average velocity to zero at some small height above the actual surface. Since z_0 is a measure of actual roughness, a typical value for a sandy surface might be $z_0 = 10^{-3}$ m.

Prandtl's mixing length theory provides a motivation for (3.130). If we suppose the motion can be represented by an eddy viscosity η, so that

$$\tau = \eta \frac{\partial u}{\partial z}, \tag{3.131}$$

then Prandtl proposed

$$\eta = \rho l^2 \left| \frac{\partial u}{\partial z} \right|, \quad l = \kappa z, \tag{3.132}$$

from which, indeed, (3.130) follows. The quantity $l = \kappa z$ is called the mixing length. Prandtl's theory works well in explaining the logarithmic layer, and in extension it explains pipe flow characteristics very well; but it has certain drawbacks. The two obvious ones are that it is not frame-invariant; however, this would be easily rectified by replacing $|\partial u/\partial z|$ by the second invariant $2\dot{\varepsilon}$, where $2\dot{\varepsilon}^2 = \dot{\varepsilon}_{ij}\dot{\varepsilon}_{ij}$, and $\dot{\varepsilon}_{ij}$ is the strain rate tensor. Also not satisfactory is the rather loosely defined mixing length, which becomes less appropriate far from the boundary, or in a closed container. Despite such misgivings, we will use a version of the mixing length theory to see how it deviates from the constant eddy viscosity assumption.

We want to see how to solve a shear flow problem in dimensionless form. To this end, suppose for the moment that we fix $u = U_\infty$ on $z = d$. Then $U_\infty = (u^*/\kappa)\ln(d/z_0)$ determines u_* (and thus τ), and we can define a parameter[5] ε by

$$\varepsilon = \frac{u_*}{U_\infty} = \frac{\kappa}{\ln(d/z_0)}. \tag{3.133}$$

For $d = 10^3$ m, $z_0 = 10^{-3}$ m, $\kappa = 0.4$, $\varepsilon \approx 0.03$. Writing u in terms of U_∞ rather than u_* yields

$$u = U_\infty \left[1 + \frac{\varepsilon}{\kappa} \ln\left(\frac{z}{d}\right)\right]. \tag{3.134}$$

Note also that the basic eddy viscosity is then

$$\eta = \varepsilon\rho U_\infty d\left(\frac{\kappa z}{d}\right), \tag{3.135}$$

and the shear stress is

$$\tau = \varepsilon^2 \rho U_\infty^2. \tag{3.136}$$

We shall use these observations in scaling the equations. Note that in reality, one prescribes the shear stress delivered by the far field flow, and thus U_∞ is determined, given d.

3.7.2 Turbulent flow model

Again we assume a mean two-dimensional flow $(u, 0, w)$ with horizontal coordinate x and vertical coordinate z over a surface topography given by $z = s$. The basic equations are

$$\begin{aligned} u_x + w_z &= 0, \\ \rho(uu_x + wu_z) &= -p_x + \tau_{1x} + \tau_{3z}, \\ \rho(uw_x + ww_z) &= -p_z + \tau_{3x} - \tau_{1z}, \end{aligned} \tag{3.137}$$

where $\tau_1 = \tau_{11}$ and $\tau_3 = \tau_{13}$ are the deviatoric Reynolds stresses, and are defined, we suppose, by

$$\begin{aligned} \tau_1 &= 2\eta u_x, \\ \tau_3 &= \eta(u_z + w_x). \end{aligned} \tag{3.138}$$

---

[5] Note that this definition of ε is unrelated to its previous definition and use, as for example in (3.110).

Our choice of the eddy viscosity η will be motivated by the Prandtl mixing length theory (3.132), but we postpone a precise specification for the moment.

The basic flow then dictates how we should non-dimensionalise the variables. We choose scales

$$\begin{gathered} u = U_\infty(1+\varepsilon u^*), \quad w \sim \varepsilon U_\infty, \quad x, z \sim d, \\ \tau_1, \tau_3 \sim \varepsilon^2 \rho U_\infty^2, \quad \eta \sim \varepsilon\rho d U_\infty, \quad p \sim \varepsilon\rho U_\infty^2, \end{gathered} \tag{3.139}$$

and then the dimensionless equations are (dropping the asterisk on u^*)

$$\begin{aligned} u_x + w_z &= 0, \\ u_x + p_x &= \varepsilon[\tau_{1x} + \tau_{3z} - \{uu_x + wu_z\}], \\ w_x + p_z &= \varepsilon[\tau_{3x} - \tau_{1z} - \{uw_x + ww_z\}], \\ \tau_3 &= \eta(u_z + w_z), \\ \tau_1 &= 2\eta u_x. \end{aligned} \tag{3.140}$$

3.7.3 Boundary conditions

The depth scale of the flow d is, we suppose, the depth of the atmospheric boundary layer, of the order of hundreds of metres to a kilometre. Above the boundary layer, there is an atmospheric shear flow, and we suppose that $u \to u_0(z)$, $w \to 0$, $p \to 0$ as $z \to \infty$.[6] The choice of u_0 is determined for us by the choice of η, as is most easily seen from the case of a uniform flow where $\partial u/\partial z = \tau/\eta$. The correct boundary condition to pose at large z is to prescribe the shear stress delivered by the main atmospheric flow, and this can be taken to be $\tau_3 = 1$ by our choice of stress scale. Thus we prescribe

$$\tau_3 \to 1, \quad w \to 0, \quad p \to 0 \quad \text{as} \quad z \to \infty. \tag{3.141}$$

Next we need to prescribe conditions at the surface. This involves two further length scales, the length L and amplitude H of the surface topography. Since we observe dunes often to have lengths in the range 100–1000 m, and heights in the range 2–100 m, we can see that there are two obvious distinguished limits, $L = d$, $H = \varepsilon d$, and it is most natural to use these in scaling the surface s. In fact since dunes are self-evolving it seems most likely that they will select length scales already present in the system. Thus, we suppose that in dimensionless terms the surface is $z = \varepsilon s(x)$, and longer, shorter, taller or smaller dunes can always be introduced as necessary later, by rescaling s. The surface boundary conditions are then taken to be (recalling the definition of the roughness length)

$$u = -\frac{1}{\varepsilon}, \quad w = 0 \quad \text{on} \quad z = \varepsilon s + z_0^*, \tag{3.142}$$

where

$$z_0^* = \frac{z_0}{d} = e^{-\kappa/\varepsilon}. \tag{3.143}$$

For completeness, we need to specify horizontal boundary conditions, for example at $x = \pm\infty$. We keep these fiarly vague, beyond requiring that the variables remain bounded. In particular, we do not allow unbounded growth of velocity or pressure.

---

[6] The modelling alternative is to specify velocity conditions on a lid at $z = 1$.

3.7.4 Eddy viscosity

Prandtl's mixing length theory in scaled units would imply

$$\eta = \kappa^2 (z - \varepsilon s)^2 \left| \frac{\partial u}{\partial z} \right|, \tag{3.144}$$

and we assume this, although other choices are possible. In particular, (3.144) is not frame indifferent, but this is hardly of significance since the eddy viscosity itself is unreliable away from the surface. (We comment further on this in the notes at the end of the chapter.)

To convert to the constant eddy viscosity model of the preceding section, equation (3.80), we would rescale $u, w, p \sim 1/\varepsilon$, and choose $\eta = \varepsilon$: thus $\varepsilon^2 = 1/R$.

3.7.5 Surface roughness layer

The basic shear flow near a flat surface $z = 0$ is given by (3.134), and in dimensionless terms is

$$u = \frac{1}{\kappa} \ln z; \tag{3.145}$$

we will require similar behaviour when the flow is perturbed. Suppose, more generally, that as $z \to \varepsilon s$,

$$u \sim a + b \ln(z - \varepsilon s) + O(z - \varepsilon s), \tag{3.146}$$

which we shall find describes the solution away from the boundary. We put

$$z = \varepsilon s + \nu Z, \tag{3.147}$$

where

$$\nu = e^{-\kappa/\varepsilon}. \tag{3.148}$$

Additionally, we write

$$u = -\frac{1}{\varepsilon} + U, \quad w = \varepsilon s_x U + \nu W, \quad \tau_1 = \varepsilon T_1, \quad \eta = \nu N. \tag{3.149}$$

Then we find that

$$\begin{aligned}
U_x + W_Z &= 0,\\
\frac{\partial \tau_3}{\partial Z} - \varepsilon^2 s_x \frac{\partial T_1}{\partial Z} + s_x \frac{\partial p}{\partial Z} &\approx 0,\\
\frac{\partial p}{\partial Z} &\approx -\varepsilon^2 \left[s_x \frac{\partial \tau_3}{\partial Z} + \frac{\partial T_1}{\partial Z} \right],\\
N &\approx \kappa^2 Z^2 \frac{\partial U}{\partial Z},\\
\tau_3 &\approx N(1 - \varepsilon^2 s_x^2) \frac{\partial U}{\partial Z},\\
T_1 &= -2\kappa^2 s_x Z^2 \left(\frac{\partial U}{\partial Z} \right)^2,
\end{aligned} \tag{3.150}$$

where we have neglected transcendentally small terms proportional to ν.

Correct to $O(\varepsilon^2)$, τ_3 is constant through the roughness layer, and equal to its surface value τ, and

$$\tau \approx \kappa^2 Z^2 \left(\frac{\partial U}{\partial Z}\right)^2, \tag{3.151}$$

again correct to $O(\varepsilon^2)$. The boundary conditions on $Z = 1$ (i.e., $z - \varepsilon s = z_0^* = \nu$) are $U = W = 0$, thus

$$U = \frac{\sqrt{\tau}}{\kappa} \ln Z, \tag{3.152}$$

and this must be matched to the outer solution (3.146). Rewriting (3.152) in terms of u and z, we have

$$u \sim \frac{\sqrt{\tau} - 1}{\varepsilon} + \frac{\sqrt{\tau}}{\kappa} \ln(z - \varepsilon s), \tag{3.153}$$

and this is in fact the matching condition that we require from the outer solution. We see immediately that variations of $O(1)$ in u yield small corrections of $O(\varepsilon)$ in τ.

Solving for W, we have

$$W = -\frac{(\sqrt{\tau})'}{\kappa}[Z \ln Z - Z], \tag{3.154}$$

where $(\sqrt{\tau})' = \partial\sqrt{\tau}/\partial x$, and in terms of w and z, this is written

$$w = s_x + \varepsilon s_x u - \frac{(\sqrt{\tau})'}{\kappa}\left[\ln(z - \varepsilon s) - 1 + \frac{\kappa}{\varepsilon}\right](z - \varepsilon s). \tag{3.155}$$

Hence the outer solution must satisfy (correct to $O(\varepsilon^2)$)

$$w \approx s_x + \varepsilon s_x u \quad \text{as} \quad z \to \varepsilon s. \tag{3.156}$$

3.7.6 Outer solution

We turn now to the solution away from the roughness layer, in the presence of surface topography of amplitude $O(\varepsilon)$ and length scale $O(1)$. The topography has two effects. The $O(1)$ variation in length scale causes a perturbation on a height scale of $O(1)$, but the vertical displacement of the logarithmic layer by $O(\varepsilon)$ causes a shear layer of this thickness to occur. Thus the flow away from the surface consists of an outer layer of thickness $O(1)$, and an inner shear layer of thickness $O(\varepsilon)$. We begin with the outer layer.

We expand the variables as

$$u = u^{(0)} + \varepsilon u^{(1)} + \cdots, \tag{3.157}$$

etc., so that to leading order, from (3.140),

$$\begin{aligned} u_x^{(0)} + w_z^{(0)} &= 0, \\ u_x^{(0)} + p_x^{(0)} &= 0, \\ w_x^{(0)} + p_z^{(0)} &= 0. \end{aligned} \tag{3.158}$$

Notice that, at this leading order, the precise form of η in (3.144) is irrelevant, as this outer problem is inviscid. We have

$$u^{(0)} + p^{(0)} = u_0(z), \tag{3.159}$$

and

$$p^{(0)}_x = w^{(0)}_z, \quad p^{(0)}_z = -w^{(0)}_x, \tag{3.160}$$

which are the Cauchy-Riemann equations for $p^{(0)}+iw^{(0)}$, which is therefore an analytic function, and $p^{(0)}$ and $w^{(0)}$ both satisfy Laplace's equation. The matching conditions as $z \to \varepsilon s$ can be linearised about $z = 0$, and if $w^{(0)} = w_0$ and $p^{(0)} = p_0$ on $z = \varepsilon s$, then from (3.156) we have

$$w^{(0)} = s_x \quad \text{on} \quad z = 0. \tag{3.161}$$

Assuming also that $w^{(0)}, p^{(0)} \to 0$ as $z \to \infty$, we can write the solutions in the form

$$w^{(0)} = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{z s_\xi \, d\xi}{[(x-\xi)^2 + z^2]}, \quad p^{(0)} = -\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{(x-\xi) s_\xi \, d\xi}{[(x-\xi)^2 + z^2]}, \tag{3.162}$$

and in particular, $p^{(0)}$ on $z = \varepsilon s$ is given to leading order by p_0, where

$$p_0 = \frac{1}{\pi} ⨍_{-\infty}^{\infty} \frac{s_\xi \, d\xi}{\xi - x} = H(s_x); \tag{3.163}$$

the integral takes the principal value, and H denotes the Hilbert transform.

The shear velocity profile $u_0(z)$ is undetermined at this stage, although we would like it to be the basic shear flow profile; but to justify this, we need to go to the $O(\varepsilon)$ terms. At $O(\varepsilon)$, we have

$$\begin{aligned}
u^{(1)}_x + w^{(1)}_z &= 0, \\
u^{(1)}_x + p^{(1)}_x &= \tau^{(0)}_{1x} + \tau^{(0)}_{3z} - \{u^{(0)}u^{(0)}_x + w^{(0)}u^{(0)}_z\}, \\
w^{(1)}_x + p^{(1)}_z &= \tau^{(0)}_{3x} - \tau^{(0)}_{1z} - \{u^{(0)}w^{(0)}_x + w^{(0)}w^{(0)}_z\}, \\
\tau^{(0)}_3 &= \eta^{(0)}[u^{(0)}_z + w^{(0)}_x], \\
\tau^{(0)}_1 &= 2\eta^{(0)}u^{(0)}_x, \\
\eta^{(0)} &= \kappa^2 z^2 \left|\frac{\partial u^{(0)}}{\partial z}\right|.
\end{aligned} \tag{3.164}$$

We can use the zero-th order solution to write $(3.164)_2$ in the form

$$u^{(1)}_x + p^{(1)}_x = \frac{\partial \tau^{(0)}_3}{\partial z} + \frac{\partial}{\partial x}\left[\tau^{(0)}_1 - \tfrac{1}{2}(u^{(0)2} + w^{(0)2}) + u_0'(z)\psi^{(0)}\right], \tag{3.165}$$

where $\psi^{(0)}$ is the stream function such that $w^{(0)} = -\psi^{(0)}_x$, and specifically, we have

$$\psi^{(0)} = -\frac{1}{2\pi}\int_{-\infty}^{\infty} \ln[(x-\xi)^2 + z^2] p_0(\xi)\, d\xi, \tag{3.166}$$

which can be found (as can the formulae in (3.162)) by using a suitable Green's function; (this is explained further below when we find $p^{(1)}$).

On integrating (3.165), we have to avoid secular terms which grow linearly in x, and we therefore require the integral of the right hand side of (3.165) with respect to x, from $-\infty$ to ∞, to be bounded. The integral of the derivative term is certainly bounded; thus the secularity condition requires $\int_{-\infty}^{\infty} \tau_{3z}^{(0)}\, dx$ to be bounded, and it is this condition that determines the function of integration $u_0(z)$ in (3.159).

For the particular choice of $\eta^{(0)}$ in (3.164), we have

$$\eta^{(0)} = \kappa^2 z^2 \left(u_0' + w_x^{(0)}\right) \tag{3.167}$$

(assumed positive), so that

$$\tau_3^{(0)} = \kappa^2 z^2 \left(u_0' + w_x^{(0)}\right)\left(u_0' + 2w_x^{(0)}\right). \tag{3.168}$$

The condition that $\partial\tau_3^{(0)}/\partial z$ have zero mean is then

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial z}[\kappa^2 z^2 (u_0'^2 + 2w_x^{(0)2})]\, dx = 0, \tag{3.169}$$

and thus $\partial\overline{\tau_3^{(0)}}/\partial z = 0$, where the overbar denotes the horizontal mean. Thus (with $\overline{\tau_3^{(0)}} = 1$ from the condition at $z = \infty$), u_0 is determined via

$$u_0'^2 + 2\overline{w_x^{(0)2}} = \frac{1}{\kappa^2 z^2}. \tag{3.170}$$

The non-zero quantity $2\overline{w_x^{(0)2}}$ represents the form drag due to the surface topography. Note that the logarithmic behaviour of u_0 near $z = 0$ is unaffected by this extra term, and we can take

$$u_0 = \frac{1}{\kappa} \ln z + O(z^2) \quad \text{as} \quad z \to 0. \tag{3.171}$$

In particular, since $p^{(0)} \approx p_0 + p_z^{(0)}|_{\varepsilon s}(z - \varepsilon s)$ as $z \to \varepsilon s$, and $p_z^{(0)}|_{\varepsilon s} = -w_x^{(0)}|_{\varepsilon s} = -s_{xx}$, we have

$$u^{(0)} \sim -p_0 + \frac{1}{\kappa}\ln z + s_{xx}(z - \varepsilon s) + O(z^2) \quad \text{as} \quad z \to \varepsilon s. \tag{3.172}$$

From (3.168), we now have

$$\tau_3^{(0)} = 1 + 3\kappa^2 z^2 u_0' w_x^{(0)} + \frac{\partial \Phi}{\partial x}, \tag{3.173}$$

where we define

$$\Phi = \int_{-\infty}^{x} 2\kappa^2 z^2 \left\{ w_x^{(0)2} - \overline{w_x^{(0)2}} \right\} dx. \tag{3.174}$$

Hence from (3.165),

$$u^{(1)} + p^{(1)} = \frac{\partial}{\partial z}\left[3\kappa^2 z^2 u_0' w^{(0)} + \Phi\right] + \tau_1^{(0)} - \tfrac{1}{2}\left(u^{(0)2} + w^{(0)2}\right) + u_0'(z)\psi^{(0)} + u_1(z), \tag{3.175}$$

where u_1 must be determined at $O(\varepsilon^2)$.

Now $u_0 \sim \dfrac{1}{\kappa}\ln z + O(z^2)$, $\Phi = O(z^2)$, $\tau_1^{(0)} = O(z)$, $w^{(0)} = s_x + O(z)$, $\psi^{(0)} = -s - zp_0 + O(z^2)$ (this last follows from manipulation of (3.166)). Therefore, as $z \to 0$,

$$u^{(1)} = -p_{10} + 3\kappa s_x - \tfrac{1}{2}\left[s_x^2 + \left\{-p_0 + \frac{1}{\kappa}\ln z\right\}^2\right] + \frac{1}{\kappa z}(-s - zp_0) + u_1 + O(z), \quad (3.176)$$

where $p_{10} = p^{(1)}|_{z=0}$.

3.7.7 Determination of p_{10}

Define the Green's function

$$K(x,z;\xi,\zeta) = -\frac{1}{4\pi}\left[\ln\{(x-\xi)^2 + (z-\zeta)^2\} + \ln\{(x-\xi)^2 + (z+\zeta)^2\}\right]. \quad (3.177)$$

We then have, for example,

$$\begin{aligned} p^{(0)} &= \iint_{\zeta>0} \left[K\nabla^2 p^{(0)} - p^{(0)}\nabla^2 K\right] d\xi\, d\zeta \\ &= \oint \left(K\frac{\partial p^{(0)}}{\partial n} - p^{(0)}\frac{\partial K}{\partial n}\right) ds \\ &= -\int_0^\infty K\frac{\partial p^{(0)}}{\partial \zeta}\, d\xi = \int_0^\infty K\frac{\partial w^{(0)}}{\partial \xi}\, d\xi = -\int_0^\infty w^{(0)}\frac{\partial K}{\partial \xi}\, d\xi, \end{aligned} \quad (3.178)$$

whence we derive (3.162) for example; the integrals with respect to ξ are taken along $\zeta = 0$.

Next, expanding (3.156) about $z = 0$, we find

$$w^{(1)} \sim (su^{(0)})_x \quad \text{as} \quad z \to 0. \quad (3.179)$$

Putting

$$w^{(1)} = s_x u_0 + W, \quad (3.180)$$

we deduce the condition

$$W = -(sp_0)_x \quad \text{on} \quad z = 0, \quad (3.181)$$

and from (3.164)

$$\begin{aligned} p_x^{(1)} - W_z &= R, \\ p_z^{(1)} + W_x &= S, \end{aligned}$$

where

$$\begin{aligned} R &= \tau_{1x}^{(0)} + \tau_{3z}^{(0)} - \{u^{(0)}u_x^{(0)} + w^{(0)}u_z^{(0)} - s_x u_0'\}, \\ S &= \tau_{3x}^{(0)} - \tau_{1z}^{(0)} - \{u^{(0)}w_x^{(0)} + w^{(0)}w_z^{(0)} + s_{xx}u_0\}. \end{aligned}$$

Also

$$\nabla^2 p^{(1)} = R_x + S_z, \tag{3.182}$$

and it follows from using the Green's function as in (3.178) that, after some manipulation involving Green's theorem,

$$p^{(1)} = \int\int_{\zeta>0} (RK_\xi + SK_\zeta)\, d\xi\, d\zeta - \int_0^\infty K_\xi W\, d\xi, \tag{3.183}$$

and therefore

$$p_{10} = \frac{1}{\pi}\int\int_{\zeta>0} \frac{[(x-\xi)R(\xi,\zeta) - \zeta S(\xi,\zeta)]}{(x-\xi)^2+\zeta^2}\, d\xi\, d\zeta - \frac{1}{\pi} ⨍_{-\infty}^{\infty} \frac{(sp_0)_\xi\, d\xi}{\xi - x}. \tag{3.184}$$

3.7.8 Matching

Overall, then, the outer solution can be written, as $z \to 0$, in the form (using (3.172))

$$\begin{aligned} u \;\sim\; & -p_0 + \frac{1}{\kappa}\ln z + s_{xx}(z - \varepsilon s) + \varepsilon\left[-p_{10} + 3\kappa s_x - \tfrac{1}{2}s_x^2 - \tfrac{1}{2}p_0^2 + \frac{p_0}{\kappa}\ln z \right. \\ & \left. - \frac{1}{2\kappa^2}\ln^2 z - \frac{s}{\kappa z} - \frac{p_0}{\kappa} + u_1\right]. \end{aligned} \tag{3.185}$$

If we define

$$\sqrt{\tau} = 1 + \varepsilon A_1 + \varepsilon^2 A_2 + \ldots, \tag{3.186}$$

then (3.153) takes the form

$$u \sim A_1 + \frac{1}{\kappa}\ln z + \varepsilon\left[-\frac{s}{\kappa z} + \frac{A_1}{\kappa}\ln z + A_2\right] + \ldots, \tag{3.187}$$

and the leading order term can be matched directly to that of (3.185) by choosing

$$A_1 = -p_0. \tag{3.188}$$

Using (3.188), (3.186) and (3.163), we have

$$\tau \approx 1 + \frac{2\varepsilon}{\pi} ⨍_{-\infty}^{\infty} \frac{s_\xi\, d\xi}{x - \xi}, \tag{3.189}$$

and this can be compared with (3.115). Whereas in (3.115) $K(x) = 0$ for $x < 0$, the kernel $K(x)$ in (3.189) is proportional to $1/x$ for all x, and thus non-zero for $x < 0$. Consequently, there is no instability, and to find this we need to progress to the next order term.

Unfortunately, the $O(\varepsilon)$ terms do not match because the terms $\pm\frac{p_0}{\kappa}\ln z$ in the two expansions are not equal, and also because of the linear term in (3.185). In order to match the expansions to $O(\varepsilon)$, we have to consider a further, intermediate layer: this is the shear layer we alluded to earlier.

3.7.9 Shear layer

A distinguished limit exists when $z = O(\varepsilon)$, and thus we put

$$\begin{aligned} z &= \varepsilon s + \varepsilon\zeta, \quad w = s_x + \varepsilon[u s_x + W], \\ \eta &= \varepsilon N, \quad \tau_1 = \varepsilon T_1, \\ u &= -p_0 + \frac{1}{\kappa}\ln(z - \varepsilon s) + \varepsilon v, \end{aligned} \tag{3.190}$$

and from (3.185) and (3.153) (using (3.186) and (3.188)), we require

$$\begin{aligned} v &\sim s_{xx}\zeta - p_{10} + 3\kappa s_x - \tfrac{1}{2}s_x^2 - \tfrac{1}{2}p_0^2 - \frac{p_0}{\kappa} + \frac{p_0}{\kappa}\ln\varepsilon\zeta + \left(u_1 - \frac{1}{2\kappa^2}\ln^2\varepsilon\zeta\right) \quad \text{as} \quad \zeta \to \infty, \\ v &\sim A_2 - \frac{p_0}{\kappa}\ln\varepsilon\zeta \quad \text{as} \quad \zeta \to 0. \end{aligned} \tag{3.191}$$

It follows from (3.190) that

$$\begin{aligned} N &= \kappa^2\zeta^2\frac{\partial u}{\partial\zeta}, \\ T_1 &= 2N[u_x - s_x u_\zeta], \\ \tau_3 &= N[u_\zeta + \varepsilon s_{xx} + O(\varepsilon^2)], \\ u_x + W_\zeta &= 0, \\ (u+p)_x - s_x p_\zeta &= \tau_{3\zeta} - \varepsilon[u u_x + W u_\zeta] + O(\varepsilon^2), \\ p_\zeta &= -\varepsilon s_{xx} + O(\varepsilon^2). \end{aligned} \tag{3.192}$$

Since we have $p = p_0 + \varepsilon p_{10}$ and $W = 0$ on $\zeta = 0$, then

$$\begin{aligned} p &= p_0 + \varepsilon(p_{10} - s_{xx}\zeta) + O(\varepsilon^2), \\ W &= p_0'\zeta + O(\varepsilon), \end{aligned} \tag{3.193}$$

and thus v satisfies

$$v_x + p_{10}' - s_{xxx}\zeta + s_x s_{xx} = \frac{\partial}{\partial\zeta}[2\kappa\zeta v_\zeta + \kappa\zeta s_{xx}] - \left[-p_0'\left(-p_0 + \frac{1}{\kappa}\ln\varepsilon\zeta\right) + \frac{p_0'}{\kappa}\right] + O(\varepsilon), \tag{3.194}$$

together with (3.191).

