1 Fundamentals of the Earth surface system

If God had consulted me before embarking on the Creation,
I would have suggested something simpler.
Alfonso of Castille (15th century)

Chapter summary

The Earth surface system is dominated by the interaction of climate and topography, the energy for which comes from solar radiation and from the Earth's internal convection. Solar radiation drives a global hydrological cycle which stabilizes the climatic zones of the Earth and makes the planet habitable for humans. Processes deep within the interior of the Earth are responsible for the relative motion of lithospheric plates whose interaction generates much of the surface topography.

The solar radiation received by the Earth has a short wavelength which enables most of it to penetrate the atmosphere to reach the Earth's surface. However, the radiation emitted back from the Earth is of a longer wavelength, a substantial proportion of which is trapped within the atmosphere, a phenomenon known as the greenhouse effect. The greenhouse effect is sensitive to concentrations of certain gases in the atmosphere such as carbon dioxide.

The global climate system is mediated by the transfer of water between the reservoirs in the oceans, atmosphere and the upper part and surface of the lithosphere. The abundance of water on Earth and the coexistence of water in gaseous, liquid and solid states is crucial to the efficient stabilization of the climate system. This is a unique situation in the solar system. Each of the stores in the global hydrological cycle involves particular residence times of water molecules, and characteristic fluxes between them. The oceans have long residence times, and polar ice caps even longer. This buffers the climate system against rapid change, but also causes sudden perturbations to have medium- to long-term effects that are not easily predicted.

Heat is transferred through hydrological processes by atmospheric motions, wind drift of the surface layer of the oceans, and by deeper, slow oceanic circulation. These atmospheric and oceanic systems work synergetically and interactively. The main driving forces in this complex global system are intense convection in the tropical atmosphere over the high sea surface temperatures of warm pools, in the polar ocean where cold, dense water sinks, and in the subtropical ocean where evaporation causes the sinking of saline surface waters.

The imbalances between precipitation and evapotranspiration result in a surface runoff on the continents which is crucial for the denudation of the land surface and the transfer of particulate and dissolved matter into the oceans. Runoff reflects the interplay between topography and climate, with a rough latitudinal zonation disturbed by topographic effects and the influences of the distribution of land and sea.

The mass of the biosphere is minute within the Earth surface system yet its impact is profound. Chemicals pass through the pathways of the biosphere in biogeochemical cycles. The flux of nutrients such as nitrogen and phosphorus is critical in explaining the activity of the biosphere. The carbon cycle is driven by the forces of respiration and photosynthesis. Although the mass of carbon (in the form of CO_2) in the atmosphere is small compared to that of the ocean, it exchanges carbon very rapidly with ocean water, the

soil and with terrestrial biota. Carbon dioxide is also a greenhouse gas, so plays a major feedback role in global temperatures and precipitation.

The topography of the Earth's surface is the result of the interaction of primary tectonic mechanisms and the modifying agents of erosion. Tectonic and erosive processes act at a variety of spatial scales. Topography is produced by the tectonic processes leading to heterogeneities in the lithosphere related to its density and/or thickness which result primarily from the horizontal motions of plates. Isostasy is the principle explaining how the Earth compensates for these heterogeneities. Topography is also produced by the response to processes involving viscous flow in the mantle, such as convection and mantle plume activity. This is termed dynamic topography.

Application of isostasy to sites of continental collision where the lithosphere is substantially thickened demonstrates that at certain wavelengths of topographic load, such as that of the Himalayan mountain chain, the lithosphere responds by a broad-wavelength flexure, causing sedimentary basins to form. At longer wavelengths of topographic load, such as the Tibetan plateau, a vertical, local, isostatic balance can be assumed. The topography at sites of extension can be explained by the isostatic effect of lithospheric thinning, or by the effect of the impingement on the base of the lithosphere of a hot plume of mantle material, as in East Africa. Part of the subsidence following extension can be modelled as similar to the cooling of newly formed oceanic lithosphere at mid-ocean ridges.

Dynamic topography may result from a number of mechanisms involving the dynamic effect of the mantle. Continental and oceanic surfaces are known to be uplifted over mantle plumes (hotspots) and their tracks. More speculatively, some broad sags may be related to mantle downwelling. Dynamic effects causing far-field tilting of continental plates may result from the onset and evolution of subduction of a cool plate beneath the overriding continent. Any signal produced by this mechanism is, however, difficult to decipher from other processes producing topography. Dynamic effects due to the insulation of supercontinental assemblies of plates in the geological past may be important in explaining long-term cycles of continental aggregation and supercontinent dispersal.

1.1 Introduction

The Earth's surface is constantly changing. The energy for this change comes from a number of sources. At the largest scale, the Earth has a planetary energy because of its position within a solar system powered by gravitational forces. It rotates on its axis as it moves around the Sun, the water in its oceans is affected by the gravitational pull of the Moon and to a lesser extent the Sun, and it is bathed in solar radiation. This solar radiation drives atmospheric circulation, controls the Earth's water budget and provides the energy for biological activity. These energy sources are all external. But the Earth's surface is also affected by processes from within. Heat generated in the interior of the Earth, largely from radioactive decay, drives a slow, deep convection. Thermal convection in turn is responsible for the very long-wavelength topography of the Earth's surface, and indirectly for the relative motion of a number of rigid plates across the Earth's surface. These plates, which comprise the relatively cool lithosphere, collide and override each other, forming ocean trenches, island arcs, mountains and plateaux. Where they separate, mid-ocean ridges, oceanic basins, continental margins and continental rifts are formed. Relative plate motion also generates earthquakes and volcanoes. This internal energy is therefore primarily responsible for the Earth's topography at various scales, providing potential energy for a host of denudational, transport and depositional processes.

The processes operating on the surface of the Earth (or in its seas) can be divided into those that are due to processes originating externally to the Earth (*exogenic*) and those caused by processes within the Earth (*endogenic*). Exogenic processes therefore include river, wind and glacial action on land, and tides, currents and waves in the ocean. Endogenic processes include volcanic activity and earthquakes, and horizontal and vertical motions of the Earth's surface caused by plate tectonics and mantle convection.

Solar radiation heats the atmosphere and surface of the Earth, resulting in a global average surface temperature of 288 K (15°C). Solar radiation drives the *hydrological cycle*, representing the continuous exchange of water in its solid, liquid and gaseous states between the atmosphere, oceans and land surface. The portion of the total incoming solar radiation reaching the Earth's surface as light supplies the energy for

photosynthesis of plants and is therefore critical for the operation of the biosphere. The net excess of solar radiation in low latitudes and net deficit in higher latitudes results in a poleward energy transfer. This takes the form of a general atmospheric and oceanic circulation that controls and stabilizes the climatic zones on the Earth. The interaction of climate, topography and geology controls the weathering, transport and deposition of sediment on the Earth's surface (Chapter 3).

1.2 The Earth's energy balance

The primary energy source for the Earth is solar radiation, being 99.98% of all energy received (heat flow from the interior of the Earth accounts for 0.018% and tidal energy 0.002% [1]).

The energy radiated from a body such as the Sun or the Earth is proportional to its radiating temperature. The Sun's energy output includes radiation in the visible spectrum (*c.* 50%, 0.4–0.7 μm in wavelength), a great deal at longer wavelengths (nearly 50%, infrared, 0.7–1000 μm wavelength), and a small proportion of about 1% at short wavelengths (ultraviolet, 0.1–0.4 μm) (Fig. 1.1). The relation between the relative energy emitted and the wavelength is known as a *Planck curve*. Both the total energy and the wavelength of the maximum emission are determined by the radiating temperature. The total energy is given by the *Stefan–Boltzmann law*

$$E = \sigma T^4 \quad \textbf{(1.1)}$$

where σ is the Stefan–Boltzmann constant (5.5597×10^{-8} W m^{-2} K^{-4}) and T is temperature. The peak emission is given by *Wien's law*, which states that the higher the radiating temperature, the shorter the wavelength of the maximum emission. Since the Sun has a surface temperature of 6000 K, the peak emission is 2897/T = 48 μm, within the visible part of the spectrum. However, the Earth is much cooler than the Sun, with an average surface temperature of 288 K (15°C). Consequently, it radiates with a longer peak emission wavelength of about 10 μm—well into the infrared part of the spectrum. This marked difference between the incoming solar radiation and the outgoing radiation from Earth has a profound effect on the climate system.

The total energy emitted by the Sun is obviously not the same as the total solar radiation received by the Earth. The energy flux (in watts per square metre) received at the orbit of the Earth depends inversely on the distance between the two bodies. Consequently, if the Stefan–Boltzmann law (equation (1.1)) gives

$$E_{sun} = (5.559 \times 10^{-8})\,(6000)^4 = 7.20 \times 10^7 \text{ W m}^{-2} \quad \textbf{(1.2)}$$

then the *solar constant* is

$$S = E_{sun}\left(\frac{R_{sun}}{R_{orbit}}\right)^2 = 1557 \text{ W m}^{-2} \quad \textbf{(1.3)}$$

where R_{sun} is the radius of the Sun, and R_{orbit} is the radius from the Sun to the orbit of the Earth. This gives a maximum envelope for the solar radiation received at the orbit of the Earth, making the assumption that the sun is a so-called *blackbody radiator*, or perfect emitter (Fig. 1.1). The energy received by the Earth now depends simply on the size of the Earth and the amount reflected back into space. If the radius of the Earth is R and the reflectivity of the Earth is A, otherwise known as *albedo* (typically about 0.3), the total solar energy received by Earth is

$$E_{in} = \pi R^2 S(1-A) \quad \textbf{(1.4)}$$

The Earth can also be treated ideally as a blackbody radiator. The total energy emitted from the entire surface area of the spherical Earth is then

$$E_{out} = (4\pi R^2)\,(\sigma T^4) \quad \textbf{(1.5)}$$

The energy balance between incoming solar radiation and outgoing radiation is therefore

$$\pi R^2 S(1-A) = 4\pi R^2 (\sigma T^4)$$

$$\frac{S(1-A)}{4} = \sigma T^4 \quad \textbf{(1.6)}$$

We can now substitute into equation (1.6) reasonable values for the solar constant (1295 W m^{-2}) and albedo (0.3) in order to calculate the average surface temperature of the Earth,

$$\frac{1295 \times 0.7}{4(5.5597 \times 10^{-8})} = 265 \text{ K}$$

which is 23 K below the known average surface temperature of the planet. There must be another mechanism or set of mechanisms that are warming the Earth's surface. What has been neglected from the analysis is the blanketing effect of the Earth's atmosphere. We need to pause here and consider in a little more detail the fate of incoming solar radiation impinging on the outer surface of the atmosphere (Fig. 1.2).

Molecules which are common in the atmosphere, such as H_2O and CO_2 absorb radiation at particular wavelengths. Water vapour absorbs strongly at 12 μm, and CO_2 absorbs strongly at 15 μm. Ozone (O_3) and

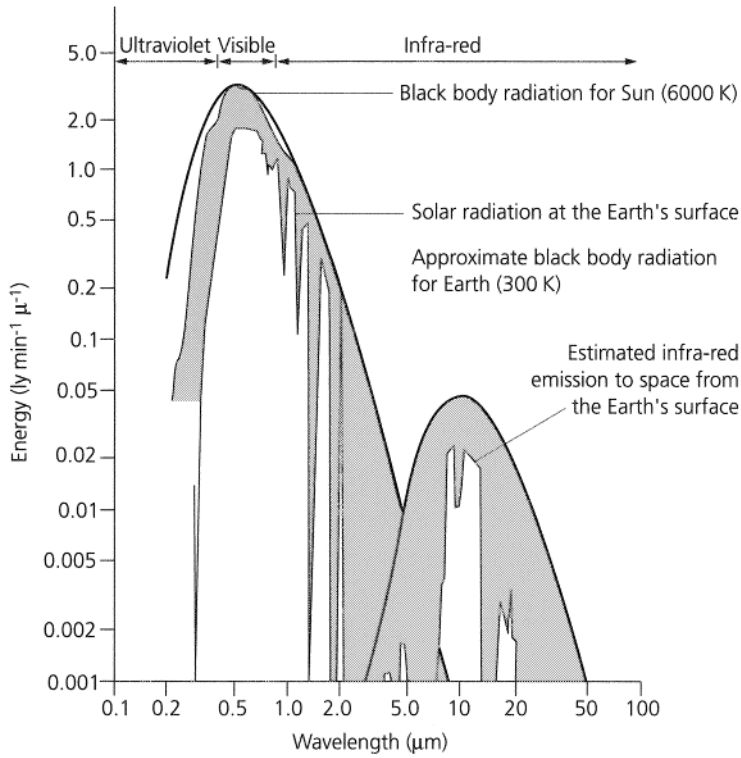


Fig. 1.1 Electromagnetic spectra for solar and terrestrial radiation (Planck curves) showing blackbody emittance as envelopes. After Sellers (1965) [2].

O_2 absorb radiation at shorter wavelengths. Much of the short-wavelength (<4 μm) incoming solar radiation therefore penetrates the atmosphere and reaches the ground as light. However, most of the shorter-wavelength radiation (ultraviolet, <0.4 μm) is absorbed in the upper atmosphere by O_3 and O_2 thereby protecting life on the surface of the Earth from its harmful effects. The bulk of the small amount of incoming longer-wavelength radiation (>4 μm) is absorbed by water vapour and CO_2 in the atmosphere. Consequently, only 51% of solar radiation reaches the Earth's surface. The roles of backscattering, absorption and reflection in the atmosphere are shown in Fig. 1.2.

The radiation absorbed in the atmosphere and received by the surface of the Earth is emitted back into the atmosphere as a long-wavelength (>4 μm, with a maximum in the infrared at 10 μm) radiation. This long-wavelength radiation is particularly susceptible to being absorbed by water vapour and CO_2 in the atmosphere. This property of allowing short-wavelength radiation from the Sun through the atmosphere to reach the Earth, but absorbing and retaining the outgoing long-wavelength radiation, is termed the *atmospheric greenhouse effect*. Apart from radiation, the Earth also gives off a heat flux through conduction from its hot surface (sensible heat), and in the form of latent heat flux through the processes of evaporation and condensation (Section 1.3).

The energy balance can be rewritten to take into account the opacity to outgoing radiation or *emissivity*, e, of the Earth's atmosphere

$$e\sigma T^4 \frac{S}{4} = (1 - A) \qquad \textbf{(1.7)}$$

This represents the fundamental relation between incoming radiation, reflection from the Earth's surface and absorbtion in the atmosphere that together control climate and climate change. It is a simplification for what is an extremely complex climate system, involving a number of important feedbacks. For example, an increase in temperature increases water vapour content in the atmosphere, driving a greenhouse effect and further warming. Another positive feedback effect from increased temperatures is the melting of snow and ice, decreasing the Earth's albedo and causing further warming.

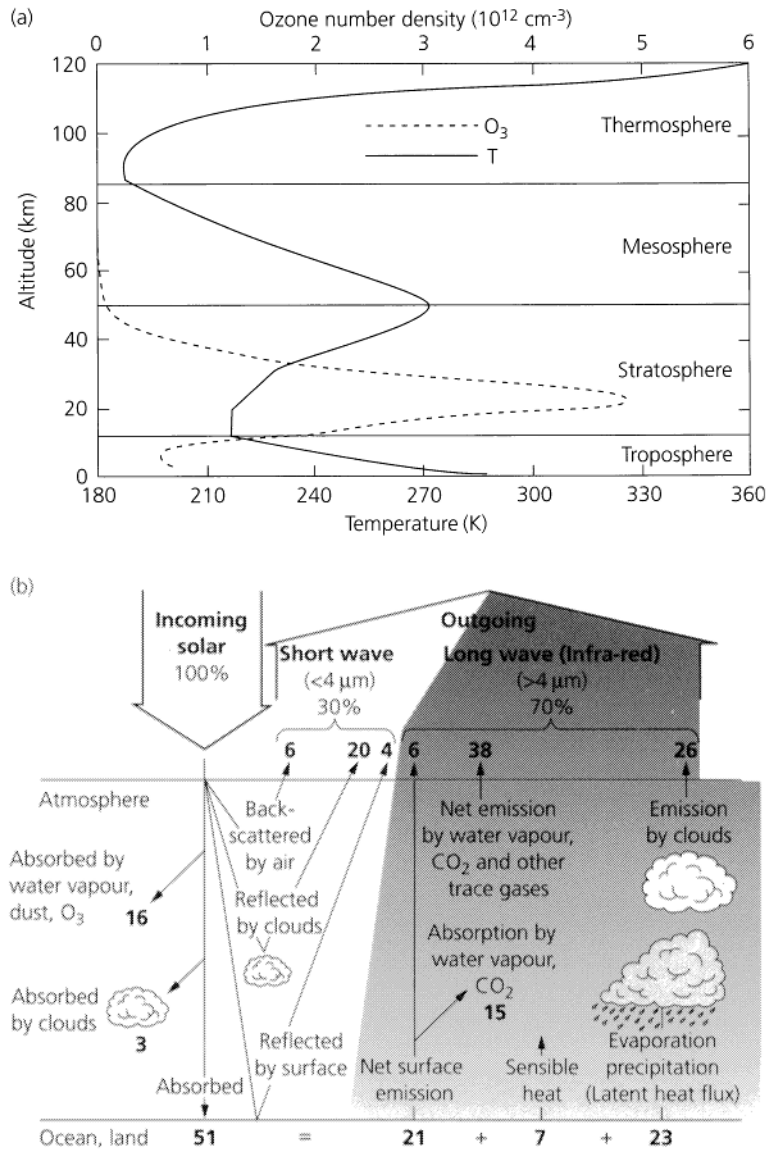


Fig. 1.2 The vertical stratification and heat budget of the atmosphere and Earth. (a) Vertical profile of temperature (solid line) and ozone density (dashed line). After Harrison *et al.* (1993) [3]. (b) Mean annual radiation and heat balance, with units assigned so that the total incoming radiation is 100. Data from the US Committee for the Global Atmospheric Research Program (1975).

The incoming radiation is thus balanced by the outgoing radiation. Over time, net gains in radiation, or net losses, should result in a global warming or cooling respectively. About 70% of the incoming solar radiation (the portion not reflected back into space) is used in interacting with the hydrological cycle.

1.3 The hydrological cycle

1.3.1 Role of the hydrological cycle in the global climate system

The hydrological cycle plays a crucial role in the climate system of the Earth. Water permeates all of the major components of the Earth's climate system—the ocean, the atmosphere, the lithosphere (or at least its upper part), the cryosphere (realm of snow and ice) and the biosphere. Any investigation of the global climate system must be based on a knowledge of hydrological processes. The abundance of water on Earth and its occurrence in multiple forms is unique in the solar system. The presence of water in the form of vapour, liquid and ice, and the relative ease with which water may transform from one phase to the other allows it to play a strongly stabilizing role in the Earth's climate.

Water stores and fluxes

Over 97% of water is stored in the oceans. Of the remainder, most is held in ice sheets and glaciers, some in groundwater, and minute percentages in lakes, the atmosphere, the soil and in rivers (Fig. 1.3). The transfers of water between these different stores or reservoirs have profound impact on Earth surface processes and require enormous inputs of energy. Water is transfered from the ocean to the atmosphere by evaporation and from the land by a combination of evaporation and transpiration from plants (*evapotranspiration*). A return flux of water is achieved by precipitation in the form of rain and snow. Over the continents the precipitation in general exceeds evapotranspiration, which leads to a flux of water from the continents to the oceans in the form of surface *runoff*. This runoff, although small in volume, is crucial in the physical evolution of the land surface, in the transfer of particulate sediment from erosional areas to depositional sites, and in the fluxing of chemical species in solution from the weathered land surface to the oceanic reservoir.

Since the surface of the Earth is dominated by the saline waters of the oceans, the human occupation of the land surface is dependent on the distillation of seawater to fresh water and its precipitation on the land. The relatively small percentage of fresh water in the hydrological cycle demands that it is continuously recycled. The fluxes of water between the different storages are large, a molecule of water having a characteristic residence time in each store (Table 1.1). For example, the turnover of water in the atmosphere is very rapid (with a residence time of only 8–10 days), the rivers only slightly less rapid (up to 2 weeks), whereas the oceans have long residence times of over 4000 years and water may remain in ice caps for tens of thousands of years [5]. The residence times of water resources are also of great relevance to the impacts of pollution, damage to ocean, lake and groundwater being much longer-lasting than to rivers. The extremely rapid turnover of the atmosphere requires the energy of a large proportion of the solar radiation received by the Earth. This can be appreciated by considering the

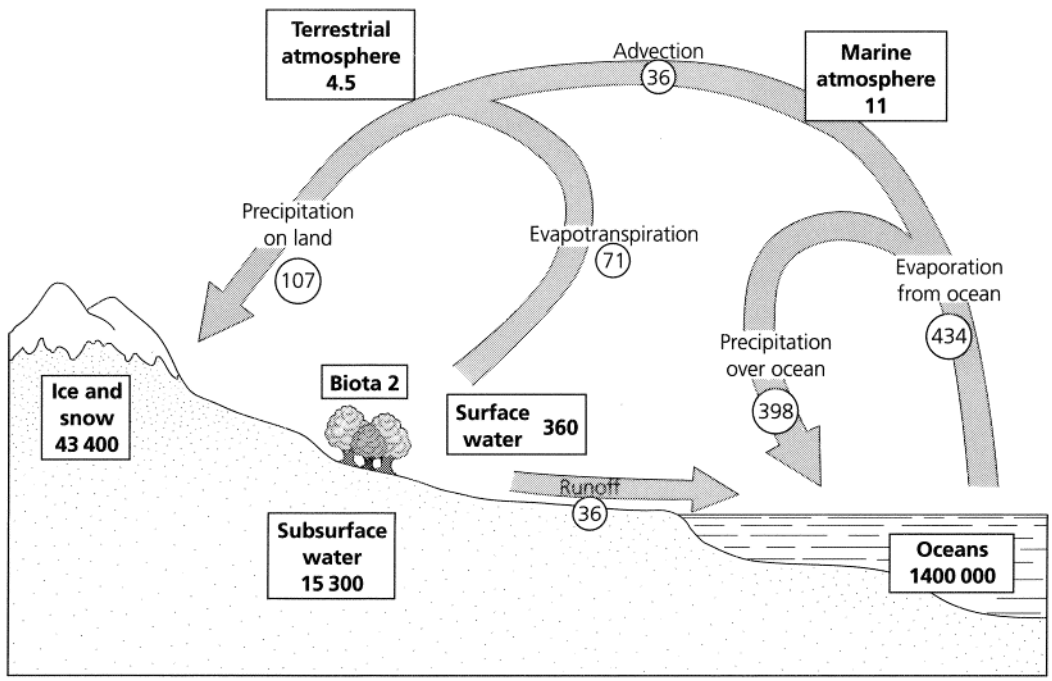


Fig. 1.3 The global water cycle, with its storages in 10^{15} kg (boxes) and fluxes in 10^{15} kg y^{-1} (figures circled). Water in the atmosphere constitutes 0.0001% of the total water in the hydrological cycle but is crucial for the efficient functioning of the system. Data from Chahine (1992) [4].

Table 1.1 Storages, fluxes and residence times in the global hydrological cycle. From Berner & Berner (1987) [1].