The solution of (3.194) is

$$v \approx -p_{10} - \tfrac{1}{2}s_x^2 - \frac{p_0}{\kappa} - \tfrac{1}{2}p_0^2 + \frac{p_0}{\kappa}\ln\varepsilon\zeta + s_{xx}\zeta + 3\kappa s_x + V, \tag{3.195}$$

where

$$\frac{\partial V}{\partial x} = \frac{\partial}{\partial\zeta}\left[2\kappa\zeta\frac{\partial V}{\partial\zeta}\right], \tag{3.196}$$

and (3.191) implies

$$V \to 0 \quad \text{as} \quad \zeta \to \infty,$$

$$V \sim A_2^* - \frac{2p_0}{\kappa} \ln \varepsilon\zeta \quad \text{as} \quad \zeta \to 0, \tag{3.197}$$

where

$$A_2 = A_2^* - p_{10} - \tfrac{1}{2}s_x^2 - \frac{p_0}{\kappa} - \tfrac{1}{2}p_0^2 + 3\kappa s_x. \tag{3.198}$$

The solution of (3.196) which tends to zero as $\zeta \to \infty$ is

$$V = \int_{-\infty}^{\infty} \hat{V}(\zeta, k) e^{ikx}\, dk, \tag{3.199}$$

where the Fourier transform $\hat{V}$ (as thus defined) is given by

$$\hat{V} = BK_0 \left[\left(\frac{2ik\zeta}{\kappa} \right)^{1/2} \right], \tag{3.200}$$

the square root is chosen so that $\operatorname{Re}(ik)^{1/2} > 0$, and K_0 is a modified Bessel function of order zero. Evidently we require

$$\hat{V} \sim \hat{A}_2^* - \frac{2\hat{p}_0}{\kappa} \ln(\varepsilon\zeta) \quad \text{as} \quad \zeta \to 0, \tag{3.201}$$

where the overhat defines the Fourier transform, in analogy to (3.199). Now $K_0(\xi) \sim -\ln \frac{1}{2}\xi - \gamma$ as $\xi \to 0$, where $\gamma \approx 0.5772$ is the Euler-Mascheroni constant. Also

$$\left(\frac{2ik\zeta}{\kappa} \right)^{1/2} = \left(\frac{2|k|\zeta}{\kappa} \right)^{1/2} \exp\left[\frac{i\pi}{4} \operatorname{sgn} k \right]; \tag{3.202}$$

therefore (3.200) implies

$$\hat{V} \sim -B \left[\gamma + \tfrac{1}{2}\ln|k| - \tfrac{1}{2}\ln 2\kappa + \tfrac{1}{2}\ln\zeta + \frac{i\pi}{4}\frac{k}{|k|} \right], \tag{3.203}$$

and matching this to (3.201) implies

$$B = \frac{4\hat{p}_0}{\kappa}, \tag{3.204}$$

whence

$$\hat{A}_2^* = \frac{2\hat{p}_0}{\kappa} \ln \varepsilon - \frac{4\hat{p}_0}{\kappa} \left[\gamma + \tfrac{1}{2}\ln|k| - \tfrac{1}{2}\ln 2\kappa + \frac{i\pi k}{4|k|} \right]. \tag{3.205}$$

We have $\hat{s}_x = ik\hat{s}$, $\widehat{H(s_x)} = -|k|\hat{s}$, and $\widehat{J * s_x} = |k|\hat{s}\ln|k|$, where $J * s_x$ is the convolution of J with s_x, and $\hat{J} = -(i/2\pi)\ln|k| \operatorname{sgn} k$. (The convolution theorem here takes the form $\widehat{f * g} = 2\pi\hat{f}\hat{g}$.) It follows from this that

$$J(x) = -\frac{1}{\pi x}\left[\gamma + \ln|x|\right]. \tag{3.206}$$

Thus

$$A_2^* = \frac{2}{\kappa}\left(\ln 2\varepsilon\kappa - 2\gamma\right)p_0 + \frac{\pi}{\kappa}s_x + \frac{1}{\pi\kappa}J * s_x, \tag{3.207}$$

and, from (3.198),

$$\begin{aligned} A_2 &= \frac{2}{\kappa}\left(\ln 2\varepsilon\kappa - 2\gamma - \tfrac{1}{2}\right)p_0 + \left(\frac{\pi}{\kappa} + 3\kappa\right)s_x + \frac{1}{\pi\kappa}J * s_x \\ &\quad -p_{10} - \tfrac{1}{2}p_0^2 - \tfrac{1}{2}s_x^2, \end{aligned} \tag{3.208}$$

where J is given by (3.206), $p_0 = H(s_x)$ ((3.163)), and p_{10} is given by (3.184).

3.7.10 Linear stability

The Exner equation is, in appropriate dimensionless form,[7]

$$\varepsilon s_t + q_x = 0, \tag{3.209}$$

and since $q = q(\tau)$,

$$q = q_1 - 2\varepsilon q_1' p_0 + \varepsilon^2\left[(2A_2 + p_0^2)q_1' + 2p_0^2 q_1''\right] + \ldots, \tag{3.210}$$

where $q_1 = q(1)$, $q_1' = q'(1)$, $q_1'' = q''(1)$.

Thus s satisfies the nonlinear evolution equation

$$\frac{\partial s}{\partial t} - 2q_1'\frac{\partial p_0}{\partial x} + \varepsilon\frac{\partial}{\partial x}\left[(2A_2 + p_0^2)q_1' + 2p_0^2 q_1''\right] \approx 0. \tag{3.211}$$

This is

$$\begin{gathered} \frac{\partial s}{\partial t} - \alpha\frac{\partial p_0}{\partial x} + \varepsilon\frac{\partial}{\partial x}\left[q_1'\left(2\omega s_x + 2\lambda J * s_x - 2p_{10} - s_x^2\right) + 2q_1'' p_0^2\right] = 0, \\ p_0 = H(s_x), \end{gathered} \tag{3.212}$$

where

$$\begin{aligned} \alpha &= 2q_1'\left[1 - \frac{2\varepsilon}{\kappa}\left(\ln 2\varepsilon\kappa - 2\gamma - \tfrac{1}{2}\right)\right], \\ \omega &= \frac{\pi}{\kappa} + 3\kappa, \\ \lambda &= \frac{1}{\pi\kappa}. \end{aligned} \tag{3.213}$$

We linearise (3.212) for small s by neglecting the terms in s_x^2 and p_0^2. Taking the Fourier transform (as defined here in (3.199)), we have

$$\hat{s}_t = ik\alpha\widehat{p_0} - ik\varepsilon q_1'\left(2\omega ik\hat{s} + 4\pi\lambda ik\hat{J}\hat{s} - 2\widehat{p_{10}}\right). \tag{3.214}$$

From (3.184),

$$p_{10} = \int_0^\infty (a * R + b * S)\,d\zeta - H\{(sp_0)_x\}, \tag{3.215}$$

---

[7] Note that the definition of ε here is that pertaining to the mixing length theory, i. e., (3.133) and not (3.47).

where

$$a(x,\zeta)=\frac{x}{\pi(x^2+\zeta^2)},\quad b(x,\zeta)=-\frac{\zeta}{\pi(x^2+\zeta^2)}. \tag{3.216}$$

Hence, neglecting the quadratic Hilbert transform term,

$$\hat{p}_{10}=2\pi\int_0^\infty(\hat{a}\hat{R}+\hat{b}\hat{S})d\zeta. \tag{3.217}$$

Calculation of $\hat{a}$ and $\hat{b}$ gives

$$\hat{a}=-\frac{i}{2\pi}e^{-|k|\zeta}\operatorname{sgn}k,\quad \hat{b}=-\frac{1}{2\pi}e^{-|k|\zeta}, \tag{3.218}$$

so that

$$\hat{p}_{10}=-\int_0^\infty[i\hat{R}\operatorname{sgn}k+\hat{S}]e^{-|k|\zeta}\,d\zeta. \tag{3.219}$$

Now

$$\begin{aligned}
\tau_3^{(0)} &= 1+3\kappa z w_x^{(0)}+2\kappa^2z^2w_x^{(0)2},\\
\tau_1^{(0)} &= -2\kappa z w_z^{(0)}-2\kappa^2z^2w_x^{(0)}w_z^{(0)},\\
u^{(0)} &= u_0(z)-p^{(0)},\\
u_x^{(0)} &= -w_z^{(0)},\\
u_z^{(0)} &= u_0'+w_x^{(0)},
\end{aligned} \tag{3.220}$$

thus, retaining only the perturbed linear (in s) terms, we have from (3.182)

$$\begin{aligned}
\hat{R} &\approx ik\hat{t}_1+\hat{t}_{3z}+u_0\hat{w}_z-u_0'[\hat{w}-ik\hat{s}],\\
\hat{S} &\approx ik\hat{t}_3-\hat{t}_{1z}-iku_0[\hat{w}+ik\hat{s}],
\end{aligned} \tag{3.221}$$

where $\hat{w}=\widehat{w^{(0)}}$, and

$$\hat{t}_1=-2ik\kappa z\hat{p},\quad \hat{t}_3=3ik\kappa z\hat{w}, \tag{3.222}$$

where $\hat{p}=\widehat{p^{(0)}}$.

Finally, from (3.162),

$$w^{(0)}=-b(x,z)*s_x,\quad p^{(0)}=-a(x,z)*s_x, \tag{3.223}$$

whence using (3.218),

$$\begin{aligned}
\hat{w} &= ik\hat{s}e^{-|k|z},\\
\hat{p} &= -|k|\hat{s}e^{-|k|z},
\end{aligned} \tag{3.224}$$

and we eventually obtain

$$\hat{p}_{10}=-\hat{s}\int_0^\infty\left[k^2u_0(1+2e^{-|k|\zeta})-|k|u_0'(1-e^{-|k|\zeta})-5i\kappa k|k|e^{-|k|\zeta}\right]e^{-|k|\zeta}\,d\zeta. \tag{3.225}$$

Simplification of this, using the fact that $\int_0^\infty e^{-t}\ln t\,dt = -\gamma$, where $\gamma \approx 0.5772$ is the Euler-Mascheroni constant, yields

$$\hat{p}_{10} = \hat{s}\left[\frac{2|k|}{\kappa}\left(\ln 2|k| + \gamma\right) + \tfrac{5}{2}i\kappa k\right]. \tag{3.226}$$

Solutions of (3.214) are $\hat{s} = e^{\sigma t}$, where $\sigma = r - ikc$, and after some simplification, we find that the growth rate ρ is

$$r = 2k^2\varepsilon q_1'\left(\frac{\pi}{\kappa} + \tfrac{1}{2}\kappa\right), \tag{3.227}$$

and the wave speed c is

$$c = 2q_1'|k|\left[1 + \frac{2\varepsilon}{\kappa}\left\{-\left(1 - \frac{1}{2\pi}\right)\ln|k| - \ln 4\varepsilon\kappa + \gamma + \tfrac{1}{2}\right\}\right]. \tag{3.228}$$

Thus dunes grow, as $r > 0$, on a time scale of $O(1/\varepsilon)$, while the wave forms move downstream at a speed $c \approx 2q_1'|k| = O(1)$.

This more realistic theory for dune-forming instability is less satisfactory than the constant eddy viscosity theory, because the growth rate $r \propto k^2$, and the basic model is again ill posed. As before, we can stabilise the model by including the downslope force, thus replacing the stress by the effective stress defined using (3.120). The effect of this is to add a term to the stress definition in (3.186), which can then be written as

$$\tau_e = 1 - 2\varepsilon p_0 + 2\varepsilon^2\left(A_2 + \ldots\right) - \hat{\beta}s_x, \tag{3.229}$$

where the definition of $\hat{\beta}$ differs from that in (3.122) because of the different scaling used in the aeolian model. Using (3.136), $x \sim d$ and $s \sim \varepsilon d$, we find

$$\hat{\beta} = \frac{\Delta\rho}{\rho}\frac{D_s}{d}\frac{1}{\varepsilon F^2}, \tag{3.230}$$

where the Froude number is

$$F^2 = \frac{U_\infty^2}{gd}. \tag{3.231}$$

Using values $\Delta\rho/\rho = 10^3$, $D_s/d = 10^6$, $\varepsilon \sim 0.1$, we find $\hat{\beta} \sim \varepsilon^2/F^2$.

If we consult (3.208), we see that the destabilising term arises from that proportional to s_x in A_2. Effectively we can write

$$\tau_e = 1 + \ldots + \left[\varepsilon^2\left(\frac{2\pi}{\kappa} + \kappa\right) - \hat{\beta}\right]\frac{\partial s}{\partial x} + \ldots, \tag{3.232}$$

where the modification of the coefficient ω reflects the effect of the terms in J and p_{10}, as indicated by (3.227). We see that the downslope term stabilises the system if $\hat{\beta} > O(\varepsilon^2)$, and thus practically if $F^2 < 1$. On the Earth, a typical value is $F^2 = 0.04$, so that the instability is removed, at all wave numbers. This is distinct from the constant eddy viscosity case, because the stabilising term has the same wave number dependence as the destabilising one.

This situation is somewhat reminiscent of the rill-forming instability of chapter 2. There the instability is regularised at long wavelength by inclusion of singularly perturbed terms. The most obvious modification to make here in a similar direction is to allow for a finite thickness of the moving sand layer. It seems likely that this will make a substantial difference, because the detail of the mixing length model relies ultimately on the existence of an exponentially small roughness layer through which the wind speed drops to zero. It is noteworthy that the constant eddy viscosity model does not share this facet of the problem.

3.8 Separation at the wave crest

The constant eddy viscosity model can produce a genuine instability, with decay at large wave numbers. If pushed to a nonlinear régime, it allows shock formation, although it also allows unlimited wavelength growth. The presumably more accurate mixing length theory actually fares somewhat worse. It can produce a very slow instability via an effective negative diffusivity, but this is easily stabilised by downslope drift. It is possibly the case that specific consideration of the mobile sand layer will alleviate this result.

A complication arises at this point. Aeolian sand dunes inevitably form slip faces. There is a jump in slope at the top of the slip face, and the wind flow separates, forming a wake (or cavity, or bubble). One authority is of the opinion that no model can be realistic unless it includes a consideration of separation. In this section we will consider a model which is able to do this. Before doing so, it is instructive to consider how such separation arises.

If the constant eddy viscosity model has any validity, it suggests that the uniform flat bed is unstable, and that travelling waves grow to form shocks. If the slope within the shocks is steep enough to exceed the angle of repose of sand grains (some 34°), then a slip face will occur, with the sand resting on the slip face at this angle. The turbulent flow over the dune inevitably separates at the cusp of the dune, forming a separation bubble, as indicated in figure 3.10. The formation of a separation bubble makes the model fundamentally nonlinear, and it provides a possible mechanism for length scale selection. It is thus an attractive way out of the conundrums concerning instability alluded to above.

It is simplest to treat the separation bubble in the context of the mixing length theory, and this we now do, despite our misgivings about its applicability for small amplitude perturbations. We suppose that there is a periodic sequence of dunes, with period chosen to be 2π. We suppose that there is a slip face, as shown in figure 3.10, and we suppose the corresponding separation bubble occupies the interval (a, b). We denote the bubble interval as B, and the corresponding attached flow region as B'.

Because our method will use complex variables, it is convenient to rechristen the space coordinates as x and y, and the corresponding velocity components as u and v.

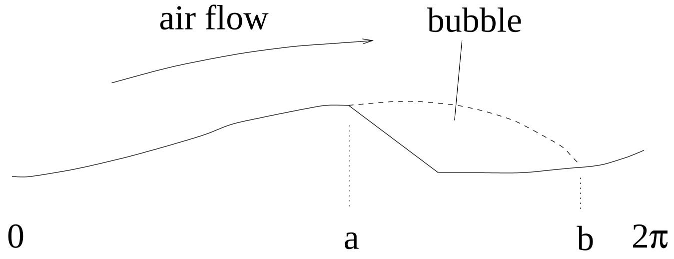

Figure 3.10: Separation behind a dune.

At leading order, the inviscid flow is described by the outer equations (3.158):

$$\begin{aligned} u_x + v_y &= 0, \\ u_x + p_y &= 0, \\ v_x + p_y &= 0, \end{aligned} \tag{3.233}$$

and these are valid in $y > \varepsilon s$. From these it follows that p and v satisfy the Cauchy-Riemann equations, and thus

$$p + iv = f(z) \tag{3.234}$$

is an analytic function, where $z = x + iy$.

The boundary conditions for p and v are that both tend to zero as $y \to \infty$, and v satisfies the no flow through condition (3.156), $v = s_x + \varepsilon u s_x$ on $y = \varepsilon s$. These completely specify the problem in the absence of a separation bubble.

If we suppose that a separation bubble occurs, as shown in figure 3.10, then its upper boundary is unknown, and must be determined by an extra boundary condition. We let $y = \varepsilon s(x)$ denote this unknown upper boundary, and define the ground surface to be $y = \varepsilon s_0(x)$; thus $s(x) = s_0(x)$ for $x \in B'$.

There are various ways to provide the extra condition. Two such are that the pressure, or alternatively the vorticity, are constant in the bubble. We shall suppose the former, and therefore we prescribe

$$p = p_B \quad \text{for} \quad y = \varepsilon s, \quad x \in B. \tag{3.235}$$

The bubble pressure p_B is an unknown constant, and must be prescribed as part of the solution.

Separation occurs because the viscous boundary layer (here, the roughness layer) detaches from the surface, forming a free shear layer at the top of the bubble, which rapidly thickens to form a more diffuse upper boundary. The assumption of constant pressure in the bubble is essentially a consequence of this shear layer, implying that mean fluid velocities in the bubble are small.

For small ε, we can expand the boundary conditions at $y = \varepsilon s$ about $y = 0$, so that to leading order, the problem becomes that of finding an analytic function

$f(z) = p + iv$ in the upper half plane $\operatorname{Im} z > 0$, satisfying

$$\begin{aligned} f &\to 0 \quad \text{as} \quad z \to \infty, \\ v = s_x &\quad \text{on} \quad y = 0, \\ p = p_B &\quad \text{on} \quad y = 0, \quad x \in B. \end{aligned} \tag{3.236}$$

The extra pressure condition should help determine s in B, but the endpoint locations are not necessarily known. Specification of the behaviour of the solution at the endpoints is necessary to determine these. Firstly, we expect s to be continuous at the end points:

$$s(a) = s_0(a), \quad s(b) = s_0(b). \tag{3.237}$$

A difference now arises depending on whether a slip face occurs or not. If not, then the bed slope is continuous, and at the upstream end point $x = a$, we might surmise that boundary layer separation is associated with the skin friction dropping to zero. Now from (3.186) and (3.188), we have the surface stress defined by

$$\sqrt{\tau} = 1 - \varepsilon p_0, \tag{3.238}$$

where p_0 is the surface pressure. The only apparent interpretation of this which we can make in our simplified model is to require that

$$p \to +\infty \quad \text{on} \quad y = 0 \quad \text{as} \quad x \to a- \in B'; \tag{3.239}$$

more detailed consideration of the boundary layer structure near the separation point would be necessary to be more precise than this. We do not pursue this possibility here, mainly because the more relevant situation is when a slip face is present.

If we suppose a slip face is present, then we can presume that separation occurs at its top, and this determines the point $x = a$. In addition, it is natural to suppose that boundary layer detachment occurs smoothly, in the sense that we suppose the slope of s is continuous at a:

$$s'(a+) = s_0'(a-); \tag{3.240}$$

this implies that v is continuous at $x = a$. If possible, we would like to have smooth reattachment at b, and in addition (and in fact, because of this) continuity of pressure also:

$$[p]_{b-}^{b+} = [p]_{a-}^{a+} = 0, \quad s'(b-) = s_0'(b+). \tag{3.241}$$

We shall in fact find that all these conditions can be satisfied. This is not always the case in such problems, and sometimes (worse) singularities have to be tolerated. The choice of the behaviour of the solution at the end points actually constitutes the most subtle part of solving Hilbert problems.

3.8.1 Formulation of Hilbert problem

The first thing we do is to analytically continue $f(z)$ into the lower half plane. Specifically, we define

$$G(z) = \begin{cases} \frac{1}{2}\left[f(z) - p_B\right], & \operatorname{Im} z > 0, \\ \\ -\frac{1}{2}\left[\overline{f(\bar{z})} - p_B\right], & \operatorname{Im} z < 0. \end{cases} \tag{3.242}$$

G is analytic in both the upper and lower half planes, and if G_+ and G_- denote the limiting values of G as $z \to x$ from above and below, then

$$G_+ + G_- = is',$$
$$G_+ - G_- = p - p_B, \tag{3.243}$$

everywhere on the real axis.

Because of the assumed periodicity in x, we make the following transformations:

$$\zeta = e^{iz}, \quad \xi = e^{ix}, \quad G(z) = H(\zeta). \tag{3.244}$$

The geometry of the problem is then illustrated in figure 3.11. The problem to solve is identical to (3.243), replacing G by H, and thus we have the standard Hilbert problem

$$\begin{aligned} H_+ - H_- &= 0 \quad \text{on} \quad B, \\ H_+ + H_- &= i\sigma_0 \quad \text{on} \quad B', \end{aligned} \tag{3.245}$$

where $\sigma_0(\xi) = s_0'(x)$. We have to solve this subject to the supplementary conditions

$$H(0) = -\tfrac{1}{2}p_B, \quad H(\infty) = \tfrac{1}{2}p_B\,; \tag{3.246}$$

the first of these in fact implies the second automatically. We seek to apply the conditions that both $\frac{1}{2}(p - p_B) = \operatorname{Re} H$ and $\frac{1}{2}v = \operatorname{Im} H$ are continuous (thus H is continuous) at both endpoints $\xi = \xi_a = e^{ia}$ and $\xi = \xi_b = e^{ib}$. Given H satisfying (3.245), then the separation bubble boundary is given by the solution of

$$s' = -2iH \quad \text{on} \quad B, \quad s(a) = s_0(a), \tag{3.247}$$

and the pressure on B' is given by

$$p = p_B + H_+ - H_- \quad \text{on} \quad B'. \tag{3.248}$$

Solution

The solution to (3.245), given the location of a and b, is as follows. Define a function $\chi(\zeta)$ such that

$$\chi_+ + \chi_- = 0 \quad \text{on} \quad B' \tag{3.249}$$

(and χ is analytic away from B'); then

$$\left(\frac{H}{\chi}\right)_+ - \left(\frac{H}{\chi}\right)_- = \frac{i\sigma_0}{\chi_+}, \tag{3.250}$$

and by the discontinuity theorem, we have

$$H = \frac{\chi(\zeta)}{2\pi i}\int_{B'} \frac{i\sigma_0(t)\,dt}{\chi_+(t)(t-\zeta)} + \chi P, \tag{3.251}$$

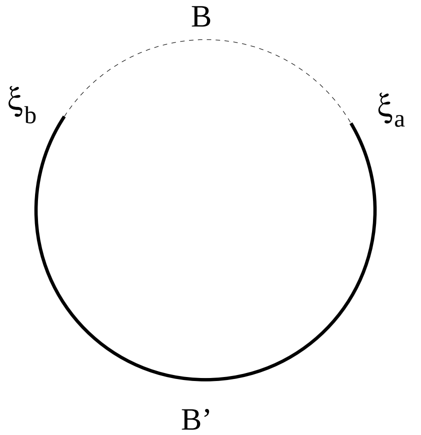

Figure 3.11: B and B' on the unit circle in the complex ζ plane. B' is a branch cut for the solution of the Hilbert problem (3.245).

where P is an as yet undetermined polynomial. To find P, we must specify χ, and this in turn depends on the required singularity structure of the solution.

The smoothness of H is essentially that of χ, and so we will choose the function

$$\chi = [(\zeta - \xi_a)(\zeta - \xi_b)]^{1/2}. \tag{3.252}$$

The most general choice is $\chi = (\zeta - \xi_b)^{m_b + \frac{1}{2}}(\zeta - \xi_a)^{m_a + \frac{1}{2}}$, where m_a and m_b are integers, but most of these possibilities can in general be eliminated by requirements either of continuity or at least integrability of the solution.

We consider the behaviour of the Cauchy integral

$$\Phi(\zeta) = \frac{1}{2\pi i}\int_{B'} \frac{\phi(t)\,dt}{t - \zeta} \tag{3.253}$$

near the end points of integration. Note that in the present case,

$$\phi(t) = \frac{i\sigma_0(t)}{\chi_+(t)}. \tag{3.254}$$

First suppose that $\phi(t)$ is continuous at an end point.[8] Then we have

$$\Phi(\zeta) = \pm\frac{\phi(c)}{2\pi i}\ln(\zeta - c) + O(1), \tag{3.255}$$

where c denotes either end point of B', and the upper and lower signs apply at the right (ξ_a) and left (ξ_b) hand ends of B', respectively. (3.255) applies as $\zeta \to c$, with $\zeta \notin B'$.

---

[8] More precisely, ϕ should be *Hölder continuous*, that is $|\phi(t_1) - \phi(t_2)| < K|t_1 - t_2|^\gamma$, for some positive γ.

Similarly, for $\xi \in B'$,

$$\Phi(\xi) = \pm \frac{\phi(c)}{2\pi i} \ln(\xi - c) + O(1), \tag{3.256}$$

where $\Phi(\xi)$ denotes the principal value of the integral (and $\Phi(\xi) = \frac{1}{2}\left[\Phi_+(\xi) + \Phi_-(\xi)\right]$). Bearing in mind (3.254), we see that if χ is unbounded at c, and specifically goes algebraically to infinity, then the corresponding Cauchy integral is bounded, and thus H will be unbounded (unless the choice of P can be chosen to remove the singularity). Using the definition

$$H = \chi(\zeta)\left[\Phi(\zeta) + P\right], \tag{3.257}$$

we have from (3.247) and (3.248) that

$$\begin{aligned} s' = -2iH(\xi) = -2i\chi(\xi)\left[\Phi(\xi) + P\right] \quad &\text{on} \quad B, \\ p - p_B = 2\chi_+(\xi)\left[\Phi(\xi) + P\right] \quad &\text{on} \quad B'. \end{aligned} \tag{3.258}$$

The implication of this is that if χ is unbounded at an end point, then in general both p and s' will also be unbounded, unless the choice of P removes the singularity. The worst singularity we can tolerate is an integrable one, thus $\chi \sim (\zeta - c)^{-1/2}$.

Now suppose that χ is bounded at an end point, and specifically $\chi \sim (\zeta - c)^{1/2}$. (Any higher power causes the Cauchy integral to be undefined, because then ϕ is not integrable.) If we define $\tilde{\phi}$ via

$$\phi(t) \sim \frac{\tilde{\phi}(t)}{(t - c)^{1/2}} \quad \text{as} \quad t \to c, \tag{3.259}$$

then

$$\begin{aligned} \Phi(\zeta) &= \frac{\tilde{\phi}(c)}{2(\zeta - c)^{1/2}} + o\left(\frac{1}{(\zeta - c)^{1/2}}\right), \quad \zeta \in B, \\ \Phi(\xi) &= o\left(\frac{1}{(\xi - c)^{1/2}}\right), \quad \xi \in B'. \end{aligned} \tag{3.260}$$

It then follows from (3.258) that s' is bounded (and in fact continuous) and p is continuous at c. It is because of this that we choose χ as defined in (3.252), in order to satisfy the smoothness conditions (3.240) and (3.241).

In this case, the polynomial P must be zero in order to satisfy the condition at $\zeta = \infty$, and we have

$$H = \frac{\chi(\zeta)}{2\pi i} \int_{B'} \frac{i\sigma_0(t)\,dt}{\chi_+(t)(t - \zeta)}. \tag{3.261}$$

We define the integrals

$$I_0 = \frac{\chi_0}{2\pi i} \int_{B'} \frac{i\sigma_0(t)\,dt}{t\chi_+(t)}, \quad I_\infty = \frac{1}{2\pi i} \int_{B'} \frac{i\sigma_0(t)\,dt}{\chi_+(t)}; \tag{3.262}$$

we thus have $H(0) = I_0$, $H(\infty) = -I_\infty$, and the conditions In (3.246) correspond to prescribing

$$I_0 = I_\infty = -\tfrac{1}{2}p_B. \tag{3.263}$$

It is a straightforward exercise in contour integration to show that $\bar{I}_0 = I_\infty$, where the overbar denotes the complex conjugate, therefore (3.263) is tantamount to the single condition $I_0 = -\frac{1}{2}p_B$. Because this is a complex-valued integral, (3.263) actually comprises two conditions for the two unknown quantities p_B and b.

It remains to be seen whether s is continuous at b. Since (3.263) determines b, and s is fully determined by (3.247), it is not obvious that this will be the case. (If it were not, we would have to allow for a singularity in the solution at one of the end points.)

In fact, it is easy to show that (3.263) automatically implies that s is continuous at b. To show this, it is sufficient to show that s is continuous over the periodic domain $[0, 2\pi]$. Equivalently, we need to show that

$$I = \int_0^{2\pi} s'\,dx = \int_{B\cup B'} -i(H_+ + H_-)\,\frac{d\xi}{i\xi} = 0, \tag{3.264}$$

using (3.247) and (3.245). Denoting contours just inside and outside the unit circle as C_+ and C_- (see figure 3.12), we see that

$$I = -\left[\int_{C_+} \frac{H\,d\xi}{\xi} + \int_{C_-} \frac{H\,d\xi}{\xi}\right]. \tag{3.265}$$

H is analytic inside and outside the unit circle. The integral over C_+ is thus just $2\pi i H(0)$ using the residue theorem, while the integral over C_- can be extended by deforming the contour out to infinity, whence we obtain the integral $2\pi i H(\infty)$. Thus

$$I = -2\pi i[I_0 - I_\infty] = 0, \tag{3.266}$$

and continuity of s at b is assured. We have thus obtained a solution in which the separated streamline leaves and rejoins the surface smoothly, and the pressure is continuous at the end points.