Storages	*Fluxes*
Atmosphere	Precipitation
Oceans and seas	Runoff slopes and channels
Lakes and reservoirs	Evaporation
Rivers	Horizontal vapour flux
Wetlands	Infiltration
Biological water	Percolation
Soil water	Groundwater flows
Groundwater	
Ice	
Residence times	
Atmosphere	8–10 days
Oceans and seas	4000 years +
Lakes and reservoirs	up to 2 weeks
Rivers	up to 2 weeks
Wetlands	years
Biological water	1 week
Soil water	2 weeks to 1 year
Groundwater	days to thousands of years
Ice	tens to thousands of years

behaviour of water vapour in the atmosphere. In its short residence time of 8–10 days, a molecule of water vapour is likely to have travelled an average distance of 1000 km.

Quantifying the global water balance is a formidable task requiring coordinated international efforts at gauging precipitation and runoff and estimating water volumes in all of the zones of the hydrological cycle. The fact that almost 70% of all fresh water is locked up in polar ice and in glaciers [6] has major implications for the effects of climate change, since large-scale warming or cooling causing shrinkage or expansion of the ice store will have profound effects on the entire hydrological cycle. Although some investigations have been made into the changes in the hydrological cycle over the last post-glacial period (approximately the last 10 000 years) (Chapter 2), remarkably little is known of the likely hydrological changes on a future warmer greenhouse Earth. This is now a major focus of research.

Why is the Earth unique in the solar system?

Among our solar system neighbours (Table 1.2), only

Table 1.2 Water in the solar system. Physical properties of the inner planets, their atmospheric compositions and disposition of water. Modified from Webster (1994) [7].

	Mercury	Venus	Earth	Mars
Specifications				
Planetary mass, 10^{23} kg	3.4	48.7	59.8	6.43
Planetary radius, km	2439	6049	6371	3390
Gravitation, m s^{-2}	3.8	8.9	9.8	3.7
Solar distance, 10^6 km	58	108	150	228
Solar irradiance, W m^{-2}	9200	2600	1393	596
State				
Mean surface temperature, K	442	700	288	210
Mean planetary albedo, %	6	71	33	17
Surface pressure, hPa	0	7900	1013	6
Atmospheric composition %				
CO_2	0	95	3×10^{-2}	>50
N_2, A	0	<5	79	<50
O_2	0	$<4\times10^{-3}$	21	1×10^{-1}
H_2O	0	1×10^{-2}	1	$\leq1\times10^{-1}$
HCl	0	1×10^{-4}	0	0
HF	0	2×10^{-6}	0	0
CO	0	2×10^{-2}	1×10^{-5}	1×10^{-1}
Water disposition, kg				
Atmospheric mass	0	4.2×10^{20}	5.3×10^{18}	2.4×10^{10}
Liquid	0	0	1.4×10^{21}	0
Ice	0	0	4.3×10^{19}	1×10^{17}
Gas	0	4.2×10^{16}	1.6×10^{16}	2×10^{13}

the Earth has large reservoirs of liquid water and regions of ice over both continents and oceans [7]. The atmosphere contains water in all three forms: as water vapour comprising 1% of the entire atmospheric mass, and as suspended ice and water droplets in clouds. The bulk of the water on the Earth is, however, in liquid form. On Mars there is no known liquid water, though water vapour is an important constituent of the atmosphere and ice occurs in the polar ice cap and perhaps also as permafrost. On Venus water vapour exists in abundance in the atmosphere but there is no liquid water or ice, so that the total mass of water on the planet is far smaller than on the Earth. Without liquid water, the climate of Mars (and also of Mercury) is strongly coupled to changes in its radiation balance with outer space, giving rise to extremes in temperature between day and night. The large store of liquid water in the oceans dampens external fluctuations in supply of heat, releasing heat gradually and allowing a latitudinal transport of heat because of the liquid form of the water (Section 1.3.3).

The Clausius–Clapeyron relation

Why is it that water on the Earth exists close to its triple junction? This can be appreciated by considering the phase transitions of water as a function of temperature and water vapour partial pressure (Fig. 1.4a). The likely trajectories of the Earth, Venus and Mars [8] indicate clearly that the triple point where all three phases exist in equilibrium is very close to the Earth's present conditions, so that all three forms of water may coexist at virtually any point on the Earth's surface (with the exception perhaps of the polar regions during winter) (Fig. 1.4b).

The phase transition lines are nonlinear and can be written in the form of the *Clausius–Clapeyron equation*

$$\frac{\mathrm{d}\ln e_s}{\mathrm{d}T} = \frac{L_e, L_s}{R_v T^2} \qquad (1.8)$$

where e_s is the saturation vapour pressure (in newtons per square metre) as a function of temperature T (Kelvin), L_e, L_s are the latent heats of evaporation ($2.5\times10^6\,\mathrm{J\,kg^{-1}}$) and sublimation (direct from solid to vapour) ($2.84\times10^6\,\mathrm{J\,kg^{-1}}$) respectively, and R_v is the specific gas constant for water vapour ($0.462\,\mathrm{J\,kg^{-1}\,K^{-1}}$). If it is assumed that the latent heats of evaporation and sublimation are constant with temperature, the Clausius–Clapeyron equation (1.8) can be solved to give the vapour pressure over liquid water or solid ice as a function of temperature. Integrating the Clausius–Clapeyron equation gives

$$e_s(T) = e_s(T_0)\exp\left\{\frac{L_e, L_s}{R_v}\left(\frac{1}{T_0}-\frac{1}{T}\right)\right\} \qquad (1.9)$$

where $e_s(T_0)$ is the saturation vapour pressure at temperature T_0. Inspection of (1.9) shows the saturation vapour pressure to depend exponentially on absolute temperature (in Kelvin). For example, the vapour pressure of the atmosphere in the tropics is more than an order of magnitude greater than that over the poles.

During the evolution of the Earth, outgassing of water vapour would have caused an increase in the vapour pressure with time, leading to greater absorbtion of outgoing radiation, creating an early

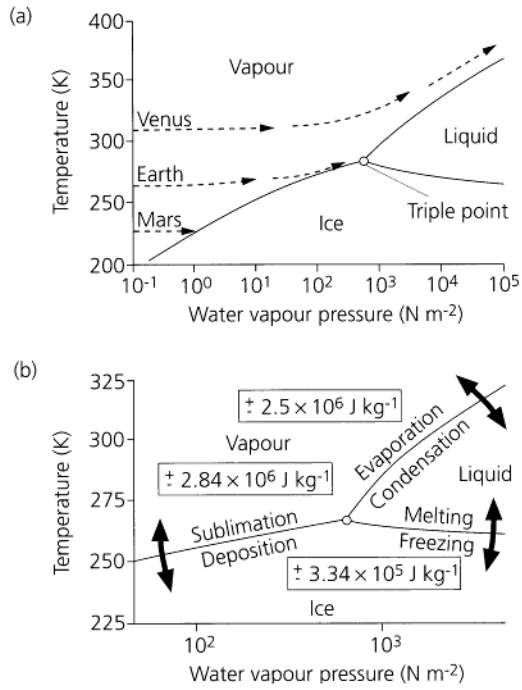


Fig. 1.4 (a) The phases of water as a function of water vapour pressure and absolute temperature. The phase transition curves between vapour, liquid and ice are given by the Clausius–Clapeyron equation (1.8). The dashed lines are hypothetical trajectories for the climate of Venus, the Earth and Mars through time. After Rasool & de Bergh (1970) [8]. (b) The water phase diagram close to the triple junction with the physical parameters pertaining to the Earth, with the latent energies for the phase transitions. After Webster (1994) [7].

greenhouse effect and leading to warming of the planetary surface. Surface warming in turn affects the phase transition equilibria because of the Clausius–Clapeyron effect. On Venus, the temperature–vapour pressure trajectory caused it to miss the vapour–liquid phase transition, leading to a 'runaway' greenhouse effect. On Mars, however, the trajectory intersected the sublimation–deposition transition, so that transfer is only possible between vapour and ice. This takes place, along with the deposition and sublimation of CO_2, at the Martian winter pole. The larger number of other thermodynamic possibilities on the Earth (Fig. 1.4b) explains the complexity and variability of the hydrological system.

We know from the Clausius–Clapeyron relation that saturation vapour pressure depends on temperature. It follows, therefore, that the global distribution of water in its three phases is determined by the global temperature structure. Before looking at this problem in further detail, it is important to stress the time-scales over which the various stores of water in the hydrosphere modulate climate. The ability of the ocean to mix vertically (see Chapter 9), rather than acting as a static pond, causes it to release and absorb heat on long time-scales. It takes of the order of hundreds to thousands of years for the global ocean to mix vertically throughout its entire depth. The modulation of climate by the oceans therefore involves the deep, slow circulation of its waters as well as the more rapid mixing at its surface (Section 1.3.3).

The Clausius–Clapeyron relation also fundamentally affects atmospheric dynamics. This is because of the following two main processes:

- Radiative absorption is a strong function of water vapour in the atmosphere (Section 1.2); this shows important geographical variations controlled principally by the Earth's temperature variations. The net radiative impact of water is a trade-off between the infrared absorption of water vapour in the atmosphere causing warming and the cooling caused by reflection of incoming solar radiation by the liquid water and ice in clouds. Consequently, the different phases of water affect climate in different ways.
- The global transfer of heat includes that due to the latent heat flux caused by evaporation at one locality and condensation elsewhere. The latent heat released by a saturated parcel of air depends on its initial temperature, so there are major variations between equator and poles determined by the Clausius–Clapeyron relationship. The latent heat released may be several times greater in the case of tropical air. This powers vigorous convection in low latitudes.

1.3.2 Global heat transfer

The circulation of the atmosphere and oceans is fundamentally caused by the fact that the amount of incoming solar radiation varies from a maximum at the equator to a minimum at the poles. This is caused by a number of factors:

- The angle of incidence of the Sun's rays changes from 90° at the equator to 0° at the poles. Less energy is therefore received at the poles because the energy is spread over a larger surface area at high latitudes.
- More reflection and absorption of incoming radiation takes place in high latitudes because of the greater thickness of atmosphere that must be penetrated.
- Variations in the amount of daylight are caused by the tilt of the Earth's axis relative to the plane of the Earth's orbit around the Sun, producing the seasons. The shortness of the daylight hours causes less annual radiation to be received per unit surface area in high-latitude than in low-latitude regions.

Since the long-wavelength radiation leaving the Earth does not vary greatly with latitude, radiation imbalances are set up, with surplus radiation between latitudes of about 50°N and 50°S and a deficit elsewhere. It is necessary to spread this heat imbalance over the surface of the Earth to prevent the tropics from getting increasingly warm and the poles increasingly cold. This heat transfer is accomplished through a strongly coupled circulation of the atmosphere and oceans.

1.3.3 Ocean–atmosphere interaction: driving mechanisms

The temperature gradient from equator to poles drives the atmospheric and oceanic circulations. The total heat transport is roughly equally balanced between ocean and atmosphere, but the types of flux of heat are somewhat different.

- Atmospheric motions are produced by heat fluxes at the atmosphere's lower boundary with land and ocean, and at its upper boundary by radiative cooling into outer space. The majority of heating takes place at the hot land and sea surface in the tropics, and in the low to middle troposphere of the tropics through the release of latent heat. This distribution forces a direct thermal circulation of the atmosphere shown in schematic form in Fig. 1.5.
- Ocean water motion is driven by wind stress at its upper surface, and horizontal density gradients arising from lateral variations in temperature, fresh water influx from precipitation, ice melting and river runoff from land. Whereas the atmospheric circulation is relatively efficient, the tendency for the upper tropical

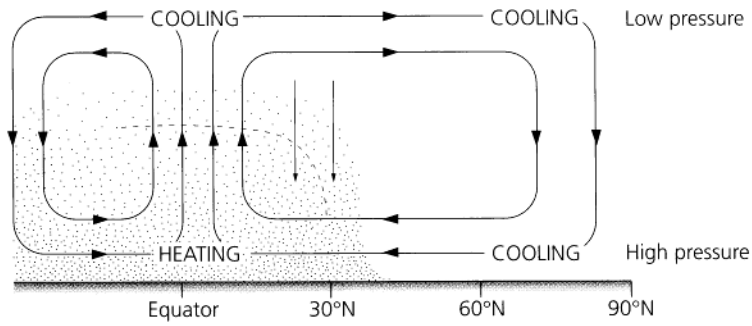


Fig. 1.5 The rudimentary thermal circulation of the atmosphere driven by heating in low latitudes.

ocean to stratify causes the global ocean circulation to be slow and inefficient. Despite this, it appears to have a major control on long-term climate variation.
We now examine these driving mechanisms in oceans and atmosphere in more detail.

Forcing of ocean circulation

The average temperature of the oceans is just 4°C and mean sea surface temperature 19°C. The mean ocean salinity is 35.5 parts per thousand, with mean sea surface salinity of 35.2 parts per thousand.

There is therefore a warmer and less saline layer of water at the surface of the ocean as a thin veneer over a deeper, colder and more saline body of water. This thin veneer is concentrated in the tropics. This correlates with a net excess of precipitation over evaporation over the warm tropical 'pools'. However, it is important to recognize that although the warm tropical oceans associated with vigorous atmospheric convection and resulting precipitation have a major impact on global climate, these warm (>28°C) waters constitute a minute part of the total water mass. Within the deeper water mass, the temperature and salinity distributions with depth are remarkably similar between the equatorial and subtropical regions. The main differences within the deeper water mass are found in higher latitudes where deep cold water is formed.

It is therefore a priority to understand the reasons for the thermal and salinity variations in the oceans. The emphasis here is on global patterns, but the mechanics of ocean circulation are discussed in some detail in Chapter 9.

In the open ocean, the net flux of fresh water is the balance between precipitation (P) and evaporation (E) (and at high latitudes by the balance between freezing and melting processes). The fresh water flux shows a latitudinal variation (Fig. 1.6):

- in the tropics and at high latitudes precipitation exceeds evaporation, leading to a fresh, stable upper layer in the ocean;
- in the subtropics evaporation exceeds precipitation, causing an unstable surface saline layer.

The thermal effects and saline effects may oppose or reinforce each other (Fig. 1.7). In the subtropics, for example, dense saline surface water sinks and underflows the stable, warm equatorial pool created by the positive fresh water flux and high temperatures. On the other hand, in middle latitudes (20–60°) the thermal and haline circulations are opposed. In high latitudes summer melting of ice and solar heating generate a light surface layer; winter cooling and ice formation causes a dense surface layer to sink to great depths.

Ocean–atmosphere interaction: buoyancy The relation between water temperature, salinity and density is complex and nonlinear (Fig. 1.8). In warm water, a given density change $\Delta\rho$ is associated with a small temperature change ΔT_W and salinity change ΔS_W, where the subscript refers to 'warm' water. The same density change in cold water is associated with a similar salinity change ΔS_C but a much larger temperature change ΔT_C, where the subscript refers to 'cold' water. Since for the same $\Delta\rho$, $\Delta T_W < \Delta T_C$, whereas $\Delta S_W \approx \Delta S_C$, we can assume that circulation in the tropical ocean is forced by processes causing temperature changes, but that in high-latitude oceans the effects of salinity and temperature are comparable. There are therefore different regional responses to temperature and salinity variations [9].

This can be formalized by defining the *buoyancy* as the relative density of a parcel of ocean water compared to a neighbouring parcel

$$B = \alpha T - \beta S \qquad \textbf{(1.10)}$$

where α and β are the thermal and salinity expansion coefficients respectively, and T and S are temperature

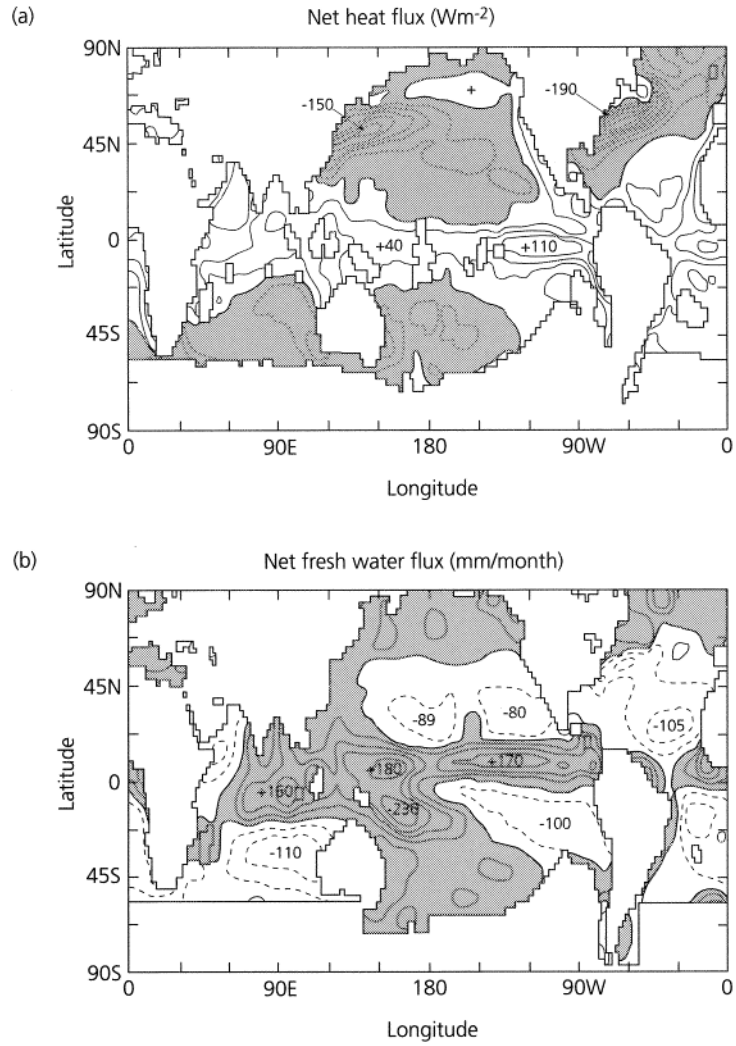


Fig. 1.6 (a) The annual mean net heat flux into the ocean (Q_T), showing regions of net heat loss (shaded) and net heat gain (blank). The tropical oceans have a net positive Q_T, the subtropics and high latitudes suffer moderate heat loss (negative Q_T), and the western parts of northern hemisphere ocean basins experience high net heat loss to the atmosphere (strongly negative Q_T). (b) Annual mean net fresh water flux into the ocean (F_W). Positive fluxes, where $P-E>0$, are shown as shaded, and dominate the tropical regions where sea surface temperatures are highest. Negative fluxes, where $P-E<0$, occur especially in the subtropics. Data are from Oberhuber (1988) in Webster (1994) [7].

and salinity. If the ocean receives a total heat flux Q_T, and the flux of fresh water into the ocean is denoted by F_W (equal to precipitation minus evaporation, $F_W = P - E$), then the flux between the atmosphere and ocean can be thought of as a *buoyancy flux*

$$F_B = \alpha g Q_T + \beta g S(P-E) \qquad \textbf{(1.11)}$$

(where g is acceleration due to gravity) written so as to make clear the thermal (first term on the right-hand side) and haline (second term) effects. Now it can be seen that if the fresh water flux is small compared to the heat flux, as in the tropics, buoyancy is driven by temperature variations, whereas in high-latitude oceans the melting and freezing of ice as well as temperature variations control buoyancy.

The global picture is of a small number of domains

Practical exercise 1.1: Regional sensitivity of water to changes in salinity and temperature

Water at the surface of the warm tropical pools is at about 29°C and has a salinity of 35 parts per thousand, whereas surface water in high-latitude regions has a temperature of about 7°C and salinity of 33 parts per thousand. What would be the effect on the density of these two different water masses of (a) a 2.5 parts per thousand increase in salinity; (b) a 5°C decrease in temperature?

Solution

Using Fig. 1.8, the warm tropical surface water has a

Continued on p. 14.

Practical exercise 1.1: *Continued*

density of 1022 kg m^{-3}. The cool high-latitude water has a density of 1026 kg m^{-3}. A 2.5 parts per thousand decrease in salinity causes a change in density of just less than 20 kg m^{-3} for both water masses. However, a 5°C temperature change causes a density change of less than 4 kg m^{-3} for the cold water compared to over 10 kg m^{-3} for the warm water. Warm water masses are therefore controlled by processes which influence temperature, whereas cold water masses are equally influenced by factors causing temperature and salinity changes. These regional sensitivities are controlled by the thermal expansion and salinity expansion coefficients for water. Although the thermal coefficient of expansion varies greatly with changes in water temperature between 2.5°C and 30°C (over a factor of 4 from 781×10^{-7} to 3413×10^{-7} K^{-1}), the salinity expansion coefficient is relatively constant over the same temperature range (8010×10^{-7} to 7490×10^{-7} K^{-1}).

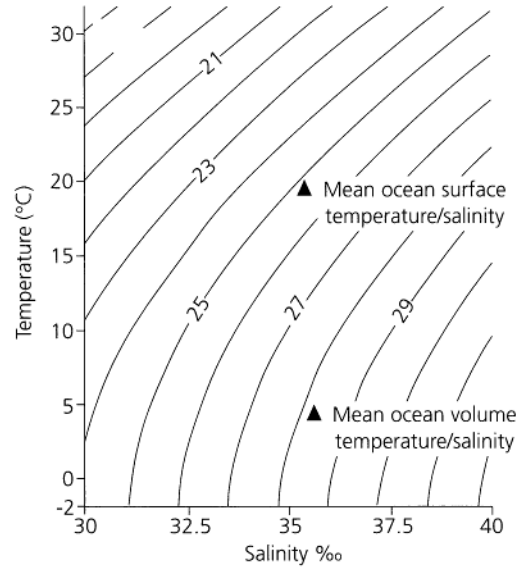


Fig. 1.8 The sensitivity of the density of seawater to changes in temperature and salinity. Density shown in kilograms per cubic metre in excess of 1000. Water masses in tropical and high-latitude regions have markedly different sensitivity to a given change in temperature or salinity. See Practical Exercise 1.1.

for the distribution of both fresh water flux and total heat flux. The heat flux, Q_T, is positive in equatorial–tropical oceans; negative in the subtropics and high–latitude oceans, where there is moderate heat loss to the atmosphere; and strongly negative in the western parts of the northern hemisphere, where there is major heat loss. The fresh water flux, F_W, is positive in the tropical oceans, especially where the sea surface temperatures are highest; and negative in the subtropics.

With this modicum of theory we can now make more sense of the observed global pattern of atmospheric and oceanic circulation.