3.8.2 Calculation of the free boundary

In order to solve (3.247) for s, we need to evaluate H on B'. There are various ways to do this. One simple one, which may be convenient for subsequent evolution of the bed using spectral methods, is to use a Fourier series representation.[9] Let us suppose that

$$s(x) = \sum_{k=-\infty}^{\infty} a_k e^{ikx}, \tag{3.267}$$

so that

$$i\sigma_0(\xi) = \sum_{k=-\infty}^{\infty} d_k \xi^k, \tag{3.268}$$

---

[9] One would suppose that a bed with discontinuous derivative at the slip face would make this an inefficient method. However, since evaluation of s in B only requires the evaluation of an integral over B', it is possible to eliminate this problem by artificially smoothing s_0 near $x = a$.

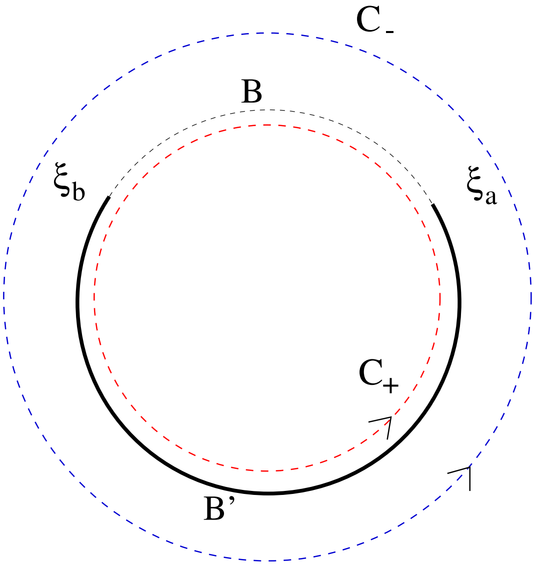

Figure 3.12: The contours C_+ and C_- lie just inside and outside the unit circle, respectively.

where

$$d_k = -k a_k. \tag{3.269}$$

We suppose that the Laurent expansion for $i\sigma_0$ extends to the complex plane as an analytic function with singularities only at 0 and ∞. (This is automatically true for any finite such series.) Then we can write the solution for H as

$$H = \tfrac{1}{2}[i\sigma_0(\zeta) - q(\zeta)\chi(\zeta)], \tag{3.270}$$

where q has a Laurent expansion

$$q = \sum_{k=-\infty}^{\infty} l_k \zeta^k. \tag{3.271}$$

Then we obtain s by solving

$$s' = s_0' + iq\chi \tag{3.272}$$

on $[a, b]$, with $s(a) = s_0(a)$. In practice, we obtain b by shooting.

Suppose that

$$\frac{1}{\chi(\zeta)} = \sum_{r=0}^{\infty} \frac{f_r}{\zeta^{r+1}}, \quad |\zeta| > 1 \tag{3.273}$$

(see question 3.8 for one way to calculate the coefficients); then we can write

$$\frac{H}{\frac{1}{2}\chi} = \sum_{m=-\infty}^{\infty} d_m \zeta^m \sum_{r=0}^{\infty} \frac{f_r}{\zeta^{r+1}} - \sum_{j=-\infty}^{\infty} l_j \zeta^j. \tag{3.274}$$

As $\zeta \to \infty$, $\chi \sim \zeta$ and $H \to \frac{1}{2}p_B$; equating coefficients of ζ^j in (3.274 for $j \geq 0$ yields

$$l_j = \sum_{r=0}^{\infty} d_{j+r+1} f_r, \quad j \geq 0, \tag{3.275}$$

and for $j = 0$ we have

$$p_B = \sum_{r=0}^{\infty} d_r f_r - l_{-1}. \tag{3.276}$$

For $|\zeta| < 1$, we find

$$\frac{1}{\chi(\zeta)} = \frac{1}{\chi_0} \sum_{r=0}^{\infty} \bar{f}_r \zeta^r, \tag{3.277}$$

and thus

$$\frac{H}{\frac{1}{2}\chi} = \frac{1}{\chi_0} \sum_{r=0}^{\infty} \bar{f}_r \zeta^r \sum_{m=-\infty}^{\infty} d_m \zeta^m - \sum_{j=-\infty}^{\infty} l_j \zeta^j. \tag{3.278}$$

As $\zeta \to 0$, $H \to -\frac{1}{2}p_B$; equating powers of ζ^j for $j \leq -1$, we find

$$l_j = \frac{1}{\chi_0} \sum_{r=0}^{\infty} \bar{f}_r d_{j-r}, \quad j \leq -1, \tag{3.279}$$

and for $j = 0$ we have

$$\frac{p_B}{\chi_0} = \frac{1}{\chi_0} \sum_{r=0}^{\infty} \bar{f}_r d_{-r} - l_0. \tag{3.280}$$

Putting these results together, we find that (3.276) and (3.280) together give (bearing in mind that $d_{-k} = -\bar{d}_k$)

$$p_B = \sum_{r=0}^{\infty} d_r f_r + \bar{\chi}_0 \sum_{r=0}^{\infty} \bar{f}_r \bar{d}_{r+1}, \tag{3.281}$$

with the added constraint that p_B is real.

We can now use the definitions of l_j in (3.275) and (3.279) to evaluate $iq\chi$ in (3.272). Being careful with the arguments, we find that on B,

$$\chi = 2\xi^{1/2}\chi_0^{1/2} R, \tag{3.282}$$

where

$$\chi_0 = \exp\left[\tfrac{1}{2}(a+b)\right], \quad R = \left[\sin\left(\frac{x-a}{2}\right)\sin\left(\frac{b-x}{2}\right)\right]^{1/2}, \tag{3.283}$$

and after some algebra, we have the differential equation for s on B:

$$s' = s_0' - 4R \,\mathrm{Im}\left[\chi_0^{1/2} \sum_{j=0}^{\infty} \sum_{r=0}^{\infty} f_r d_{j+r+1} \exp\left\{i\left(j + \tfrac{1}{2}\right)x\right\}\right], \tag{3.284}$$

with initial condition $s(a) = s_0(a)$. To solve this, guess b; we can then calculate the right hand side. Solving for s, we adjust b by decreasing it if s reaches s_0 for $x < b$, and increase it if s remains $> s_0$ for all $x \leq b$.

3.9 Notes and references

3.9.1 Patterns in rivers

Bars

Braiding

Meanders

3.9.2 Linear stability

3.9.3 Flow over a hill

Separation

Vosper *et al.* (2002) O'Malley *et al.* (1991) Parsons *et al.* (2004)

3.9.4 Computational dune theory

The Herrmann model

CFD results

Exercises

3.1 Generation of secondary flow by small perturbation of a straight river.

3.2 Write down the Exner equation for bedload transport, and show how it can be used to study the onset of bedform instability, assuming a suitable bedload transport law. Show that in conditions of slow flow, the resultant equation for the bed profile $s(x,t)$ is a first order hyperbolic equation, and deduce that the profile is neutrally stable. Show also that bed waves will form shocks which propagate downstream.

Now suppose that the bedload transport $q_b(x,t)$ is a function of the basal stress τ evaluated at $x-\delta$. Show that instability can occur if $\delta < 0$, i.e. the stress *leads* the bed profile.

Can you think of a physical reason why such a lead should occur?

Do you think such a model would be a good *nonlinear* model?

3.3 Parker (J. Fluid Mech. **89**, p. 109 (1978)) suggests the following expressions for the erosion and deposition rates in a stream:

$$v_E = \beta u_*^3/v_s^2, \quad v_D = \frac{\gamma v_s^2 c}{\rho_s u_*},$$

where c is the sediment concentration (mass per unit volume), v_s is the settling velocity, u_* is the 'friction' velocity $(\tau/\rho_w)^{1/2}$, and β and γ are constants ($\approx$ 0.007 and 13, respectively).

Find appropriate scales for x, t and c if $h \sim h_0$ and q is the fluid flux per unit width. Hence derive the dimensionless model for slow flow

$$\varepsilon \frac{\partial}{\partial t}(hc) + \frac{\partial c}{\partial x} = h^{3/2} - \frac{c}{h^{1/2}} = \frac{\partial h}{\partial t},$$

where

$$\varepsilon = \frac{c_0}{\rho_s(1-n)}.$$

By analysing the stability of the basic state $h = c = 1$, show that, for ε small, an erosional instability will occur.

More generally, derive a stability criterion (when ε is small) if $v_E = E(h)$, $v_D = cV(h)$. How is the result affected if ε is *not* small?

3.4 A simple model of bed erosion based on the St. Venant equations can be written in dimensionless form as

$$\varepsilon h_t + (uh)_x = 0,$$

$$F^2(\varepsilon u_t + uu_x) = -\eta_x + \delta\left(1 - \frac{u^2}{h}\right),$$

$$h(\varepsilon c_t + uc_x) = E^*(u) - c = -s_t,$$

where $h = \eta - s$. Explain a plausible basis for the derivation of this model. By considering the stability of the steady state $u = h = c = 1$ on a time scale t of $O(1)$, and assuming that $\delta \ll 1$, $\varepsilon \ll 1$, show that instability can occur depending on the size of $E^{*\prime}(1)$. Show also that η and s are out of phase if $F < 1$, and in phase if $F > 1$; interpret this in terms of dune and anti-dune formation.

3.5 The Exner equation for bed evolution is written in the form

$$(1-n)s_t + q_x = 0,$$

and the bedload transport is given by

$$q_x = K\left[q_0(\tau) - q\right],$$

where τ is the bed shear stress.

Explain in physical terms why such an equation should be appropriate to describe bedload transport.

Show that if the bed stress is $\tau = f\rho u^2$, then the momentum equation of St. Venant can be written in the approximate form

$$\frac{h}{2f}\tau_x + \tau = \rho g h(S_0 - s_x).$$

Supposing that the depth h is approximately constant, show how to non-dimensionalise these equations to obtain the set

$$s_t + q_x = 0,$$

$$\delta q_x = q_0(\tau) - q,$$

$$\tau_x + \tau = 1 - s_x,$$

and identify the parameter δ.

Write down a suitable steady state solution, and show that if $q_0(\tau)$ is a monotonically increasing function of τ, then the steady state is linearly unstable if $K > 0$. Show also that the corresponding waves move upstream. Show that the growth rate remains positive as the wavenumber $k \to \infty$. [*This is an indication of ill-posedness.*]

3.6 Suppose that

$$s = s(u) = \tfrac{1}{2}F^2(1-u^2) + 1 - \frac{1}{u},$$

and that

$$s'(u)\frac{\partial u}{\partial t} = cD^*(u) - E^*(u) = -\frac{\partial c}{\partial x}.$$

Assume $D^* = 1$ and $E^* = u^3$. Simplify the equations to the form

$$\frac{\partial u}{\partial t} = f(u,c), \quad \frac{\partial c}{\partial x} = g(u,c),$$

giving expressions for f and g.

Suppose that $c = 1$ at $x = 0$ and $u = u_0(x)$ at $t = 0$, and that $u_0(\infty) = 1$. Derive an ordinary differential equation for $U(t) = u(0,t)$ in the form $\dfrac{dU}{dt} = h(U)$, and by consideration of the graphical form of $h(U)$ in the two cases $F < 1$ and $F > 1$, determine the behaviour of U for $t > 0$, explaining in particular how it depends on $U(0)$.

Why is it inadvisable to prescribe $c \to 1$ as $x \to \infty$ instead of the boundary condition at $x = 0$?

3.7 Large τ behaviour of (3.128), maybe outline of pseudo-spectral method, and a simpler example with algebraic instability rather than integral.

3.8 Suppose that

$$\psi(t) = \sum_{r=0}^{\infty} f_r t^r = (1-\xi_a t)^{-1/2}(1-\xi_b t)^{-1/2},$$

and it is desired to calculate the coefficients f_r numerically. By consideration of the power series for ψ^2 (or otherwise!) show that an iterative recipe for f_n is

$$f_0 = 1,$$

$$2f_n = \left(\frac{\xi_a^{n+1} - \xi_b^{n+1}}{\xi_a - \xi_b}\right) - \sum_{s=1}^{n-1} f_s f_{n-s}.$$

Chapter 4

Glaciers and ice sheets

Glaciers are huge and slow moving rivers of ice which exist in various parts of the world: Alaska, the Rockies, the Alps, Spitzbergen, China, for example. They drain areas in which snow accumulates, much as rivers drain catchment areas where rain falls. Glaciers also flow in the same basic way that rivers do. Although glacier ice is solid, it can deform by the slow creep of dislocations within the lattice of ice crystals which form the fabric of the ice. Thus, glacier ice effectively behaves like a viscous material, with, however, a very large viscosity: a typical value of ice viscosity is 5 bar year (in the metre-bar-year system of units!). Since 1 bar $= 10^5$ Pa, 1 year $\approx 3 \times 10^7$ s, this is a viscosity of some 10^{13} Pa s, about 10^{16} times that of water. As a consequence of their enormous viscosity, glaciers move slowly — a typical velocity would be in the range 10–100 m y^{-1} (metres per year), certainly measurable but hardly dramatic. More awesome are the dimensions of glaciers. Depths of hundreds of metres are typical, widths of kilometres, lengths of tens of kilometres. Thus glaciers can have an important effect on the human environment in their vicinity. They are also indirect monitors of climate; for example, many lithographs of Swiss glaciers show that they have been receding since the nineteenth century, a phenomenon thought to be due to the termination of the 'Little Ice Age' which lasted from about 1500 to about 1900.

Where glaciers are the glacial equivalent of rivers, i.e. channelled flow, *ice sheets* are the equivalent of droplets, but altogether on a grander scale. When an entire continent, or at least a substantial portion thereof, has a polar climate, then snow accumulates on the uplands, is compressed to form ice, and flows out to cover the continent, much as a drop of fluid on a table will spread under the action of gravity. However, whereas droplets can reach a steady state through the contractile effect of surface tension, this is not relevant to large ice sheets. In them, equilibrium can be maintained through a balance between accumulation in the centre and *ablation* at the margins. This can occur either through melting of the ice in the warmer climate at the ice margin, or through calving of icebergs. (Indeed, the same balance of accumulation at higher elevations with ablation at lower elevations is responsible for the normal quasi-steady profile of valley glaciers.)

There are two major ice sheets on the Earth, namely those in Antarctica and Greenland (the Arctic is an ocean, and its ice is sea ice, rarely more than three metres thick). They are on the order of thousands of kilometres in extent, and kilometres deep (up to four for Antarctica). They are thus, in fact, *shallow* flows, a fact which greatly facilitates the solution of mathematical models for their flow. Possibly more famous are the ice sheets which covered much of North America (the Laurentide ice sheet) and northern Europe (the Fennoscandian ice sheet) during the last ice age. Throughout the Pleistocene era (that is, the last two million years), there have been a succession of ice ages, each lasting a typical period of around 90,000 years, during which global ice sheet volume gradually increases, interspersed with shorter (10,000 year) *interglacials*, when the ice sheets rapidly retreat. The last ice age finished some ten thousand years ago, so that we are about due for another now. Perhaps the Little Ice Age was indeed the start of ice sheet build-up, only to be interrupted by the Industrial Revolution and the resultant global warming: nobody knows.

Further back in Earth's geologic history, there is evidence for dramatic, large scale glaciation in Carboniferous (c. 300 My (million years) ago) and Permian (c. 600 My ago) times. In the Permian glaciation, it seems that the whole landmass of the Earth was glaciated, leading to the concept of 'snowball Earth'. It was following the shrinkage of the global ice sheet that the explosion of life on Earth started.[1]

Drainage and sliding

While the motion of ice sheets and glaciers can be understood by means of viscous theory, there are some notable complications which can occur. Chief among these is that ice can reach the melting point at the glacier bed, due to frictional heating or geothermal heat input, in which case water is produced, and the ice can *slide*. Thus, unlike an ordinary viscous fluid, slip can occur at the base, and this is determined by a sliding law which relates basal shear stress τ to sliding velocity u_b and also, normally, the *effective pressure* $N = p_i - p_w$, where p_i and p_w are ice and water pressures. The determination of p_w further requires a description of the subglacial hydrology, and thus the dynamics of ice is intricately coupled to other physical processes: as we shall see, this complexity leads to some exotic phenomena.

4.1 Dynamic phenomena

4.1.1 Waves on glaciers

Just as on rivers, gravity waves will propagate on glaciers. Because the flow is very slow, they only propagate one way (downstream), and at speeds comparable to the surface speed (but slightly faster). These waves are known as surface waves, as they are evidenced by undulations of the surface: an example is shown in figure 4.1. They are examples of *kinematic* waves, driven by the dependence of ice flux on glacier depth.

A more exotic kind of wave is the 'seasonal wave'. This has no obvious counterpart in other fluid flows. It consists of (sizeable) perturbations in the surface velocity field which propagate down glacier at speeds in the order of 20-150 times the surface

---

[1] Snowball Earth was discussed in chapter 1.

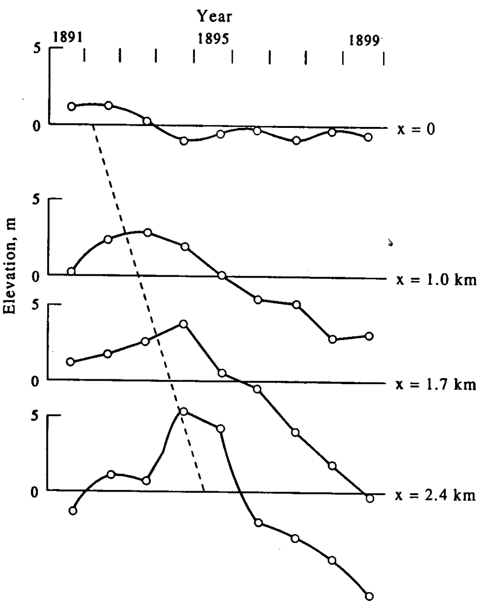

Figure 4.1: Changes of mean surface elevation of Mer de Glace, France, along four cross-profiles over a period of 9 years. The broken line corresponds to a wave velocity of 800 m/a. Reproduced from L. Lliboutry, IASH publ. **47**, 125-138.

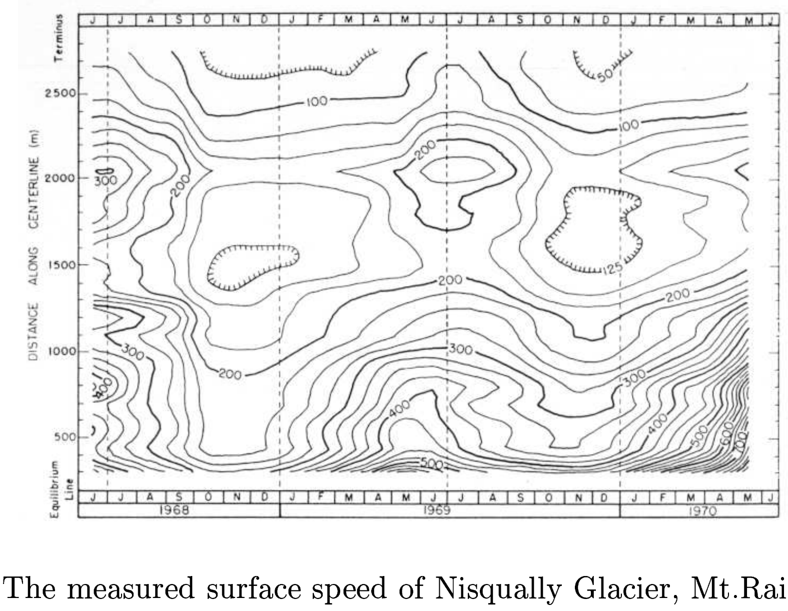

Figure 4.2: The measured surface speed of Nisqually Glacier, Mt.Rainier, as a function of time and distance. The contour interval is 25 mm d^{-1}. The maximum and minimum speeds occur progressively later with distance down-glacier; this represents a "seasonal wave" in the ice flow. Reproduced from Hodge (1974).

Figure 4.3: Variegated Glacier at the beginning of a surge, 29 August, 1964. Photograph by Austin Post, U.S. Geological Survey.

speed. There is no significant surface perturbation, and these waves must in fact be caused by variations of the basal sliding speed due to annual fluctuations in the basal water pressure. Although well-known and reported at the turn of the century, little attention has been paid to these waves in recent years. Figure 4.2 shows measurements of Hodge on Nisqually Glacier which indicates the rapid passage of a velocity wave downstream.

Mention should also be made of wave ogives, although we will not deal with them here. They are bands (also known as Forbes bands) which propagate below ice-falls, and are due to the annual ablation cycle.

4.1.2 Surges

Perhaps the most spectacular form of wave motion is the glacier *surge*. Surges are large scale relaxation oscillations of the whole length of a glacier. They are roughly periodic, with periods of the order of 20–100 years. During a long quiescent phase, the glacier is over-extended and thin. Ice accumulation causes the glacier to thicken upstream, while the over-extended snout thins and retreats. Eventually, a critical thickness is reached, and the glacier slumps rapidly downslope again. These surges will typically last only a year or two, during which time the velocity may increase a

Figure 4.4: Variegated Glacier at the end of a surge, 22 August, 1965. Photograph by Austin Post, U.S. Geological Survey.

hundred-fold. The glacier snout will then advance by several kilometres.

A typical (and much studied) example is the Variegated Glacier in Alaska. Its surge periodicity is about twenty years, while its surges last about two years. The glacier, of length twenty kilometres and depth four hundred metres, advances some six kilometres during its surge, at measured speeds of up to 65 metres per *day*. Such large velocities can only occur by basal sliding, and detailed observations during the 1982–3 surge showed that the surge was mediated by an alteration in the basal drainage system, which had the effect of raising water pressure dramatically. A dynamic model suggests, in fact, that the oscillations are caused by the competitive interaction between the basal sliding law and the hydraulics of the subglacial drainage system. When the ice is relatively thin (hence the driving shear stress is low) the drainage occurs through a network of channels incised into the ice at the glacier bed — called Röthlisberger channels. These allow effective drainage at quite low water pressures (hence high effective pressures) and thus also low ice velocities. At higher driving stresses, however, an instability forces the channel system to close down, and the basal water is forced into cavities which exist between the ice and bedrock protuberances (such cavities are well-known to exist). The water flow is reduced, and the sudden increase in water pressure causes a sudden increase in ice velocity — the surge. The transition front between the linked cavity drainage system and the channel system is nucleated near the maximum depth, and propagates rapidly both upstream and downstream, at (measured) speeds on the order of hundreds of metres per hour. At the end of the surge, the channel drainage system is re-established. Figures 4.3 and 4.4 show an aerial view of Variegated Glacier in pre- and post-surge states.

Our understanding of the Variegated surges relies on the concept of drainage switch between channelised flow and linked cavities, implicitly for ice flowing over (hard) bedrock. A rather different situation appears to operate in Trapridge Glacier, another well-studied surging glacier, in the Yukon. Here the glacier is cold in its interior (unlike the temperate (at the melting point) Variegated Glacier); and rests on a thick ($\sim$ 6 metres) layer of *till*, sometimes more graphically called *boulder clay* — a non-uniform mixture of angular rock fragments in a finer-grained, clay-rich groundmass (see figure 4.5). Till has a bimodal grain size distribution, and is produced by the erosion of brittle underlying bedrock, and is evacuated by the slow motion of the ice downstream.

The sequence of events which appears to be occurring as Trapridge thickens is that, firstly, the basal ice reaches melting point (and the till thaws). When this happens, the till becomes deformable, and the basal ice can slide over the bed by riding on the deforming till. The rate at which this occurs depends on the till rheology, where opinion is currently divided as to whether a viscous or plastic rheology is the more appropriate. [2] What does seem to be clear is that the water pressure will have a major effect. Increasing saturation causes increasing water pressure, which pushes the sediment grains apart and allows them to move more freely, so that in effect enhanced water production causes enhanced sliding. In turn, increased sliding causes increased frictional heating, so that there is a positive feedback which potentially can cause runaway and consequent surging behaviour. Whether the effect is strong enough is not obvious, but we shall examine a simple model which suggests that it may be.

4.1.3 Ice streams

Although ice sheets also flow under the horizontal pressure gradients induced by the glaciostatic pressures beneath their sloping surfaces, they rest on essentially unsloping bases, and therefore have no advective component in their dynamics. Thus ice sheets do not, at least on the large scale, exhibit wave motion: the governing equations are essentially diffusive in character. On a more local scale, however, ice sheets have interesting phenomena of their own.

Principal among these may be ice streams. Ice sheets do not tend to drain uniformly to the margin from their central accumulation zones, but rather the outflows from catchment areas are concentrated into fast-moving ice streams. Examples are the Lambert Glacier in Antarctica and Jakobshavn in Greenland, a fast-moving (8 metres per day) outlet glacier. These ice streams gain their speed by carving out deep channels through which they flow. Indeed, there is an obvious positive feedback

---

[2] Note the use of the word ‘appropriate’. As a saturated, granular material, somewhat like soil, there is little argument that a plastic rheology accommodating a yield surface is the most apposite description; such a description does not in itself provide an answer to such questions as to whether till deforms with depth (i. e., shears), or whether discrete slip occurs at the ice-till interface.

Figure 4.5: Subglacial till in a coastally exposed drumlin at Scordaun, Killough, Co. Down, Northern Ireland.

here. The deeper an ice flow, the larger the driving basal stress, and the warmer the basal ice (due to increased frictional heat and decreased conductive heat loss), and hence the softer the ice. Both of these effects contribute to enhanced ice flow, which explains the formation of such channels, since the erosive power of ice flow increases with the basal velocity and the basal shear stress. Indeed, flow of ice over a plane bed is subject to a lateral instability (much as overland flow of surface water is unstable to the formation of rills and gullies).

A similar kind of mechanism may operate when ice flows over deforming sediments, as in the Siple Coast of West Antarctica. Here, it is found that the flow is concentrated into five ice streams, A, B[3], C, D, and E, which are characterised by their heavily crevassed appearance. The flow in these ice streams is very rapid and is due to

---

[3] Ice stream B is now known as the Whillans ice stream, in memory of the glaciologist Ian Whillans.

basal sliding over the underlying sediment (except for ice stream C, which appears to have ‘switched off’ several hundred years ago). Measurements on the Whillans ice stream indicate that the basal water pressure is high (within 0.4 bar of the overburden pressure), and that it is underlain by some eight metres of saturated till. A similar instability to that concerning ice flow over hard bedrock can explain the streaming nature of the flow. Where ice flow is larger, there is increased water production. *If* the drainage system is such that increased water production leads to increased water pressure (as one might expect, e.g. for a Darcy flow), then the higher water pressure decreases the viscosity of the till, and hence enhances the ice flow further. This is an instability mechanism, and the limiting factor is that when ice flow increases, there is increased heat loss from the base, which acts to limit the increase of melt rate. Although this mechanism is viable, it has not yet been shown that it works.

4.1.4 Ice shelf instability

Where continental ice sheets are not diminished by ablation, the ice will flow to the continental margin, where it will spill into the ocean. This, for example, is the case in Antarctica, where it is so cold that ordinary mass wastage of the ice is virtually absent. As a result, *ice shelves* are formed, which are tongues of floating ice connected to the grounded ice at the *grounding line*. The grounding line is a dynamical free boundary, whose location determines the hold up of land ice, and its determination is therefore of some interest as regards sea level changes.

Over the past several decades, various arguments, none of them altogether convincing, have been put forward to suggest that ice shelves are inherently unstable and liable to collapse. This idea was originally put forward in consideration of the West Antarctic Ice Sheet (WAIS), much of the grounded part of which lies on a submarine bed. If the WAIS were to collapse completely, global sea level would rise by some six metres, inundating many coastal cities.

The basic physical mechanism for this putative catastrophic collapse is a positive feedback between grounding line retreat and ice flow rate. Since ice shelves are not resisted at their base, they can plausibly flow more rapidly, and the consequent drawdown effect will lower ice elevation, thus allowing further grounding line retreat. The debate has been fuelled by the remarkable collapse of the Larsen B Ice Shelf on the Antarctic Peninsula in 2002, which is thought to be due a climatic warming trend over recent decades. However, as we shall see below, it is by no means trivial to pose a theoretically coherent model for grounding line motion, and the issue of stability remians unresolved.

4.1.5 Tidewater glaciers

If the position of the grounding line indicates a balance between inland ice flow and ice shelf evacuation, the actual mechanism of break up involves mass wastage by calving icebergs. Indeed, in the absence of ablation, calving is the way in which marine ice sheets (i. e., those terminating in the sea) satisfy mass balance.