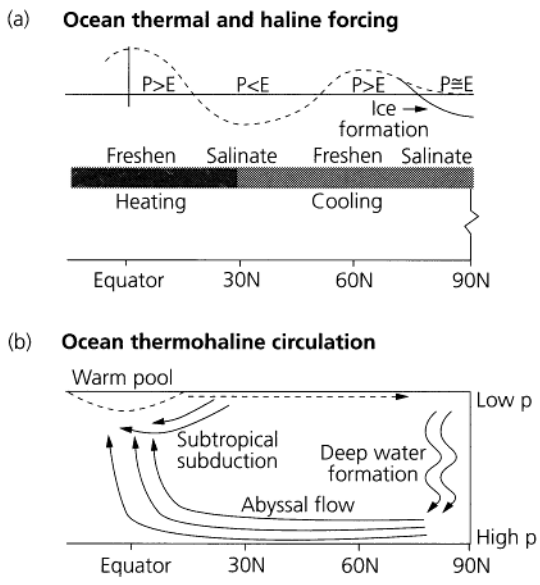
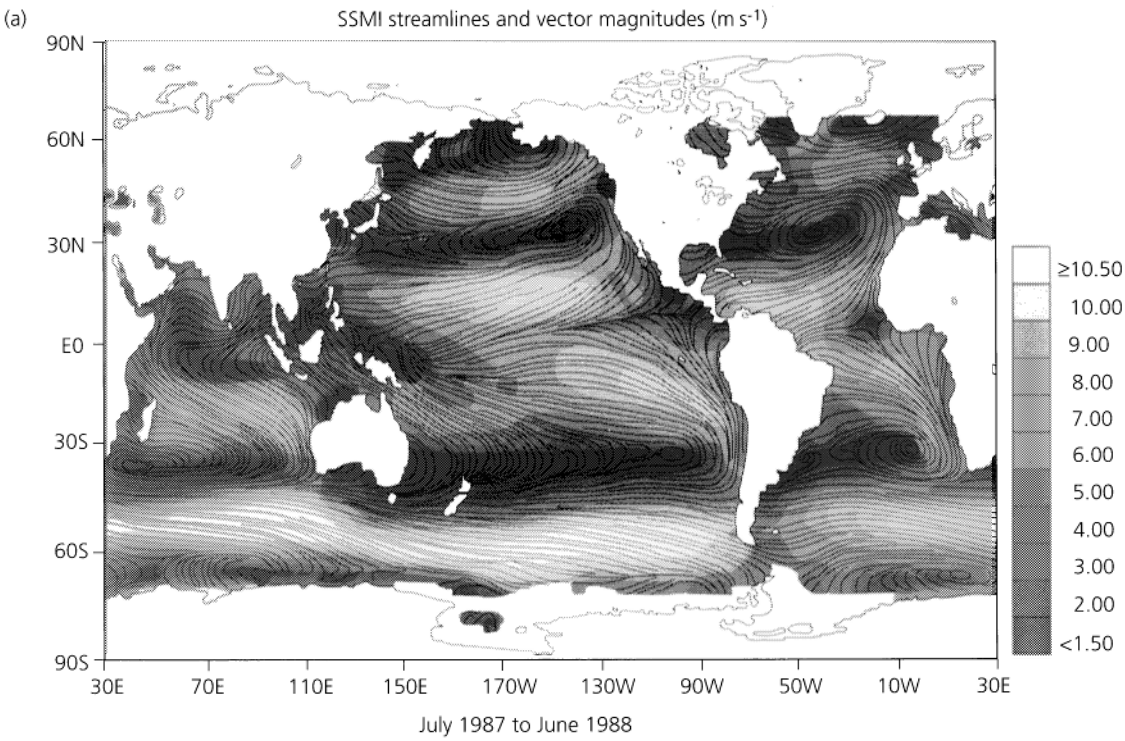


Fig. 1.7 Forcing mechanisms and circulation in the ocean. (a) Thermal and haline forcing. (b) Resulting thermohaline oceanic circulation (P = pressure). After Webster (1994) [7].

Observed oceanic and atmospheric circulation

Oceanic currents About a quarter of the heat surplus building up in low latitudes is carried by warm, wind-driven ocean currents such as the Gulf Stream, moving poleward from the region between 20°N and 20°S [2]. These warm currents heat the overlying atmosphere, moderating the climate of adjacent land masses. The water movement is caused by wind stress, the effects of rotation of the Earth, and by frictional interaction with continents (see Chapter 9). The circulation pattern that results is of *gyres* that flow clockwise in the northern hemisphere and anticlockwise in the southern (Fig. 1.9). Each gyre has



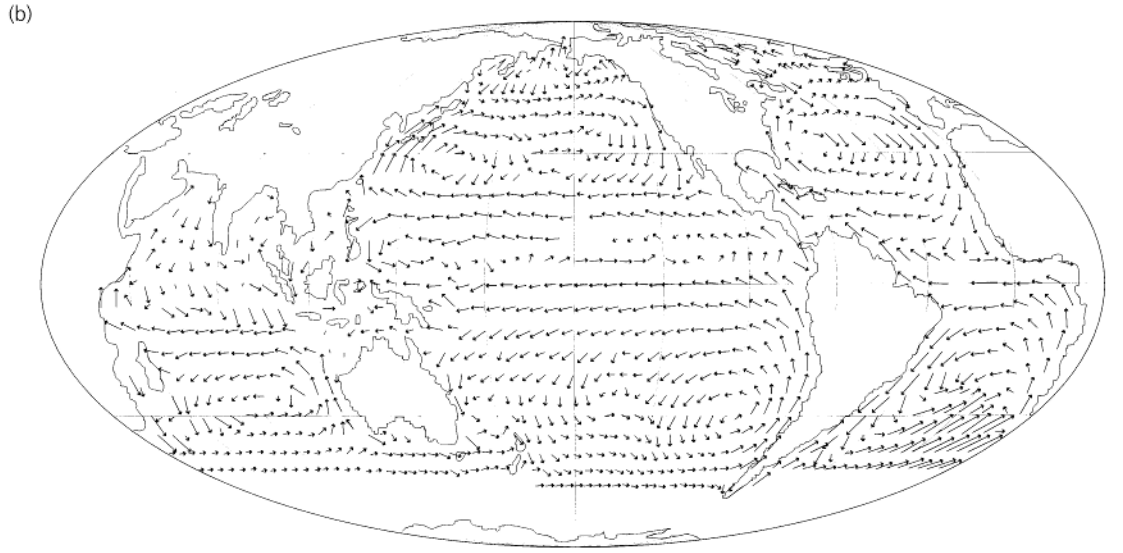


Fig. 1.9 (a) The ocean surface wind speeds (SSM/I data) as streamlines and vector magnitude coded by colour, from July 1987 to June 1988 (Atlas *et al*. (1993) [10]). (b) The mean total surface currents of the ocean derived from the average of historical ship drift observations (from Meehl (1982) [11]). The pattern is a resultant of geostrophic and wind-driven flows (see also Chapter 9).

a strong poleward current on the western side of the gyre and a weaker equatorward current on the east, a feature known as *western intensification*.

Whereas in the shallow ocean (the top few hundred metres) the circulation is driven by the wind, in the deep ocean the circulation is caused by density variations due to differences in temperature and salinity. This deep *thermohaline circulation* (Fig. 1.10), however, owes its origin to processes taking place at or near the surface, such as heating and cooling, evaporation, addition of fresh water, or abstraction of water as sea ice. At high latitudes in the Atlantic surface cooling and abstraction of sea ice (which leaves the seawater denser due to the concentration of salts not incorporated into the ice) causes the water to sink because of negative bouyancy. This cold, salty water then flows laterally as a deep oceanic circulation, diffusing slowly into surface layers and occasionally upwelling rapidly at continental edges. Where this upwelling of nutrient-rich water takes place high organic productivities result. The velocities of the deep thermohaline currents are very low, perhaps only 100 m per day, and the residence time of deep oceanic water is long, 200–500 years for the Atlantic and 1000–2000 years for the Pacific.

The actual pattern of thermohaline circulation (Fig. 1.10) is made complicated by:

- the finite size of the ocean basins;
- the interconnectedness of the different ocean basins;

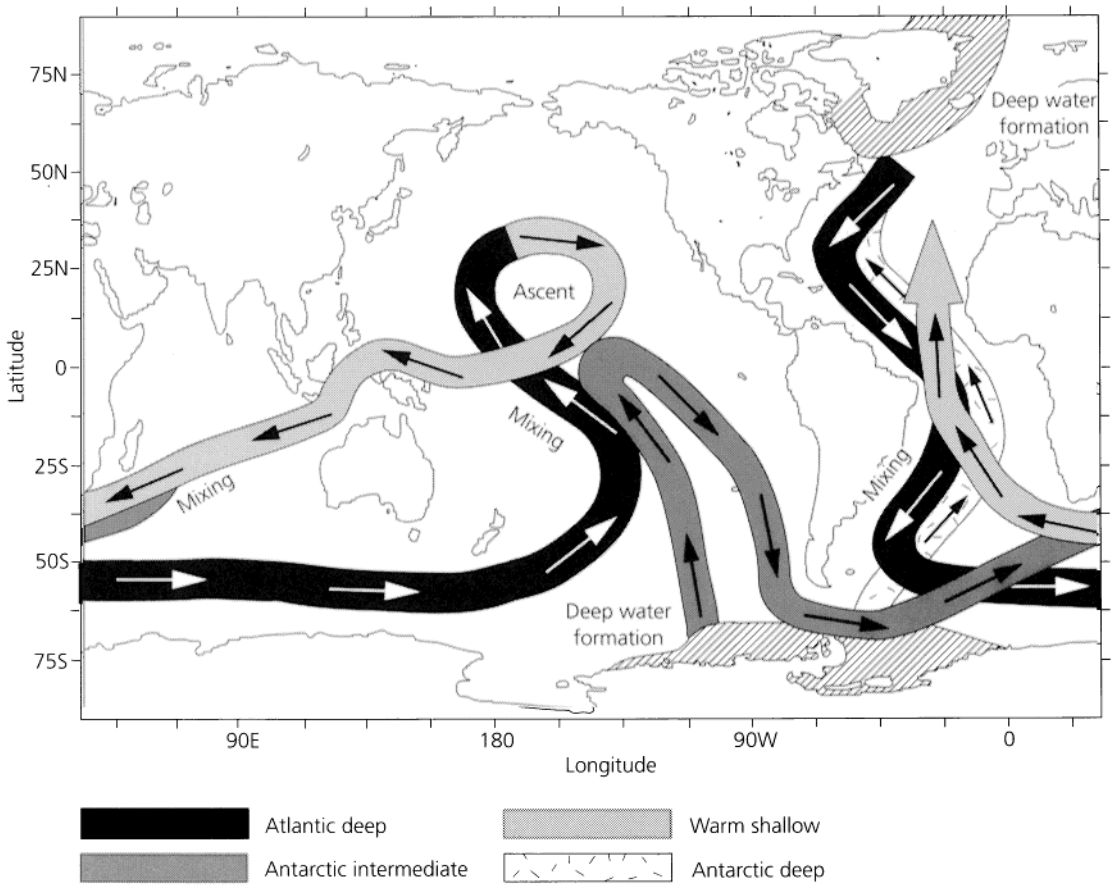


Fig. 1.10 The deep thermohaline circulation of the oceans, modified from Stommel (1958) [12]. The major sources of deep water at the present day in the north Atlantic and Weddell Sea are shown by hatching. The lack of deep water formation in the north Pacific may be due to the greater stability of the Pacific caused by the higher fresh water flux. The text describes the circulation patterns observed.

• the spatially and temporally varying forcing from the atmosphere.

The deep circulation of water through the ocean basins from regions of deep water generation in the north Atlantic and Weddell Sea area of the Antarctic Ocean was first demonstrated by Stommel in 1958. Deep water originating in the north Atlantic flows southwards along the western side of the Atlantic basin before turning east as a circumpolar current, then entering the Pacific basin. Removal of water from the Atlantic and build-up in the Pacific demands a return flow of some form. It has been suggested that the northward-moving deep water in the Pacific derived from the north Atlantic mixes with water derived from the Weddell Sea (Antarctic Intermediate Water) and leaves the Pacific basin through Drake Passage, between South America and Antarctica (Fig. 1.10). The remainder ascends in the northern Pacific and leaves the basin as a surface current through the Indonesian archipelago to the Indian Ocean. These two outlets re-enter the south Atlantic and flow northwards as a near-surface current towards the original site of the Atlantic Deep Water. There is therefore a complete circuit of oceanic circulation. The time for water to circulate in this global system is thought to be of the order of 10 000 years. Variations in the deep thermohaline circulation may be responsible for climate change on a similar time-scale, that is, the time-scale of the climatic fluctuations associated with the Pleistocene glaciations (Chapter 2). Much climate change research is now focused on the dynamics of the deep thermohaline circulation.

The episodic massive release of icebergs into the north Atlantic Ocean during the last glaciation, known as *Heinrich events* [13], is thought to have strongly perturbed global climate through the effects on sea surface temperatures and ocean–atmosphere circulation. Further information is given in Chapter 2.

Atmospheric circulation The atmospheric circulation set up by the latitudinal radiation imbalances is responsible for over half of the heat transfer in the form of warm poleward-blowing winds and latent heat transfer. More information on the elementary physics of this global atmospheric circulation is given in Chapter 10.

We have seen that the simplest way in which the atmosphere might respond to the excess of heat in the tropics and deficit at the poles is in the establishment of a simple circulation of rising air at the equator, a high-level poleward motion, sinking at the poles as the air cools, and a return flow at low level to the equator (Fig. 1.5). This simple idealized pattern of two recirculating cells, with a deflection of the winds by the rotation of the Earth (Coriolis force) (see also Chapters 9 and 10), was originally proposed in 1735 by George Hadley. The actual general circulation of the atmosphere (i.e. the mean annual winds) (Fig. 1.11) shows the Hadley circulation to be broken into several latitudinal zones. Instead of air rising at the

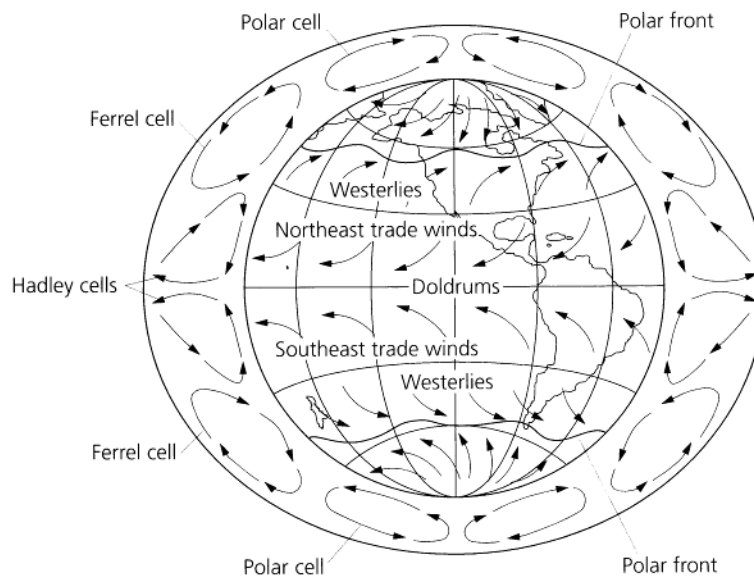


Fig. 1.11 The general circulation of the atmosphere showing the Hadley, Ferrel and Polar cells. Modified from Miller *et al.* (1983) [14].

equator and travelling all the way to the poles, it descends at about 30° latitude. The subtropical deserts are located under this descending dry air. It is dry for two main reasons: first, it has lost its moisture by condensation in the tropics; and second, it is descending and therefore warming, increasing its capacity to hold moisture. The air then travels equatorwards at low level as the trade winds. This low-latitude circulation is known as a *Hadley cell*. There is a region of low-level convergence of dry trade wind air, and moist equatorial air, known as the *intertropical convergence zone*. It is characterized by heavy rainfalls, and rainforests have developed under it, as in Amazonia and western equatorial Africa.

Air descending at about 30° latitude also flows polewards at low level as the westerlies until it meets cold polar air moving equatorwards at about 50°, forming the *polar front* of atmospheric instability. The instability generates storms and heavy precipitation. The air from low latitudes rises over the cold polar air, and returns to the subtropics at high level, completing the *Ferrel cell*. The polar air warms by mixing with the westerlies and by condensation at the polar front, rises, and flows back to the polar region where it cools and sinks. This is the *polar cell*. There are thus three major atmospheric cells per hemisphere.

In addition to the three latitudinal cells per hemisphere, high sea surface temperatures in the warm pools of the Indian and Pacific Oceans cause ascending air patterns, resulting in a series of east–west cells along the equator known as the *Walker circulation*. The maximum upward motion in the Walker circulation is associated with the highest sea surface temperatures, emphasizing the crucial link between ocean and atmosphere in determining global climate. The distribution and intensity of warm pools [15] rather than a latitudinally continuous zone of high sea surface temperatures is probably related to the distribution of the continental land masses. Air rises over the Indonesian region and descends over the eastern Pacific Ocean. It fluctuates in intensity, when it is termed the *southern oscillation*. When the circulation reverses, known as *El Niño events* (see also Section 9.6), major droughts may occur in the southern hemisphere, accompanied by wet weather in deserts.

At levels in the upper troposphere there are two channels of extremely high winds which travel as westerlies around the poles. These are known as the *jet streams*. One reaches as far equatorwards as the subtropics (the *subtropical jet stream*), and the other is restricted to higher latitudes (the *polar front jet stream*). These jet streams appear to delimit the circulatory cells, the poleward limit of the upper part of the Hadley cell being marked by the vigorous ($<65\ m\ s^{-1}$) subtropical jet stream, and the weaker ($<25\ m\ s^{-1}$) polar front jet stream being located in the zone of high meridional (longitudinal) pressure gradients at the tropospheric continuation of the polar front.

The distribution of pressure in the upper troposphere at high latitudes indicates a number of waves, whose motion facilitates the poleward transfer of heat. These *Rossby waves* appear to grow as disturbances in the jet stream, varying in their latitudinal position over weeks or months. Rossby waves may play an important role in the global circulation of heat in the atmosphere.

The *monsoon* is a large-scale circulation pattern which is asymmetrical about the equator. The forcing mechanism for the monsoonal circulation is the distribution of land and sea in the eastern hemisphere. The monsoonal circulation is characterized by extreme seasonality. In the winter Asia is typified by descending air. This reverses in southern Asia to a strong rising motion of air in the summer in response to intense heating, and is accompanied by heavy precipitation. Compensation for the rising of air over southern Asia causes a drawing in of air from the Middle East and north Africa (Fig. 1.12), where air descends, causing a strengthening of summer desert conditions. The vigorous upward motion forms part of a circuit involving a reverse flow in the upper troposphere.

The interannual variability of the monsoon can be correlated with reversals of the Walker cell and the occurrence of El Niño, causing drought over Indonesia and north Australia. Further discussion of monsoonal and Walker circulations is deferred to Section 10.2.3.

The tropical oceans and overlying atmosphere are extremely important in the global hydrological system. They are the sites of high sea surface temperatures and associated deep convection in the atmosphere. Consequently, the Walker cell and monsoonal circulations have a major impact on the global hydrological cycle.

Latent heat transfer Latent heat results from the change of state of water. It is given out when condensation takes place and consumed when evaporation takes place. Heat is transferred when latent heat stored in evaporated water (water vapour) is released elsewhere during condensation. Tropical air (between

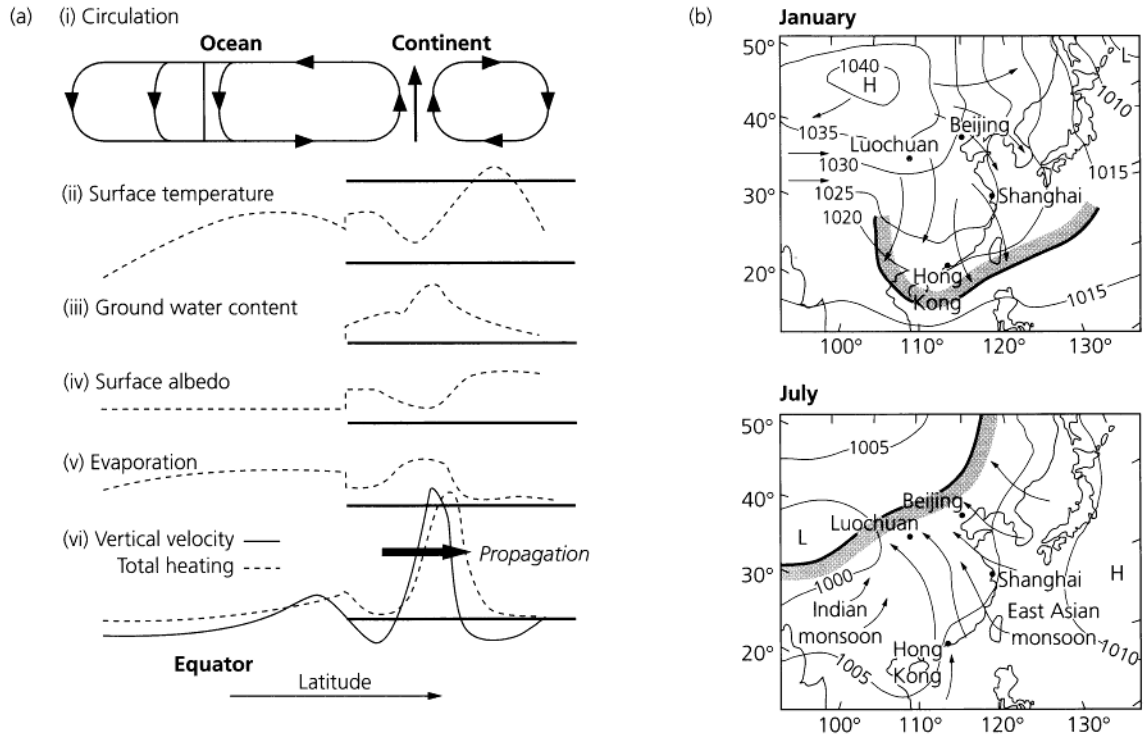


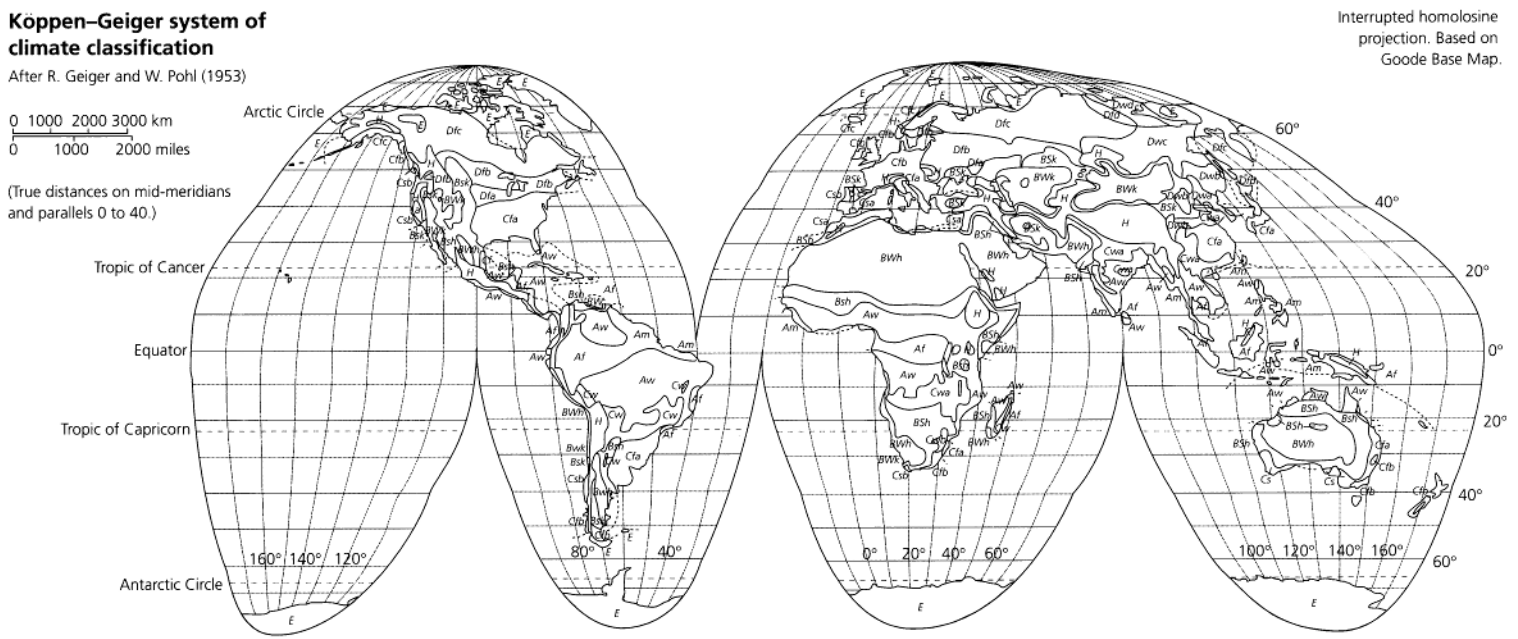
Fig. 1.12 The monsoonal circulation. (a) Interaction of atmospheric and surface hydrological systems. (i)–(vi) show the function of latitude during the poleward propagation of a monsoon convective event. After Webster (1994) [7]. (b) The winter and summer monsoon regimes affecting China, showing surface pressures (in millibars), dominant wind vectors (arrows) and polar front (January) and monsoonal front (July) (bold line). After Porter & Zhisheng (1995) [16].