Glaciers which terminate in the sea are called *tidewater glaciers*, and are susceptible to a similar kind of catastrophic retreat to that which may be important for ice shelves. They also lose mass by calving, but are distinguished from ice shelves by the fact that the ice is grounded right to the margin. Instability is promoted by the fact that the calving rate increases with depth of water. If a tidewater glacier advances (in a cold climate), it will push a ridge of moraine ahead of it, snowplough style. In a stationary state, the water depth at the calving front will then be less than it is away from the margin, because of this moraine. Then, if the glacier snout retreats due, for example, to a warming trend, the snout will suddenly find itself in deeper water. The resultant increased calving can then lead to a catastrophic retreat. Just such a rapid retreat was observed in the Columbia Glacier in Alaska, which retreated some 12 km in twenty years from 1982, and it seems such rapid retreat is a common behavioural feature of tidewater glaciers in conditions of warmer climates.

4.1.6 Jökulhlaups

It will be clear by now that basal water is tremendously important in determining the nature of ice flow. Equally, the basal water system can fluctuate independently of the overlying ice dynamics, most notably in the outburst floods called jökulhlaups. In Iceland, in particular, these are associated with volcanoes under ice caps, where high rates of geothermal heat flux in the confines of a caldera cause a growing subglacial lake to occur, Eventually this overflows, causing a subglacial flood which propagates downglacier, and whose subsequent emergence at the glacier terminus releases enormous quantities of water over the southern coastal outwash plains. These floods carry enormous amounts of volcanic ash and sediments, which create vast beaches of black ash. Despite their violence, the ice flow is hardly disturbed. Jökulhlaups are essentially internal oscillations of the basal drainage system. They are initiated when the rising subglacial lake level causes leakage over a topographic rim, and the resultant water flow leads to an amplifying water flow by the following mechanism. Water flow through a channel in ice enlarges it by meltback of the walls due to frictional heating. The increased channel size allows increased flow, and thus further enlargement. The process is limited by the fact that the ice tends to close up the channel due to the excess overburden pressure over the channel water pressure, and this is accentuated when the channel is larger. In effect, the opening of the valve by the excess lake pressure is closed by the excess ice pressure. These floods occur more or less periodically, every five to ten years in the case of one of the best known, that of Grímsvötn under Vatnajökull in South-east Iceland.

4.2 The shallow ice approximation

4.2.1 Glaciers

We consider first the motion of a glacier in a (linear) valley. We take the x axis in the direction of the valley axis, z upwards and transverse to the mean valley slope, and y

across stream. The basic equations are those of mass and momentum conservation, which for an incompressible ice flow (neglecting inertial terms) are

$$\begin{aligned} \boldsymbol{\nabla}.\mathbf{u} &= 0, \\ 0 &= -\boldsymbol{\nabla} p + \boldsymbol{\nabla}.\boldsymbol{\tau} + \rho \mathbf{g}, \end{aligned} \tag{4.1}$$

where $\mathbf{g}$ is the gravity vector, p is the pressure, and $\boldsymbol{\tau}$ is the deviatoric part of the stress tensor. The stress and strain rate are related by

$$\tau_{ij} = 2\eta\, \dot{\varepsilon}_{ij}, \tag{4.2}$$

where η is the effective viscosity, and $\dot{\varepsilon}_{ij}$ is the strain rate

$$\dot{\varepsilon}_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right). \tag{4.3}$$

The most common choice of flow law is known as Glen's law, that is

$$\varepsilon_{ij} = A(T)\tau^{n-1}\tau_{ij}, \tag{4.4}$$

where the second stress invariant is given by $2\tau^2 = \tau_{ij}\tau_{ij}$ (using the summation convention)[4] and $A(T)$ is a temperature dependent rate factor which causes A to vary by about three orders of magnitude over a temperature range of 50 K: variation of A is thus significant for ice sheets (which may be subject to such a temperature range), but less so for glaciers. If we adopt the configuration shown in figure 4.6, then $\mathbf{g} = (g\sin\alpha, 0, -g\cos\alpha)$, where α is the mean valley slope downhill.

Boundary conditions for the flow are conditions of no stress at the top surface $z = s(x, y, t)$:

$$\begin{aligned} (-p + \tau_{11})s_x + \tau_{12}s_y - \tau_{13} &= 0, \\ \tau_{12}s_x + (-p + \tau_{22})s_y - \tau_{23} &= 0, \\ \tau_{13}s_x + \tau_{23}s_y - (-p + \tau_{33}) &= 0. \end{aligned} \tag{4.5}$$

At the base $z = b(x, y, t)$, we prescribe the velocity:

$$u = u_b, \quad v = v_b, \quad w = ub_x + vb_y; \tag{4.6}$$

(u_b, v_b) is the (horizontal) sliding velocity, whose form is discussed later. Finally, the kinematic condition on the free surface $z = s$ is

$$w = s_t + us_x + vs_y - a, \tag{4.7}$$

where a is the prescribed surface accumulation: positive for ice accumulation from snowfall, negative for ice ablation by melting.

---

[4] The summation convention is essentially a device for omitting summation signs; it asserts that summation is implied over repeated suffixes; thus $\tau_{ij}\tau_{ij}$ means $\sum_{i,j} \tau_{ij}\tau_{ij}$.

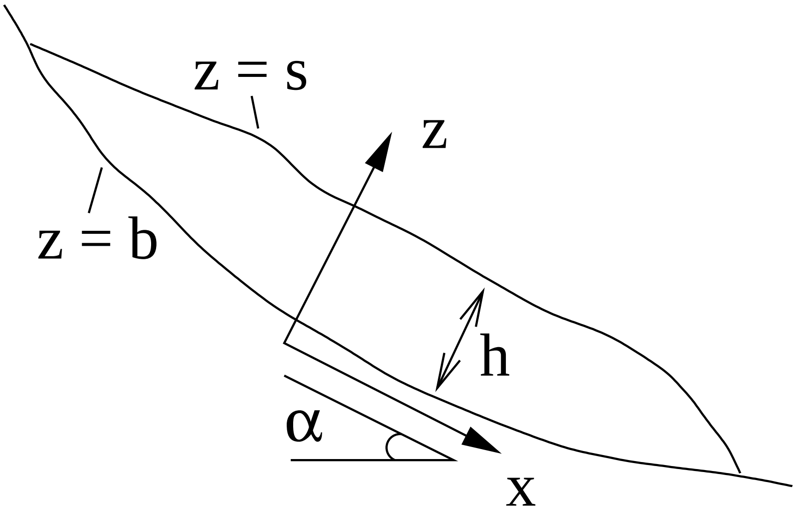

Figure 4.6: Typical profile of a valley glacier

A major simplification ensues by adopting what has been called the *shallow ice approximation.* It is the lubrication theory idea that the depth $d \ll$ the glacier length l, and is adopted as follows. We scale the variables by putting

$$\begin{aligned}
&u \sim U;\ v, w \sim \varepsilon U;\\
&x \sim l;\ y, z, b, s \sim d; \quad t \sim l/U;\\
&\tau_{13}, \tau_{12} \sim [\tau];\ A \sim [A];\ a \sim [a];\\
&p - p_a - (\rho g \cos\alpha)(s - z) \sim \varepsilon[\tau];\\
&\tau_{11}, \tau_{22}, \tau_{33}, \tau_{23} \sim \varepsilon[\tau],
\end{aligned} \tag{4.8}$$

where

$$\varepsilon = d/l \tag{4.9}$$

is the aspect ratio and we anticipate $\varepsilon \ll 1$. The choice of d and $[\tau]$ has to be determined self-consistently; we choose l from the given spatial variation of accumulation rate, and we choose U via $\varepsilon U = [a]$, which balances vertical velocity with accumulation rate. If we choose $[\tau] = \rho g d \sin\alpha$, and define

$$\mu = \varepsilon \cot\alpha, \tag{4.10}$$

then the scaled momentum equations are

$$\begin{aligned}
\frac{\partial \tau_{12}}{\partial y} + \frac{\partial \tau_{13}}{\partial z} &= -1 + \mu \frac{\partial s}{\partial x} + \varepsilon^2 \left(\frac{\partial p}{\partial x} - \frac{\partial \tau_{11}}{\partial x} \right),\\
\frac{\partial s}{\partial y} &= \frac{\varepsilon^2}{\mu} \left[-\frac{\partial p}{\partial y} + \frac{\partial \tau_{12}}{\partial x} + \frac{\partial \tau_{22}}{\partial y} + \frac{\partial \tau_{23}}{\partial z} \right],\\
\frac{\partial p}{\partial z} &= \frac{\partial \tau_{13}}{\partial x} + \frac{\partial \tau_{23}}{\partial y} + \frac{\partial \tau_{33}}{\partial z}.
\end{aligned} \tag{4.11}$$

The boundary conditions (4.6) and (4.7) are unchanged in form, and the stress conditions at the top surface $z = s(x, y, t)$ (4.5) become

$$\begin{aligned} \varepsilon^2(-p + \tau_{11})s_x + \tau_{12}s_y - \tau_{13} &= 0, \\ \tau_{12}s_x + (-p + \tau_{22})s_y - \tau_{23} &= 0, \\ \tau_{13}s_x + \tau_{23}s_y - (-p + \tau_{33}) &= 0. \end{aligned} \tag{4.12}$$

To get some idea of typical magnitudes, use values $d \sim 100$ m, $l \sim 10$ km, $\tan\alpha \sim 0.1$; then $\varepsilon \sim 10^{-2}$, $\mu \sim 10^{-1}$, so that to leading order $s = s(x, t)$ and

$$\frac{\partial \tau_{12}}{\partial y} + \frac{\partial \tau_{13}}{\partial z} = -1 + \mu \frac{\partial s}{\partial x}; \tag{4.13}$$

we retain the μ term for the moment.

The final relation to choose d (and hence also $[\tau]$) is determined by effecting a balance in the flow law. If the viscosity scale is $[\eta]$, then we choose

$$[\tau] = [\eta]U/d. \tag{4.14}$$

For example, for Glen's law, we can choose $[\eta] = 1/\{2[A][\tau]^{n-1}\}$, from which we find

$$d = \left[\frac{[a]l}{2[A](\rho g \sin\alpha)^n}\right]^{1/(n+2)}, \tag{4.15}$$

which leads, with sensible choices of $[A]$, l, $[a]$ and n, to values of d comparable to those observed ($d \sim 100$ m).

The two important shear stresses are given by

$$\tau_{13} = \eta\left(\frac{\partial u}{\partial z} + \varepsilon^2 \frac{\partial w}{\partial x}\right), \quad \tau_{12} = \eta\left(\frac{\partial u}{\partial y} + \varepsilon^2 \frac{\partial v}{\partial x}\right), \tag{4.16}$$

and the second stress invariant τ is given by

$$\tau^2 = \tau_{13}^2 + \tau_{12}^2 + \varepsilon^2\left[\tfrac{1}{2}\left(\tau_{11}^2 + \tau_{22}^2 + \tau_{33}^2\right) + \tau_{23}^2\right]; \tag{4.17}$$

for Glen's flow law, the dimensionless viscosity is

$$\eta = [A(T)\tau^{n-1}]^{-1}, \tag{4.18}$$

where $A(T)$ is the scaled (with $[A]$) temperature dependent rate factor. If we now put $\varepsilon = 0$ (the shallow ice approximation) we see that

$$\tau \approx \eta|\boldsymbol{\nabla} u|, \tag{4.19}$$

where $\boldsymbol{\nabla} = (\partial/\partial y, \partial/\partial z)$, and for Glen's law,

$$\eta = A^{-1/n}|\boldsymbol{\nabla} u|^{-(n-1)/n} \tag{4.20}$$

(note $n = 1$ for a Newtonian flow; Glen's law usually assumes $n = 3$); the determination of velocity distribution in a cross section S of a glacier then reduces to the elliptic equation for u in S (putting $\mu = 0$ and $\varepsilon = 0$ in $(4.11)_1$):

$$\boldsymbol{\nabla}.[\eta\{A, |\boldsymbol{\nabla} u|\}\boldsymbol{\nabla} u] = -1 \quad \text{in} \quad S, \tag{4.21}$$

with appropriate boundary conditions for no slip at the base being $u = 0$ on $z = b$, and the no stress condition at $z = s$ is, from $(4.12)_1$ and (4.16) with $s_y \approx 0$ and $\varepsilon = 0$, $\partial u/\partial z = 0$. The scalar s (independent of y) is determined through prescription of the downslope ice volume flux, $\displaystyle\int_S u\,dy\,dz = Q$, which will depend on x and t, but can be presumed to be known. In general, this problem requires numerical solution. Analytic solutions are available for constant A and a semi-circular cross section, but the free boundary choice of s cannot then be made.

Most studies of wave motion and other dynamic phenomena ignore lateral variation with y, and in this case (with $\tau_{13} = 0$ on $z = s$) (4.13) gives

$$\tau_{13} = (1 - \mu s_x)(s - z), \tag{4.22}$$

and Glen's law is, approximately,

$$\frac{\partial u}{\partial z} = A(T)\tau_{13}^n = A(T)[1 - \mu s_x]^n (s - z)^n. \tag{4.23}$$

If $A = 1$ is constant, then two integrations of (4.23) give the ice flux $Q = \displaystyle\int_b^s u\,dz$ as

$$Q = u_b H + (1 - \mu s_x)^n \frac{H^{n+2}}{n+2}, \tag{4.24}$$

where $H = s - b$ is the depth, and u_b is the sliding velocity. Integration of the mass conservation equation, together with the basal velocity condition (4.6) and the kinematic surface boundary condition (4.7), then leads to the integral conservation law

$$\frac{\partial H}{\partial t} + \frac{\partial Q}{\partial x} = a, \tag{4.25}$$

where a is the dimensionless accumulation rate. (4.25) is an equation of convective diffusion type, with the diffusive term being that proportional to μ.

Note that if transverse variations were to be included, we should solve

$$\frac{\partial S}{\partial t} + \frac{\partial Q}{\partial x} = a, \tag{4.26}$$

where S is the cross sectional area, and Q would be given by $Q = \displaystyle\int_S u\,dS$, where u solves (4.21) in the cross section S, together with appropriate boundary conditions.

4.2.2 Ice sheets

A model for ice sheets can be derived in much the same way — typical aspect ratios are 10^{-3} — but there is no 'downslope' gravity term $\rho g \sin\alpha$ (effectively $\alpha = 0$), and the appropriate balance determines the driving shear stress at the base in terms of the surface slope. Effectively, the advection term is lost and $\mu = 1$. Another difference is that $x \sim y \sim l$ (~ 3000 km) while $z \sim 3$ km is the only small position variable.

We will illustrate the scaling in two dimensions; the three dimensional version is relegated to the exercises (question 4.4). In two dimensions, we write the deviatoric stresses as

$$\tau_{11} = -\tau_{33} = \tau_1, \quad \tau_{13} = \tau_3. \tag{4.27}$$

Then the governing equations are

$$\begin{aligned} u_x + w_z &= 0, \\ 0 &= -p_x + \tau_{1x} + \tau_{3z}, \\ 0 &= -p_z + \tau_{3x} - \tau_{1z} - \rho g, \\ \dot{\varepsilon}_{ij} &= A\tau^{n-1}\tau_{ij}. \end{aligned} \tag{4.28}$$

The surface stress boundary conditions are, on $z = s$,

$$\begin{aligned} (-p + \tau_1)s_x - \tau_3 &= 0, \\ \tau_3 s_x - (-p - \tau_1) &= 0; \end{aligned} \tag{4.29}$$

at the base $z = b(x, y, t)$, we prescribe the velocity:

$$u = u_b, \quad w = ub_x; \tag{4.30}$$

and on $z = s$, the kinematic condition is

$$w = s_t + us_x - a. \tag{4.31}$$

We scale the variables by putting

$$\begin{aligned} &u \sim U;\ w \sim [a]; \\ &x \sim l;\ z, b, s \sim d; \quad t \sim l/U; \\ &\tau_3 \sim [\tau];\ A \sim [A];\ a \sim [a]; \\ &p - p_a - \rho_i g(s - z) \sim \varepsilon[\tau]; \\ &\tau_1 \sim \varepsilon[\tau], \end{aligned} \tag{4.32}$$

where

$$\varepsilon = \frac{d}{l}. \tag{4.33}$$

An appropriate balance of terms is effected by choosing $U = 2[A][\tau]^n d = [a]/\varepsilon$, $[\tau] = \rho g d\varepsilon$, and this leads to

$$d = \left(\frac{[a] l^{n+1}}{2[A](\rho g)^n} \right)^{\frac{1}{2(n+1)}}, \tag{4.34}$$

and thus

$$\varepsilon = \left(\frac{[a]}{2[A](\rho g)^n l^{n+1}}\right)^{\frac{1}{2(n+1)}}, \tag{4.35}$$

and typical values of the constants do lead to depths of the correct order of magnitude, kilometres, so that $\varepsilon \sim 10^{-3}$.

The corresponding dimensionless equations are

$$\begin{aligned} u_x + w_z &= 0, \\ 0 &= -s_x + \tau_{3z} + \varepsilon^2\left[-p_x + \tau_{1x}\right], \\ 0 &= -p_z + \tau_{3x} - \tau_{1z}, \\ \left(u_z + \varepsilon^2 w_x\right) &= A\tau^{n-1}\tau_3, \\ 2u_x &= A\tau^{n-1}\tau_1, \\ \tau^2 &= \tau_3^2 + \varepsilon^2\tau_1^2, \end{aligned} \tag{4.36}$$

and the boundary conditions are, on $z = s$:

$$\begin{aligned} \tau_3 + \varepsilon^2(p - \tau_1)s_x &= 0, \\ \tau_3 s_x + p + \tau_1 &= 0, \\ w &= s_t + us_x - a; \end{aligned} \tag{4.37}$$

at the base $z = b(x, y, t)$:

$$u = u_b, \quad w = ub_x. \tag{4.38}$$

The shallow ice approximation puts $\varepsilon = 0$, and then we successively find

$$\tau_3 = -s_x(s - z), \quad \tau = |s_x|(s - z), \tag{4.39}$$

whence

$$p + \tau_1 = \left[ss_x\right]_x(s - z) + \tfrac{1}{2}(z^2 - s^2)s_{xx}, \tag{4.40}$$

and, if we assume that $A = 1$ is constant, the ice flux is

$$\int_b^s u\,dz = u_b H + \frac{H^{n+2}}{n+2}|s_x|^{n-1}(-s_x), \tag{4.41}$$

so that conservation of mass leads to

$$\frac{\partial H}{\partial t} = \frac{\partial}{\partial x}\left[\frac{H^{n+2}}{n+2}|s_x|^{n-1}s_x - Hu_b\right] + a, \tag{4.42}$$

a nonlinear diffusion equation for the depth H, since $s = H + b$.

The three-dimensional version of this equation is (with $\boldsymbol{\nabla} = (\partial/\partial x,\ \partial/\partial y)$)

$$H_t = \boldsymbol{\nabla}.\left[\left\{\frac{|\boldsymbol{\nabla} s|^{n-1}H^{n+2}}{n+2}\boldsymbol{\nabla} s\right\} - H\mathbf{u}_b\right] + a. \tag{4.43}$$

The sliding velocity $\mathbf{u}_b$ is apparently a convective term, but in fact the sliding law usually has $\mathbf{u}_b$ in the direction of shear stress, whence $\mathbf{u}_b \propto -\boldsymbol{\nabla} s$, and this term also is diffusive.

Boundary conditions

Normally one would expect a boundary condition to be applied for (4.43) at the margin of the ice sheet, whose location itself may not be known. For an ice sheet that terminates on land, this condition would be $H = 0$ at the margin, but since the diffusion equation (4.43) is degenerate, in the sense that the diffusion coefficient vanishes where $H = 0$, no extra condition to specify the margin location is necessary, other than requiring that the ice flux also vanish where $H = 0$.

A different situation pertains for a marine ice sheet which terminates (and is grounded), let us suppose, at the edge of the continental shelf. Then the margin position is known, and the ice thickness and flux are finite. In this case, a calving boundary condition must be applied. The typical form of such conditions relates calving ice velocity to water depth, and this would provide a mixed boundary condition for H at the margin of the general form

$$-\frac{\partial H}{\partial n} = g(H). \tag{4.44}$$

A more representative condition for marine ice sheets occurs when the grounded ice extends into an ice shelf, so that the ice is floating. Extended ice shelves occur in the Antarctic ice sheet, two notable examples being the Ronne-Filchner ice shelf and the Ross ice shelf. The grounding line where the ice changes from grounded ice to floating ice is a free boundary whose location must be determined. The appropriate boundary condition for the ice sheet at the grounding line is bound up with the mechanics of the ice shelf, whose behaviour is altogether different; the mechanics of ice shelves is studied in section 4.2.5.

4.2.3 Temperature equation

Although the isothermal models are mathematically nice, they are not quantitatively very realistic. For a glacier, probably the neglect of variation of the rate parameter $A(T)$ in the flow law is as important as the assumption of a two-dimensional flow, although the possible coupling of temperature to water production and basal sliding is also significant. For ice sheets, temperature variation is unquestionably significant, and cannot in practice be neglected.

With variables scaled as in the previous section for a shallow shear flow, the temperature equation for an ice sheet may be written approximately as

$$\frac{dT}{dt} = \alpha\tau^2/\eta + \beta T_{zz}, \tag{4.45}$$

where $T - T_m$ is scaled with ΔT (a typical surface temperature below melting point). The derivative dT/dt is a material derivative. The stress invariant τ is related to the horizontal velocity $\mathbf{u} = (u, v)$ by

$$\tau \approx \eta \left|\frac{\partial \mathbf{u}}{\partial z}\right| = (s - z)|\boldsymbol{\nabla} s|, \tag{4.46}$$

since the horizontal stress vector $\boldsymbol{\tau} = (\tau_{13}, \tau_{23})$ satisfies

$$\boldsymbol{\tau} = \eta \frac{\partial \mathbf{u}}{\partial z} = -(s-z)\boldsymbol{\nabla} s. \tag{4.47}$$

(For an ice sheet, this relation is derived as for (4.22), but the downslope term 1 is absent, scales are chosen so that $\varepsilon = 1$, and (4.47) represents the two (horizontal) dimensional version.)

The parameters α and β are given by

$$\alpha = \frac{gd}{c_p \Delta T}, \; \beta = \frac{\kappa}{d[a]}, \tag{4.48}$$

where d is the depth scale, c_p is specific heat, g is gravity, κ is thermal diffusivity, $[a]$ is accumulation rate. Typical sorts of value for Antarctica are $\alpha \sim 0.3$, $\beta \sim 0.12$. We see that viscous heating (the α term) is liable to be significant, while thermal conduction is small or moderate. In addition, a scaled geothermal heat flux condition at the base (cf. $T < 0$ there) is $\partial T/\partial z \approx -\Gamma$, where $\Gamma \sim 1.5$ is a typical value. Temperature variation is likely to be significant, while the rate factor in the flow law can be modelled as $A \sim \exp(\gamma T)$, with $\gamma \sim 11$ for a temperature range of 50 K.

The temperature equation for a valley glacier is the same as (4.47), although with the previous scalings, (4.46) is corrected by simply replacing $|\boldsymbol{\nabla} s|$ in the last expression by $(1 - \varepsilon s_x)$. Although the scales are different, typical values of α and β are $\alpha \sim 0.25$, $\beta \sim 0.33$, and thus of significance. On the other hand, geothermal heat is of less importance.

4.2.4 Using the equations

Nonlinear diffusion

For flow over a flat base, $h = 0$, with no sliding, the isothermal ice sheet equation (4.43) is just

$$H_t = \boldsymbol{\nabla}. \left[\frac{H^{n+2}|\boldsymbol{\nabla} H|^{n-1}}{n+2} \boldsymbol{\nabla} H \right] + a, \tag{4.49}$$

which for Glen's flow law would have $n = 3$. This is a degenerate nonlinear diffusion equation, and has singularities at ice margins ($H = 0$) or divides (where $\boldsymbol{\nabla} H/|\boldsymbol{\nabla} H|$ is discontinuous). In one space dimension, we have near a margin $x = x_m(t)$ where $a < 0$ (ablation),

$$\begin{aligned} H &\sim (a/|\dot{x}_m|)(x_m - x) \text{ if } \dot{x}_m < 0 \text{ (retreat)}, \\ H &\sim \left(\frac{2n+1}{n}\right)^{\frac{n}{2n+1}} [(n+1)\dot{x}_m]^{\frac{1}{2n+1}} (x_m - x)^{\frac{n}{2n+1}} \text{ if } \dot{x}_m > 0 \text{ (advance)}. \end{aligned} \tag{4.50}$$

This is the common pattern for such equations: margin retreat occurs with finite slope, while for an advance, the slope must be infinite. Consequently, there is a waiting time between a retreat and a subsequent advance, while the front slope grows.

Near a divide $x = x_d$, where $H_x = 0$ and $a > 0$, H is given by

$$H \sim H_0(t) - \left(\frac{n}{n+1}\right)\left[\frac{(n+2)(a-\dot{H}_0)}{H_0^{n+2}}\right]^{1/n} |x - x_d|^{(n+1)/n}, \tag{4.51}$$

and thus the curvature is infinite. Singularities of these types need to be taken into account in devising numerical methods.

Thermal runaway

One of the interesting possibilities of the thermomechanical coupling between flow and temperature fields is the possibility of thermal runaway, and it has even been suggested that this may provide an explanation for the surges of certain thermally regulated glaciers. The simplest model is that for a glacier, with exponential rate factor, thus

$$\frac{dT}{dt} = \alpha\tau^{n+1}e^{\gamma T} + \beta T_{zz}, \tag{4.52}$$

where the stress is given by

$$\tau = s - z. \tag{4.53}$$

The simplest configuration is the parallel sided slab in which s = constant, $\mathbf{u} = (u(z), 0, 0)$, so that

$$\frac{\partial T}{\partial t} = \alpha(s-z)^{n+1}e^{\gamma T} + \beta\frac{\partial^2 T}{\partial z^2}, \tag{4.54}$$

with (say)

$$T = -1 \text{ on } z = s,\ T_z = -\Gamma \text{ on } z = 0. \tag{4.55}$$

For given s, (4.54) will exhibit thermal runaway for large enough α, and $T \to \infty$ in finite time. As the story goes, this leads to massive melting and enhanced sliding, thus 'explaining' surges. The matter is rather more complicated than this, however. For one thing, s would actually be determined by the criterion that the flux $\int_0^s u\,dz$ is prescribed, $= s$ say, where s would be the integrated ice accumulation rate from upstream ($= \int a\,dx$).

Thus even if we accept the unrealistic parallel slab 'approximation', it would be appropriate to supplement (4.54) and (4.55) by requiring s to satisfy

$$\int_0^s u\,dz = s. \tag{4.56}$$

Since the flow law gives

$$\frac{\partial u}{\partial z} = (s-z)^n e^{\gamma T}, \tag{4.57}$$

we find, if $u = 0$ on $z = 0$, that (4.56) reduces to

$$\int_0^s (s-z)^{n+1}e^{\gamma T}dz = s. \tag{4.58}$$

Thermal runaway is associated with multiple steady states of (4.54), in which case we wish to solve

$$\begin{aligned} 0 &= \alpha(s-z)^{n+1}e^{\gamma T}+\beta T_{zz}, \\ T &= -1 \text{ on } z=s, \\ T_z &= -\Gamma \text{ on } z=0, \\ T_z &= -[\Gamma+(\alpha s/\beta)] \text{ on } z=s. \end{aligned} \tag{4.59}$$

Putting $\xi = s - z$, we solve

$$\begin{aligned} T_{\xi\xi} &= -(\alpha/\beta)\xi^{n+1}e^{\gamma T}, \\ T &= -1,\ T_\xi = \Gamma+(\alpha s/\beta) \text{ on } \xi=0, \end{aligned} \tag{4.60}$$

as an initial value problem. T_ξ is monotone decreasing with increasing ξ, and thus there is a unique value of s such that $T_\xi = \Gamma$ there. It follows that there is a unique solution to the free boundary problem, and in fact it is linearly stable. It then seems that thermal runaway is unlikely to occur in practice.