10°N and 10°S) is heated both by latent heat released during condensation and by conduction from the warm land surface (sensible heat). Latent heat is consumed in the subtropical zone (15–30°N and S) where evaporation exceeds precipitation. It is released in mid-latitude zones (30–50°N and S) by condensation associated with storms. The tropical rainforests (see also Section 2.5) are extremely important in transferring latent heat high into the atmosphere. Very high rates of evaporation take place from the surfaces of rainforest plants, the water vapour produced condensing in large clouds extending to high altitudes, thereby releasing latent heat.

1.3.4 Summary: a global interactive model

The result of the heat transfers through oceanic and atmospheric circulations is the climatic zones of the Earth (Fig. 1.13). It is important to emphasize that the ocean and atmospheric systems do not work in isolation but rather operate synergetically to maintain climatic stability. The global system can be viewed as a highly interactive entity operating at a number of different time-scales. This interactive system and its latitudinal structure are summarized in Fig. 1.14. Hydrological processes act as the linkages of the different interacting parts of the coupled atmosphere–ocean system, shown schematically in Fig. 1.15. The key elements of the interactive system are as follows.

- **Major convection** takes place:
  (i) in the tropical atmosphere (upward motion of moist air over high sea surface temperatures);
  (ii) in the polar ocean (downward motion of negatively buoyant water caused by radiative cooling in winter and abstraction of sea ice producing an increase in salinity);



Key to letter code designating climate regions:

First letter:

A C D Sufficient heat and precipitation for growth of high-trunked trees.
A Tropical climates. All monthly mean temperatures over 64.4°F (18°C).
B Dry climates. Boundaries determined by formula using mean annual temperature and mean annual precipitation.
C Warm temperature climates. Mean temperature of coldest month: 64.4°F (18°C) down to 26.6°F (-3°C).
D Snow climates. Warmest month mean over 50°F (10°C). month mean under 26.6°F (-3°C).
E Ice climates. Warmest month mean under 50°F (10°C).

Second letter:

S Steppe climate.
W Desert climate.
f Sufficient precipitation in all months.
m Rainforest despite a dry season. (i.e., monsoon cycle.)
s Dry season in summer of the respective hemisphere.
w Dry season in winter of the respective hemisphere.

Third letter:

a Warmest month mean over 71.6°F (22°C).
b Warmest month mean under 71.6°F (22°C). At least four months have means over 50°F (10°C).
c Fewer than four months with means over 50°F (10°C).
d Same as c, but coldest month mean under -36.4°F (-38°C).
h Dry and hot. Mean annual temperature over 64.4°F (18°C).
k Dry and cold. Mean annual temperature under 64.4°F (18°C).

H Highland climates.

Fig. 1.13 Climatic zones of the Earth, according to the Köppen–Geiger classification system. Originally developed by Vladimir Köppen (1846–1940), a German climatologist and botanist, it forms the basis for most morphoclimatic schemes such as that of Tricart & Cailleux (1972) [17].

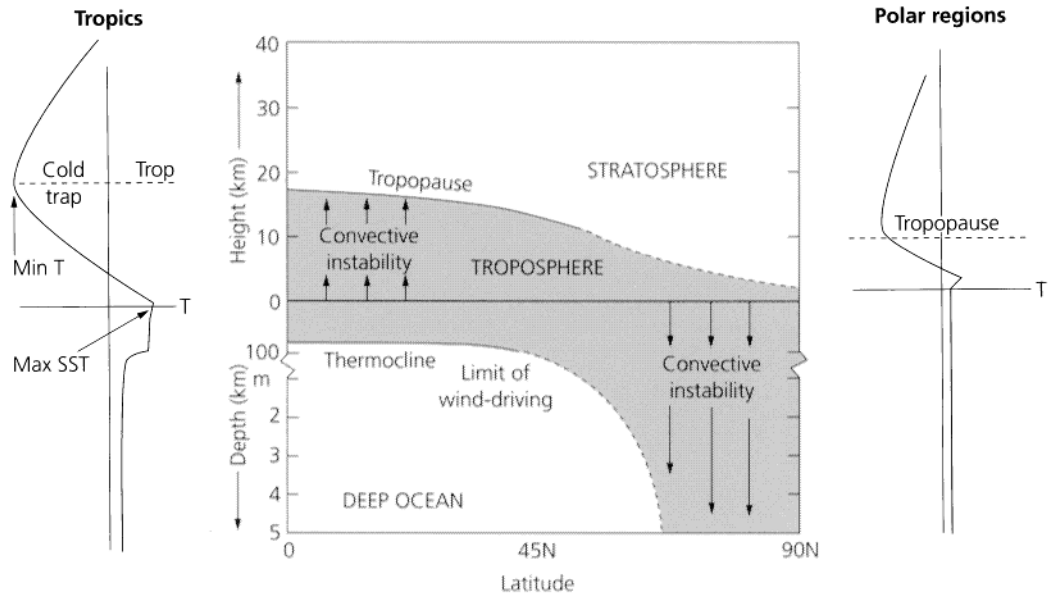


Fig. 1.14 The interactive zones of the atmosphere and ocean, encompassing the troposphere, upper ocean, and at high latitudes the entire ocean where deep water formation is involved. Interactive zones are those which operate at similar time-scales and thus directly affect each other. The stratosphere and deep ocean in lower latitudes operate at longer time-scales. After Webster (1994) [7].

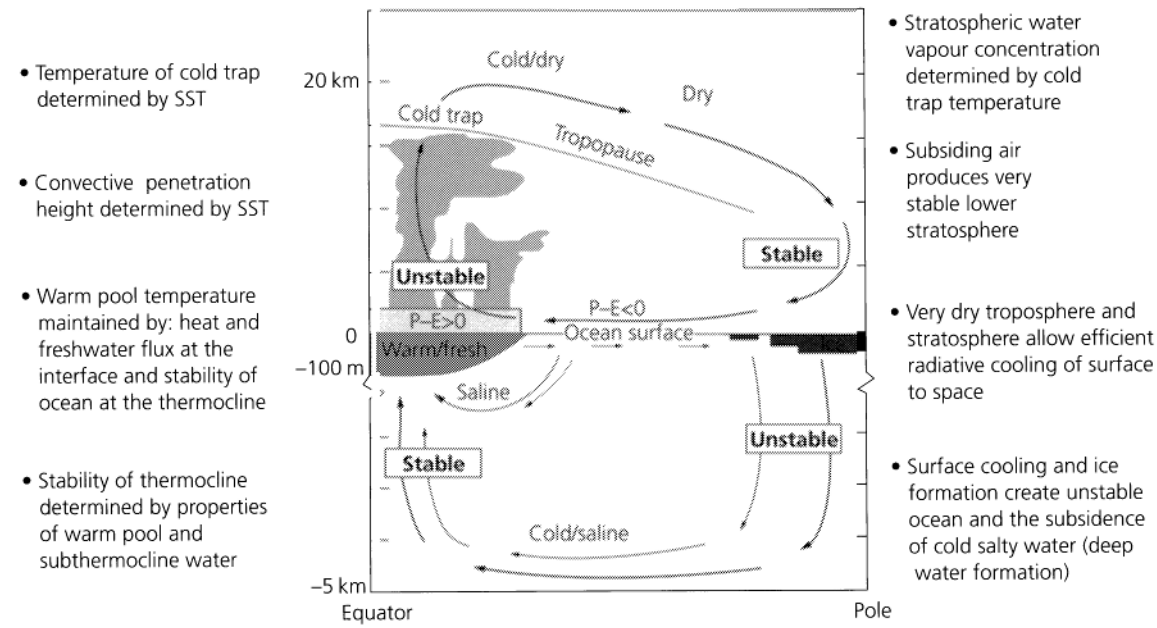


Fig. 1.15 Role of hydrological processes in the coupled atmosphere–ocean system. Deep atmospheric convection in the tropics produces a very cold tropopause which acts as a 'cold trap' for moist air passing through to the stratosphere. This ensures that the stratosphere remains dry, which in turn promotes the rapid cooling in polar regions, the formation of sea ice and of saline deep ocean water. The stable tropical ocean water ensures a surface warm pool which in turn drives atmospheric convection above. After Webster (1994) [7].

(iii) in the subtropical ocean intense evaporation at the surface causes sinking of saline waters.

- In contrast, **gravitational stability** characterizes:
  (i) the tropical ocean, with a warm upper layer overlying rising cold water;
  (ii) the polar atmosphere with its descending air flow.

Although this model does not take into account the complicated effects of the distribution of land and sea, it demonstrates the effectiveness of the linkages between the ocean and atmosphere circulations. The engine for the entire coupled system might be thought to lie in the warm pools of the tropical ocean.

1.3.5 Runoff

The surface water balance

We have seen that fresh water runoff plays a role in the coupled atmosphere–ocean system. It is also crucial in landscape development since most sediment transport on land is caused by runoff. The factors controlling denudation in the world's drainage basins, and the resulting sediment fluxes through and out of them, are discussed in Chapter 3. Although the quantity of water in the continental portion of the hydrological cycle is minute (Fig. 1.3), it has profound importance for the activities of humans and in sustaining natural ecosystems. Here, however, it is necessary to retain our global focus and to place runoff in its broad context in the hydrological cycle.

The runoff can be viewed in terms of its role in the surface water balance,

$$P = E + T + \Delta S + \Delta G + R \qquad \textbf{(1.12)}$$

where P is precipitation, E is evaporation, T is transpiration, ΔS is the change in the storage of water in the soil, ΔG is the change in the storage of water as groundwater, and R is the overland flow across the land surface as rills and gullies, streams and rivers (Fig. 1.16). This surface water balance differs between open water and land. In the former the balance is a simple relationship between the incident precipitation, the evaporation and the resultant change of volume of the water body. An excess of evaporation over precipitation results in a shrinkage of the water body, the situation common in evaporating pans. However, fresh water runoff from rivers may be important in both lakes and the open sea. In this case

$$P = E - \Delta F \qquad \textbf{(1.13)}$$

where ΔF is the inflow of water from land to ocean. Globally, this relationship must balance, but there are considerable regional variations in rates of precipitation (Fig. 1.17), evaporation (Fig. 1.18) and fresh water inflow (Tables 1.3 and 1.4).

For a land surface the full surface water balance equation (1.12) is required. If we regard the soil and groundwater storages as not fluctuating significantly over long periods of time, the difference between the long-term precipitation and long-term evaporation is reflected in runoff. Since the global average depth of precipitation over land exceeds the global average depth of evapotranspiration, there is a surplus representing the surface runoff. Once again, there are important regional variations which will be examined below.

Runoff, together with its sediment load, is routed through river drainage systems or catchments (Chapter 3). The surface water balance equation shows that a number of components require measurement in order to assess the water balance of a catchment: streamflow, precipitation, evaporation, soil moisture and groundwater (Fig. 1.16). The relative importance of these components depends on the climatic, topographic and geological setting. The ratio of streamflow measured from the river and precipitation falling on its drainage basin is termed the *runoff coefficient*. It varies in both space and time between 0 and 100%. Evaporation losses resulting from evaporation *sensu stricto* plus infiltration losses are the difference between catchment precipitation and streamflow. Soil moisture and groundwater losses may be considerable in certain climates, particularly at times of drought in a humid-temperate climate when groundwater levels are normally high and soils relatively saturated. Soil moisture and groundwater therefore also have a major role to play in controlling runoff processes.

Global patterns of runoff

The annual evaporation from the oceans is equivalent to a layer of water 1.4 m thick over the ocean surface of the Earth. Most (90%) of this is returned to the oceans directly by precipitation on the oceanic surface. However, the remainder falls as precipitation on land, equivalent to an annual layer with average thickness of 1 m. Of this amount, about 35% is returned to the oceans in the form of runoff, the remainder being lost by evapotranspiration. The regional patterns in the relative importance of precipitation, evaporation and runoff, and the controls on these patterns, are briefly discussed below.

For rain or snow to form, there is a dual requirement of sufficient water vapour in the atmosphere and

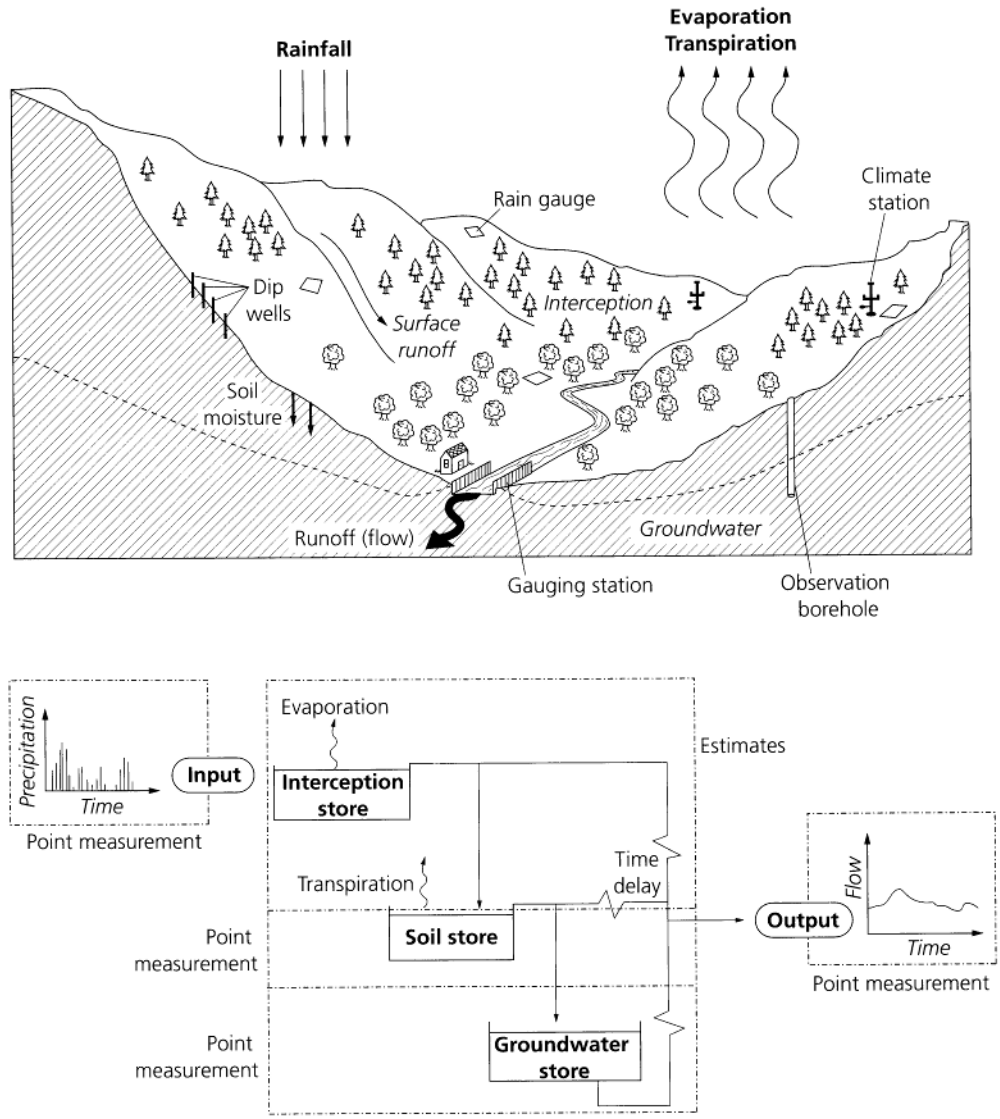


Fig. 1.16 The components of the hydrological cycle for a catchment. The individual storages (boxes) exist along a route from an input precipitation through to a river flow as an output from the catchment. After Newson (1994) [18].

rising air that can cause the vapour to cool, condense and cause precipitation. The geographical variation of mean annual precipitation (Fig. 1.17) shows it to be highest in tropical zones where humid air rises convectively through heating, and in regions where mountains force moisture-laden air to rise, so-called *orographic* precipitation. There is also a high amount of precipitation between 35° and 60°N and S where atmospheric instability associated with storms is common (Section 1.3.3).

Evaporation also varies greatly over the surface of the Earth (Fig. 1.18). High rates of evaporation require a large heat source (solar radiation), low moisture contents in the air, and a source of liquid water. Consequently, actual evaporation rates may be lower than expected in dry continental settings because of the lack of available water. The highest evaporation rates are found over subtropical oceans, from where currents such as the Gulf Stream transport warm water northwards into cooler, drier air conditions.

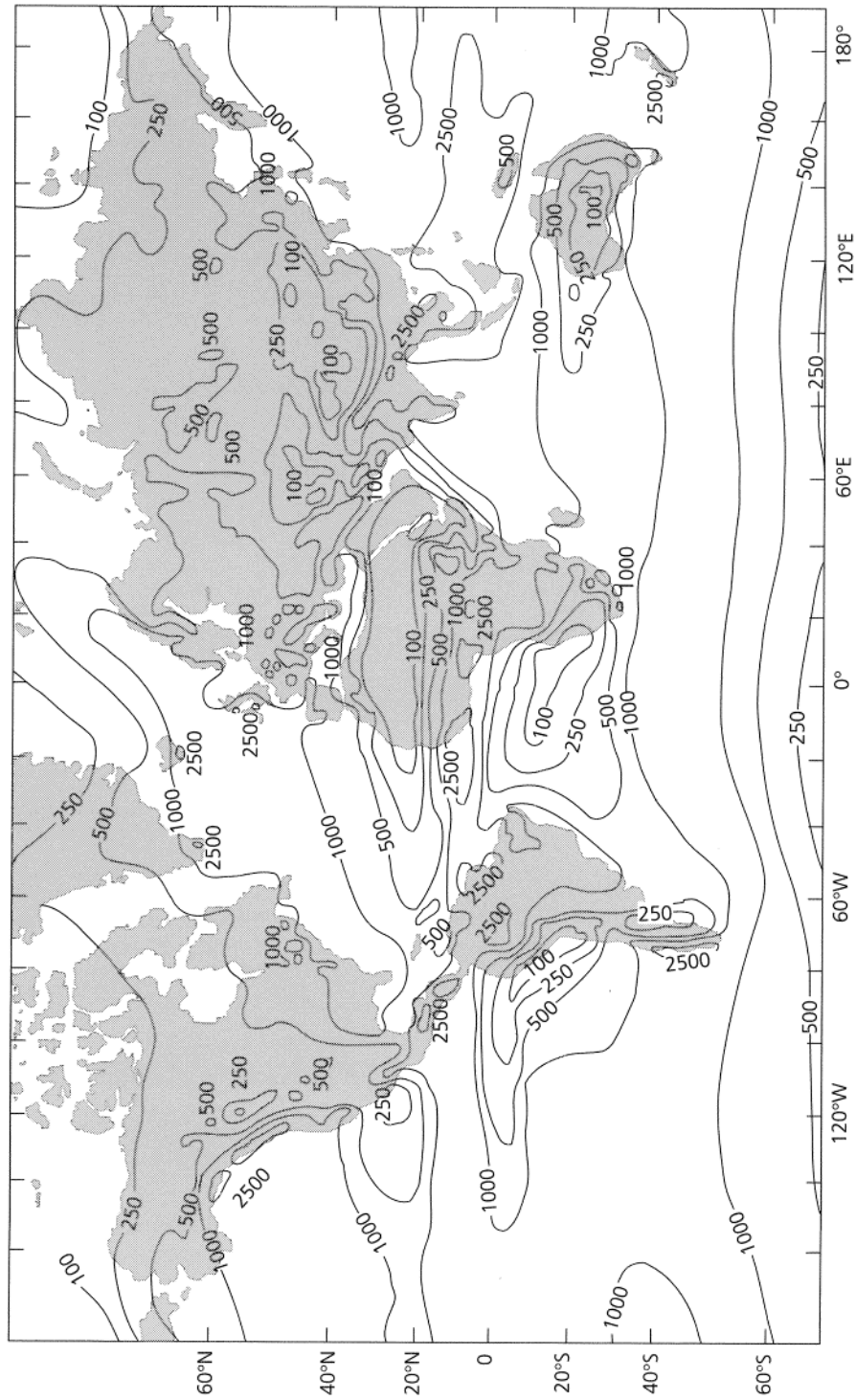


Fig. 1.17 The geographical variation of mean annual precipitation, in millimetres. After Lamb (1972) [19].

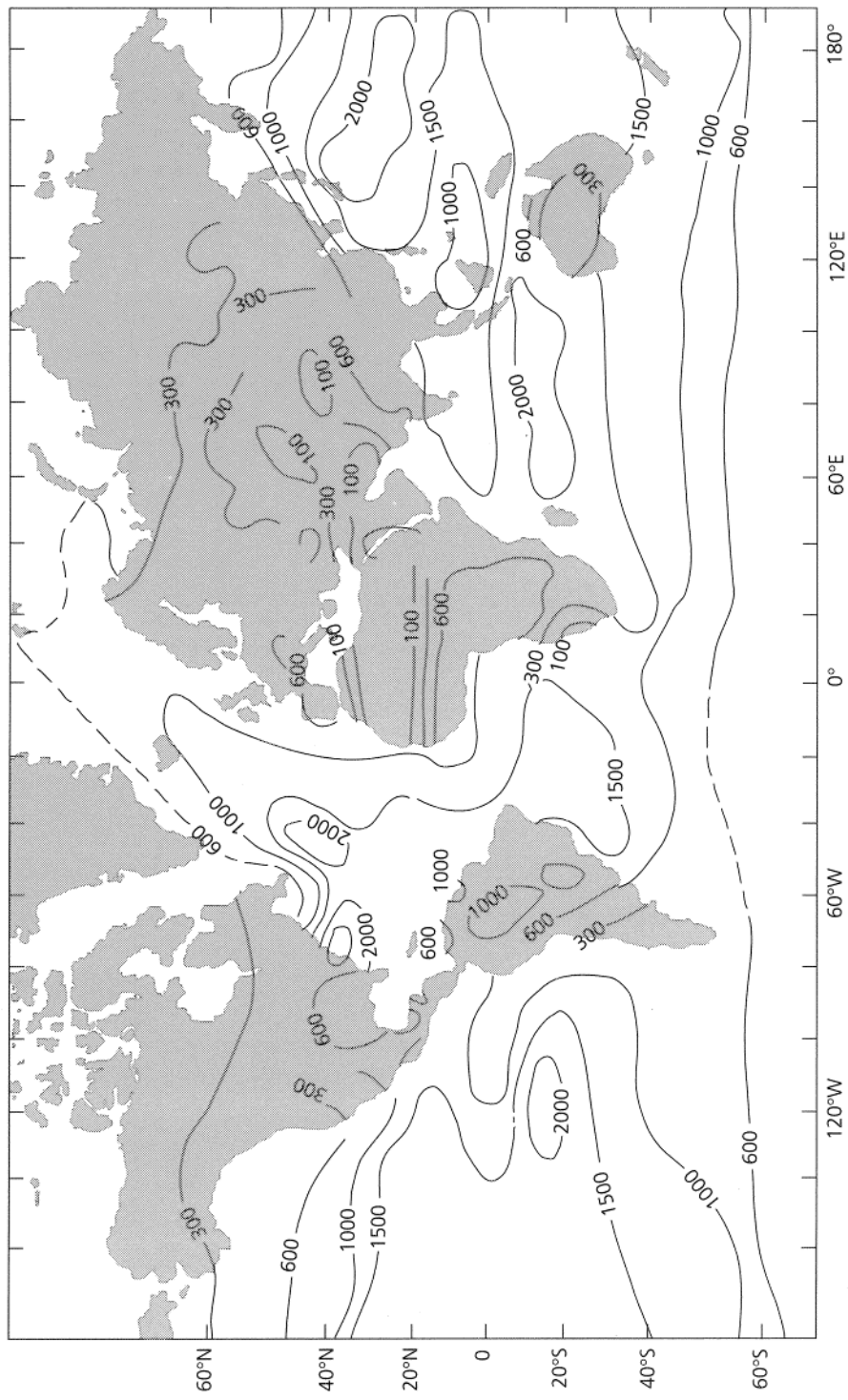


Fig. 1.18 Geographical variation of annual evaporation from ocean and evapotranspiration from land surface, in millimetres. After Barry (1970) [20].

Continent	Precipitation (mm)	Precipitation (km^3)	Evaporation (mm)	Evaporation (km^3)	Runoff (mm)	Runoff (km^3)
Europe	790	8 290	507	5 320	283	2 970
Asia	740	32 200	416	18 100	324	14 100
Africa	740	22 300	587	17 700	153	4 600
North America	756	18 300	418	10 100	339	8180
South America	1600	28 400	910	16 200	685	12 200
Australia and Oceania	791	7 080	511	4 570	280	2 510
Antarctica	165	2 310	0	0	165	2 310
Land as a whole	800	119 000	485	72 000	315	47 000
Areas of external runoff	924	110 000	529	63 000	395	47 000*
Areas of internal runoff	300	9 000	300	9 000	34	1 000†

Table 1.3 Water balance of the Earth's land surface in terms of precipitation, evaporation and runoff. From Shiklomanov (1993) [21].

* Including underground water not drained by rivers.
† Lost in the region through evaporation.

River	Average runoff ($km^3 y^{-1}$)	Area of basin ($10^3 km^2$)	Length (km)	Continent
Amazon	6930	6915	6280	South America
Congo	1460	3820	4370	Africa
Ganges (with Brahmaputra)	1400	1730	3000	Asia
Yangzijiang	995	1800	5520	Asia
Orinoco	914	1000	2740	South America
Paraná	725	2970	4700	South America
Yenisei	610	2580	3490	Asia
Mississippi	580	3220	5985	North America
Lena	532	2490	4400	Asia
Mekong	510	810	4500	Asia
Irrawaddy	486	410	2300	Asia
St Lawrence	439	1290	3060	North America
Ob	395	2990	3650	Asia
Chutsyan	363	437	2130	Asia
Amur	355	1855	2820	Asia
Mackenzie	350	1800	4240	North America
Niger	320	2090	4160	Africa
Columbia	267	669	1950	North America
Magdalena	260	260	1530	South America
Volga	254	1360	3350	Europe
Indus	220	960	3180	Asia
Danube	214	817	2860	Europe
Salween	211	325	2820	Asia
Yukon	207	852	3000	North America
Nile	202	2870	6670	Africa

Table 1.4 Mean annual runoff of the world's largest rivers. From Shiklomanov (1993) [21].

The vast bulk of runoff (nearly 98%) enters the oceans directly, and only a small fraction is routed into closed interior basins such as the Caspian and Aral Seas in southern Asia. In such cases the entire runoff is evaporated. Substantial use by man of precious fresh water runoff in interior basins can result in fundamental changes to the water–salt balance of the water bodies into which the rivers flow. Catastrophic changes to the

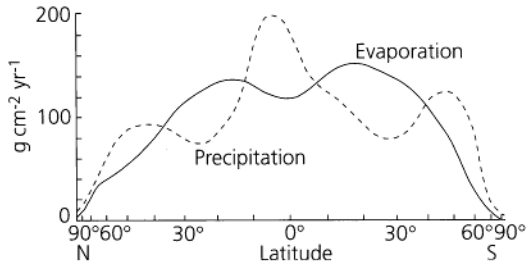


Fig. 1.19 Zonally averaged precipitation and evaporation as a function of latitude.

aquatic ecosystems and to the human societies that depend on them may result, as in the Aral Sea.

The water balance, measured in terms of the relative importance of precipitation, evaporation and runoff, varies according to latitudinal position or climatic zone (Table 1.3) [21]. The net precipitation, or precipitation minus evaporation (Fig. 1.19), is lowest in the subtropics (15–30°N and S) where the air is stable, and in the polar regions which have both stable air and very low moisture content because of the low temperatures. In some polar regions, the evaporation is so low that there is a positive net precipitation, allowing ice caps to form (Greenland, Antarctica). Very high percentages of precipitation are consumed by evapotranspiration in continents with large tropical and subtropical zones such as Africa (84%) and Australia (94%), lower percentages in continents with extensive temperate zones such as North America (62%) and Europe (57%), and a very low percentage in Antarctica (17%) with its very low evapotranspiration rates [22].

Table 1.3 shows that more than half of the global runoff occurs in Asia and South America (31% and 25% respectively), where it is concentrated in the equatorial zone. The Amazon River is outstanding in its average runoff of 6930 km^3 per year, more than 15% of the annual global runoff.

Chemistry of water in the global hydrological cycle

A comprehensive treatment of the chemistry of water in the different stores of the hydrological cycle is beyond the scope of this text. However, as attention in hydrological research is increasingly focused on problems at large spatial and long temporal scales, there is renewed interest in how chemical tracers can illuminate the dynamics of the global system. This is particularly true of the use of new methods of isotope geochemistry. The chemistry of water in the hydrosphere also determines nutrient supply and therefore impacts strongly on the biosphere (Section 1.4).

Precipitation falling over the land surface is modified in its chemical composition by reactions taking place with rock and soil, with plants and with decomposing organic matter. Land-based water is diluted or concentrated by additions of precipitation or losses through evaporation, but the most important factor determining the chemical composition of surface and soil water and the fresh water of lakes and rivers is the input from weathering of rock. This is dealt with in more detail in Chapter 3.

The global cycling of chemicals is strongly affected by the release of dissolved and particulate products from the weathering of rock and soil. In outline, waters from calcareous catchments contain high amounts of total dissolved solids (TDS), particularly Ca^{2+} and HCO^{3-}, but small particulate loads. Waters draining varied igneous, metamorphic and sedimentary rock types have a correspondingly more varied chemical composition and higher particulate loads. This total flux from the land surface is partially responsible for the steady-state composition of the oceans.

The observed differences in the chemical composition of river water have been investigated in terms of both the chemistry of rainwater and the dissolved weathering products as a function of TDS [23]. The cationic ratio

$$\frac{Na^+}{Na^+ + Ca^{2+}}$$

and the anionic ratio

$$\frac{Cl^-}{Cl^- + HCO_3^-}$$

show actual water compositions to lie along two diverging arms (Fig. 1.20) controlled by three main mechanisms.

1 Control by precipitation: where the chemistry is dominated by rainwater rich in Na^{2+}, producing a cationic ratio of nearly 1, and low TDS since evaporated seawater falling as rain is dilute. Rivers draining areas of low relief with well-weathered bedrock and plentiful rainfall, such as the tropical rivers of Africa and South America, fall into this group. The low TDS values may also be affected by the large biomass of the lowland tropical forests effectively immobilizing chemicals from release into rivers.

2 Control by weathering: where the chemistry of river water is controlled by weathering reactions within their catchments. The cationic ratio very

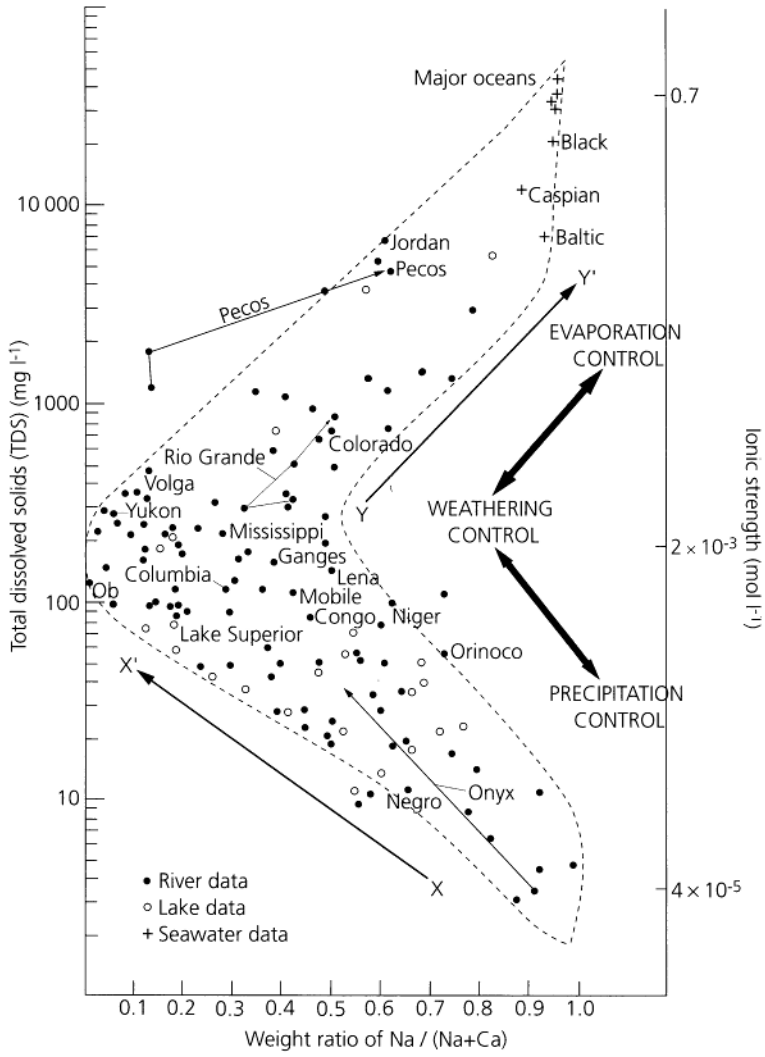


Fig. 1.20 Modified from Gibbs's (1970) [23] scheme for global water chemistry, wherein the total dissolved solids and ionic strength of surface waters are plotted against the cationic weight ratio of Na/(Na + Ca). The arrows connecting data points show the geochemical evolution of river waters from source downstream. Rivers plotting along the trend from X to X′, such as the Mackenzie (Arctic Canada) and Ganges (southern Himalayas), occur in regions with highly active weathering processes. Rivers falling along the trend from Y to Y′, such as the Jordan (Middle East), Rio Grande and Colorado (arid south-western North America) occur in areas experiencing high amounts of evaporation, and evolve towards the composition of seawater.

much depends on the type of rock being weathered, limestone bedrock producing the lowest ratio. TDS values are higher, reflecting the increased role of weathering (see Chapter 3). In the Amazon River basin, 85% of the solute load is derived from a relatively small area of weathering control in the Andes, whereas in the lowland part of the catchment the control is by precipitation.

3 Control by evaporation and crystallization: this is common in hot and arid climates where precipitation of calcite ($CaCO_3$) increases the cationic ratio, and evaporative losses of river water increase the TDS.

The chemistry of the river input to the ocean therefore varies from continent to continent in response to variations in the control of precipitation, weathering and evaporation. The total flux of dissolved matter into the ocean (Fig. 3.31) is determined by both the TDS and the discharge.

Naturally occurring isotopes in water can be used as tracers of the residence times, mixing ratios and fluxes of key elements of the hydrological cycle. They have an especially important role in assessing underground water masses of different isotopic composition and their mixing. The most commonly used isotopes are the stable isotopes deuterium (^2H), oxygen-18 (^{18}O) and carbon-13 (^{13}C). In addition, the radioisotopes tritium (^3H), carbon-14 (^{14}C) and radon-222 (^{222}Rn) are used.

Oxygen 18 and deuterium reside in the water molecule and are not affected greatly by reactions of water with rock, soil or vegetation. Since oxygen 18 has a different mass than oxygen 16, and deuterium likewise has a different mass than hydrogen, changes in phase of water from liquid to vapour to liquid to ice and back to liquid (representing evaporation, precipitation as ice and then melting) result in enrichment or depletion (fractionation) in one isotope compared to the other. Fractionation is a function of temperature. Consequently, winter precipitation is depleted in ^{18}O and depleted in ^{3}H compared to summer precipitation. Distinct differences also occur between the stable isotopic composition of precipitation at different latitudes and altitudes. Consequently, one can start to appreciate that stable isotopic compositions can be used to assess the contribution of melting of polar ice caps or of freshwater continental runoff, for example, to the oceans.

Other isotopes have been introduced into the hydrological cycle by humans, such as the large amounts of ^{3}H and ^{14}C discharged into the upper atmosphere by testing of thermonuclear weapons in the period 1957–64. Since the half-life of tritium is known (12.3 y), components of the hydrological cycle such as shallow groundwater can be 'tagged' and dated. Many other isotopes can be used.

1.4 Role of the biosphere

Although the relative mass of living (or once living) matter in the Earth surface system is very small, it has a profound effect on the upper veneer of the lithosphere, the hydrosphere and the inner part of the atmosphere, and mediates large parts of the hydrological cycle. The biosphere, the realm of life, is exceedingly complex in its components and its interactions. Although the role of the biosphere in Earth surface processes is referred to on numerous occasions in this book, the reader must look elsewhere for treatments beyond an outline of the global functioning of the biosphere dealt with here.

The basic functional organization of the biosphere is of solar radiation providing the energy for photosynthesis in the presence of chlorophyll in green plants, their growth and assimilation by herbivorous animals, and consumption by carnivorous animals. This organization represents a chemical energy flux. Energy is also transferred to a store of dead organic matter from the primary producers, herbivores and carnivores (Fig. 1.21). These layers within the functional organization of the biosphere are known as *trophic levels*. Any trophic level may experience a net gain or net loss of energy. For example, an increase in the biomass of a forest is indicative of net primary production. An increase in the number of grazing herbivores represents secondary production, also termed conversion.

The primary production of the land and ocean areas is estimated to be about 100×10^{12} kg y^{-1} and 55×10^{12} kg y^{-1} respectively (Fig. 1.22). This

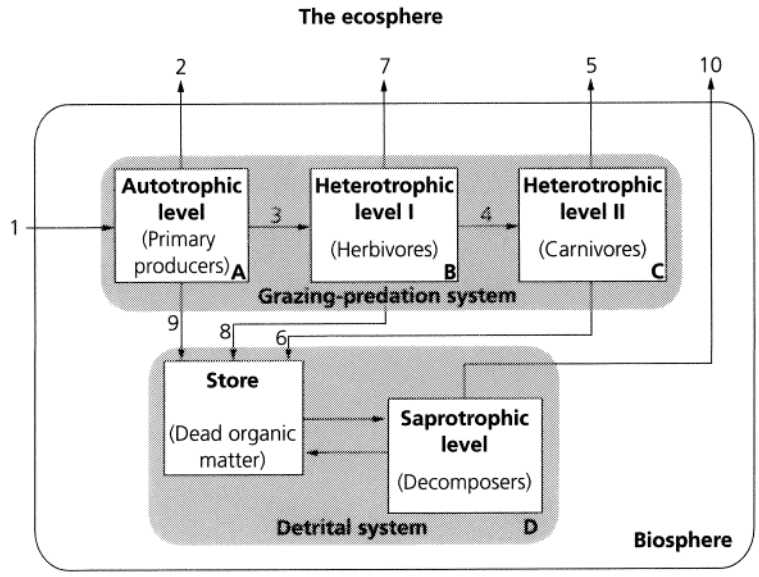


Fig. 1.21 Model of the trophic levels shown as compartments (A to D) of the biosphere, with energy transfers into and out of these compartments represented by arrows numbered from 1 to 10. After White *et al.* (1992) [24].

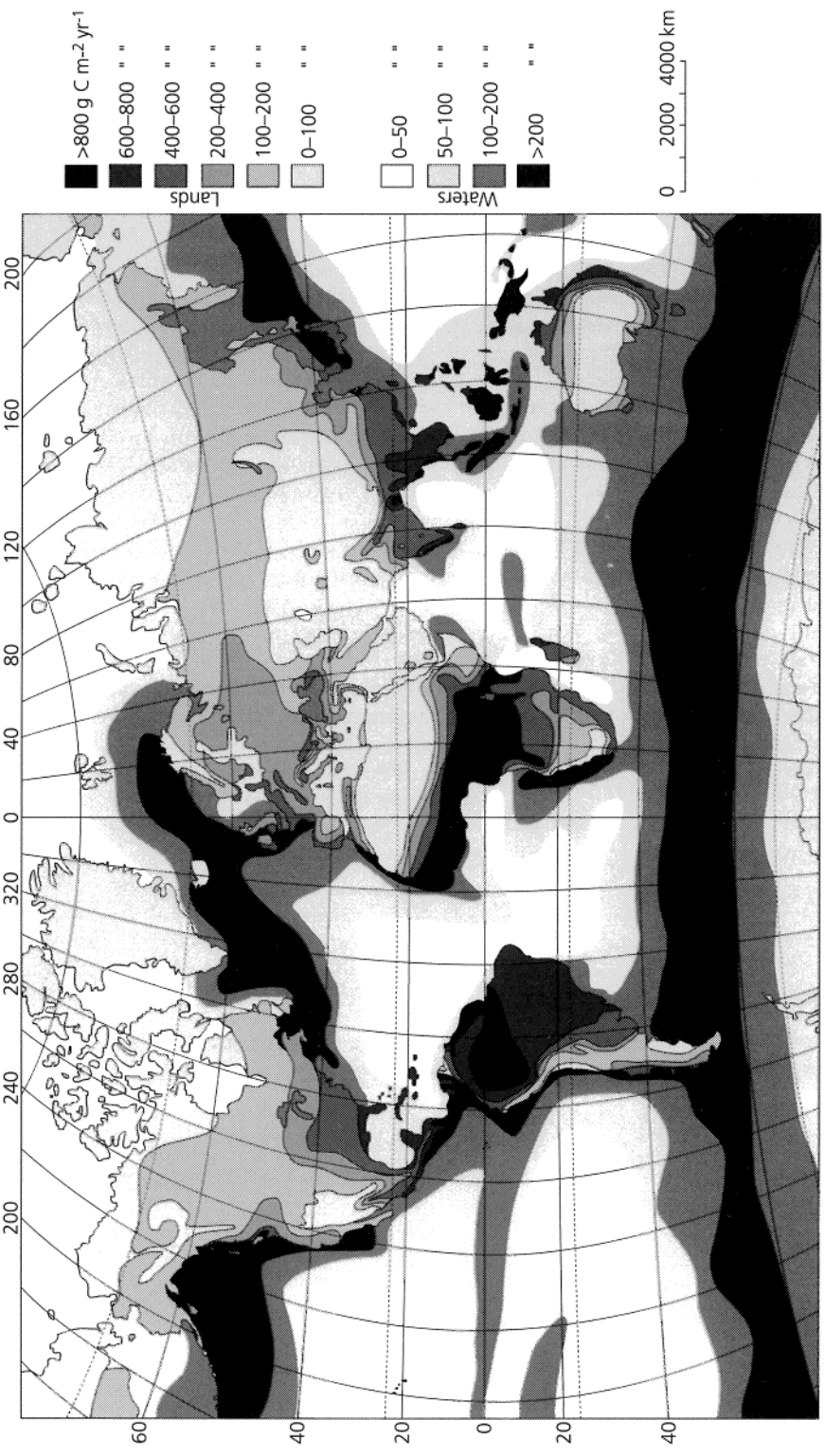


Fig. 1.22 Global net primary productivity in grams of carbon per square metre per year. After Reichle (1970) [25].

represents an extremely large quantity of energy fixed annually. The chemicals passing through the various pathways of the biosphere are involved in global geochemical cycles. Analysis of these cycles allows an estimation to be made of the fluxes involved, but a further significance is that the fluxes act as rate-limiting mechanisms. For example, the flux of nitrogen and phosphorus controls the rate of primary production of plankton and thereby of fish in the world's oceans. The flux involved in the weathering of rock controls the take-up by plant roots and therefore the production of grassland and forest. The pathways may be broken by long periods of residence of a chemical in a store, such as a deep sea sediment. This causes a slowing of the rate of geochemical cycling.

1.4.1 The carbon cycle

It is possible to study the geochemical cycles of a number of important elements such as phosphorus, nitrogen and sulphur. The approach taken here, however, is to focus briefly on the cycle driven by the opposing forces of respiration and photosynthesis involving carbon and oxygen. The main fluxes take place through the gaseous form of CO_2, but carbon is also fluxed through the activity of methane, CH_4. Methane is generated from wetlands and waterlogged soils. It is present in the atmosphere at very low concentrations (1.7 ppm), with a residence time of about 10 years.

There are a number of stores of carbon in the global carbon cycle, with fluxes between the reservoirs (Fig. 1.23). By far the greatest reservoir of carbon is the deep ocean (34 Gt), together with its shallow surface waters (0.9 Gt). The atmosphere contains just 0.74 Gt, and similar amounts are stored in the terrestrial biomass (0.55 Gt) and soil (1.2 Gt). A substantial amount of carbon (5 Gt) is locked away in fossil fuels. However, the fluxes of carbon between the various reservoirs portray a rather different picture, because although the deep ocean is an immense reservoir, it does not exchange its carbon rapidly with surface waters, whereas the atmosphere exchanges freely and rapidly with the oceanic surface waters, soil and terrestrial biota (Fig. 1.23).