A slightly different perspective may allow runaway, if we admit non-steady ice fluxes. Formally, we can derive a suitable model if $\lambda = \alpha/\beta = O(1)$, $\beta \to \infty$. In this case, we can expect T to tend rapidly to equilibrium of (4.54), and then s reacts more slowly via mass conservation, thus

$$\begin{aligned} &s_t + q_x = a, \\ &q = \frac{1}{\lambda}[T_z]_0^s. \end{aligned} \tag{4.61}$$

An x-independent version of (4.61), consistent with the previous discussion, is

$$\frac{\partial s}{\partial t} = s - q(s), \tag{4.62}$$

and this *will* allow relaxation oscillations if $q(s)$ is multivalued as a function of s — which will be the case. Surging in this sense is conceivable, but the limit $\beta \to \infty$ is clearly unrealistic, and unlikely to be attained. The earlier conclusion is the more likely.

4.2.5 Ice shelves

When an ice sheet flows to the sea, as mostly occurs in Antarctica, it starts to float at the *grounding line*, and continues to flow outwards as an *ice shelf*. The dynamics of ice shelves can be described by an approximate theory, but this is very different from appropriate to ice sheets.

We begin with the equations in the form (4.36) and (4.37), as scaled for the ice sheet. These must be supplemented by conditions on the floating base $z = b$. To be specific, we take the level $z = 0$ to be sea level. The water depth at $z = b$ is thus $-b$, and the resulting hydrostatic pressure must balance the normal stress in the ice.

In addition, there is no shear stress. The general form of the (vector) stress balance condition at an interface of this type which supports only a pressure p_i is

$$\boldsymbol{\sigma}.\,\mathbf{n} = -p_i\mathbf{n}, \tag{4.63}$$

and in addition to this there is a kinematic boundary condition. When written in terms of the ice sheet scales, these boundary conditions become

$$\begin{aligned} -\tau_3 + \varepsilon^2(-p + \tau_1)b_x &= (s + \delta b)b_x, \\ s &= -\delta b - \varepsilon^2\left[\tau_3 b_x + p + \tau_1\right], \\ w &= b_t + ub_x - m, \end{aligned} \tag{4.64}$$

where m is bottom melting rate, and the parameter δ is given by

$$\delta = \frac{\rho_w - \rho_i}{\rho_i}, \tag{4.65}$$

and ρ_i and ρ_w are ice and water densities. The second of these conditions, the flotation condition, essentially says that 90% of the ice is below the surface, as in Archimedes' principle.

Whereas the dominant force balance in the ice sheet is between shear stress and horizontal pressure gradient, and longitudinal stresses are negligible, this is not true in the ice shelf, where the opposite is true: shear stress is small, and the primary balance is between longitudinal stress and horizontal pressure gradient. Therefore the equations must be rescaled in order to highlight this fact. The issue is complicated by the presence of two small parameters $\delta \sim 0.1$ and $\varepsilon \sim 10^{-3}$.

We suppose that the length scale for the ice shelf is $x \sim \lambda$ (relative to the horizontal ice sheet scale), and that the depth scale is $z \sim \nu$, and we anticipate that $\nu \ll 1$. We then find that a suitable balance of terms reflecting the dominance of longitudinal stresses is given by writing

$$\begin{gathered} x \sim \lambda, \quad z, b \sim \nu, \quad u \sim \frac{1}{\nu}, \quad w \sim \lambda, \\ p, \tau_1 \sim \frac{\delta\nu}{\varepsilon^2}, \quad \tau_3 \sim \frac{\delta\nu^2}{\lambda}, \quad \tau \sim \frac{\delta\nu}{\varepsilon}, \quad s \sim \delta\nu. \end{gathered} \tag{4.66}$$

The governing equations become

$$\begin{aligned} u_x + w_z &= 0, \\ 0 &= -s_x + \tau_{3z} - p_x + \tau_{1x}, \\ 0 &= -p_z - \tau_{1z} + \omega^2\tau_{3x}, \\ \left(u_z + \omega^2 w_x\right) &= \omega^2\tau^{n-1}\tau_3, \\ 2u_x &= \tau^{n-1}\tau_1, \\ \tau^2 &= \omega^2\tau_3^2 + \tau_1^2, \end{aligned} \tag{4.67}$$

and the appropriate boundary conditions are, on the top surface $z = \delta s$:

$$\begin{aligned} \tau_3 + \delta(p - \tau_1)s_x &= 0, \\ p + \tau_1 + \delta\omega^2\tau_3 s_x &= 0, \\ w &= \lambda\delta\nu s_t + \delta u s_x - \lambda a; \end{aligned} \tag{4.68}$$

and on the base $z = b$:

$$\begin{aligned} \tau_3 + (p - \tau_1)b_x &= (s + b)b_x, \\ s + b &= -\left(p + \tau_1 + \omega^2\tau_3 b_x\right), \\ w &= \nu\lambda b_t + u b_x - \lambda m; \end{aligned} \tag{4.69}$$

in these equations,

$$\omega = \frac{\nu\varepsilon}{\lambda} \ll 1. \tag{4.70}$$

The length scale is essentially arbitrary; observations suggest $\lambda \lesssim 1$. The parameter ν is defined by the constraint that longitudinal stress balances longitudinal strain rate, and this determines

$$\nu = \frac{\varepsilon}{\delta}\left(\frac{\delta}{\lambda A}\right)^{1/(n+1)}, \tag{4.71}$$

where, if A varies with temperature, it is the surface (lowest) value that should be used.

We let $\omega \to 0$ in these equations; it follows that $u \approx u(x,t)$, $\tau \approx |\tau_1|$, whence $\tau_1 \approx \tau_1(x,t)$; $p + \tau_1 \approx 0$, so that $\tau_{3z} \approx s_x - 2\tau_{1x}$, and thus

$$\tau_3 \approx (s_x - 2\tau_{1x})(z - \delta s) + 2\delta\tau_1 s_x. \tag{4.72}$$

Applying the boundary conditions at $z = b$, we have

$$s = -b, \quad 2\tau_1 b_x = (s_x - 2\tau_{1x})(b - \delta s) + 2\delta\tau_1 s_x, \tag{4.73}$$

whence, integrating, we find

$$\tau_1 = -\tfrac{1}{4}b, \tag{4.74}$$

and the integration constant (for $(4.73)_2$) is taken to be zero on applying an averaged force balance at the ice shelf front (see question 4.6). Thus we finally obtain the stretching equations, noting that the ice thickness $H \approx -b$ to $O(\delta)$,

$$u_x = \tfrac{1}{2}\left(\tfrac{1}{4}H\right)^n, \quad \nu\lambda H_t + (uH)_x = \lambda(a - m); \tag{4.75}$$

the second equation is that of mass conservation, and is derived by integrating the mass continuity equation. Note that the time scale for mass adjustment is $O(\nu\lambda) \ll 1$, so that the ice shelf responds rapidly to changes in supply.

Suitable initial conditions for H and u would follow from continuity of ice flux and depth across the grounding line, but the position of the grounding line $x = x_G$

is not apparently determined. Let us anticipate that suitable conditions on H and u are that

$$u \to 0, \quad Hu \to q_I \quad \text{as} \quad x \to x_G; \tag{4.76}$$

assuming steady conditions, it follows from (4.75) that

$$Hu = q_I + \int_{x_G}^{x} (a - m)\, dx. \tag{4.77}$$

The solution for u follows by quadrature. In the particular case that $a = m$ (and in any case as $x \to x_G$), we have $Hu = q_I$, and thus

$$u = \left\{ \frac{(n+1)}{2} \left(\frac{q_I}{4} \right)^n \right\}^{1/(n+1)} (x - x_G)^{1/(n+1)}. \tag{4.78}$$

In order to find a condition for q_I and for the position of x_G, we need to consider the region near the grounding line in more detail, and this is done in the following subsection.

4.2.6 The grounding line

In the transition region, we need to retain terms which are of importance in both ice sheet and ice shelf approximations. This requires us to rescale the ice sheet scaled variables in the following way:

$$x - x_G \sim \gamma, \quad z, b \sim \beta, \quad s \sim \delta\beta, \quad u \sim \frac{1}{\beta}, \quad w \sim \frac{1}{\gamma},$$

$$t \sim \beta, \quad p, \tau_1, \sim \frac{\delta\beta}{\varepsilon^2}, \quad \tau_3 \sim \frac{\delta\beta^2}{\gamma}, \quad \tau \sim \frac{\delta\beta}{\varepsilon}, \tag{4.79}$$

where x_G is the grounding line position; the parameters β and γ are defined by

$$\beta = \left(\frac{\varepsilon}{\delta} \right)^{\frac{n}{n+2}} \frac{1}{A^{\frac{1}{n+2}}}, \quad \gamma = \beta\varepsilon. \tag{4.80}$$

This rescaling reintroduces the full Stokes equations. Denoting the rescaled variables (except time) by capitals, and writing

$$x - x_G(t) = \gamma X, \quad t = \beta t^*, \tag{4.81}$$

we derive the model

$$\begin{aligned} U_X + W_Z &= 0, \\ 0 &= -S_X + T_{3Z} - P_X + T_{1X}, \\ 0 &= -P_Z - T_{1Z} + T_{3X}, \\ U_Z + W_X &= T^{n-1} T_3, \\ 2U_X &= T^{n-1} T_1, \\ T^2 &= T_3^2 + T_1^2. \end{aligned} \tag{4.82}$$

The boundary conditions are the following. On the surface $Z = \delta S$,

$$\begin{aligned} T_3 + \delta(P - T_1)S_X &= 0, \\ P + T_1 + \delta T_3 S_X &= 0, \\ W &= \delta(\gamma S_{t^*} - \dot{x}_G S_X) + \delta U S_X - \gamma a, \end{aligned} \tag{4.83}$$

where $\dot{x}_G = \dfrac{dx_G}{dt^*}$. On the base $Z = B$, when $X > 0$,

$$\begin{aligned} -T_3 + (-P + T_1)B_X &= (S + B)B_X, \\ S + B &= -\left[P + T_1 + T_3 B_X\right], \\ W &= \gamma B_{t^*} - \dot{x}_G B_X + U B_X - \gamma m, \end{aligned} \tag{4.84}$$

and when $X < 0$,

$$W = 0, \quad U = 0, \tag{4.85}$$

where we assume that the sliding velocity is zero for grounded ice.

To leading order, we can approximate the top surface boundary conditions as $\gamma \to 0$ and also $\delta \to 0$ as

$$T_3 = P + T_1 = W = 0 \quad \text{on} \quad Z = 0. \tag{4.86}$$

The kinematic condition at the shelf base is approximately

$$W = -\dot{x}_G B_X + U B_X \quad \text{on} \quad Z = B. \tag{4.87}$$

In addition, the solution must be matched to the outer (sheet and shelf) solutions. We consider first the ice sheet behaviour as $x \to x_G$. We suppose that the ice sheet is described in one dimension by (4.43), thus

$$h_t = -q_x + a, \tag{4.88}$$

where the ice flux is (in ice sheet scaled variables)

$$q = \frac{(s - b)^{n+2}(-s_x)^n}{n + 2}. \tag{4.89}$$

We can carry out a local analysis near x_G similar to those in section 4.2.4. As $x \to x_G$, $s - b \to 0$ (since $s - b \sim \beta \ll 1$), but the ice flux is non-zero; in this case we find that always

$$H = s - b \sim C(x_G - x)^{\frac{n}{2(n+1)}}, \quad q \sim q_G = \left(\frac{n}{2(n+1)}\right)^n \frac{C^{2n+2}}{n + 2}. \tag{4.90}$$

When the surface slope is computed from this, we find that the requisite matching condition for the slope written in terms of the transition zone scalings is that

$$S_X \sim -\frac{nC}{2(n+1)} \frac{1}{(-\delta X)^{\frac{n+2}{2(n+1)}}} \quad \text{as} \quad X \to -\infty. \tag{4.91}$$

Clearly the presence of the small parameter δ does not allow direct matching of the transition zone to the ice sheet.

The problem is easily resolved, however. There is a 'joining' region in which $X = \tilde{X}/\delta$, $S = \tilde{S}/\delta$, and then also $P, T_1, W \sim \delta$; the resultant set of equations is easily solved (it is a shear layer like the ice sheet), and we find

$$\tilde{S} = B_G + \left[(-B_G)^{2(n+1)/n} - \frac{2(n+1)}{n}\{(n+2)q_G\}^{1/n}\tilde{X}\right]^{n/(2n+2)}, \tag{4.92}$$

where $B = B_G$ at $x = x_G$. Expanding this as $\tilde{X} \to 0$, we find that the matching condition for S in the transition zone as $X \to -\infty$ is

$$S \sim -\Lambda X, \tag{4.93}$$

where

$$\Lambda = \frac{\{(n+2)q_G\}^{1/n}}{(-B_G)^{(n+2)/n}}. \tag{4.94}$$

A final simplification to the transition zone problem results from defining

$$\Pi = P + S; \tag{4.95}$$

to leading order in γ and δ, the transition problem is then

$$\begin{aligned} U_X + W_Z &= 0, \\ \Pi_X &= T_{3Z} + T_{1X}, \\ \Pi_Z &= -T_{1Z} + T_{3X}, \\ U_Z + W_X &= T^{n-1}T_3, \\ 2U_X &= T^{n-1}T_1, \\ T^2 &= T_3^2 + T_1^2, \end{aligned} \tag{4.96}$$

together with the boundary conditions

$$T_3 = W = 0 \quad \text{on} \quad Z = 0, \tag{4.97}$$

$$\begin{aligned} B &= -(\Pi + T_1 + T_3 B_X), \\ T_3(1 - B_X^2) &= 2T_1 B_X, \\ W &= (-\dot{x}_G + U)B_X \quad \text{on} \quad Z = B, \quad X > 0, \end{aligned} \tag{4.98}$$

and

$$W = U = 0 \quad \text{on} \quad Z = B_G, \quad X < 0. \tag{4.99}$$

The matching conditions to the ice sheet may be summarised as

$$\Pi_X \to -\Lambda, \quad W \to 0, \quad T_3 \sim -\Lambda Z \quad \text{as} \quad X \to -\infty, \tag{4.100}$$

with the flow becoming the resultant pressure gradient driven shear flow at $-\infty$. Towards the ice shelf, a comparison of orders of magnitude shows firstly that

$$\frac{\nu}{\beta}=\left(\frac{\gamma}{\lambda}\right)^{1/(n+1)}\ll 1, \tag{4.101}$$

and that in the ice shelf, the transition scaled variables are

$$S,B\sim\frac{\nu}{\beta},\quad W\sim\frac{\gamma}{\lambda},\quad P,T_1\sim\frac{\nu}{\beta},\quad T_3\sim\left(\frac{\nu}{\beta}\right)^2\frac{\gamma}{\lambda}; \tag{4.102}$$

note also that the ice shelf time scale $\nu\lambda$ is much less than the transition zone time scale β, so that it is appropriate in the transition zone to assume that the far field ice shelf is at equilibrium, and thus described by (4.77) and (4.78). Bearing in mind (4.101), it follows from this that suitable matching conditions for the transition region are

$$\begin{aligned} T_1\sim -\tfrac{1}{4}B,\qquad & U\sim MX^{1/(n+1)},\\ W\to 0,\qquad & B\sim -\frac{q_I}{U}\quad\text{as}\quad X\to\infty, \end{aligned} \tag{4.103}$$

where

$$M=\left\{\frac{(n+1)}{2}\left(\frac{q_I}{4}\right)^n\right\}^{1/(n+1)}, \tag{4.104}$$

and the flow becomes an extensional flow as $X\to\infty$. It follows from integration of the continuity equation between B and S that the ice flux to the ice shelf, q_I, is given by

$$q_I=q_G+\dot{x}_G B_G. \tag{4.105}$$

The top surface is defined by

$$S=\left.(\Pi+T_1)\right|_{Z=0}, \tag{4.106}$$

and uncouples from the rest of the problem. The extra condition on $Z=B,\ X>0$, should determine B providing $\dot{x}_G$ is known. This is the basic conundrum of the grounding line determination, since there appears to be no extra condition to determine $\dot{x}_G$.

The determination of the location of the grounding line is a matter of practical concern, because it is a necessary ingredient of any theory of marine ice sheet instability, as discussed previously.

4.3 Sliding and drainage

The sliding law relates the basal shear stress τ_b to the basal sliding velocity u_b. The classical theory, enunciated by Lliboutry, Weertman, Nye, Kamb, and others, considers ice flowing at the base of a glacier over an irregular, bumpy bedrock. The ice is lubricated at the actual interface by the mechanism of *regelation*, or melting-refreezing, which allows a thin film (microns thick) to exist at the ice-rock interface, and allows the ice to slip. The drag is then due to two processes; regelation itself, and the viscous flow of the ice over the bedrock. Regelation is dominant for small wavelength roughness, while viscous drag is dominant for large wavelengths, and early work emphasised the importance of a controlling (intermediate) wavelength (of several centimetres). More recently, the emphasis has been away from regelation and has been put on consideration of the viscous flow.

A suitable model for discussion is that of a Newtonian fluid over a rough bedrock of 'wavelength' $[x]$ and amplitude $[y]$, given, in coordinates scaled with $[x]$, by $y = \nu h(x)$, where y is now the vertical coordinate, and

$$\nu = [y]/[x] \tag{4.107}$$

is a measure of corrugation. The governing equation for slow, two-dimensional flow is the biharmonic equation

$$\nabla^4 \psi = 0 \tag{4.108}$$

for a suitably scaled stream function. Appropriate boundary conditions for no flow through the bed, and no shear stress there, are

$$\begin{gathered}\psi = 0,\\ (1-\nu^2 h'^2)(\psi_{yy}-\psi_{xx}) - 4\nu h' \psi_{xy} = 0,\end{gathered} \tag{4.109}$$

on $y = \nu h$. As $y \to \infty$, the local basal flow must match to a far field flow with 'basal' velocity u_b and 'basal' stress τ_b; thus the main body of the ice flow sees the bedrock flow as a boundary layer, and u_b and τ_b are then the appropriate basal limits of the 'outer' ice flow. Specifically, we find that the correct matching condition is (in terms of correctly scaled 'outer' velocities and stresses)

$$\psi \sim u_b y + \tfrac{1}{2}\nu^2 \tau_b y^2 \text{ as } y \to \infty. \tag{4.110}$$

A convenient solution method can be presented if ν is small. In this case, we subtract $u_b y$ from ψ and divide by ν (so $\psi \to (\psi - u_b y)/\nu$); then to leading order in ν, the new ψ satisfies (4.108), with

$$\begin{gathered}\psi \to 0 \text{ as } y \to \infty,\\ \psi = -u_b h(x),\ \psi_{yy} - \psi_{xx} = 0 \text{ on } y = 0.\end{gathered} \tag{4.111}$$

The shear stress is uncoupled from the determination of ψ, but can be determined by an integrated force balance, whence (e.g. if h is periodic with period 2π)

$$\tau_b = \frac{1}{2\pi}\int_0^{2\pi} (p + 2\psi_{xy})|_{y=0} h' dx; \tag{4.112}$$

more generally a spatial average would be used. Notice that the expression in brackets in (4.112) is simply (minus) the normal stress, and therefore is equal to the water pressure p_w in the lubricating film. We come back to this below.

A nice way to solve this problem is via complex variable theory. We define the complex variable $z = x+iy$, and then the general solution of the biharmonic equation is

$$\psi = (\bar{z} - z)f(z) - B(z) + (cc), \tag{4.113}$$

where f and B are analytic functions and (cc) denotes the complex conjugate. The zero stress condition (4.111) requires $f = -\frac{1}{2}B'$, and also $B \to 0$ as $z \to \infty$ (with Im $z > 0$), and the last condition is then

$$B + \bar{B} = u_b h \text{ on } \operatorname{Im} z = 0. \tag{4.114}$$

If h is periodic, with a Fourier series

$$h = \sum_{-\infty}^{\infty} a_k e^{ikx}, \tag{4.115}$$

then B is simply given by

$$B = u_b \sum_{1}^{\infty} a_k e^{ikz} \tag{4.116}$$

(we can assume $a_0 = 0$, i.e. the mean of h is zero). However, it is also convenient to formulate this problem as a Hilbert problem. We define $L(z) = B''(z)$, which is analytic in Im $z > 0$, and then $L(z) = \overline{B''(\bar{z})}$ is analytic in Im $z < 0$. It then turns out that, with the usual notation for the values on either side of the real axis,

$$\begin{aligned} L_+ + L_- &= u_b h'', \\ L_+ - L_- &= \tfrac{1}{2} ip, \end{aligned} \tag{4.117}$$

relate the values either side of Im $z = 0$; here p is ice pressure ($p = -2i(B'' - \bar{B}'')$ on $y = 0$, since $p + i\nabla^2\psi$ is analytic), and in fact $p = p_w$ on $y = 0$, since ψ_{xy} is found to be zero there. The drag (i.e. the sliding law) is then computed as (for a 2π-periodic h)

$$\tau_b = \frac{1}{i\pi} \int_0^{2\pi} (L_+ - L_-) h' dx, \tag{4.118}$$

and turns out to be

$$\tau_b = 4u_b \sum_{1}^{\infty} k^3 |a_k|^2. \tag{4.119}$$

For a linear model such as this, τ_b is necessarily proportional to u_b. For Glen's flow law, variational principles can be used to estimate

$$\tau_b \approx R u_b^{1/n}. \tag{4.120}$$

Weertman's original sliding law drew a balance between (4.120) and the linear dependence due to regelation, and the heuristic 'Weertman's law' $\tau_b \propto u_b^{1/m}$, with $m \approx (n+1)/2$ was often used.

Simplistic sliding laws such as the above have been superceded by the inclusion of cavitation. When the film pressure behind a bump decreases to a value lower than the local subglacial water pressure, a cavity must form, and indeed, such cavities are plentifully observed. An appropriate generalisation of (4.117) is then

$$\begin{aligned} L_+ + L_- &= u_b h'' \quad \text{in } C', \\ L_+ - L_- &= -\tfrac{1}{2} i p_c \quad \text{in } C, \end{aligned} \tag{4.121}$$

where the bed is divided into cavities (C) where p is known ($= -p_c$), and attached regions where h is known. One can solve this problem to find the unknown cavity shapes, and for a bed consisting of isolated bumps, $\tau_b(u_b)$ increases monotonically for small u_b, reaches a maximum, and then decreases for large u_b, as shown in figure 4.7. The decreasing portion of the curve is unstable (increasing velocity decreases drag) and is caused by the roofs of the cavities from one bump reaching the next bump.

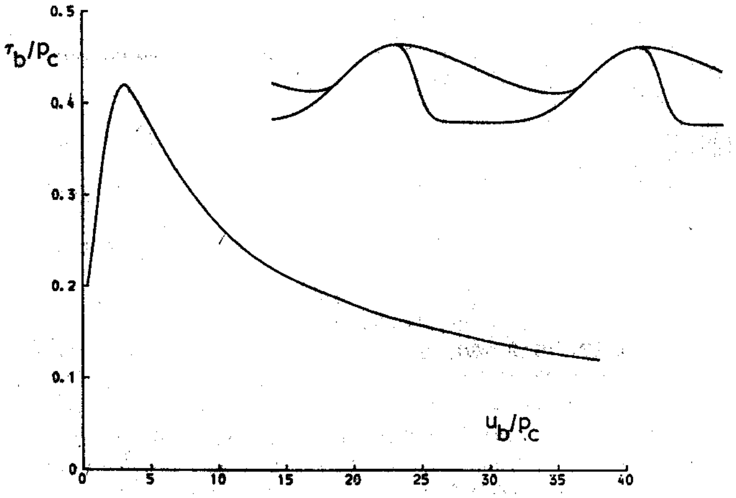

Figure 4.7: Stress versus velocity for a bed of isolated bumps

Since p in the scaled ice flow model is measured relative to ice overburden pressure, it follows that p_c in (4.121) is proportional to the effective pressure N, and in fact the sliding law has the specific form $\tau_b = N f(u_b/N)$. For a nonlinear Glen's law, the suggested generalisation is

$$\tau_b = N f(u_b/N^n). \tag{4.122}$$

The multivaluedness of $u_b(\tau_b)$ is very suggestive of surging — but is it realistic? Consideration of more realistic beds suggests that in fact $f(\cdot)$ in (4.122) will be an increasing function of its argument, since when smaller bumps start to be drowned, larger ones will take up the slack. A plausible sliding law then has $f(\xi)$ increasing as a power of ξ, or (for example)

$$\tau_b = c u_b^r N^s, \tag{4.123}$$

where we would expect $r, s > 0$. More specifically, (4.122) would suggest $s = 1 - rn$, and also that $r \approx 1/n$ would be appropriate at low u_b, where cavitation is absent. When cavitation occurs, one would then expect lower stresses, so that $r < 1/n$. There is in fact some experimental and field evidence consistent with laws of this type, with $r \approx s \approx 1/3$, for example.

An apparently altogether different situation occurs when ice slides over wet, deforming till. If the till is of thickness h_T and has (effective) viscosity η_T, then an appropriate sliding law would be

$$\tau_b = \eta_T u_b / h_T. \tag{4.124}$$

In fact, till is likely to have a nonlinear rheology, and also in accordance with Terzaghi's principle of soil mechanics, one would expect η_T to depend on effective pressure N. One (measured) rheology for till gives the strain rate as

$$\dot{\varepsilon} = A_T \tau^a N^{-b}, \tag{4.125}$$

in which case the sliding law would be again of the form (4.123), with $c = (A_T h_T)^{-1/a}$, $r = 1/a$, $s = b/a$. Thus there are some good reasons to choose (4.123) as an all purpose sliding law, and this points up the necessity of a subglacial hydraulic theory to determine N.

4.3.1 Drainage

Subglacial water is generated both by basal melt (of significance in ice sheets) and from run-off of surface melt or rainfall through crevasses and *moulins*, which access the glacier bed. Generally the basal water pressure p_w is measured to be below the overburden ice pressure p_i, and the resulting positive effective pressure $N = p_i - p_w$ tends to cause any channels in the ice to close up (by creep of the ice). In fact, water is often seen to emerge from outlet streams which flow through large tunnels in the ice, and the theory which is thought to explain how such channels remain open asserts that the channel closure rate is balanced by melt back of the channel walls by frictional heating due to the water flow.

The classical theory of subglacial drainage is due to Röthlisberger, and is described below. We willdiscuss only the determination of effective pressure in steady state conditions. We consider a single channel of cross sectional area S, through which there is a water flux Q. We take Q as being determined by external factors such as surface meltwater runoff; this is appropriate for glaciers, but not for ice sheets, where

Q must be determined by subglacial melting (we come back to this later). If the flow is turbulent, then the Manning law for flow in a straight conduit is

$$\rho_w g \sin\alpha - \frac{\partial p}{\partial s} = f_1 Q^2/S^{8/3}, \tag{4.126}$$

where ρ_w is water density, g is gravity, s is distance down channel, α is the local bed slope, p is water pressure, and f_1 is a roughness coefficient related to the Manning friction factor. If we suppose that the frictional heat dissipated by the turbulent flow is all used to melt the walls, then

$$mL = Q\left[\rho_w g \sin\alpha - \frac{\partial p}{\partial s}\right], \tag{4.127}$$

where L is the latent heat, and m is the mass of ice melted per unit length per unit time.

The last equation to relate the four variables S, Q, p and m stems from a kinematic boundary condition for the ice, and represents a balance between the rate at the which the ice closes down the channel, and the rate at which melting opens it up:

$$\frac{m}{\rho_i} = KS(p_i - p)^n; \tag{4.128}$$

here m/ρ_i is the rate of enlargement due to melt back, while the term on the right hand side represents ice closure due to Glen's flow law for ice; the parameter K is proportional to the flow law parameter A.