Carbon is transferred to and from the ocean surface and atmosphere by diffusion. It is transferred in the form of CO_2 from the land surface by a combination of weathering of rocks, burning of forest and grassland, volcanic emissions, burning of fossil fuels and other industrial outputs. Atmospheric CO_2 is consumed over the land surface by the photosynthesis of terrestrial biota.

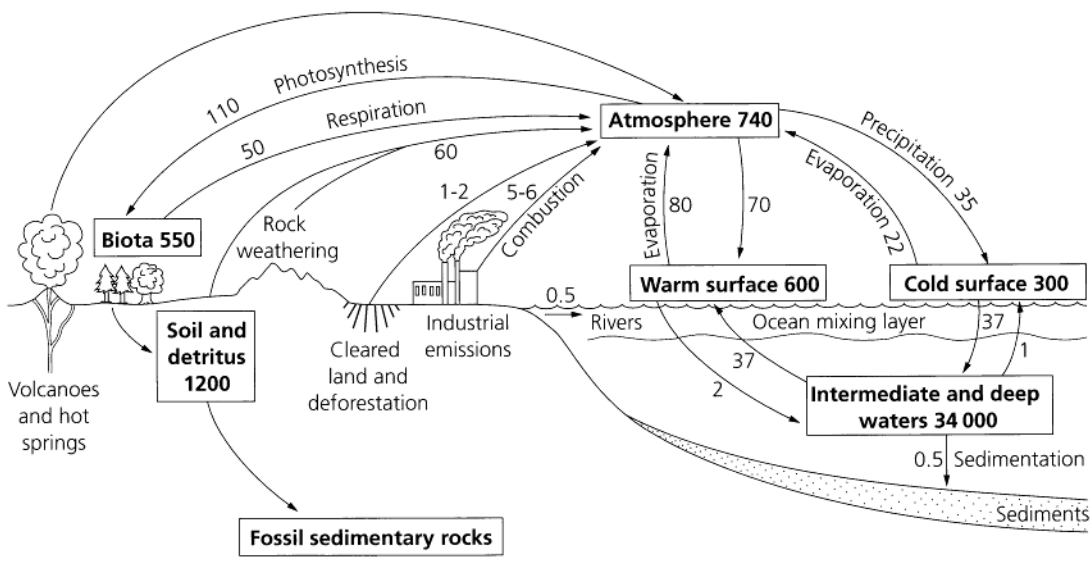


Fig. 1.23 The global carbon cycle with estimates of storages (in gigatonnes, boxes) and fluxes (in gigatonnes per year, arrows). From various sources, including Berner & Berner (1987) [1] and McClain *et al.* (1993) [26].

The major source of carbon for the biological system on Earth is therefore from atmospheric CO_2. Despite its small quantity (0.03% by volume) it is extremely important in its role in the radiation balance, especially in absorbing the long-wavelength radiation from Earth (Section 1.2). It may therefore be called a greenhouse gas. The concentration of CO_2 in the atmosphere has been rising steadily over recorded periods (the last 120 years), due mainly to the anthropogenic release of the gas from the burning of fossil fuels (see also Chapter 2). Some of the additional CO_2 released from the burning of fossil fuels has stayed in the atmosphere, but 40% has been taken up in two reservoirs in the carbon cycle which exchange rapidly with the atmosphere on a human time-scale—the surface waters of the oceans, and the terrestrial biosphere plus soils.

Carbon is present in the oceans as bicarbonate and carbonate ions. When atmospheric CO_2 is added to ocean surface water, it is converted to bicarbonate. This exchange occurs rapidly. The surface waters of the ocean mix only slowly with the waters below the thermocline (depth of about 1000 m), so the exchange time of atmospheric CO_2 with the deep sea is very long. The activities of carbon-bearing organisms in the surface waters of the oceans, principally plankton, affect the coupling between atmosphere and ocean. Marine plankton fix in their tissues and skeletons a high proportion of the carbon flux. The size of the marine plankton biomass therefore affects the ability of the biomass to draw out of the carbon cycle (or sequester) the carbon present in the atmosphere.

Availability of important biolimiting nutrients such as nitrogen and phosphorus exerts a major control on the global carbon cycle through its effect on ocean productivity. Nutrient availability therefore has impact on ocean–atmosphere coupling and global climate change through the build-up or abstraction of the greenhouse gas CO_2.

It has been stated that the biosphere is 'hungry' for phosphorus and nitrogen [27], since the levels potentially used by the biosphere are far higher than average concentrations in the hydrosphere and lithosphere. Where light and temperature are optimal, the availability of these ions therefore limits primary production. Where P and N are plentiful, as in coastal and shelfal waters, primary productivity is also high.

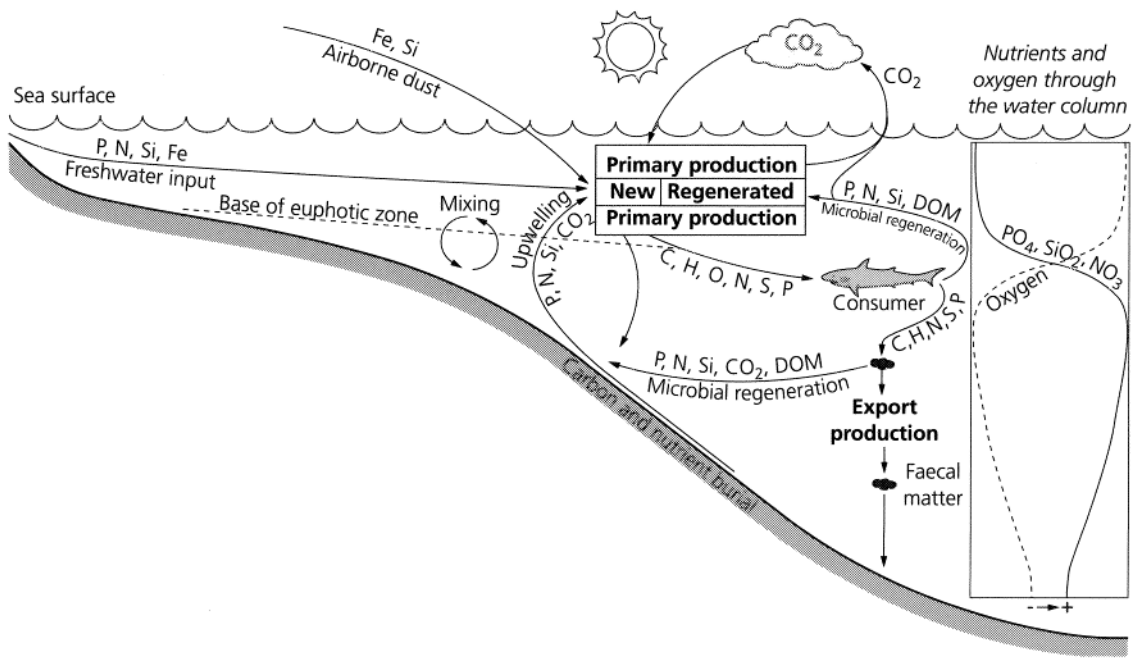


Fig. 1.24 Nutrient cycling in the ocean, showing the importance of the biological carbon 'pump'. Variations with depth of the biolimiting nutrients PO_4, SiO_2 and NO_3, and oxygen concentration are shown on the right. DOM is dissolved organic matter. After Brasier (1995) [27].

Nutrients are made available by runoff from the land or by mixing with deeper waters. Part of the organic matter involved in gross primary productivity is removed or 'exported' from surface waters by a biological 'pump' driven by grazing by zooplankton and fish, and sinking to the sea bed of faecal pellets (Fig. 1.24). On the sea floor the organic material may be used for further biological activity (deposit feeding, bacterial respiration) or locked into the sedimentary record as a black shale, phosphorite or biogenic chert, all of which are common under regions of high rates of export of productivity. Below the productive surface waters, nutrients and CO_2 are returned to the water column by microbial regeneration.

The most likely way of bringing about large changes in atmospheric CO_2 is, therefore, by varying the rate of carbon fixation by primary productivity in the ocean under the influence of biolimiting nutrients. Zones of upwelling may be crucial in this process, since their high productivities cause a removal of CO_2 from the atmosphere, and carbon burial in sea bed sediments. The conditions leading to upwelling, such as the intensity of wind stress (see Chapter 9), are important elements in the larger picture of global climate change through the impact on nutrient supply, productivity and sequestering of atmospheric CO_2.

There is also a rapid exchange of atmospheric CO_2 with the terrestrial biosphere. Increased storage in the terrestrial biosphere, however, is counteracted by the effects of *deforestation* which leads to an addition of CO_2 in the atmosphere (see also Section 2.5). The problem of net loss or net uptake by the biosphere is presently unsettled, but the flux is thought to be much smaller than the uptake of CO_2 by the oceans.

The increase in CO_2 (and other greenhouse gases) in the atmosphere has climatic and hydrological effects. Climatic greenhouse models predict that a doubling of the CO_2 content of the atmosphere would result in a worldwide increase in the mean annual temperature of between 1.5 and 3°C [28]. As we have previously seen (Section 1.2), some of these processes act in a positive feedback.

1.5 Topography and bathymetry

The topography of the Earth's surface and the bathymetry of its oceans (Plate 1.1, facing p. 204) are determined by the interplay of the primary 'tectonic' mechanisms and the modifying exogenic forces causing denudation and deposition. Clearly, topography can be viewed at a range of scales. With our global perspective, the topography at small scales, such as that of an individual hillslope, is not of great importance. At the medium scale, topography may be dominated by endogenic processes causing, for example, the formation of volcanic edifices, or fault scarps at the Earth's surface. At a still larger scale, however, the primary mechanisms for the topography of continents and the bathymetry of ocean basins are related to the horizontal and vertical motions of the lithospheric plates. Plate tectonic theory proposes that the lithospheric plates are sufficiently rigid to move over the surface of the Earth as reasonably coherent entities, their relative motion causing deformation to be concentrated at their edges.

However, plate convergence may result in thickening of the lithosphere very far from plate boundaries, suggesting that the continental plates may act at large time and spatial scales more like viscous sheets undergoing a continuum of deformation. Whether concentrated at plate boundaries or widely distributed into plate interiors, the crustal and lithospheric thickness changes or lateral variations in density caused by the relative motion of plates result in the major features of the Earth's topography. The driving forces for the relative motion of the lithospheric plates are not fully understood, but are probably dominated by the negative bouyancy of cold oceanic lithosphere being subducted into the Earth's interior. Such deep subduction is part of a slow and very large-scale convection of the mantle. Other large-scale topography appears not to be related to relative plate motion *per se*, but to the direct interaction between the underlying mantle and the overlying plates. The uplift of the oceanic and continental lithosphere over hotspots is an example. The mechanisms for the support of topography can therefore be broadly classified as:

- **isostatic**—the response to density and thickness variations in the lithosphere. Such variations result primarily from the horizontal motions of plates.
- **dynamic**—the direct response to processes involving flow in the viscous mantle. Large-scale three-dimensional variations in mantle density are most likely caused by whole-mantle thermal convection. Smaller-scale heterogeneity is probably caused by mantle plume activity.

1.5.1 The shape of the Earth

The Earth is not a perfect sphere. If it were a perfect sphere, the gravitational acceleration would be the same at every point on its surface, but it is known that this is not the case. Instead, when the effects of rotation are taken into account, gravity observations show the Earth to be approximated by an *oblate*

spheroid, the shape generated by rotating an ellipse around its shorter axis. There is a bulge at the equator and a flattening at the poles. The oblate spheroid is the theoretical gravitational *equipotential surface*. However, at a smaller, subglobal scale, the Earth shows important deviations from the oblate spheroid. The actual equipotential surface for the Earth, which can be thought of as the level to which still water accumulates in the ocean, is the so-called *geoid*. The geoidal surface can be viewed as the ultimate base level for terrestrial denudation.

Lateral variations in the density distribution within the Earth give rise to deviations of the geoid height from the reference oblate spheroid; these are called *geoid height anomalies*. At long wavelength, the geoidal surface shows some impressive departures from the reference oblate spheroid, with a deep (−80 m) geoid low in the Indian Ocean and a geoid high (+60 m) over the western Pacific (Plate 1.2, facing p. 204). Geoid height anomalies are accompanied by gravity anomalies, troughs in the geoid being associated with negative gravity anomalies (mass deficit at depth) and peaks in the geoid being associated with positive gravity anomalies (mass excess). Geoid height anomalies can therefore be used to calculate the variation in density with depth.

1.5.2 Isostatic topography

Hypotheses of Airy and Pratt

Isostasy is a principle requiring that the rigid surface layer of the Earth, the lithosphere, is able to 'float' on a fluid substratum, the aesthenosphere. It is assumed that below a certain depth, the *depth of compensation*, all the pressures are equal and hydrostatic (the same in all directions). The excess mass represented by the topography of a mountain is therefore compensated for by a mass deficiency beneath it at a depth above the compensation depth. Conversely, the mass deficit of a deep trough at the surface may be compensated for by a dense mass at depth above the compensation depth. Two competing hypotheses were proposed in the mid-1850s. The *Airy hypothesis* involves compensation by variations in the thickness of the upper layer, rather like blocks of wood floating in a tank of water (Fig. 1.25a) so that

$$h_{root} = \frac{h_{mt}\rho_1}{\rho_2 - \rho_1} \qquad \textbf{(1.14)}$$

where ρ_1 and ρ_2 are the densities of the upper and lower layers respectively, h_{root} and h_{mt} are the depth of the root below the level of the base of the upper layer associated with a surface at sea level, and the

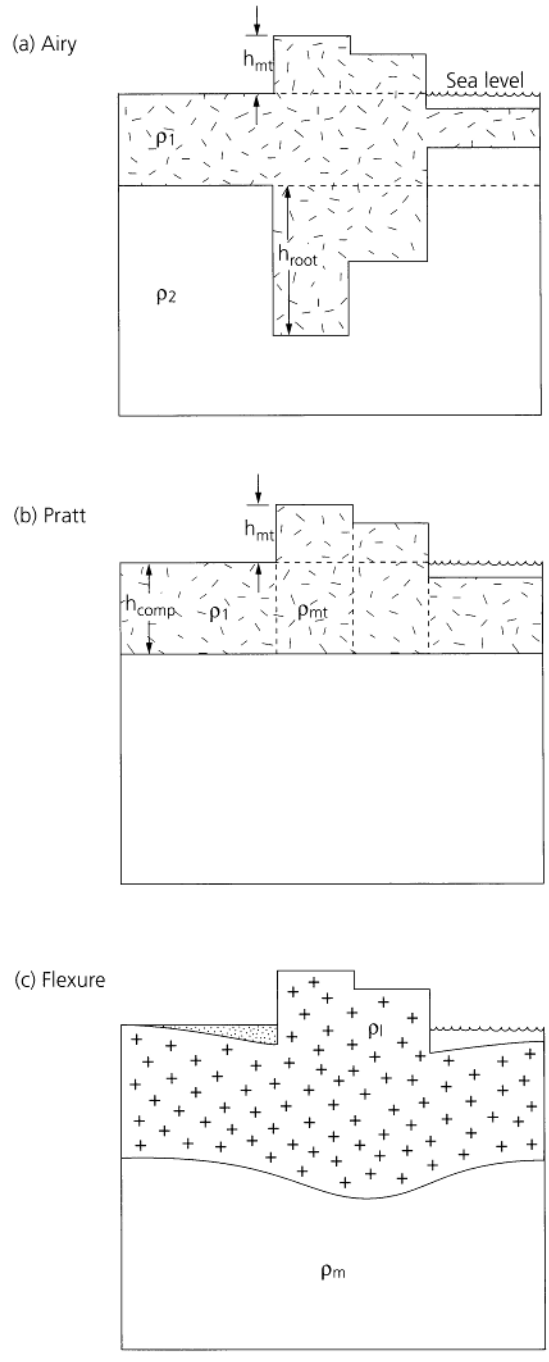


Fig. 1.25 Isostasy. Contrasting (a) Airy, (b) Pratt and (c) flexural models. The notation is explained in the text.

height of the mountain above sea level, respectively. The upper layer is customarily taken as the crust and the lower layer as the mantle lithosphere. Since the aesthenosphere is weak and can tolerate only small lateral pressure gradients, the depth of compensation must lie at relatively shallow depths within or at the base of the lithosphere.

In the *Pratt hypothesis* the base of the upper layer is assumed to be constant, isostatic equilibrium being achieved by density variations within the upper layer (Fig. 1.25b). In this hypothesis mountains are not compensated by possessing a root, but by being underlain by light material compared to adjacent regions. In such a situation

$$\rho_{mt} = \rho_1 \left(\frac{h_{comp}}{h_{mt} + h_{comp}} \right) \quad \textbf{(1.15)}$$

where h_{comp} is the depth of compensation (below sea level) and ρ_{mt} is the density of material underlying the mountain. In the Pratt hypothesis oceans are underlain by denser material than adjacent regions.

Flexural isostasy

In both the Airy and Pratt hypotheses the pressures at the depth of compensation under imaginary, narrow, vertical columns of rock are constant. Compensation can be thought of as local. However, there are instances where the upper part of the lithosphere appears to behave elastically by a process of long-wavelength flexure (Fig. 1.25c). A mountain range or an oceanic seamount chain acts as a load, causing a regional flexure, the wavelength of which depends on the rheological properties of the plate, specifically its *flexural rigidity* (or elastic thickness). The characteristic response of an elastic lithosphere to an applied load system is of a maximum deflection under the centre of gravity of the load, the deflection decreasing laterally and passing into a region of relative uplift, the *flexural forebulge*. The sea floor bathymetry around oceanic islands such as the Emperor-Hawaiian seamount chain clearly shows the expression of the flexural moat and flanking bulge.

The way in which the lithosphere compensates flexurally for the excess mass of a mountain belt (or any other load) is determined by the rigidity of the plate, but also by the wavelength of the load (Fig. 1.26). Consider a sinusoidally varying load of wavelength λ at the Earth's surface with a maximum amplitude h_0 which produces a sinusoidal deflection beneath it of the same wavelength and amplitude w_0. If the flexural rigidity of the elastic lithosphere is D, then

$$w_0 = \frac{h_0}{\frac{\rho_m}{\rho_\ell} - 1 + \frac{D}{\rho_\ell g}\left(\frac{2\pi}{\lambda}\right)^4} \quad \textbf{(1.16)}$$

where ρ_m is the density of the mantle displaced by the downward deflection of the elastic lithosphere, ρ_ℓ is the density of the material comprising the load, and $2\pi/\lambda$ is termed the *wavenumber*. If the wavelength of the load is short, the denominator becomes very large, indicating that the deflection is very small compared

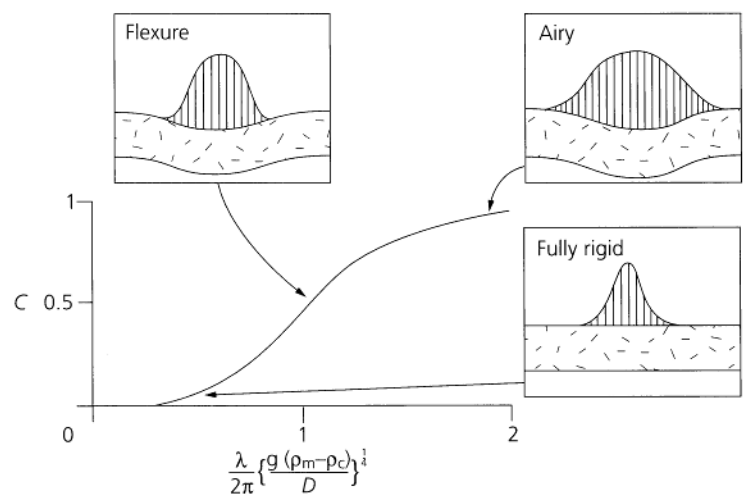


Fig. 1.26 Dependence of degree of compensation C on the dimensionless wavelength of periodic (sinusoidal) topography. Modified from Turcotte & Schubert (1982) [29]. D is the flexural rigidity, λ the wavelength of the load, ρ_m and ρ_c the mantle and crustal densities respectively.

to the load. The lithosphere appears to behave very rigidly to loads under these circumstances. However, if the wavelength of the load is large, the equation for the maximum deflection simplifies to

$$w_0 = \frac{\rho_\ell h_0}{(\rho_m - \rho_\ell)} \quad (1.17)$$

which is recognizable as the solution for Airy (local) isostasy. At long wavelengths, therefore, the lithosphere appears to have little flexural strength. We can define the *compensation* as the ratio of the actual deflection to the maximum (Airy) deflection. It is illustrated in Fig. 1.26.

In summary, loads of very long wavelength are highly compensated, the lithosphere appearing in such cases to be very weak in its ability to support loads. Consequently, the load represented by the water in the Pacific Ocean is compensated locally by the underlying oceanic lithosphere. Loads of small wavelength, however, may be uncompensated, the lithosphere appearing to be extremely rigid in supporting the load. At intermediate wavelengths the lithosphere deforms by flexure under the load. Flexural depressions and bulges are commonplace in the ocean, either at the trenches of subduction zones or the moats surrounding seamounts, and on the continents as foreland basins adjacent to mountain belts.

Apart from mountain belts, some of the greatest topography on the surface of the Earth is associated with continental rifting (e.g. E Africa) and with continental margins (e.g. SW Africa) (Plate 1.4). Both of these settings involve stretching, and therefore thinning, of the lithosphere. It is therefore important to note that considerable topography may be associated with both mountain building at zones of collision and extension at sites of rifting or continental margin development.

Flexural uplift of rift flank and ocean margins Studies of the elastic thickness (flexural rigidity) of regions undergoing active extension such as East Africa suggest that the lithosphere has a finite rigidity rather than behaving according to local (Airy) isostasy. The unloading of the lithosphere by extension along major rift-bounding crustal faults may therefore cause flexural uplift (Fig. 1.27). This may explain some rift flank and continental margin uplifts, such as the South African Atlantic margin.

The wavelength of the rift flank uplifts should depend on the flexural rigidity of the lithosphere being unloaded, the amplitude depending on the magnitude of the unloading. In the case of small grabens in Tibet, flexural isostatic compensation is thought to be facilitated by flow in a viscous lower crust, giving a short-wavelength rift margin uplift, rather than in the aesthenosphere, which would produce the longer-wavelength feature recognized in the flank uplifts of the Baikal Rift (Siberia) and Shanxi Rift (China) (Fig. 1.27). If so, it would lend support to the idea that the upper crust and mantle are capable of decoupling along a low-viscosity lower crust.

Practical exericse 1.2: Flexural isostasy

1 The Alps of western Europe have a width of about 150 km and are responsible for flexing down the underlying lithosphere. Assuming the density of the load to be 2800 kg m^{-3}, the density of mantle to be 3200 kg m^{-3}, and flexural rigidity 10^{23} N m, what is the degree of compensation of a sinusoidal load represented by the Alpine mountain belt?

2 The topographic load represented by the Tibetan plateau has an average width of 1400 km. Assuming the same density terms as above, and a flexural rigidity of 5×10^{23} N m, what is the degree of compensation for the Tibetan load?