Elimination of m and S yields a second order ordinary differential equation for the effective pressure $N = p_i - p$, which can be solved numerically. However, it is also found that typically $\partial p/\partial s \ll \rho_w g \sin\alpha$ (in fact, we expect $\partial p/\partial s \sim \rho_w g d/l$, so that in the notation of (4.10), the ratio of these terms is of $O(\varepsilon)$); the neglect of the $\partial p/\partial s$ term in (4.126) and (4.127) is singular, and causes a boundary layer of size $O(\varepsilon)$ to exist near the terminus in order that p decrease to atmospheric pressure. Away from the snout, then

$$S \approx \left[\frac{f_1 Q^2}{\rho_w g \sin\alpha}\right]^{3/8}, \; KSN^n \approx \frac{Q\rho_w g \sin\alpha}{\rho_i L}, \tag{4.129}$$

thus

$$N \approx \beta Q^{1/4n}, \tag{4.130}$$

where β is a material parameter which depends (inversely) on roughness. Typical values give $N = 30$ bars when $Q = 10$ m^3 s^{-1}. Since $p_i = 9$ bars for a 100 metre deep glacier, it is clear that the computed N may exceed p_i. In this case, p must be atmospheric and there will be open channel flow. It is likely that seasonal variations are important in adjusting the hydraulic régime.

4.4 Waves, surges and mega-surges

4.4.1 Waves on glaciers

Waves on glaciers are mostly easily understood by considering an isothermal, two-dimensional model. We suppose the base is flat ($h = 0$), so that equations (4.24) and (4.25) give

$$H_t + \left[\{1 - \varepsilon H_x\}^n \frac{H^{n+2}}{n+2} + u_b H\right]_x = s'(x), \tag{4.131}$$

where $s'(x)$ is the accumulation rate, and $\varepsilon \sim 0.1$. If we firstly put $\varepsilon = 0$ and also $u_b = 0$, then

$$H_t + H^{n+1} H_x = s'(x), \tag{4.132}$$

which has the steady state

$$\frac{H_0^{n+2}}{n+2} = s(x). \tag{4.133}$$

With $s' > 0$ in $x < 0$ (say) and $s' < 0$ in $x > 0$ ($x = 0$ is then the *firn line*) (4.133) defines a concave profile like that in figure 4.6. (4.132) is clearly hyperbolic, and admits wave like disturbances which travel at a speed H^{n+1}, which is in fact $(n+1)$ (≈ 4) times the surface speed. For an arbitrary initial condition $H = \tilde{H}(x)$ at $t = 0$, the solution by characteristics is

$$\begin{aligned} \frac{H^{n+2}}{n+2} &= s(x) - s_1(\sigma), \\ t &= \int_\sigma^x \frac{dx}{[(n+2)\{s(x) - s_1(\sigma)\}]^{(n+1)/(n+2)}}, \end{aligned} \tag{4.134}$$

where s_1 is defined by

$$\frac{\tilde{H}^{n+2}(\sigma)}{n+2} = s(\sigma) - s_1(\sigma). \tag{4.135}$$

Thus, for small perturbations, s_1 is small.

The characteristics of (4.132) propagate downstream and reach the snout (where $H = 0$) in finite time. If we wish to approximate the characteristic solution where s_1 is small, straightforward linearisation is invalid near the snout where $H_0 = 0$; rather, a uniformly valid approximation can be obtained by linearising the characteristics:

$$H_t + H_0^{n+1} H_x \approx s'(x) \tag{4.136}$$

for $H \approx H_0$, where the general solution is

$$H = H_0(x) + \phi(\xi - t), \tag{4.137}$$

where

$$\xi = \int_0^x \frac{dx}{H_0^{n+1}(x)} \tag{4.138}$$

is a characteristic spatial coordinate (note ξ is finite at the snout). (4.137) clearly reveals the travelling wave characteristic of the solution.

Margin response

If H is increased locally (e.g. due to the surge of a tributary glacier) then a shock travels forward. The rôle of the term in ε is then to diffuse such shocks. A shock at $x = x_s$ will propagate at a rate

$$\dot{x}_s = \frac{[H^{n+2}]_-^+}{(n+2)[H]_-^+}, \tag{4.139}$$

where $[\ \ ]_-^+$ denotes the jump across x_s. When the shock reaches the snout, it then propagates at a speed $H_-^{n+1}/(n+2)$, which is *slower* than the surface speed.

In the neighbourhood of a shock (with $u_b = 0$), we put

$$x = x_s + \nu X, \tag{4.140}$$

so that

$$\frac{\partial H}{\partial t} - \frac{\dot{x}_s}{\nu}\frac{\partial H}{\partial X} + \frac{1}{\nu}\left[\left\{1 - \frac{\varepsilon}{\nu}H_X\right\}^n \frac{H^{n+2}}{n+2}\right]_X = s'(x_s + \nu X); \tag{4.141}$$

if ν is small, the profile rapidly relaxes to the steady travelling wave described by

$$\dot{x}_0 H_X = \left[\{1 - H_X\}^n \frac{H^{n+2}}{n+2}\right]_X, \tag{4.142}$$

providing we choose $\nu = \varepsilon$, which thus gives the width of the shock structure. (4.142) can be solved by quadrature.

Seasonal waves

There is no explanation of seasonal waves available. On the face of it, we might seek waves of amplitude of velocity of $O(1)$ propagating at a speed $O(1/\mu)$, where μ is the ratio of one year to the convective time scale, so $\mu \lesssim 0.05$. Apparently we should associate the variations in u with variations in water supply, so that a natural model would involve only sliding, so

$$H_t + (Hu_b)_x = s'(x), \tag{4.143}$$

and if $u_b = \phi(t/\mu)H^m/(m+1)$, where $\phi(t/\mu)$ represents the seasonal variation of water supply and hence of N, then

$$H_t + \phi(t/\mu)H^m H_x = s'(x). \tag{4.144}$$

Unfortunately, while the surface speed will indeed oscillate seasonally in this model, the kinematic wave equation propagates waves at $(m+1)$ times the surface speed, so there seems to be no mechanism for the rapid propagation which has been observed. Another possibility is then that the variations in N force a wave passage in the hydraulic system itself, but this has not been explored.

4.4.2 Surges

It has long been suggested that the fast velocities during surges could only be caused by rapid sliding. Therefore it is sufficient to analyse the mass conservation equation in the form

$$H_t + (Hu)_x = s'(x), \tag{4.145}$$

where u is the sliding velocity. Also, it has been thought that if the sliding velocity were a multi-valued function of basal stress τ_b (i.e. $\tau_b(u)$ has a decreasing portion)) then since $\tau_b = H(1 - \varepsilon s_x) \approx H$, this would cause the ice flux $Q = uH$ to be multi-valued as shown in figure 4.8, in which case we might expect relaxation oscillations to occur for values of s intermediate between the two noses of $Q(H)$. Two fundamental questions arise. Firstly, is there any genuine reason why $\tau_b(u_b)$ should be non-monotone, and secondly, how would such a relaxation oscillator work in the spatially dependent case?

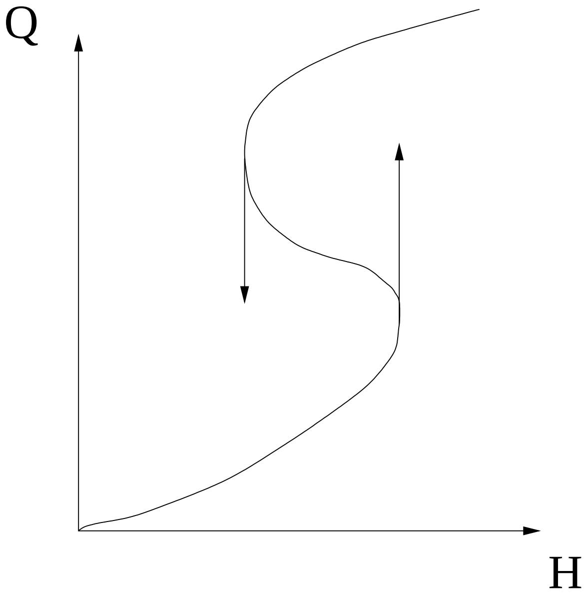

Figure 4.8: A multi-valued flux-depth relation can cause oscillatory surges.

The discussion in section 4.2 suggested the possibility of non-monotone $\tau_b(u)$ for flow over a periodic bedrock. However, more realistic bedrocks probably do not have this feature, and τ increases with both u and N. What observations of the 1982–3 surge of Variegated Glacier showed, however, was that there is a switch in drainage pattern during a surge. There are two possible modes of drainage: the Röthlisberger channel described in section 4.2 with the value of N determined by the water flow, N_R, say; however, if no channels are present, then water will fill cavities at the bed and leak from one to another. This is called the linked cavity régime and operates at a higher water pressure and thus lower effective pressure, N_c, than in the channel drainage. The crucial factor which enables surges to take place is the switch mechanism, and this depends on the ice flow over the cavities. If the sliding law is, as discussed in section 4.2, of the form $\tau_b = Nf(u/N^n)$, then in fact the stresses in the ice are actually determined by u/N^n, and in particular the water stored by cavities depends on this parameter.

It turns out that a simple model of combined water flow through both cavities and a channel system exhibits instability (the channels close down) if the cavity storage volume is large enough, and thus the instability occurs at a critical value of $\Lambda = u/N^n$, denoted Λ_c. It follows from this that a combined model of the drainage system is

$$\begin{aligned} N &= N_R, \qquad u/N^n < \Lambda_c; \\ N &= N_c, \qquad u/N^n > \Lambda_c; \end{aligned} \tag{4.146}$$

and if this is written as a function $N(u)$, it is multi-valued, as shown in figure 4.9. As a consequence of this, the sliding law is indeed multi-valued, and hence $Q(H)$ has the form shown in figure 4.10.

There are two critical values of Q in figure 4.10, denoted Q_+, Q_-: these are the values at the noses of the curve (where also $H = H_+, H_-$). If $s(x) < Q_+$, then an equilibrium glacier profile exists in which $Q = s(x)$. However, if the maximum value of s, $s_{\max}$, is greater than Q_+, then such a stable equilibrium cannot occur, and the glacier surges.

The sequence of events in a surge is then as follows. The glacier grows from a quiescent state in which $Q < Q_+$ on the lower (slow) branch everywhere. When the maximum depth reaches H_+, there is a *reservoir* zone where $H > H_-$. The ice flux at H_+ jumps to the upper (fast) branch by switching drainage pattern, and this switch propagates upstream and downstream to where $H = H_-$. These *activation waves* propagate at rates of hundreds of metres per *hour* (and in effect have been observed). Once the activation waves have propagated to the boundaries of the reservoir zone, the ice flow is described by the fast mode on the upper branch, and the activated reservoir zone propagates rapidly downstream, overriding the stagnant snout and propagating forwards as a front. In terms of figure 4.10, the surge terminates when H reaches H_- everywhere, and deactivation waves propagate inwards from the boundaries of the exhausted reservoir zone to re-establish the channel drainage system. There then follows another quiescent phase where the maximum value of H increases from H_- to H_+ before the next surge is initiated.

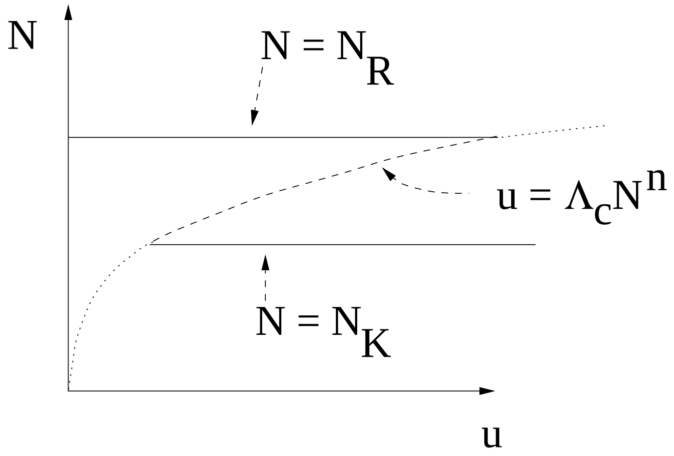

Figure 4.9: N is a multi-valued function of u.

4.4.3 Sliding and ice streams

It is not known why the ice flow on the Siple Coast of Antarctica, which flows out to the floating Ross ice shelf, segregates itself into the five distinct ice streams A to E. The picture which one has of this region is of a gently sloping (slope $\alpha \sim 10^{-3}$) kilometer thick ice sheet which flows in the ice streams at typical rates of 500 m y^{-1}. Such rapid velocity can only be due to basal sliding, and the seismic evidence indicates that the ice is underlain by several metres of wet till. One can expect that a sliding law of the form advocated previously is appropriate, that is

$$\tau_b = c u_b^r N^s, \tag{4.147}$$

with r and s positive. The issue then arises as to how to prescribe N. Recall from section 4.2 that for drainage through Röthlisberger channels, an appropriate law is $N = \beta Q_w^{1/4n}$, where Q_w is water flux. When ice flows over till, an alternative flow route is possible, that is, that water excavates 'canals' in the subglacial till. A theoretical description of this drainage system suggests that it is more likely for gently sloping ice flow, and also that the relation between N and Q_w is of the opposite sense, that is, that $\partial N/\partial Q_w < 0$.

In this case an interesting feedback exists. In Antarctic ice streams, there is little, if any, surface melt reaching the bed, and the basal water flow is due to melting there. The quantity of meltwater produced per unit area per unit time is given by the melt velocity

$$v_m = \frac{G + \tau_b u_b - g}{\rho_w L}, \tag{4.148}$$

where ρ_w is water density, L is latent heat, G is geothermal heat flux, and g is the basal heat flux into the ice. This assumes the base is at the melting point. Thus we expect the basal water flux $Q_w \sim G + \tau_b u_b - g$, and so Q_w increases with u_b (the

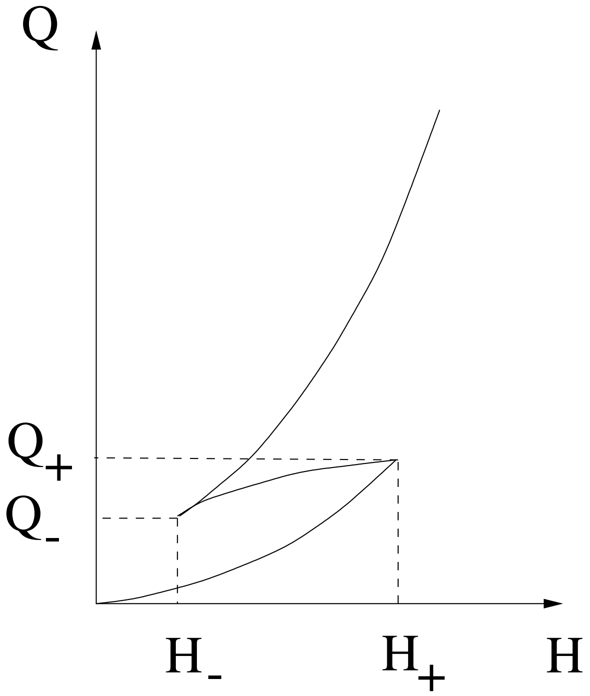

Figure 4.10: Q is a multi-valued function of H

dependence of g on u_b is likely to be weaker — boundary layer theory would suggest $g \sim u_b^{1/2}$). If also N decreases with Q_w, then N decreases as u_b increases. But this causes further increase of u_b via the sliding law. This positive feedback can lead to a runaway phenomenon which we may call hydraulic runaway.

To get a crude idea of how this works, we denote the ice thickness as h and slope $\sin \alpha$. If the velocity is u, then the ice flux is

$$Q = hu, \tag{4.149}$$

the basal shear stress is

$$\tau = \rho g h \sin \alpha = c u^r N^s, \tag{4.150}$$

we suppose

$$N = \gamma Q_w^{-p}, \tag{4.151}$$

and that

$$Q_w = b[G + \tau u - g], \tag{4.152}$$

with

$$g = a u^{1/2}. \tag{4.153}$$

Consequently

$$h = \frac{fu^r}{[G + (\rho g \sin\alpha)hu - au^{1/2}]^m}, \tag{4.154}$$

where

$$m = ps, \; f = \frac{c\gamma^s}{(\rho g \sin\alpha)b^m}. \tag{4.155}$$

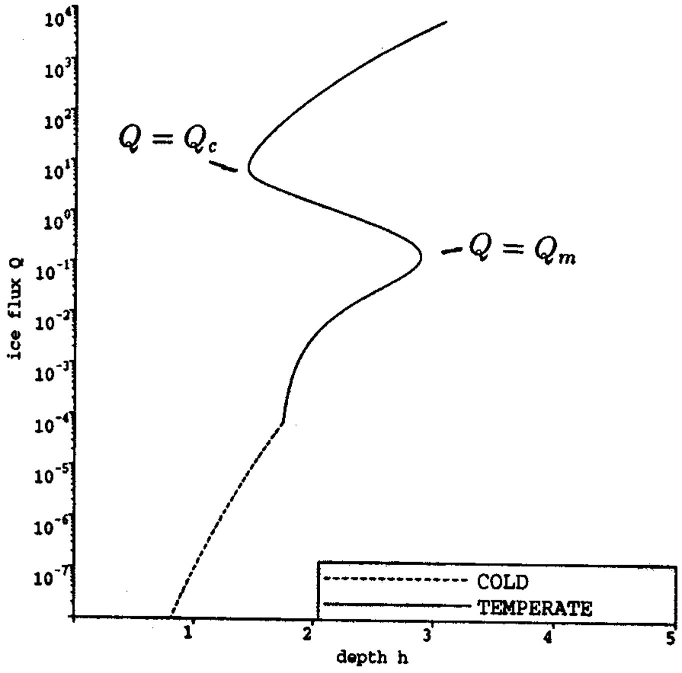

Figure 4.11: Thermal feedback causes a multi-valued ice flux.

It is not difficult to see from (4.154) if f is low enough (equivalently, the friction coefficient c is low enough), that u and hence the ice flux Q will be a multivalued function of h, as shown in figure 4.11. In fact, application of realistic parameter values suggests that such multi-valued flux laws are normal.

What then happens in a region such as the Ross ice shelf area? We suppose that the ice flux is determined by conditions upstream, so that if the ice flux per unit width is q, and the width of the discharge region is W, then

$$Wq = s, \tag{4.156}$$

where s is the volume flux of ice discharged. Now if $s/W < Q_c$ (see figure 4.15), then a uniform slow moving ice flow is possible. Similarly, if $s/W > Q_m$, a uniform fast moving ice stream is possible. What if $Q_m < s/W < Q_c$? A uniform stream is now unstable, and we may expect an instability to occur, whereby ice streams spontaneously occur, as observed. Such an instability would be mediated by transitions in water pressure, since basal water will flow from fast streams at high water pressure to slower ice at low water pressure. This generates a lateral enthalpy flux, and in a steady state this can be balanced by a heat flux in the ice in the opposite direction, since cooling (g) is less effective at lower u, therefore the slow ice is warmer near the base than the ice streams. No analysis of this idea has yet been done, though it is an intriguing mathematical problem.

4.4.4 Heinrich events and the Hudson strait mega-surge

What if the drainage channel of an ice sheet over deforming till is relatively narrow? By analogy of the pattern formation mechanism in reaction diffusion equations, one would expect that a multivalued flux-depth relation would not allow separate streams to form if the channel width is too small, and in this case we would expect periodic surges to occur down the channel, if the prescribed mass flux lies on the unstable position of figure 4.15.

A situation of this type appears to have occurred during the last ice age. The Laurentide ice sheet which existed in North America drained the ice dome which lay over Hudson Bay out through the Hudson Strait, a 200 km wide trough which discharged the ice (as icebergs) into the Labrador sea and thence to the North Atlantic.

Hudson Bay is underlain by soft carbonate rocks, mudstones, which can be mobilised when wet. It is suggested that the presence of these deformable sediments, together with the confined drainage channel, led to the occurrence of semi-periodic surges of the Hudson Strait ice stream. The evolution of events is then as follows. When ice is thin over Hudson Bay, the mudstones are frozen at the base, there is no sliding and very little ice flow. Consequently, the ice thickens and eventually the basal ice warms. The basal muds thaw, and sliding is initiated. If the friction is sufficiently low, then the multi-valued sliding law of figure 4.15 is appropriate, and if the accumulation rate is large enough, cyclic surging will occur. During a surge, the flow velocity increases dramatically (calculations suggest a velocity of 2 centimetres a *second*!), and there results a massive iceberg flux into the North Atlantic. On the lower branch of figure 4.15, water production is virtually absent, Q_w is low in (4.152) since the flow is slow and the geothermal and viscous heat at the base can be conducted away by the ice. The low value of Q_w gives high N, consistent with low u. On the upper branch, however, viscous heat dominates, and Q_w is large, N is small, also consistent with a high u.

At the end of a surge, the rapid ice drawdown causes the water production to drop, and the rapid velocities switch off. This may or may not also be associated with re-freezing of the basal mudstones.

When water saturated soils freeze, frost heave occurs by sucking up water to the freezing front via capillary action, and this excess water freezes (at least for fine grained clays and silts) in a sequence of discrete ice lenses. Heaving can occur at a typical rate of perhaps a metre per year, though less for fine grained soils, and the rate of heave is suppressed by large surface loads. Calculations suggest a surge period of perhaps a hundred years, with a drawdown of a thousand metres, and a recovery period on the order of 5,000—10,000 years. During the surge, the rapidly deforming basal muds will dilate (in the deforming horizon, likely to be only a metre or so thick). At the termination of a surge, this layer re-consolidates, and we can expect the total heave to be a certain (small) fraction of the frost penetration depth. In effect, the ice lenses freeze the muds into the ice stream, so that when the next surge phase is initiated, some of this frozen-in basal sediment will be transported downstream, and thence rafted out into the North Atlantic in iceberg discharge.

In fact, there is evidence that this rather glamorous sequence of events actually occurs. Heinrich events are layers of ice-rafted debris in deep-sea sediment cores from the North Atlantic which indicate (or are consistent with) massive iceberg discharges every 7000 years or so. In addition, oxygen isotope concentrations in ice cores from Greenland indicate that severe cooling cycles occurred during the last ice age. One theory has it that such cooling events can be caused by a switch-off of North Atlantic deep water (NADW) circulation — effectively switching off the convective heat transport from equatorial latitudes and thus cooling the atmosphere. It seems that bunches of these cooling cycles are terminated by Heinrich events, in the sense that following Heinrich events the climate warms suddenly. There are two reasons why this should be so. On the one hand, the sudden reduction in ice thickness should warm the air above, and also it can be expected that a massive iceberg flux to the North Atlantic acts as a source of negative thermal buoyancy, which can re-initiate an otherwise stagnant circulation. Rather than being lumbering beasts, glaciers and ice sheets show every sign of being dynamically active agents in shaping the climate and the earth's topography.

4.5 Notes and references

The best source for general information about glaciers and ice sheets is the book by Paterson (1994).

4.5.1 Waves on glaciers

An early discussion of surface waves is by Finsterwalder (1907). The modern theory is largely due to Nye (1960), who analyses linear waves; a nonlinear analysis is given by Fowler and Larson (1980b). Seasonal waves are discussed by Deeley and Parr (1914) and more recently by Hodge (1974). Wave ogives are lucidly discussed by Waddington (1986).

4.5.2 Surges

Surging glaciers are located in various places round the world, including Alaska and Svalbard. Famously, there are no surging glaciers in the European Alps, but it is thought that there used to be at least one, Vernagtferner, in the Austrian Alps, which last surged in about 1900. Early paintings, documented by Nicolussi (1990) indicate surges occurring in about 1600 and 1680, to judge from the jagged surface of the glacier in the images.

The surge on Variegated Glacier is discussed by Kamb et al. (1985), and theoretical descriptions are given by Kamb (1987) and Fowler (1987a). Observations of Trapridge Glacier are described by Clarke et al. (1984). The issue of the Journal of Geophysical Research in which Fowler's (1987a) article appears is a collection of articles on fast glacier flow, including both ice streams, surging glaciers, and tidewater glaciers. For a discussion of the dynamics of ice stream B in Antarctica, see Engelhardt et al. (1990). The basic theory of jökulhlaups is due to Nye (1976).

The theory of surface waves on glaciers was really worked out by several early authors, and is discussed by Lliboutry (1965) in his voluminous treatise. The modern linear theory is worked out by Nye (1960), and is expanded on by Fowler and Larson (1980b). Fowler (1982) obtained a theory of seasonal waves on the rather dubious basis of a sliding law with $\partial u_b/\partial \tau_b \gg 1$, but this is unlikely to be correct.

Surges are discussed by Kamb et al. (1985) and Clarke et al. (1984). The present discussion is based on work by Fowler (1987a), the mathematical details of which are worked out in Fowler (1989).

4.5.3 Streams, shelves, sheets, grounding line

The dynamics of ice streams are reviewed by Bentley (1987), while the theory of Hudson Strait mega-surges is due to MacAyeal (1993). Heinrich events are discussed by Bond et al. (1992), while the present discussion is based on a paper by Fowler and Johnson (1995).

The basic scaling in the shallow ice approximation is due to Fowler and Larson (1978): it is elaborated in the book by Hutter (1983). For ice sheets, similar derivations have been done by Morland (1984), Hutter et al. (1986) and Fowler (1992), of whom we follow the latter. The concept of thermally induced instability was enunciated by Robin (1955) and taken up by Clarke et al. (1977) and Yuen and Schubert (1979), but more or less scotched by Fowler and Larson (1980a).

The issue of the extra condition which describes the position of the grounding line is a thorny one, which is as yet not completely resolved. At a formal level, the most detailed studies are those of Wilchinsky and co-workers (Chugunov and Wilchinsky 1996, Wilchinsky and Chugunov 2000, 2001), but these papers are severely impenetrable, even to initiates. Chugunov and Wilchinsky (1996) consider the transition zone in a similar manner to that presented here. They assume Newtonian flow and a steady state, and claim to deduce the grounding line position. Two key assumptions are apparent in their reasoning. The first is the arbitrary assumption that the horizontal length scale for the ice shelf is comparable to that for the ice sheet. This allows then to deduce that (with present notation) $H_G = \beta(\varepsilon/\delta)^{1/3}$ for some $O(1)$ coefficient β; the origin of this (realistic) scale is, however, mysterious. The deduction of a numerical value of $\beta \approx 1.5$ from a numerical calculation involves (in the appendix to the paper) the assumption that the bed $B(X)$ (in present notation) is smooth. While this may be correct, it is not immediately obvious why it should be so.

Wilchinsky and Chugunov (2000) extend this analysis to the junction between a rapidly moving ice stream, where shear is less important, and an ice shelf. They now state that the grounding line position is determined by the requirement of continuity of the lower ice surface at the grounding line, but they do not carry through the calculation. The scaling analysis involved is rather different than for the shear-dominated sheet/shelf transition. Finally, Wilchinsky and Chugunov (2001) extend the scaling of the 1996 paper to the nonlinear rheology of Glen's law. The flow is still steady, and it is stated that the condition $B(0+) = 0$ determines the grounding line position, and that the depth at the grounding line is, in present notation,

$$q_G = \left(\frac{\beta\varepsilon}{\delta}\right)^n H_G^{n=2}; \qquad (4.157)$$

this can be compared with (4.94). Numerical evaluation of β is again only done for the Newtonian case $n = 1$, under the additional assumptions of $B_X = B_{XX} = 0$ at $X = 0+$, which are geometric constraints associated with the geometrical transformation of the domain used in solving the problem numerically. Like its predecessors, this paper is hard to fathom.

4.5.4 Sliding and drainage

The theory of basal sliding over hard beds stems from Weertman (1957) and Lliboutry (1968). Two reviews of progress by the end of the seventies are in Lliboutry (1979) and Weertman (1979). The linear theory is primarily due to Nye (1969, 1970) and Kamb (1970), while the material presented here is based largely on Fowler (1986, 1987b). Till rheology is discussed by Boulton and Hindmarsh (1987). An up to date theoretical discussion of sub-glacial cavitation is given by Schoof (2005).

The classical theory of drainage is due to Röthlisberger (1972), while the development for jökulhlaups follows Nye (1976).

A recent review of subglacial processes of current interest is by Clarke (2005).

Exercises

4.1 Solve the cross section velocity equation (4.21) with (4.20), assuming the rate factor $A = 1$ and a semi-circular profile for the ice cross section I. Hence find the ice flux Q as a function of the cross sectional area of the flow.

4.2 Use lubrication theory to derive an approximate model for two-dimensional flow of a valley glacier, assuming Glen's flow law with a rate constant independent of temperature, and no sliding at the base. Non-dimensionalise the model, and show that for typical lengths of 10 km, accumulation rates of 1 m y^{-1}, and if the rate constant in Glen's law is 0.2 $\mathrm{bar}^{-3}\ \mathrm{y}^{-1}$ (with the Glen exponent being $n = 3$), a typical glacier depth is 100 m. Show that the dimensionless model depends on the single dimensionless parameter $\mu = d \cot\alpha / l$, where d is the depth scale, l is the length scale, and α is the valley slope. What are typical values of μ?

Show that if $\mu \ll 1$, the model takes the form of a first order hyperbolic wave equation. Write down the solution for small perturbations to the steady state, and show that the perturbations grow unboundedly near the glacier snout. Why is this? Write an alternative linearisation which allows a bounded solution to be obtained.

More generally, an exact characteristic solution of the model allows shocks to form (and thus for the glacier snout to advance). Discuss the rôle of μ in shock formation.

4.3 A glacier is subject to a balance function a whose amplitude varies sinusoidally in time about a mean (space dependent) value; specifically

$$a = a_0(x) + a_1 e^{i\omega t},$$

where a_1 is constant (the real part may be assumed). Use an appropriate linearised wave theory to determine the resultant form of the perturbed surface. What can you say about the effect of millennial scale climate changes? about annual balance changes?

How would you generalise your result to a general time dependent amplitude variation?

4.4 Write down the equations governing three-dimensional flow of an ice sheet, and show how they can be non-dimensionalised to obtain (4.43), assuming Glen's flow law and a temperature independent rate coefficient. Show that the dimensionless basal shear stress is $\boldsymbol{\tau} = -H\boldsymbol{\nabla} s$.

4.5 A marine ice sheet terminates on the continental shelf during an ice age, in water of fixed depth d_w. The calving ice velocity is cd_w, where c is a constant. Show how this, together with an expression for the ice flux which can be inferred from (4.43), yields a boundary condition for H of the form (4.44).

(Assume firstly that $\mathbf{u}_b = \mathbf{0}$. Then find the corresponding boundary condition assuming a Weertman sliding law of the form $\mathbf{u}_b = C|\boldsymbol{\tau}|^{m-1}\boldsymbol{\tau}$, where $m > 1$, using $\boldsymbol{\tau} = -H\boldsymbol{\nabla} s$.)

4.6 The horizontal force balance at the calving front of an ice shelf can be written in the form

$$\sigma_{11} = -p_w,$$

where p_w is the (hydrostatic) water pressure. Show that, when written in terms of the ice sheet scales, this condition takes the form

$$-s + \varepsilon^2(-p + \tau_1) = \delta z,$$

where $\delta = (\rho_w - \rho_i)/\rho_i$ and z is measured from sea level. Hence show that in terms of the ice shelf scales, the condition is

$$-s - p + \tau_1 = z.$$

Use the approximate results $p \approx -\tau_1$, $s + b \approx 0$, to show that the vertically averaged deviatoric longitudinal stress at the calving front is

$$\bar{\tau}_1 = -\tfrac{1}{4} b \left(1 + \delta\right)^2 ,$$

and if this is taken as the boundary condition for the ice shelf stress τ_1, show that

$$\tau_1 \approx -\tfrac{1}{4} b$$

everywhere.

4.7 The drainage pressure in a subglacial channel is determined by the Röthlisberger equations

$$\begin{aligned} \rho_w g \sin\alpha_c &= \frac{f_1 Q^2}{S^{8/3}}, \\ mL &= \rho_w g Q \sin\alpha_c, \\ \frac{m}{\rho_i} &= KSN^n. \end{aligned}$$

Explain the meaning of these equations, and use them to express the effective pressure N in terms of the water flux Q. Find a typical value of N, if $Q \sim 1$ m^3 s^{-1}, and $\sin\alpha_c \sim 0.1$, $f_1 = f\rho_w g$, $f = 0.05$ m$^{-2/3}$ s^2, $n = 3$, $L = 3.3 \times 10^5$ J kg^{-1} and $K = 0.1$ bar^{-3} y^{-1}.

Use a stability argument to explain why Röthlisberger channels can be expected to form an arterial network.

4.8 The relation between ice volume flux and depth for a surging glacier is found to be a multivalued function, consisting of two monotonically increasing parts, from $(0,0)$ to (H_+, Q_+) and from (H_-, Q_-) to (∞, ∞) in (H, Q) space, where $H_+ > H_-$ and $Q_+ > Q_-$, with a branch which joins (H_-, Q_-) to (H_+, Q_+). Explain how such a flux law can be used to explain glacier surges if the balance function $s(x)$ satisfies $\max s > Q_+$, and give a rough estimate for the surge period.

What happens if $\max s < Q_-$? $\max s \in (Q_-, Q_+)$?

4.9 The depth of a glacier satisfies the equation

$$H_t + \frac{\partial}{\partial x}\left[\left(1 - \mu H_x\right)^n \frac{H^{n+2}}{n+2}\right] = s'(x),$$

where $\mu \ll 1$. Suppose first that $\mu \ll 1$, so that the diffusion term can be neglected. Write down the characteristic solution for an arbitrary initial depth profile. What is the criterion on the initial profile which determines whether shocks will form?

Now suppose $s = 1/(n+2)$ is constant, so that a uniform steady state is possible. Describe the evolution of a perturbation consisting of a uniform increase in depth between $x = 0$ and $x = 1$, and draw the characteristic diagram.

Shock structure. By allowing $\mu \neq 0$, the shock structure is described by the local rescaling $x = x_s(t) + \mu X$. Derive the resulting leading order equation for H, and find a first integral satisfying the boundary conditions $H \to H_\pm$ as $X \to \pm\infty$, where $H_- > H_+$ are the values behind and ahead of the shock. Deduce that the shock speed is

$$\dot{x}_s = \frac{[H^{n+2}]_-^+}{[H]_-^+},$$

and that $\phi = H/H_+$ satisfies the equation

$$\phi_\xi = -[g(\phi)^{1/n} - 1],$$

where $\xi = X/H_+$, $\phi \to r$ as $\xi \to -\infty$, $\phi \to 1$ as $\xi \to \infty$, and

$$g(\phi) = \frac{(r^{n+2}-1)(\phi-1)+(r-1)}{(r-1)\phi^{n+1}},$$

with $r = H_-/H_+ > 1$. Show that $g(1) = g(r) = 0$, and that $g(\phi) > 1$ for $1 < \phi < r$, and deduce that a monotonic shock structure solution joining H_- to H_+ does indeed exist.

Suppose that $\delta = \Delta H/H_+$ is small, where $\Delta H = H_- - H_+$. By putting $r = 1+\delta$ and $\phi = 1 + \delta\Phi$, show that

$$g = 1 + \frac{\delta^2(n+1)(n+2)}{2}\Phi(1-\Phi) + \ldots,$$

and deduce that

$$\Phi_\Xi \approx -\Phi(1-\Phi),$$

where

$$\Xi = \frac{\delta^2(n+1)(n+2)}{2n}\xi.$$

Deduce that the width of the shock structure is of dimensionless order

$$x - x_s \sim \frac{2n\mu H_+}{\delta^2(n+1)(n+2)},$$

or dimensionally

$$\frac{2n}{(n+1)(n+2)}\frac{d_+^3}{(\Delta d)^2\tan\alpha},$$

and that for a glacier of depth 100m, slope ($\tan\alpha$) 0.1, with $n = 3$, a wave of height 10 m has a shock structure of width 30 km! (*This is the monoclinal flood wave for glaciers, analogous to that for rivers discussed in chapter 2.*)[5]

---

[5] The observation that the smallness of surface slope diffusion is offset by the smallness of surface amplitude is made, for example, by Gudmundsson (2003) (see paragraph 16).

References

Aagaard, P. and H. Helgeson 1983 Activity/composition relations among silicates and aqueous solutions: II. Chemical and thermodynamic consequences of ideal mixing of atoms on homological sites in montmorillonites, illites, and mixed-layer clays. Clays and Clay Minerals **31**, 207–217.

Abbott, M. R. 1956 A theory of the propagation of bores in channels and rivers. Proc. Camb. Phil. Soc. **52**, 344–362.

Abramowitz, M. and I. Stegun 1964 Handbook of mathematical functions. Dover Publications, New York.

Agustin, L. C. Barbante, P. R. F. Barnes, J. M. Barnola, M. Bigler, E. Castellano, O. Cattani, J. Chapellaz, D. Dahl-Jensen, B. Delmonte, G. Dreyfus, G. Durand, S. Falourd, H. Fischer, J. Flückiger, M. E. Hansson, P. Huybrechts, G. Jugie, S. J. Johnsen, J. Jouzel, P. Kaufmann, J. Kipfstuhl, F. Lambert, V. Y. Lipenkov, G. C. Littot, A. Longinelli, R. Lorrain, V. Maggi, V. Masson-Delmotte, H. Miller, R. Mulvaney, J. Oerlemans, H. Oerter, G. Orombelli, F. Parrenin, D. A. Peel, J.-R. Petit, D. Raynaud, C. Ritz, U. Ruth, J. Schwander, U. Siegenthaler, R. Souchez, B. Stauffer, J. P. Steffensen, B. Stenni, T. F. Stocker, I. E. Tabacco, R. Udisti, R. S. W. van de Wal, M. van den Broeke, J. Weiss, F. Wilhelms, J.-G. Winther, E. W. Wolff and M. Zucchelli 2004 Eight glacial cycles from an Antarctic ice core. Nature **429**, 623–628.

Allen, J. R. L. 1985 Principles of physical sedimentology. Chapman and Hall, London.

Alley, R. B., S. Anandakrishnan and P. Jung 2001 Stochastic resonance in the North Atlantic. Paleoceanography **16** (2), 190–198.

Alley, R. B., P. A. Mayewski, T. Sowers, M. Stuiver, K. C. Taylor and P. U. Clark 1997 Holocene climatic instability: A prominent, widespread event 8200 years ago. Geology **25**, 483–486.

Andrews, D. G. 2000 An introduction to atmospheric physics. C. U. P., Cambridge.

Angevine, C. L. and D. L. Turcotte 1983 Porosity reduction by pressure solution: a theoretical model for quartz arenites. Geol. Soc. Amer. Bull. **94**, 1129–1134.

Aris, R. 1956 On the dispersion of a solute in a fluid flowing through a tube. Proc. R. Soc. A**235**, 67–78.

Aris, R. 1975 Mathematical theory of diffusion and reaction in permeable catalysts. Two volumes, O. U. P., Oxford.

Arrhenius, S. 1896 On the influence of carbonic acid in the air upon the temperature of the ground. Phil. Mag. **41**, 237–275.

Athy, L. F. 1930 Density, porosity, and compaction of sedimentary rocks. Amer. Assoc. Petrol. Geol. Bull. **14**, 1–22.

Audet, D. M. and A. C. Fowler 1992 A mathematical model for compaction in sedimentary basins, Geophys. J. Int. **110**, 577–590.

Baker, V. R. 2001 Water and the martian landscape. Nature **412**, 228–236.

Baker, V. R. and D. J. Milton 1974 Erosion by catastrophic floods on Mars and Earth. Icarus **23**, 27–41.

Baldwin, P. 1985 Zeros of generalized Airy functions. Mathematika **32**, 104–117.

Balmforth, N. J. and S. Mandre 2004 Dynamics of roll waves. J. Fluid Mech. **514**, 1–33.

Barcilon, V. and D. R. MacAyeal 1993 Steady flow of a viscous ice stream across a no-slip/free-slip transition at the bed. J. Glaciol. **39** (131), 167–185.

Barry, R. G. and R. J. Chorley 1998 Atmosphere, weather and climate, 7th ed. Routledge, London.

Baú, D., G. Gambolati and P. Teatini 2000 Residual land subsidence near abandoned gas fields raises concern over Northern Adriatic coastland. EOS **81**, no. 22, pp. 245, 248–9.

Bear, J. 1972 Dynamics of fluids in porous media. Elsevier. (Dover reprint, 1988).

Bear, J. and Y. Bachmat 1990 Introduction to modelling of transport phenomena in porous media. Kluwer, Dordrecht.

Bear, J. and A. Verruijt 1987 Modelling groundwater flow and pollution. Reidel, Dordrecht.

Bender, C. M. and S. A. Orszag 1978 Advanced mathematical methods for scientists and engineers. McGraw-Hill, New York.

Benjamin, T. B. 1959 Shearing flow over a wavy boundary. J. Fluid Mech. **6**, 161–205.

Bentley, C. R. 1987 Antarctic ice streams: a review. J. Geophys. Res. **92**, 8843–8858.

Berner, R. A., A. C. Lasaga and R. M. Garrels 1983 The carbonate–silicate geochemical cycle and its effect on atmospheric carbon dioxide over the past 100 million years. Amer. J. Sci. **283**, 641–683.

Best, J. 1992 On the entrainment of sediment and initiation of bed defects: insights from recent developments within turbulent boundary-layer research. Sedimentol. **39**, 797–811.

Birchwood, R. A. and D. L. Turcotte 1994 A unified approach to geopressuring, low-permeability zone formation, and secondary porosity generation in sedimentary basins. J. Geophys. Res. **99**, 20,051–20,058.

Björnsson, H. 1974 Explanation of jökulhlaups from Grímsvötn, Vatnajökull, Iceland. Jökull **24**, 1–26.

Björnsson, H. 1988 Hydrology of ice caps in volcanic regions. Societas Scientarium Islandica, University of Iceland, Reykjavik.

Björnsson, H. 1992 Jökulhlaups in Iceland: prediction, characteristics and simulation. Ann. Glaciol. **16**, 95–106.

Bond, G., and 13 others 1992 Evidence for massive discharges of icebergs into the North Atlantic ocean during the last glacial period. Nature **360**, 245–249.

Bond, G. C., W. Showers, M. Cheseby, R. Lotti, P. Almasi, P. deMenocal, P. Priore, H. Cullen, I. Hajdas and G. Bonani 1997 A pervasive millennial-scale cycle in North Atlantic Holocene and glacial climates. Science **278**, 1,257–1,266.

Bond, G. C., W. Showers, M. Elliott, M. Evans, R. Lotti, I. Hajdas, G. Bonani and S. Johnson 1999 The North Atlantic's 1–2 kyr climate rhythm: relation to Heinrich events, Dansgaard/Oeschger cycles and Little Ice Age. In: Mechanisms of global climate change, eds. P. U. Clark, R. S. Webb and L. D. Keigwin, Geophys. Monog. **112**, A. G. U., Washington, D. C., pp. 35–58.

Boulton, G. S. and R. C. A. Hindmarsh 1987 Sediment deformation beneath glaciers: rheology and geological consequences. J. Geophys. Res. **92**, 9059–9082.

Bretz, J. H. 1923 The Channeled Scablands of the Columbia Plateau. J. Geol. **31**, 617–649.

Bretz, J. H. 1969 The Lake Missoula floods and the Channeled Scabland. J. Geol. **77**, 505–543.

Broecker, W. S., G. Bond and M. Klas 1990 A salt oscillator in the glacial Atlantic? I: the concept. Paleoceanogr. **5**, 469–477.

Buckmaster, J. D. and G. S. S. Ludford 1982 Theory of laminar flames. C. U. P., Cambridge.

Budyko, M. I. 1969 The effect of solar radiation variations on the climate of the Earth. Tellus **21**, 611–619.

Burgers, J. M. 1948 A mathematical model illustrating the theory of turbulence. Adv. Appl. Mech. **1**, 171–199.

Calov, R., A. Ganopolski, V. Petoukhov, M. Claussen and R. Greve 2002 Large-scale instabilites of the Laurentide ice sheet simulated in a fully coupled climate-system model. Geophys. Res. Letts. **29** (24), 69; doi:10.1029/2002GL016078.

Carrier, G. F., M. Krook and C. E. Pearson 1966 Functions of a complex variable. McGraw-Hill, New York.

Chandler, M. A. and L. E. Sohl 2000 Climate forcings and the initiation of low-latitude ice sheets during the Neoproterozoic Varanger glacial interval. J. Geophys. Res. **105** (D10), 20,737–20,756.

Chandrasekhar, S. 1960 Radiative transfer. Dover Publications, New York.

Chapman, M. G., M. T. Gudmundsson, A. J. Russell and T. M. Hare 2003 Possible Juventae Chasma sub-ice volcanic eruptions and Maja Valles ice outburst floods, Mars: implications of Mars Global Surveyor crater densities, geomorphology, and topography. J. Geophys. Res. **108** (E10), art np. 5113.

Chapman, S. 1930 A theory of upper atmospheric ozone. Mem. Roy. Meteorol. Soc. **3**, 103–125.

Chorley, R. J. (ed.) 1969 Introduction to physical hydrology. Methuen, London.

Chow, Ven Te 1959 Open-channel hydraulics. McGraw-Hill, New York.

Christensen, U. R. 1983 Comments on "On the thermal state of the earth's mantle", by A. C. Fowler, J. Geophys. **53**, 201–202.

Chugunov, V. A. and A. V. Wilchinsky 1996 Modelling of a marine glacier and ice-sheet–ice-shelf transition zone based on asymptotic analysis. Ann. Glaciol. **23**, 59–67.

Clarke, G. K. C. 1982 Glacier outburst floods from 'Hazard Lake', Yukon Territory, and the problem of flood magnitude prediction. J. Glaciol. **28**, 3–21.

Clarke, G. K. C. 2003 Hydraulics of subglacial outburst floods: new insights from the Spring-Hutter formulation. J. Glaciol. **49**, 299–313.

Clarke, G. K. C., U. Nitsan and W. S. B. Paterson 1977 Strain heating and creep instability in glaciers and ice sheets. Revs. Geophys. Space Phys. **15**, 235–247.

Clarke, G. K. C., S. G. Collins and D. E. Thompson 1984 Flow, thermal structure, and subglacial conditions of a surge-type glacier. Can. J. Earth Sci. **21**, 232–240.

Clarke, G. K. C., D. W. Leverington, J. T. Teller and A. S. Dyke 2004 Paleohydraulics of the last outburst flood from glacial Lake Agassiz and the 8200 BP cold event. Quat. Sci. Revs. **23**, 389–407.

Clarke, G. K. C., D. W. Leverington, J. T. Teller, A. S. Dyke and S. J. Marshall 2005 Fresh arguments against the Shaw megaflood hypothesis. A reply to comments by David Sharpe on "Paleohydraulics of the last outburst flood from glacial Lake Agassiz and the 8200 BP cold event". Quat. Sci. Revs. **24**, 1,533–1.541.

Clarke, G. K. C. 2005 Subglacial processes Ann. Rev. Earth Planet. Sci. **33**, 247–276.

Colbeck, S..C. ed. 1980 Dynamics of snow and ice masses. Academic Press.

Coleman, N. M. 2003 Aqueous flows carved the outflow channels on Mars. J. Geophys. Res. **108** (E5), art. no. 5039.

Colombini, M., G. Seminara and M. Tubino 1987 Finite-amplitude alternate bars. J. Fluid Mech. **181**, 213–232.

Crank, J. 1984 Free and moving boundary value problems. Clarendon Press, Oxford.

Crowley, T. J. and S. K. Baum 1993 Effect of decreased solar luminosity on late Precambrian ice extent. J. Geophys. Res. **98** (D9), 16,723–16,732.

Cushman, J. H. (ed.) 1990 Dynamics of fluids in hierarchical porous media. Academic Press, London.

Dash, J. G., A. W. Rempel and J. S. Wettlaufer 2006 The physics of premelted ice and its geophysical consequences. Revs. Mod. Phys. **78**, 695–741.

Deeley, R. M. and P. H. Parr 1914 On the Hintereis glacier. Phil. Mag. **27** (6), 153–176.

Defant, A. 1958 Ebb and flow: the tides of Earth, air and water. University of Michigan Press, Ann Arbor, MI.

Denton, G. H. and D. E. Sugden 2005 Meltwater features that suggest Miocene ice-sheet overriding of the Transantarctic Mountains in Victoria Land, Antarctica. Geograf. Ann. **87**A, 67–85.

Dewynne, J. N., A. C. Fowler and P. S. Hagan 1993 Multiple reaction fronts in the oxidation-reduction of iron-rich uranium ores. SIAM J. Appl. Math. **53**, 971–989.

Dold, J. W. 1985 Analysis of the early stage of thermal runaway. Quart. J. Mech. Appl. Math. **38**, 361–387.

Drazin, P. G. and R. S. Johnson 1989 Solitons: an introduction. C. U. P., Cambridge.

Drazin, P. G. and W. H. Reid 1981 Hydrodynamic stability. C. U. P., Cambridge.

Dressler, R. F. 1949 Mathematical solution of the problem of roll waves in inclined open channels. Comm. Pure Appl. Maths. **2**, 149–194.

Dullien, F. A. L. 1979 Porous media: fluid transport and pore structure. Academic Press, New York.

Eberl, D. and J. Hower 1976 Kinetics of illite formation. Geol. Soc. Amer. Bull. **87**, 1326–1330.

Einstein, H. A. 1950 The bedload function for bedload transportation in open channel flows. Tech. Bull. No. 1026, U. S. D. A., Soil Conservation Service, pp. 1–71.

Engelhardt, H., N. Humphrey, B. Kamb and M. Fahnestock 1990 Physical conditions at the base of a fast moving Antarctic ice stream. Science **248**, 57–59.

Engelund, F. 1970 Instability of erodible beds. J. Fluid Mech. **42**, 225–244.

Engelund, F. and J. Fredsøe 1982 Sediment ripples and dunes. Ann. Rev. Fluid Mech. **14**, 13–37.

Erdélyi, A., W. Magnus, F. Oberhettinger and F. G. Tricomi 1953 Higher transcendental functions, Vol. 1. McGraw-Hill, New York.

Erlingsson, U. 2006 Lake Vostok behaves like a 'captured lake' and may be near to creating an Antarctic jökulhlaup. Geogr. Ann. **88**A, 1–7.

Finsterwalder, S. 1907 Die Theorie der Gletscherschwankungen. Z. Gletscherkunde **2**, 81–103.

Fisher, R. A. 1937 The wave of advance of advantageous genes. Ann. Eugenics **7**, 353–369.

Fowkes, N. D. and J. J. Mahony 1994 An introduction to mathematical modelling. John Wiley, Chichester.

Fowler, A. C. 1982 Waves on glaciers. J. Fluid Mech. **120**, 283–321.

Fowler, A. C. 1983 On the thermal state of the earth's mantle. J. Geophys. **53**, 42–51.

Fowler, A. C. 1985 Fast thermoviscous convection. Stud. Appl. Math. **72**, 189–219.

Fowler, A. C. 1986 A sliding law for glaciers of constant viscosity in the presence of subglacial cavitation. Proc. R. Soc. Lond. A**407**, 147–170.

Fowler, A. C. 1987a A theory of glacier surges. J. Geophys. Res. **92**, 9111–9120.

Fowler, A. C. 1987b Sliding with cavity formation. J. Glaciol. **33**, 255–267.

Fowler, A. C. 1989 A mathematical analysis of glacier surges. SIAM J. Appl. Math. **49**, 246–262.

Fowler, A. C. 1992 Modelling ice sheet dynamics. Geophys. Astrophys. Fluid Dyn. **63**, 29–65.

Fowler, A. C. 1993a Towards a description of convection with temperature and pressure dependent viscosity. Stud. Appl. Math. **88**, 113–139.

Fowler, A. C. 1993b Boundary layer theory and subduction. J. Geophys. Res. **98**, 21997–22005.

Fowler, A. C. 1997 Mathematical models in the applied sciences. C. U. P., Cambridge.

Fowler, A. C. and C. Johnson 1995 Hydraulic runaway: a mechanism for thermally regulated surges of ice sheets. J. Glaciol. **41**, 554–561.

Fowler, A. C. and E. Schiavi 1998 A theory of ice sheet surges. J. Glaciol. **44**, 104–118.

Fowler, A. C. and D. A. Larson 1978 On the flow of polythermal glaciers. I. Model and preliminary analysis. Proc. R. Soc. Lond. A**363**, 217–242.

Fowler, A. C. and D. A. Larson 1980a The uniqueness of steady state flows of glaciers and ice sheets. Geophys. J. R. Astr. Soc. **63**, 333–345.

Fowler, A. C. and D. A. Larson 1980b On the flow of polythermal glaciers II. Surface wave analysis. Proc. R. Soc. Lond. A**370**, 155–171.

Fowler, A. C. and X. S. Yang 1998 Fast and slow compaction in sedimentary basins. SIAM J. Appl. Math. **59**, 365–385.

Fowler, A. C. and X. S. Yang 1999 Pressure solution and viscous compaction in sedimentary basins. J. Geophys. Res. **104**, 12,989–12,997.

Fowler, A. C. and X. S. Yang 2003 Dissolution/precipitation mechanisms for diagenesis in sedimentary basins. J. Geophys. Res. **108** (B10), 2509, doi 10.1029/2002JB002269, EPM (13) 1–14.

Fredsøe, J. 1974 On the development of dunes in erodible channels. J. Fluid Mech. **64**, 1–16.

Freed, R. L. and D. R. Peacor 1989 Geopressured shale and sealing effect of smectite to illite transition. AAPG Bull. **73**, 1223–1232.

Freeze, R. A. and J. A. Cherry 1977 Groundwater. Prentice-Hall.

French, R. H. 1994 Open-channel hydraulics. McGraw-Hill, New York.

Ganopolski, A. and S. Rahmstorf 2001 Rapid changes of glacial climate simulated in a coupled climate model. Nature **409**, 153–158.

Ganopolski, A. and S. Rahmstorf 2002 Abrupt glacial climate changes due to stochastic resonance. Phys. Rev. Letts. **88** (3), 038501; doi:10.1103/PhysRevLett.88.038501.

Garcia, M. 1989 Ph. D. thesis, St. Anthony Falls Hydraulic Laboratory, University of Minnesota.

Ghil, M. and S. Childress 1987 Topics in geophysical fluid dynamics. Springer-Verlag.

Gill, A. E. 1982 Atmosphere-ocean dynamics. Academic Press, San Diego.

Goudie, A. 1993 The nature of the environment, 3rd edition. Blackwell, Oxford.

Goodwin, I. D. 1988 The nature and origin of a jökulhlaup near Casey Station, Antarctica. J. Glaciol. **34**, 95–101.

Gradshteyn, I. S. and I. M. Ryzhik 1980 Table of integrals, series and products. Corrected and enlarged edition. Academic Press, New York.

Grindrod, P. 1991 Patterns and waves. O. U. P., Oxford.

Guðmundsson, G. H. 2003 Transmission of basal variability to a glacier surface. J. Geophys. Res. **108**, B5, ETG 9-1–9-19.

Guðmundsson, M. T., F. Sigmundsson and H. Björnsson 1997 Ice-volcano interaction of the 1996 Gjálp subglacial eruption, Vatnajökull, Iceland. Nature **389**, 954–957.

Guðmundsson, M. T., F. Sigmundsson, H. Björnsson and T. Högnadóttir 2004 The 1996 eruption at Gjálp, Vatnajökull ice cap, Iceland: efficiency of heat transfer, ice deformation and subglacial water pressure. Bull. Volcanol. **66**, 46–65.

Haberman, R. 1998 Mathematical models. Society for Industrial and Applied Mathematics, Philadelphia.

Hack, J. T. 1957 Studies of longitudinal profiles in Virginia and Maryland. U. S. G. S. Prof. Paper, 294–B.

Hinch, E. J. 1991 Perturbation methods. C. U. P., Cambridge.

Hagan, P. S., R. S. Polizzotti and G. Luckman 1986 Internal oxidation of binary-alloys. SIAM J. Appl. Math. **45**, 956–971.

Heinrich, H. 1988 Origin and consequences of cyclic ice rafting in the Northeast Atlantic Ocean during the past 130,000 years. Quat. Res. **29**, 142–152.

Henderson-Sellers, A. and P. J. Robinson 1986 Contemporary climatology. Longman.

Hershenov, J. 1976 Solutions of the differential equation $u''' + \lambda^2 z u' + (\alpha - 1)\lambda^2 u = 0$. Stud. Appl. Math. **55**, 301–314.

Hodge, S. M. 1974 Variations in the sliding of a temperate glacier. J. Glaciol. **13**, 349–369.

Hoffman, A. J. Kaufman, G. P. Halverson and D. P. Schrag 1998 A Neoproterozoic snowball Earth. Science **281**, 1,342–1,346.

Hoffman, N. 2000 White Mars: a new model for Mars' surface and atmosphere based on CO_2. Icarus **146**, 326–342.