Solution

1 The wavelength of the Alpine load is approximately twice the width, $\lambda = 300$ km, giving a wavenumber of $2\pi/\lambda = 1.925 \times 10^{-5}$. Consequently, the degree of compensation is $C = 0.17$. This suggests that the lithosphere responds flexurally to the Alpine load.

2 The wavelength of the Tibetan load is approximately 2800 km, giving a wavenumber of $2\pi/\lambda = 2.24 \times 10^{-6}$. The degree of compensation is $C = 0.99$, indicating that the Tibetan Plateau can be treated as being in Airy isostasy.

Isostasy in zones of convergence

A vivid illustration of the topographic effects of mantle and lithospheric processes is provided by the site of collision of the Indian 'indenter' and the continent of Eurasia (Fig. 1.28). The date of collision of India and Eurasia is thought to be about 50 Ma, with approximately 2900 (±900) km of convergence since that time being accommodated by the thickening of crust both in the Himalaya and the Eurasian area beyond. This thickening may result from the underthrusting of the intact Indian plate beneath the

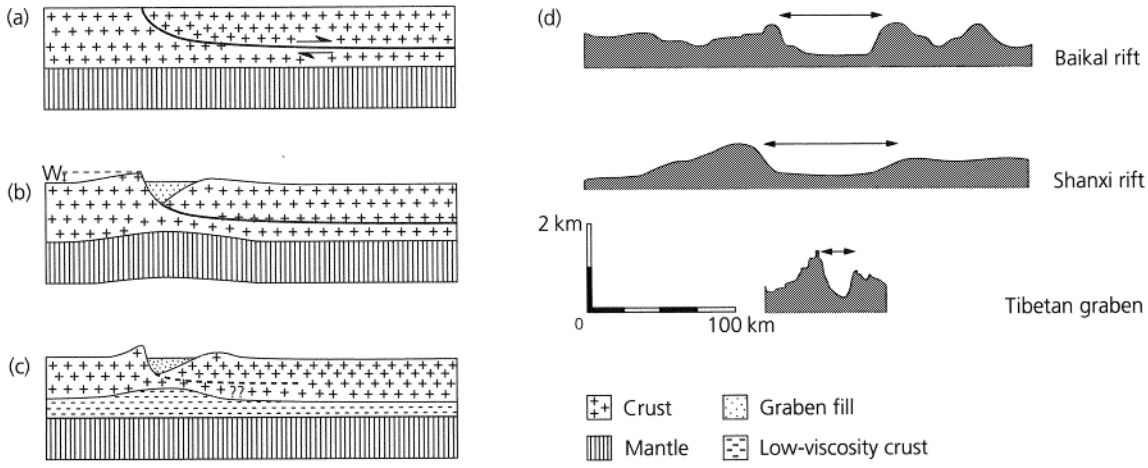


Fig. 1.27 Effect of the depth of the low-viscosity layer on the wavelength of rift flank uplift. (a), (b) and (c) illustrate basic concepts. (a) Extension along fault soling in lower crust creates graben filled with sediment. Regional compensation for the negative load of the graben (compared to normal thickness crust) produces margin uplifts of height *w*. In (b) the crustal viscosity is high and compensation takes place in the aesthenosphere, giving a long flexural wavelength. In (c) the crust has a low viscosity, compensation takes place in the crust itself and the wavelength is short. (d) Comparison of topographic profiles shows the Baikal (Siberia) and Shanxi (China) rifts to have long-wavelength flank uplifts, whereas the graben in the Tibetan plateau have very short-wavelength uplifts. After Masek *et al.* (1994) [30].

frontal edge of Eurasia, or from widespread north–south shortening of Eurasia's crust. The Tibetan plateau, at about 5.5 km average elevation, is the isostatic result of this crustal thickening.

Where convergence takes place over broad zones, it is likely that the thickening does not take place by the simple thrusting of one plate beneath the other, but by a wholesale thickening of the entire lithosphere (Fig. 1.29). In such a case, the cold, mantle lithosphere would also contribute to a thickened root. Since this root of mantle lithosphere is denser than the aesthenosphere at the same depth, it possesses negative bouyancy. An isostatic balance can be carried out as follows [31].

Consider crust of thickness h_c and density ρ_c overlying a mantle which increases in temperature linearly from the base of the crust to the top of the aesthenosphere at depth h_0 with a gradient γ. The temperature of the aesthenosphere is about 1300°C, corresponding to a density of 3200 kg m^{-3} (decompressed to surface pressures). The density of the mantle lithosphere decreases with depth below the crust because of the progressive warming at a rate determined by the coefficient of thermal expansion α,

$$\rho_\ell = \rho_a\{1 + \alpha T_\ell\} \qquad \textbf{(1.18)}$$

where ρ_a is the density of the aesthenosphere, ρ_ℓ is the density of the mantle lithosphere and T_ℓ is its temperature. Hence at a depth y

$$\rho_\ell(y) = \rho_a\{1 + \alpha\gamma(h_0 - y)\} \qquad \textbf{(1.19)}$$

Now if we instantaneously thicken this lithosphere by a factor f, we can isostatically balance the rock column of the thickened lithosphere with the initial lithosphere down to a depth equal to the depth of the thickened lithosphere. After thickening the new crustal thickness becomes $h_c f$, the new lithospheric thickness becomes $h_0 f$, the new temperature gradient becomes γ/f. However, this is accompanied by an elevation change of the overlying mountain or plateau Δh in response to the increased lithospheric root. If h is the thickness of the increased root of mantle lithosphere, then

$$\Delta h = h_0(f - 1) - h \qquad \textbf{(1.20)}$$

Performing an isostatic balance down to a depth $h_0 + h$,

$$\Delta h = (f - 1)\left\{\frac{(\rho_a - \rho_c)h_c}{\rho_a} - \frac{1}{2}\alpha\gamma(h_0 - h_c)^2\right\} \qquad \textbf{(1.21)}$$

where the first term in the curly bracket refers to the

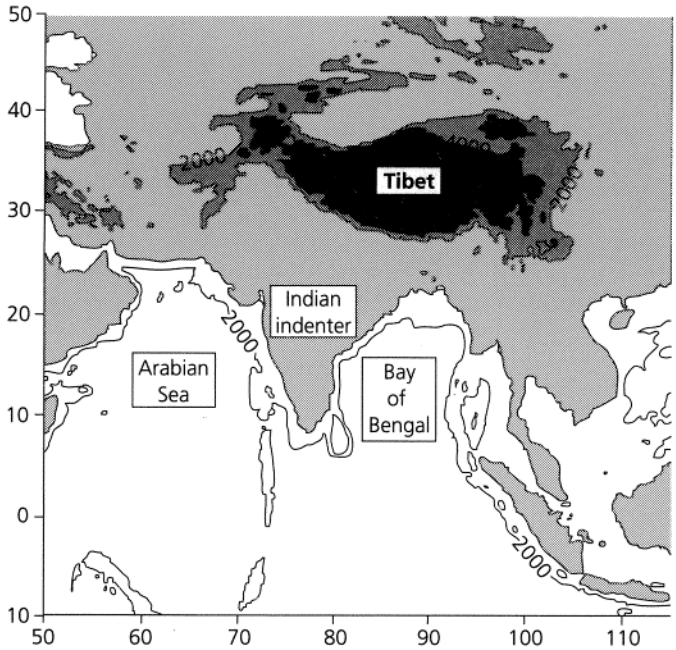


Fig. 1.28 Location of the collision zone of the Indian indenter with Asia and the resulting topography. Bathymetric contour at –2 km and topographic contours at 2 km and 4 km are shown. After Molnar *et al.* (1993) [31].

effect of crustal thickness changes (Airy isostasy), and the second gives the effect of the thickened root of mantle lithosphere.

The downward penetration of a cold lithospheric root into the aesthenosphere causes lateral temperature gradients in the aesthenospheric mantle which may cause the development of convective instabilities. Consequently, the deep lithospheric root may be swept away by detaching from overlying stronger lithosphere, and be replaced by hotter, more buoyant material. This should have the isostatic response of surface uplift. The removal of the cold lithospheric root should result in a surface elevation change (uplift) equivalent to but with opposite sign to the second term in equation (1.19), i.e. *c.* 1.7 km in the practical exercise.

This explains the present surface topography of Tibet. The time-scale of the uplift due to convective removal of the lithospheric root is thought to be relatively short compared to the time-scale of the initial thickening (20–40 My). Removal is thought to have taken place during the late Miocene (*c.* 6–9 Ma),

Practical exercise 1.3: The elevation of Tibet

The average elevation of the Tibetan plateau today is 5.5 km. The crust appears to have been thickened by a factor of 2. Assuming that the lithosphere has thickened by the same factor, what is the predicted elevation change of the Tibetan plateau resulting from the thickening? If only crustal thickening takes place, what is the prediction from Airy isostasy? Use the following parameter values:

$h_c = 35$ km

$h_0 = 135$ km

$\rho_c = 2800$ kg m^{-3}

$\rho_a = 3200$ kg m^{-3}

$\alpha = 3.5 \times 10^{-5}$ °C^{-1}

$\gamma = 15$ °C km^{-1}

Solution

The elevation change caused by thickening the entire lithosphere by a factor of 2 is 3.6 km. This is made of the components due to Airy isostasy (5.3 km) and due to the temperature differences in the mantle lithosphere (–1.7 km). Thickening of the Asian lithosphere by a factor of 2 should have resulted in a plateau with a lower elevation than observed today. The prediction from Airy isostasy, based on crustal thickening only, is much closer to the observed elevation. This is discussed further below.

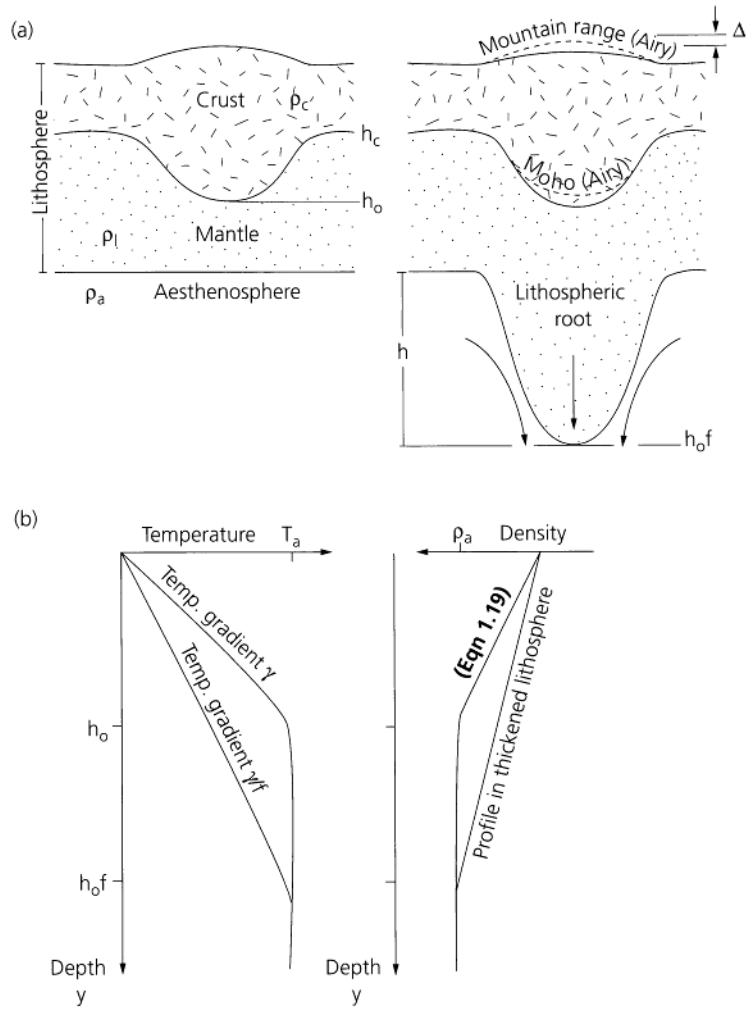


Fig. 1.29 Consequences of lithospheric thickening under Tibet, after Molnar *et al.* (1993) [31]. (a) Cross-sections to illustrate the difference between crustal thickening (left) and thickening of the entire lithosphere (right). The excess mass in the thickened lithospheric root pulls the overlying crust down and lowers the topography of the mountain belt. (b) Profiles of temperature (left) and density (right) in normal and thickened lithosphere.

and is synchronous with the onset of the Asian monsoon [31].

A similar but not identical mechanism has been invoked to explain the uplift of the Colorado plateau, western USA. The mechanism envisaged has been termed *continental delamination* to emphasize that the entire mantle lithosphere peels off from the overlying crust and sinks into the aesthenosphere [32].

Isostasy at zones of extension

The active seismicity, presence of large normal faults at the surface, high surface heat flows, generally alkaline volcanic activity, elevated Moho indicated by deep seismic investigations, and negative Bouguer gravity anomalies suggest that continental rifts are sites of lithospheric stretching. The following two models satisfy a wide range of geological and geophysical observations:

1 'Active', 'mantle-activated' or 'open system' rifting caused by thermal activity in the mantle, and characterized by large volumes of extrusive rocks. Active rifts are probably related to hotspot activity in the mantle, which leads to lithospheric heating, thinning, the release of melts by decompression and rift formation.

2 'Passive', 'lithosphere-activated' or 'closed-system' rifting, where the upwelling of aesthenosphere is a passive response to, not a cause of, mechanical stretching of the lithosphere. Passive rifts have little or no volcanism.

These two models have somewhat different dynamics, active rifting being dealt with in Section 1.5.3. Passive rifting presents the simplest scenario since we

can ignore any temperature and density variations in the aesthenosphere (Fig. 1.30). Let us assume that before stretching the crust has a thickness h_c, and lithosphere thickness h_ℓ. If the crust and lithosphere are stretched uniformly by a stretch factor β, the new crustal thickness becomes h_c/β and the new lithospheric thickness becomes h_ℓ/β. There is an elevation change Δh at the surface of the Earth above the region of stretching which depends on the amount of stretching and on the ratio h_c/h_ℓ. Let the average crustal, mantle lithosphere and aesthenosphere densities be ρ_c, ρ_m and ρ_a respectively, and the density of the material infilling the space caused by the elevation change be ρ_i. The pressure under the lithospheric column before rifting is therefore

$$h_c\rho_c g + (h_\ell - h_c)\rho_m g \qquad \textbf{(1.22)}$$

Because the acceleration due to gravity is common to all lithospheric columns we can ignore it in the following analysis. After stretching, the pressure at the depth of the original lithospheric thickness (ignoring g) is

$$\Delta h\rho_i + \left(\frac{h_c}{\beta}\right)\rho_c + \left(\frac{h_\ell - h_c}{\beta}\right)\rho_m + \left(h_\ell - \frac{h_\ell}{\beta} - \Delta h\right)\rho_a \qquad \textbf{(1.23)}$$

Equating the two columns before and after rifting and regrouping the terms,

$$\Delta h = \frac{(1-1/\beta)}{(\rho_a - \rho_i)}\left\{h_\ell\rho_a - h_c\rho_c - (h_\ell - h_c)\rho_m\right\} \qquad \textbf{(1.24)}$$

It only remains to specify the density structure in the region of stretching. Let us assume that the crust can be given a single value of ρ_c that remains constant, but that the mantle lithosphere changes in density as a function of its temperature, from the aesthenospheric temperature at its base, with a geothermal gradient given by $\gamma_0\beta$, where γ_0 is the original linear geothermal gradient. The average temperature of the mantle lithosphere is therefore

$$\rho_m = \rho_a\left\{1 + \alpha\left(T_a - \gamma_0\frac{h_\ell - h_c}{2}\right)\right\} \qquad \textbf{(1.25)}$$

where T_a is the aesthenospheric temperature. It will be

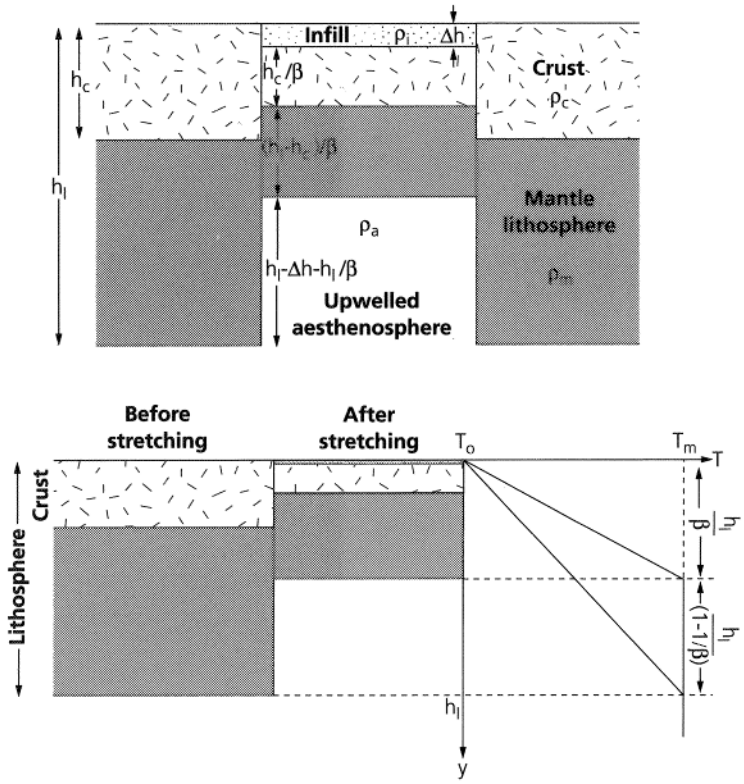


Fig. 1.30 Notation and thermal consequences of simple uniform stretching of the lithosphere.

seen in the practical example below that the effect of the temperature increase on the density of the mantle lithosphere is minimal and can in most cases be neglected.

Practical exercise 1.4: Subsidence in rifts

Calculate the elevation change in a region of passive rifting for the following parameter values:

$\beta = 2$

$h_c = 35$ km

$h_\ell = 125$ km

$\rho_c = 2800$ kg m^{-3}

$\rho_a = 3200$ kg m^{-3}

$T_a = 1333$°C

$\alpha = 3.5 \times 10^{-5}$ °C^{-1}

$\gamma_0 = 30$ °C km^{-1}

Solution

The elevation change is 3.1 km subsidence for a water filled rift basin, or 4.9 km for a rift basin filled to the brim with sediment of average bulk density 1800 kg m^{-3}.

Unless the initial crustal thickness is very small compared with the lithospheric thickness, lithospheric stretching should therefore be accompanied by large amounts of subsidence. Yet we know that active rift zones such as the East African rift are located on elevated topographic domes. The simplest way to account for this is to increase the heat flow from the mantle. In the case of active rifting, we may increase the aesthenospheric temperature by 100–200°C compared to its normal temperature, by placing a hot plume head under the lithosphere (Section 1.5.3).

Thermal subsidence following stretching Stretching of the lithosphere results in an elevation of the aesthenosphere to shallower depths determined by the stretch factor. Consequently, the entire lithospheric column is heated (Fig. 1.31). Once the source of the stretching ceases, the isotherms in the lithosphere relax to their pre-stretching position. This results in thermal contraction of the lithosphere and subsidence. Thermal subsidence may characterize a post-rift stage of continental rifts where the rift zone and margins sag after rifting is aborted, or the history of continental margins where rifting has been successful in completely splitting the continent (passive margins).

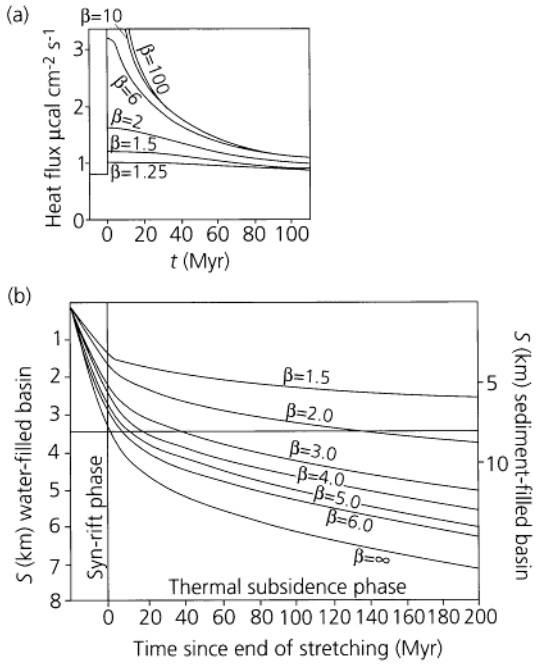


Fig. 1.31 The heat flow and thermal subsidence following stretching as a function of the stretch factor β. (a) The magnitude of the stretching-related heat flux is dependent on the stretch factor, but at long periods after the cessation of stretching, the heat fluxes are very similar irrespective of the amount of stretching. (b) The rate of thermal subsidence decreases exponentially with time. The curves shown refer to a lithosphere of thickness 125 km and crust of thickness 31.2 km. After McKenzie (1978) [33].

The form of the thermal subsidence phase of sedimentary basins is similar to that of the cooling of the oceanic lithosphere at mid-ocean ridges (Section 1.5.5). However, we should expect the stretch factor β to determine the heat added to the lithosphere during stretching, and therefore the heat lost after stretching. The form of the thermal subsidence over time must also be affected by the thermal material properties of the lithosphere.

The thermal subsidence as a function of time t is given by

$$S(t) \approx E \frac{\beta}{\pi}\left(\sin\frac{\pi}{\beta}\right)\left(1 - e^{-t/\tau}\right) \qquad \textbf{(1.26)}$$

where

$$E = \frac{4 h_\ell \rho_a \alpha T_a}{\pi^2 (\rho_a - \rho_i)}$$

and τ is the thermal time constant of the lithosphere

$$\tau = \frac{h_\ell^2}{\pi^2 \kappa} \qquad (1.27)$$

where κ is the *thermal diffusivity* (in metres squared per second).

We can now calculate the thermal subsidence history for a rift by taking the parameter values from Practical Exercise 1.4, and adding the thermal diffusivity $\kappa = 10^{-6}\,m^2\,s^{-1}$. Assuming the basin to be filled with water ($\rho_i = 1000\,kg\,m^{-3}$), the thermal subsidence follows an exponential curve (Fig. 1.31). After one lithospheric time constant, about 50 My, the thermal subsidence is 1.383 km. Failed rifts and continental margin basins are commonly fully or partially filled with sediment during the thermal subsidence phase. These sediment loads drive further subsidence which amplifies the thermal contraction signal. As a result, some continental margins such as the US eastern seaboard have up to 16 km of preserved sediment. Failed or aborted rifts such as the North Sea in NW Europe contain 3–4 km of post-rift sediment overlying the fault-bounded syn-rift sediments. Using the example above where $\beta = 2$, for a basin filled with sediment of density $1800\,kg\,m^{-3}$, the thermal subsidence after 50 My is 2.173 km.

1.5.3 The bathymetry of the ocean floor

Inspection of a bathymetric chart of the ocean shows immediately that the ocean gradually deepens away from the mid-ocean ridges. It is generally accepted that the bathymetry of the sea floor is controlled at first order by the thermal contraction of oceanic lithosphere formed at the mid-ocean ridge system.