Holmes, A. 1978 Principles of physical geology. rd edition, revised by Doris Holmes. John Wiley and sons, New York.

Hoppensteadt, F. 1975 Mathematical theories of populations: demographics, genetics and epidemics. Society for Industrial and Applied Mathematics, Philadelphia.

Horton, R. E. 1945 Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. Bull. Geol. Soc. Amer. **56**, 275–370.

Houghton, J. T. 2002 The physics of atmospheres, 3rd ed. C. U. P., Cambridge.

Houghton, J. T. 1994 Global warming. Lion Publishing, Oxford.

Houghton, J. T., L. G. Meira Filho, B. A. Callander, N. Harris, A. Kattenberg and K. Maskell, eds. 1996 Climate change 1995: the science of climate change. C. U. P., Cambridge.

Howard, A. D. 1994 A detachment-limited model of drainage basin evolution. Water Resour. Res. **30**, 2261–2285.

Howard, L. N. 1966 Convection at high Rayleigh number. Proc. 11th Int. Cong. Appl. Mech., ed. H. Görtler, pp. 1109–1115. Springer, Berlin.

Howell, P. D. 1996 Models for thin viscous sheets. Euro. J. Appl. Math. **7**, 321–343.

Howison, S. D. 2005 Practical applied mathematics: modelling, analysis, approximation. C. U. P., Cambridge.

Hunt, J. M. 1990 Generation and migration of petroleum from abnormally pressured fluid compartments. AAPG Bull. **74**, 1–12.

Hutter, K. 1983 Theoretical Glaciology. Reidel, Dordrecht.

Hutter, K., S. Yakowitz and F. Szidarovsky 1986 A numerical study of plane ice sheet flow. J. Glaciol. **32**, 139–160.

Hyde, W. T., T. J. Crowley, S. K. Baum and W. R. Peltier 2000 Neoproterozoic 'snowball Earth' simulations with a coupled climate/ice sheet model. Nature **405**, 425–429.

Ikeda, I. and K. Matsumoto 1987 High-dimensional chaotic behaviour in systems with time-delayed feedback. Physica **29**D, 223–235.

Imbrie, J. and K. P. Imbrie 1979 Ice ages; solving the mystery. Harvard University Press.

Izumi, N. and G. Parker 1995 Inception of channelization and drainage basin formation: upstream-driven theory. J. Fluid Mech. **283**, 341–363.

Izumi, N. and G. Parker 2000 Linear stability analysis of channel inception: downstream-driven theory. J. Fluid Mech. **419**, 239–262.

James, A. 1992 Mathematical modelling of river network evolution. M. Sc. dissertation, Oxford University.

Jarvis, G. T. and D. McKenzie 1982 Mantle convection as a boundary layer phenomenon. Geophys. J. R. Astr. Soc. **68**, 389–427.

Jeffrey, A. 2004 Handbook of mathematical formulas and integrals, 3rd ed. Elsevier, Amsterdam.

Jeffreys, H. and B. Jeffreys 1953 Methods of mathematical physics. C. U. P., Cambridge.

Jimenez, J. and J. A. Zufiria 1987 A boundary layer analysis of Rayleigh-Bénard convection at large Rayleigh number. J. Fluid Mech. **178**, 53–71.

Johnsen, S. J., H. B. Clausen, W. Dansgaard, K. Fuhrer, N. Gundestrup, C. U. Hammer, P. Iversen, J. Jouzel, B. Stauffer and J. P. Steffensen 1992 Irregular glacial interstadials recorded in a new Greenland ice core. Nature **359**, 311–313.

Jones, M., Mechanical principles of sediment deformation. In: *The geological deformation of sediments*, ed. A. Maltman, pp. 37–71, Chapman and Hall, London, 1994.

Julien, P. Y. 1995 Erosion and sedimentation. C. U. P., Cambridge.

Kalnay, E. 2003 Atmospheric modeling, data assimilation and predictability. C. U. P., Cambridge.

Kamb, W. B. 1970 Sliding motion of glaciers: theory and observation. Revs. Geophys. Space Phys. **8**, 673–728.

Kamb, B. *and 7 others* 1985 Glacier surge mechanism: 1982–1983 surge of Variegated Glacier, Alaska. Science **227**, 469–479.

Kamb, B. 1987 Glacier surge mechanism based on linked cavity configuration of the basal water conduit system. J. Geophys. Res. **92**, 9083–9100.

Kardar, M., G. Parisi and Y. C. Zhang 1986 Dynamic scaling of growing interfaces. Phys. Rev. Letts. **56**, 889–892.

Kargel, J. S. 2004 Mars — a warmer, wetter planet. Springer-Verlag, Berlin.

Kasting, J. F. and T. P. Ackermann 1986 Climatic consequences of very high carbon dioxide levels in Earth's early atmosphere. Science **234**, 1,383–1,385.

Kasting, J. F. 1989 Long-term stability of the Earth's climate. Palaeogeog. Palaeoclim. Palaeoecol. **75**, 83–95.

Kaye, G. W. C. and T. H. Laby 1960 Physical and chemical constants, 12th edition. Longman's.

Kemp, D. D. 1994 Global environmental issues, 2nd edition. Routledge.

Kennedy, J. F. 1963 The mechanics of dunes and anti-dunes in erodible-bed channels. J. Fluid Mech. **16**, 521–544.

Kern, R. and A. Weisbrod 1967 Thermodynamics for geologists. Translation from French by D. McKie. Freeman, Cooper and Co., San Francisco.

Kevorkian, J. and J. D. Cole 1981 Perturbation methods in applied mathematics. Springer-Verlag, Berlin.

Kirby, S. H. 1983 Rheology of the lithosphere. Revs. Geophys. Space Phys. **21**, 1458–1487.

Korteweg, D. J. and G. de Vries 1895 On the change of form of long waves advancing in a rectangular canal, and on a new type of long stationary waves. Phil. Mag. (5) **39**, 422–443.

Kramer, S. and M. Marder 1992 Evolution of river networks. Phys. Rev. Letts. **68**, 205–208.

Krauskopf, K. B. and D. K. Bird 1995 Introduction to geochemistry. McGraw-Hill, New York.

Lakin, W. D., B. S. Ng and W. H. Reid 1978 Approximations to the eigenvalue relation for the Orr-Sommerfeld problem. Phil. Trans. Roy. Soc. **289**, 347–371.

Lamb, Sir Horace 1945 Hydrodynamics, sixth ed. Dover reprint of the 1932 sixth edition. Dover Publications, New York.

Lambe, T. W. and R. V. Whitman 1979 Soil mechanics, SI version. John Wiley, New York.

Lang, C., M. Leuenberger, J. Schwander and S. Johnsen 1999 16° C rapid temperature variation in central Greenland 70,000 years ago. Science **286**, 934–937.

Lasaga, A. C., R. A. Berner and R. M. Garrels 1985 An improved geochemical model of atmospheric CO_2 fluctuations over the past 100 million years. In: The carbon cycle and atmospheric CO_2: natural variations Archaen to present, eds. E. T. Sundquist and W. S. Broecker, A. G. U., Washington, D. C., pp. 397–411.

Leuenberger, M. C., C. Lang and J. Schwander 1999 Delta ^{15}N measurements as a calibration tool for the paleothermometer and gas–ice age differences: A case study for the 8200 B. P. event on GRIP ice. J. Geophys. Res. **104** (D18), 22,163–22,170.

Lin, C. C. and L. A. Segel 1974 Mathematics applied to deterministic problems in the natural sciences. MacMillan, New York.

Liou, K. N. 2002 An introduction to atmospheric radiation, 2nd ed. Academic Press, San Diego.

Lliboutry, L. A. 1965 Traité de Glaciologie, Masson, Paris.

Lliboutry, L. A. 1968 General theory of subglacial cavitation and sliding of temperate glaciers. J. Glaciol. **7**, 21–58.

Lliboutry, L. A. 1979 Local friction laws for glaciers: a critical review and new openings. J. Glaciol. **23**, 67–95.

Loewenherz, D. S. 1991 Stability and the initiation of channelized surface drainage: a reassessment of the short wavelength limit. J. Geophys. Res. **96**, 8453–8464.

Loewenherz-Lawrence, D. S. 1994 Hydrodynamic description for advective sediment transport processes and rill initiation. Water Resour. Res. **30**, 3203–3212.

MacAyeal, D. R. 1993 Binge/purge oscillations of the Laurentide ice sheet as a cause of the North Atlantic's Heinrich events. Paleoceanography **8**, 775–784.

Manabe, S. and R. J. Stouffer 1995 Simulation of abrupt climate change induced by freshwater input to the North Atlantic Ocean. Nature **378**, 165–167.

Marsh, S. P. and M. E. Glicksman 1996 Overview of geometric effects on coarsening of mushy zones. Metall. Mater. Trans. bf 27A, 557–567.

Marshall, H. G., J. C. G. Walker and W. R. Kuhn 1988 Long term climate change and the geochemical cycle of carbon. J. Geophys. Res. **93**, 791–801.

Mason, B. and C. B. Moore 1982 Principles of geochemistry, 4th ed. John Wiley, Chichester.

Massey, B. S. 1986 Measures in science and engineering. Ellis Horwood, Chichester.

Matuszkiewicz, A., J. C. Flamand and J. A. Bouré 1987 The bubble-slug flow pattern transition and instabilities of void-fraction waves. Int. J. Multiphase Flow **13**, 199–217.

Mayer, K. U., S. G. Benner, E. O. Frind, S. F. Thornton and D. N. Lerner 2001 Reactive transport modeling of processes controlling the distribution and natural attenuation of phenolic compounds in a deep sandstone aquifer. J. Contam. Hydrol. **53**, 341–368.

McBirney, A. R. 1984 Igneous petrology. Freeman, Cooper and Co., San Francisco.

Meyer-Peter, E. and R. Müller 1948. Formulas for bed-load transport. Proc. Int. Assoc. Hydraul. Res., 3rd annual conference, Stockholm, pp. 39–64.

Moore, D. R. and N. O. Weiss 1973 Two-dimensional Rayleigh-Bénard convection. J. Fluid Mech. **58**, 289–312.

Morland, L. W. 1984 Thermo-mechanical balances of ice sheet flow. Geophys. Astrophys. Fluid Dyn. **29**, 237–266.

Morris, S. and D. Canright 1984 A boundary layer analysis of Bénard convection in a fluid of strongly temperature-dependent viscosity. Phys. Earth Planet. Int. **29**, 320–329.

Murray, J. D. 2002 Mathematical biology. 2 volumes, Springer–Verlag, Berlin.

Nataf, H.-C. and F. M. Richter 1982 Convection experiments in fluids with highly temperature-dependent viscosity and the thermal evolution of the planets. Phys. Earth Planet. Int. **29**, 320–329.

Newell, A. C. 1985 Solitons in mathematics and physics. Society for Industrial and Applied Mathematics, Philadelphia.

Nicolussi, K. 1990 Bilddokumente zur Geschichte des Vernagtferners im 17. Jahrhundert. Zeit. Gletscherkunde und Glazialgeologie Bd **26**, Heft 2, 97–119.

Noble, B. 1988 Methods based on the Wiener-Hopf technique, 2nd (unaltered) edition. Chelsea Publishing Company, New York.

North, G. R. 1975a Analytical solution to a simple climate model with diffusive heat transport. J. Atmos. Sci. **32**, 1301–1307.

North, G. R. 1975b Theory of energy-balance climate models. J. Atmos. Sci. **32**, 2033–2043.

North, G. R., J. G. Mengel and D. A. Short 1983 Simple energy balance model resolving the season and continents: applications to astronomical theory of ice ages. J. Geophys. Res. **88**, 6576–6586.

Nye, J. F. 1953 The flow law of ice from measurements in glacier tunnels, laboratory experiments and the Jungfraufirn borehole experiment. Proc. R. Soc. Lond. A**219**, 477–489.

Nye, J. F. 1960 The response of glaciers and ice sheets to seasonal and climatic changes. Proc. R. Soc. Lond. A**256**, 559–584.

Nye, J. F. 1969 A calculation on the sliding of ice over a wavy surface using a Newtonian viscous approximation. Proc. R. Soc. Lond. A**311**, 445–477.

Nye, J. F. 1970 Glacier sliding without cavitation in a linear viscous approximation. Proc. R. Soc. Lond. A**315**, 381–403.

Nye, J. F. 1976 Water flow in glaciers: jökulhlaups, tunnels, and veins. J. Glaciol. **17**, 181–207.

Olson, P. and G. M. Corcos 1980 A boundary layer model for mantle convection with surface plates. Geophys. J. R. Astr. Soc. **62**, 195–219.

Olver, F. W. J. 1974 Asymptotics and special functions. Academic Press, New York.

O'Malley, K., A. D. Fitt, T. V. Jones, J. R. Ockendon and P. Wilmott 1991 Models for high-Reynolds-number flow down a step. J. Fluid Mech. **222**, 139–155.

Ortoleva, P., 1994. Geochemical self-organisation. O. U. P., Oxford.

Parker, G. 1975 Sediment inertia as a cause of river antidunes. J. Hydraul. Div. ASCE **101**, 211–221.

Parker, G. 1978 Self-formed straight rivers with equilibrium banks and mobile bed. Part 1. The sand-silt river. J. Fluid Mech. **89**, 109–125.

Parsons, D. R., I. J. Walker and G. F. S. Wiggs 2004 Numerical modelling of flow structures over idealized transverse aeolian dunes of varying geometry. Geomorphology **59**, 149–164.

Paterson, W. S. B. 1994 The physics of glaciers, 3rd edition. Pergamon, Oxford.

Pedlosky, J. 1979 Geophysical fluid dynamics. Springer-Verlag, Berlin.

Petit, J. R., J. Jouzel, D. Raynaud, N. I. Barkov, J.-M. Barnola, I. Basile, M. Bender, J. Chappellaz, M. Davis, G. Delaygue, M. Delmotte, V. M. Kotlyakov, M. Legrand, V. Y. Lipenkov, C. Lorius, L. Pépin, C. Ritz, E. Saltzman and M. Stievenard 1999 Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. Nature **399**, 429–436.

Pierrehumbert, R. T. 2004 High atmospheric carbon dioxide necessary for the termination of global glaciation. Nature **429**, 646–648.

Pillow, A. F. 1952 The free convection cell in two dimensions. Dept. of Supply, Aeronautical Research Laboratories, Report A.79 [Dept. of Mathematics, University of Queensland, St Lucia, Queensland 4067, Australia].

Polubarinova-Kochina, P. Ya. 1962 Theory of ground water movement. Princeton University Press, Princeton.

Price, M. 1985 Introducing groundwater. George Allen and Unwin.

Prosperetti, A. and J. V. Satrape 1990 Stability of two-phase flow models. In: Two-phase flow models and waves, eds. D. D. Joseph and D. G. Schaeffer, Springer, New York, pp. 98–117.

Pugh, D. T. 1987 Tides, surges and mean sea-level. John Wiley, Chichester.

Quareni, F., D. A. Yuen, G. Sewell and U. R. Christensen 1985 High Rayleigh number convection with strongly variable viscosity: a comparison between mean field and two-dimensional solutions. J. Geophys. Res. **90**, 12633–12644.

Rahmstorf, S. 1995 Bifurcations of the atlantic thermohaline circulation in response to changes in the hydrological cycle. Nature **378** 145–149.

Rahmstorf, S. 2002 Ocean circulation and climate during the past 120,000 years Nature **419**, 207–214.

Reid, W. H. 1972 Composite approximations to the solutions of the Orr-Sommerfeld equation. Stud. Appl. Math. **51**, 341–368.

Rempel, A. W., J. S. Wettlaufer and M. G. Worster 2004 Premelting dynamics in a continuum model of frost heave. J. Fluid Mech. **498**, 227–244.

Reynolds, A. J. 1965 Waves on the erodible bed of an open channel. J. Fluid Mech. **22**, 113–133.

Richards, K. 1982 Rivers: form and process in alluvial channels. Methuen, London.

Richards, K. J. 1980 The formation of ripples and dunes on an erodible bed. J. Fluid Mech. **99**, 597–618.

Roberts, G. O. 1977 Fast viscous convection. Geophys. Astrophys. Fluid Dynamics **8**, 197–233.

Roberts, G. O. 1979 Fast viscous Bénard convection. Geophys. Astrophys. Fluid Dynamics **12**, 235–272.

Roberts, M. J. 2005 Jökulhlaups: a reassessment of floodwater flow through glaciers. Revs. Geophys. **43**, RG1002.

Robin, G. de Q. 1955 Ice movement and temperature distribution in glaciers and ice sheets. J. Glaciol. **2**, 523–532.

Robinson, J. R. 1967 Finite amplitude convection cells. J. Fluid Mech. **30**, 577–600.

Rodríguez-Iturbe, I. and A. Rinaldo 1997 Fractal river basins. C. U. P., Cambridge.

Röthlisberger, H. 1972 Water pressure in intra- and subglacial channels. J. Glaciol. **11**, 177–203.

Ruddiman, W. F. 2001 Earth's climate: past and future. W. H. Freeman, New York.

Sahimi, M. 1995 Flow and transport in porous media and fractured rock. VCH, Weinheim, Germany.

Samarskii, A. A., V. A. Galaktionov, S. P. Kurdyumov and A. P. Mikhailov 1995 Blow-up in quasilinear parabolic equations. de Gruyter Expositions in Mathematics, vol. 19. de Gruyter, Berlin.

Sass, B. M., P. E. Rosenberg, and J. A. Kittrick 1987 The stability of illite/smectite during diagenesis: an experimental study. Geochim. Cosmochim. Acta **51**, 2103–2115.

Schlichting, H. 1979 Boundary layer theory. McGraw-Hill, New York.

Schoof, C. 2005 The effct of cavitation on glacier sliding. Proc. R. Soc. Lond. A**461**, 609–627.

Selby, M. J. 1993 Hillslope materials and processes, 2nd ed. O. U. P., Oxford.

Sellers, W. D. 1969 A climate model based on the energy balance of the earth-atmosphere system. J. Appl. Meteorol. **8**, 392–400.

Sellmeijer, J. B. and M. A. Koenders 1991 A mathematical model for piping. Appl. Math. Modell. **15**, 646–651.

Seminara, G., M. Colombini and G. Parker 1996 Nearly pure sorting waves and formation of bedload sheets. J. Fluid Mech. **312**, 253–278.

Severinghaus, J. P. and E. J. Brook 1999 Abrupt climate change at the end of the last glacial period inferred from trapped air in polar ice. Science **286**, 930–934.

Sharpe, D. 2005 Comments on: “Paleohydraulics of the last outburst flood from glacial Lake Agassiz and the 8200 BP cold event” by Clarke et al. [Quaternary Science Reviews **23** (2004), 389–407]. Quat. Sci. Revs. **24, 1,529–1.532.**

Shaw, J. 1983 Drumlin formation related to inverted meltwater erosional marks. J. Glaciol. **29**, 461–479.

Shaw, J., D. Kvill and B. Rains 1989 Drumlins and catastrophic subglacial floods. Sed. Geol. **62**, 177–202.

Shields, A. 1936 Anwendung der Ähnlichkeits mechanik und der Turbulenzforschung auf die Geschiebebewegung. Mitteilung der Preussischen Versuchanstalt für Wasserbau und Schiffbau, Heft 26, Berlin.

Siegert, M. J. 2005 Lakes beneath the ice sheet: the occurrence, analysis, and future exploration of Lake Vostok and other Antarctic subglacial lakes. Ann. Rev. Earth Planet. Sci. **33**, 215–245.

Siegert, M. J., J. A. Dowdeswell, M. R. Gorman and N. F. McIntyre 1996 An inventory of Antarctic sub-glacial lakes. Antarc. Sci. **8**, 281–286.

Siegert, M. J., J. C. Ellis-Evans, M. Tranter, C. Mayer, J.-R. Petit, A. Salamatin and J. C. Priscu 2001 Physical, chemical and biological processes in Lake Vostok and other Antarctic subglacial lakes. Nature **414**, 603–608.

Smith, J. D. 1970 Stability of a sand bed subjected to a shear flow of low Froude number. J. Geophys. Res. **75**, 5928–5940.

Smith, J. D. and S. R. McLean 1977 Spatially averaged flow over a wavy surface. J. Geophys. Res. **83**, 1735–1745.

Smith, J. E. 1971 The dynamics of shale compaction and evolution in pore-fluid pressures. Math. Geol. **3**, 239–263.

Smith, T. R., B. Birnir and G. E. Merchant 1997a Towards an elementary theory of drainage basin evolution: I. The theoretical basis. Comput. Geosci. **23**, 811–822.

Smith, T. R., B. Birnir and G. E. Merchant 1997b Towards an elementary theory of drainage basin evolution: II. A computational evaluation. Comput. Geosci. **23**, 823–849.

Smith, T. R. and F. P. Bretherton 1972 Stability and the conservation of mass in drainage basin evolution. Water Resour. Res. **8**, 1506–1529.

Southard, J. B. 1991 Experimental determination of bedform stability. Ann. Rev. Earth Planet. Sci. **19**, 423–455.

Sparrow, E. M. and R. D. Cess 1978 Radiation heat transfer. Hemisphere, Belmont, CA.

Spring, U. and K. Hutter 1981 Numerical studies of jökulhlaups. Cold Reg. Sci. Technol. **4**, 227–244.

Spring, U. and K. Hutter 1982 Conduit flow of a fluid through its solid phase and its application to intraglacial channel flow. Int. J. Engng. Sci. **20**, 327–363.

Stocker, T. F. and S. J. Johnsen 2003 A minimum thermodynamic model for the bipolar seesaw. Paleoceanography **18**, art. no. 1087, 11-1—11-9.

Stocker, T. F. and D. G. Wright 1991 Rapid transitions of the ocean's deep circulation induced by changes in surface water fluxes. Nature **351**, 729–732.

Stommel, H. 1961 Thermohaline convection with two stable régimes of flow. Tellus **13**, 224–230.

Stoker, J. J. 1957 Water waves: the mathematical theory with applications. Interscience, New York.

Strahler, A. N. 1952 Hypsometric (area altitude) analysis of erosional topography. Geol. Soc. Amer. Bull. **63**, 1117–1142.

Strutt, J. W. (Lord Rayleigh) 1871 On the light from the sky, its polarization and colour. Phil. Mag. **41**, 107–120.

Sugden, D. and G. Denton 2004 Cenozoic landscape evolution of the Convoy Range to Mackay Glacier area, Transantarctic Mountains: onshorew to offshore synthesis. GSA Bull. **116**, 840–857.

Sumer, B. M. and M. Bakioglu 1984 On the formation of ripples on an erodible bed. J. Fluid Mech. **144**, 177–190.

Tayler, A. B. 1986 Mathematical models in applied mechanics. Clarendon Press, Oxford.

Taylor, G. I. 1953 Dispersion of soluble matter in a solvent flowing slowly through a tube. Proc. R. Soc. Lond. A**219**, 186–203.

Taylor, K. C., G. W. Lamorey, G. A. Doyle, R. B. Alley, P. M. Grootes, P. A. Mayewski, J. W. C. White and L. K. Barlow 1993 The 'flickering switch' of late Pleistocene climate change. Nature **361**, 432–436.

Taylor, K. C., P. A. Mayewski, R. B. Alley, E. J. Brook, A. J. Gow, P. M. Grootes, D. A. Meese, E. S. Saltzman, J. P. Severinghaus, M. S. Twickler, J. W. C. White, S. Whitlow and G. A. Zielinski 1997 The Holocene–Younger Dryas transition recorded at Summit, Greenland. Science **278**, 825–827.

Teichman, J. and L. Mahadevan 2003 The viscous catenary. J. Fluid Mech. **478**, 71–80.

Thomas, G. E. and K. Stamnes 1999 Radiativew transfer in the atmosphere and ocean. C. U. P., Cambridge.

Titchmarsh, E. C. 1939 The theory of functions, 2nd ed. O. U. P., Oxford.

Tozer, D. C. 1972 The present thermal state of the terrestrial planets. Phys. Earth Planet. Int. **6**, 182–197.

Tricker, R. A. R. 1965 Bores, breakers, waves and wakes. Elsevier, New York.

Tucker, G. E. and R. L. Slingerland 1994 Erosional dynamics, flexural isostasy, and long-lived escarpments: a numerical modeling study. J. Geophys. Res. **99**, 12,229–12,243.

Turcotte, D. L. and E. R. Oxburgh 1967 Finite amplitude convection cells and continental drift. J. Fluid Mech. **28**, 29–42.

Turcotte, D. L., W. F. Haxby and J. R. Ockendon 1977 Lithospheric instabilities. In Island Arcs, Deep Sea Trenches, and Back Arc Basins, ed. M. Talwani and W. C. Pitman III, A. G. U., Washington, D. C., pp. 63–69.

Turcotte, D. L. 1992 Fractals and chaos in geology and geophysics. C. U. P., Cambridge.

Van den Berg, J. H. and A. van Gelder 1993 A new bedform stability diagram, with emphasis on the transition of ripples to plane bed in flows over fine sand and silt. In: Alluvial sedimentation, eds. M. Marzo and C. Puigdefábregas, Spec. Publ. Int. Assoc. Sediment. **17**, pp. 11–21, Blackwell, Oxford.

Van Dyke, M. D. 1975 Perturbation methods in fluid mechanics. Parabolic Press, Stanford, CA.

Van Rijn, L. C. 1984 Sediment transport. Part II. Suspended road transport. J. Hydraul. Engng. **110**, 1613–1641.

Vosper, S. B., S. D. Mobbs and B. A. Gardiner 2002 Measurements of the near-surface flow over a hill. Quart. J. Meteorol. Soc. **128**, 2257–2280.

Waddington, E. D. 1986 Wave ogives. J. Glaciol. **32**, 325–334.

Waitt, jr, R. B. 1984 Periodic jökulhlaups from Pleistocene Glacial Lake Missoula—new evidence from varved sediment in Northern Idaho and Washington. Quat. Res. **22**, 46–58.

Walker, J. C. G., P. B. Hays and J. F. Kasting 1981 A negative feedback mechanism for the long-term stabilization of Earth's surface temperature. J. Geophys. Res. **86** (C10), 9,776–9,782.

Wakes, S. J., S. Hibberd and C. J. Baker 1993 Growth of a sand wave. In: Free boundary problems in fluid flow with applications, eds. J. M. Chadam and H. Rasmussen. Pitman Research Notes, Vol. 282, pp. 33–42. Longman Group, Harlow, Essex, England.

Ward, R. C. and M. Robinson 1990 Principles of hydrology. McGraw-Hill, New York.

Watson, G. N. 1944 A treatise on the theory of Bessel functions, 2nd ed. C. U. P., Cambridge.

Weertman, J. 1957 On the sliding of glaciers. J. Glaciol. **3**, 33–38.

Weertman, J. 1979 The unsolved general glacier sliding problem. J. Glaciol. **23**, 97–115.

Wesseling, P. 1969 Laminar convection cells at high Rayleigh number. J. Fluid Mech. **36**, 625–637.

Wettlaufer, J. S. and M. G. Worster 2006 Premelting dynamics. Ann. Rev. Fluid Mech. **38**, 427–452.

Whitham, G. B. 1974 Linear and nonlinear waves. John Wiley, New York.

Wilchinsky, A. V. and V. A. Chugunov 2000 Ice-stream–ice-shelf transition: theoretical analysis of two-dimensional flow. Ann. Glaciol. **30**, 153–162.

Wilchinsky, A. V. and V. A. Chugunov 2001 Modelling ice flow in various glacier zones. J. Appl. Math. Mech. **65**, 479–493. In Russian: Prikl. Mat. Mekh. **65**, 495–510.

Willgoose, G. 2005 Mathematical modeling of whole landscape evolution. Ann. Rev. Earth Sci. **33**, 443–459.

Willgoose, G., R. L. Bras and I. Rodríguez-Iturbe 1991 A coupled channel network growth and hillslope evolution model: I. Theory. Water Resour. Res. **27**, 1671–1684.

Wingham, D. J., M. J. Siegert, A. Shepherd and A. S. Muir 2006 Rapid discharge connects Antarctic subglacial lakes. Nature **440**, 1,033–1,037.

Winstanley, H. 2001 The formation of river networks. M. Sc. dissertation, Oxford University.

Yang, X. S. 2000 Nonlinear viscoelastic compaction in sedimentary basins. Nonlin. Process. Geophys. **7**, 1–7.