When new ocean floor is formed at mid-ocean ridges it cools as it moves away from the site of spreading. We can treat the cooling of the new oceanic crust by the cold seawater as instantaneous (Fig. 1.32). The distance over which the sudden cooling is felt after a time t is determined by the thermal diffusivity κ of the ocean crust. This thermal diffusion distance must be equal to $\sqrt{\kappa t}$. The cooling new oceanic material forms the oceanic lithosphere as a cold thermal boundary layer above the easily deformed aesthenospheric mantle. The base of the new oceanic lithosphere can therefore be defined by a

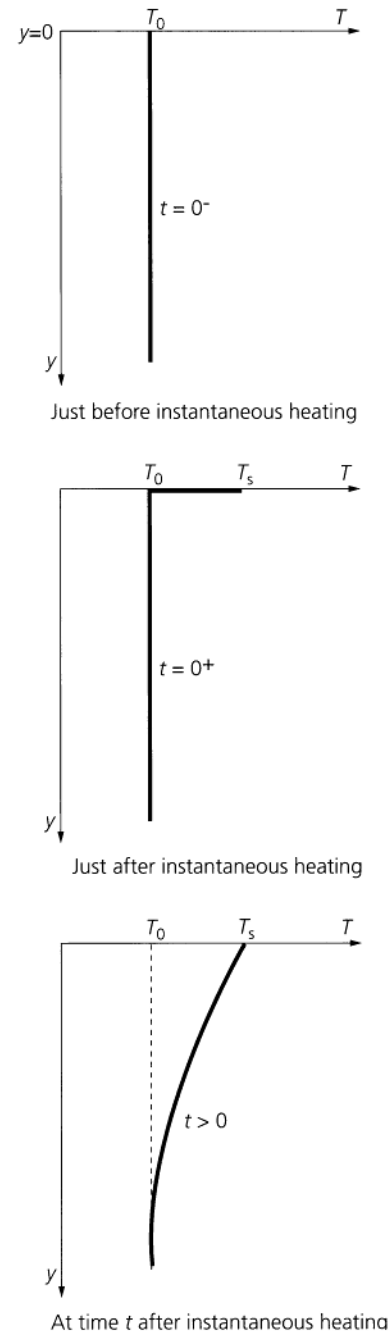


Fig. 1.32 *Right.* Heating of a half-space by an instantaneous increase of temperature from T_0 to a new surface temperature T_s. This might apply to the temperature field following the injection of a dyke into cool country rocks. A similar solution can be applied to the sudden cooling of hot mantle rock injected at a mid-ocean ridge.

characteristic isotherm (1330°C, or about 1600 K), and its thickness is clearly a function of its age. If the spreading rate is u and the horizontal distance from the ridge crest is x, the age is x/u, and the thermal diffusion distance $\sqrt{\kappa x/u}$.

Introducing a dimensionless temperature ratio

$$\theta = \frac{T - T_0}{T_s - T_0} \tag{1.28}$$

where T_0 is the initial temperature (the aesthenospheric temperature), T_s is the constant temperature of the space into which the ocean lithosphere is emplaced (the seawater temperature), and T is the temperature at time t, the solution for the temperature is then

$$\theta = \text{erfc}\left(\frac{x}{2\sqrt{\kappa t}}\right) \tag{1.29}$$

where erfc is the complementary error function.

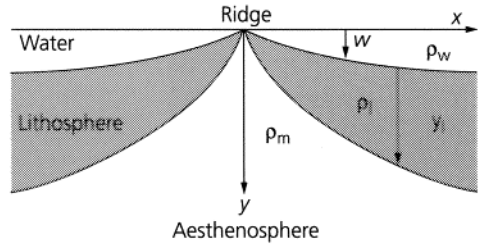
The cooled oceanic material forms a thermal boundary layer. It is a largely arbitrary choice of the definition of the boundary layer, but if we define it as the thickness to where $\theta = 0.1$, it is found that the thermal boundary layer is 2.32 times the thermal diffusion distance $\sqrt{\kappa t}$ or $\sqrt{(\kappa x/u)}$. The thickness of the oceanic lithosphere at an age of 50 My, taking the thermal diffusivity as $10^{-6}\,m^2\,s^{-1}$, is therefore approximately 92 km. The oceanic lithosphere should increase parabolically in thickness with age, and therefore also with distance from the ridge crest.

By cooling, the oceanic lithosphere contracts, increases its density and exerts a higher stress on the underlying mantle. An isostatic balance shows that the sea floor must therefore subside. Any lithospheric columns through the oceanic lithosphere can be balanced isostatically to give the depth of the ocean floor as a function of distance from the ridge crest, or time (Fig. 1.33).

The observed heat flows in the ocean, and the observed bathymetry, are in general agreement with a model of instantaneous cooling of new oceanic lithosphere and its loss of heat through time by conduction resulting in subsidence. There is therefore a simple relationship between the bathymetry of the ocean floor and its thermal age, of the form

$$h = h_{\text{ridgecrest}} + A\sqrt{t} \tag{1.30}$$

where h is the depth of the ocean floor, $h_{\text{ridgecrest}}$ is commonly in the region of 2.5 km, A is a coefficient and t is the age of the oceanic lithosphere. Since the age of the oceanic lithosphere is directly related to the distance from the ridge crest by the spreading rate, the oceanic bathymetry increases gradually away from the sites of spreading. An example from the North Atlantic is given in Figure 1.33. The isostatic balance will of course be affected by the presence of marine sediments. However, in the deep sea the cover of pelagic sediments is generally only a thin veneer.



At any point, mass per unit area = $\int_0 y_l\,\rho_l dy + w\rho_w$
At ridge crest, mass per unit area at depth $(w + y_l) = \rho_m(w + y_l)$
For equilibrium, $w(\rho_w - \rho_m) + \int_0 y_l\,(\rho_l - \rho_m)\,dy = 0$

Fig. 1.33 The principle of isostasy requires the ocean to deepen with age to offset the effects of thermal contraction of the oceanic lithosphere. The water depth below the level of the ridge crest is $w = h - h_{\text{ridgecrest}}$, the thickness of the oceanic lithosphere is y_ℓ, and ρ_m, ρ_w and ρ_ℓ are the mantle, water and lithospheric densities, respectively.

1.5.4 Dynamic topography

Dynamic topography can be defined [34] as the deflection of the solid interface of the Earth which results from buoyancy forces (due to lateral density variations) in the viscous mantle; these buoyancy forces are thought to originate primarily from thermal convection. That dynamic topography exists in the present day can be supported from two main lines of evidence.

1 The observational evidence that very long-wavelength variations in dynamic topography exist is provided primarily by an analysis of global topography and bathymetry [35]. The observed topography is first corrected for shallow-density variations in the lithosphere, including the effects of the cooling of oceanic lithosphere as a function of age (Section 1.5.3). The residual dynamic topography has an amplitude of about 300 m, and a very long-wavelength spatial pattern (dominated by degree 2, which means that two wavelengths occupy the circumference of the Earth), similar to that of the geoid. The degree 2 pattern in the geoid and dynamic

topography shows highs located over Africa and the western Pacific, and lows located over the eastern Pacific and Antarctic–Indian Ocean region (Plate 1.4). It is most likely that this very long-wavelength topographic signal results from anomalies in the deep mantle.

2 Measurements of the velocities of seismic body waves travelling through the mantle show a three-dimensional picture of relatively fast- and slow-velocity regions. These small but significant velocity anomalies are interpreted in terms of density variations caused by temperature differences due to a slow convection of mantle material. These density variations can in turn be used to generate a model of the geoid. Computer modelling of convection in the upper mantle suggests that there should be positive geoid height anomalies over the hot rising limbs of convection cells. These anomalies arise because although there is a density deficit associated with the hot rising limb, it is more than counteracted by the dynamic upward-acting force due to positive buoyancy. The seismic anomalies of the deep mantle are capable of reproducing the observed geoid, and predict large-scale (degree 2–6) variation in dynamic topography with an amplitude of 1 km.

Although the amplitudes of dynamic topography outlined above are considerably smaller than the isostatic topography of mountain ranges caused by crustal thickness changes, dynamic topography is of large wavelength, and is therefore an important control on global topography.

Plumes, hot spots and hot spot tracks

Two processes acting away from plate margins that are not readily explainable by plate tectonics are the chains of volcanic oceanic islands thought to be linked to hotspots, and large, rapidly erupted igneous provinces typified by flood basalts on land and basaltic plateaux under the sea. Both oceanic island chains and continental flood basalt provinces contain basalts whose geochemistry indicates that they come from the melting of mantle elevated above the normal temperature of the aesthenosphere. They can be explained by the rising of hot plumes from the hot thermal boundary layer at the core–mantle boundary. The lithospheric plates, on the other hand, are the cold thermal boundary layer at the surface of the Earth. Both plumes and plates are therefore associated with a particular mode of mantle convection.

Fluid dynamical work suggests that low-viscosity plumes may initiate from within the Earth and ascend as a spherical pocket of fluid (plume head) fed by a pipe-like conduit (plume tail) continuously supplying bouyant, hot material to the head region. It is likely that enlargement of the plume occurs by the entrainment of material heated by the ascent of the plume from the core–mantle boundary. The flood basalt provinces have been interpreted as originating through the melting of the heads of newly started plumes, whereas the oceanic island chains represent the trails of the relatively long-lived plume tails as the plate migrates over the mantle [36].

Fluid dynamical experiments suggest that the diameter of a new plume head varies according to the volume flux and temperature excess of the source material provided to the plume head, and the thermal and viscosity properties of the lower mantle into which the plume starts to ascend (Fig. 1.34a). The plume head then grows by entrainment as it ascends through the mantle, so that the plume head diameter grows as a function of the distance travelled (Fig. 1.34b). By the time the plume head has penetrated into the upper mantle it should have cooled to only 100–150°C above the ambient temperature.

Upon nearing the surface of the Earth, the plume head spreads out into a disc of hot material with positive buoyancy. This produces surface uplift which in some cases may lead to extension of the lithosphere and the formation of 'active' rifts. The scaling of laboratory experiments suggest that the timing and magnitude of the surface uplift depend strongly on the viscosity of the upper mantle. The results from laboratory experiments in which a plume head is sourced from the core–mantle boundary with a buoyancy flux of $3 \times 10^4\,\mathrm{N\,s^{-1}}$ and a source temperature in excess of 300°C, lower mantle dynamic viscosity of $10^{22}\,\mathrm{Pa\,s}$ (kinematic viscosity of $2.5 \times 10^{18}\,\mathrm{m^2\,s^{-1}}$) ascending into an upper mantle with a viscosity of $3 \times 10^{20}\,\mathrm{Pa\,s}$, is shown in Fig. 1.35 [37, 38]. The surface is initially weakly uplifted while the plume head is entirely within the lower mantle (–25 My). When the plume head enters the low-viscosity zone of the upper mantle (–3 My) the surface uplift takes place rapidly, reaching a maximum elevation of 600 m after further ascent to just beneath the lithosphere. At this stage the plume head has a diameter of 1300 km. Two factors may increase the maximum elevation of the topographic dome: (i) penetration of the hot plume into the cold lithosphere; (ii) a volume increase caused by melting. The release of large amounts of basalts by melting of the plume can only take place once the plume head has reached shallow depths. It should be noticed that this is only possible several millions of years after the

maximum surface uplift. The development of smaller-scale gravitational instabilities over the cap of the plume as cold, dense material is squeezed between the ascending plume head and the Earth's surface may facilitate the last-stage ascent of the plume and produce the high surface uplifts seen today in locations such as East Africa (Fig. 1.36) and the large outpourings of basalts in the geological past such as the Siberian and Deccan (India) Traps.

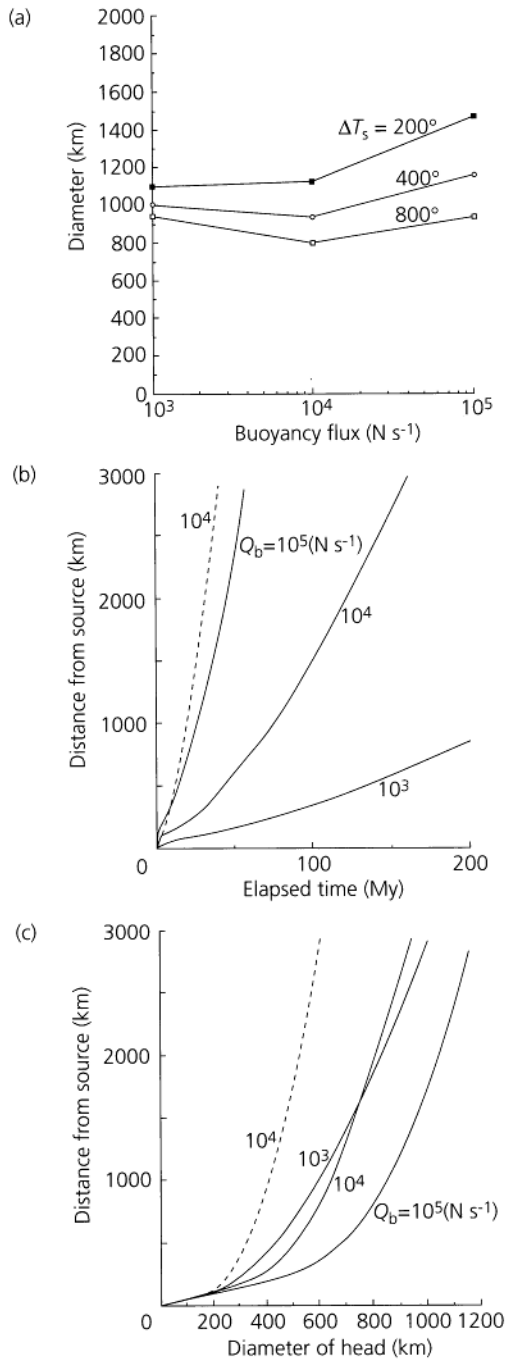


Fig. 1.34 *Left.* Evolution of a plume arising from the core–mantle boundary. (a) The predicted diameter of a starting plume from the core–mantle boundary for a mantle viscosity of 10^{22} Pa s, typical of the modern lower mantle, for a range of source temperature anomalies ΔT_s in degrees Celsius. (b), (c). The height of (b) rise and (c) plume head diameter for $\Delta T_s = 400$°C in a mantle of viscosity 10^{22} Pa s (solid lines) and viscosity 10^{21} Pa s, appropriate for the modern upper mantle and Archaean lower mantle (dashed lines). Plumes with low-bouyancy fluxes (<10^4 N s^{-1}) are likely to degenerate into a sequence of diapirs before reaching the surface after a slow ascent from the core–mantle boundary. Consequently, there is a lower bound to the buoyancy of plumes affecting surface uplift. After Griffiths & Campbell (1990) [37].

Two sorts of uplift pattern and igneous activity should result from plume activity, depending on whether the lithosphere migrates over a plume head or plume tail. Plume head provinces should be areally extensive and equant (1500–2500 km across) whereas plume tail provinces should be narrow (<300 km wide) and linear, that is, they should be hotspot tracks.

Hotspots are at their most obvious in the oceans [39], where they are associated both with volcanism and strongly elevated bathymetry. The Hawaiian Islands, Iceland and Canaries are all excellent examples (Table 1.5). The recognition of hotspots on the continents is far more problematic. However, the highspots associated with alkaline volcanism in north central Africa, such as the Hoggar and Tibesti domes, may be related to activity in the underlying mantle. Doming over hot plumes has also been interpreted as responsible for particular centrifugal palaeodrainage patterns, as well as for the eruption of vast piles of basalts [40].

If there are upwellings beneath continents producing hotspots, we should also expect there to be downwellings, which would be cold. Downwellings should therefore produce dynamic subsidence of the continental surface. The Congo basin may be situated over one such downwelling. It is adjacent to the East African dome which is thought to overlie a plume head (Fig. 1.36).

1.5.5 Continental hypsometries

A convenient way of visualizing the broad morphological characteristics of the surface of the Earth is

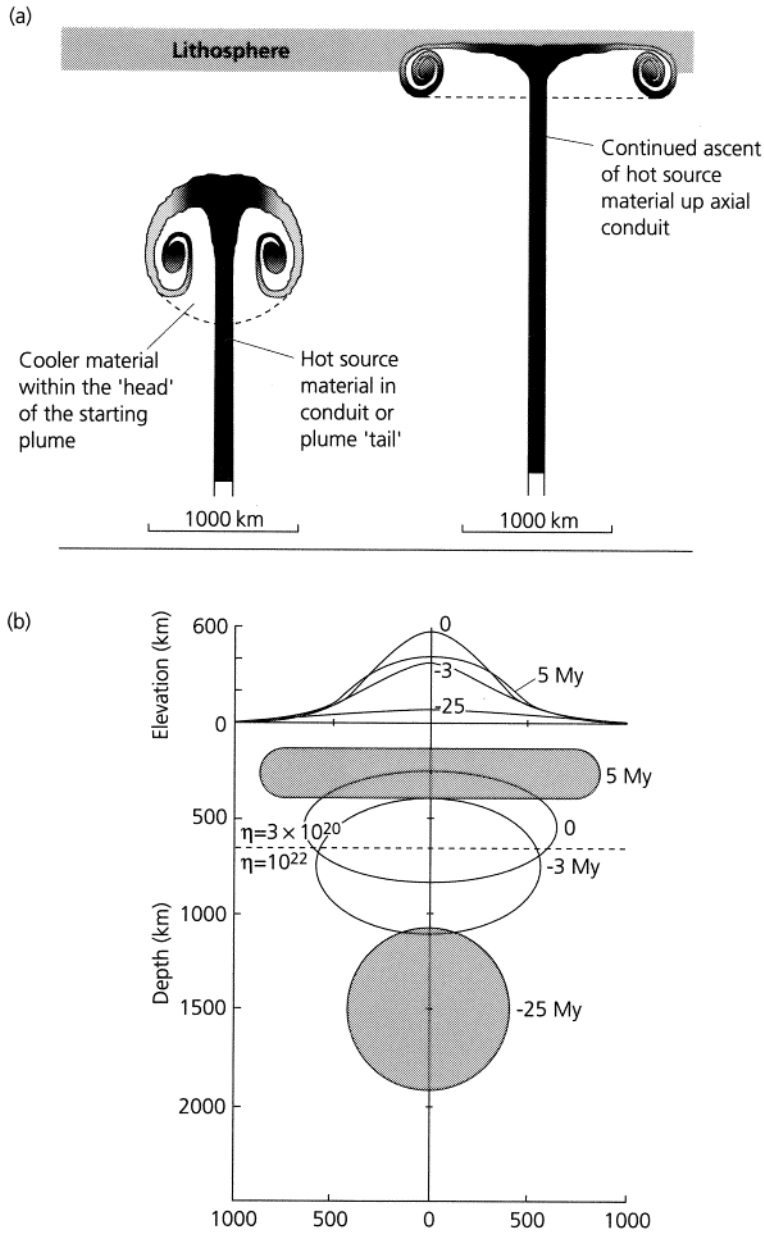


Fig. 1.35 (a) Cartoon of the ascent of a plume through the mantle and its mushrooming into a disc beneath the lithosphere, with darker shading indicating higher temperatures. After Hill (199) [41], v99, pp. 66–78, with kind permission from Elsevier Science-NL, Sara Burgerhartstraat 25, 1055 KV Amsterdam, The Netherlands. (b) The dimensions of a starting plume together with the predicted surface uplift, based on laboratory experiments. During the lateral spreading of the head the input of material from the source is discontinued, simulating the carrying away of the head from the source region by plate motion. After Griffiths & Campbell (1991) [37].

to plot the cumulative frequency distribution of elevation, or *hypsometric curve*. This reveals the proportion of the surface above a given height (Fig. 1.37). The hypsometric curve shows two important features:

1 the major change in topographic slope occurs at the edge of the continental shelf in present-day water depths of about 200 m;

2 only a very small proportion (less than 10%) of the Earth's surface is at elevations of greater than 1 km.

When the hypsometric curves for the continents are compared it is seen that Africa is unusual in having a relatively large percentage of its surface at high elevations. This is entirely accounted for by the high topographic elevation of eastern (East African Rift System) and southern Africa. East Africa is thought to

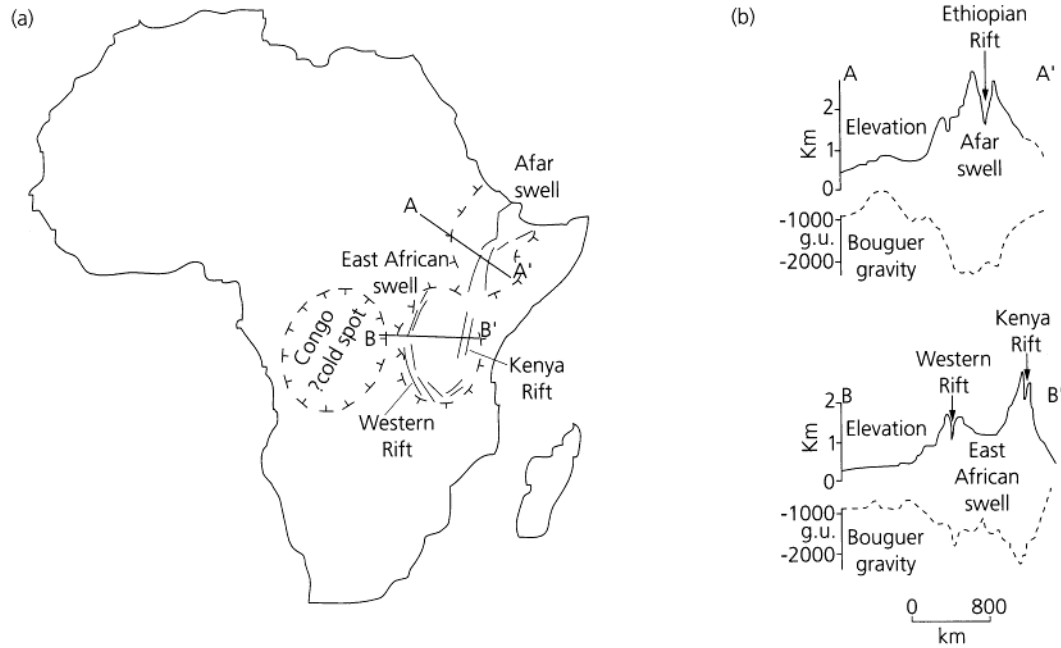


Fig. 1.36 The topographic swells of Africa (Afar and East African (Kenyan) domes) are due to uplift over hotspots. Both are rifted and are the sites of extensive volcanism. The Congo basin has been proposed to overlie a mantle coldspot (Hartley & Allen (1993) [42]. To right: topographic and Bouguer gravity profiles of Afar and East African swells. After Ebinger *et al.* (1989) [43].

Table 1.5 Buoyancy fluxes of the world's hotspots. From Sleep (1990) [44].

Hotspot	Flux* ($Mg\,s^{-1}$)	Reliability	Flux† ($Mg\ s^{-1}$)	Hotspot	Flux* ($Mg\,s^{-1}$)	Reliability	Flux† ($Mg\ s^{-1}$)
Afar	1.2	good		Juan Fernandez	1.6	poor	1.7
Australia, East	0.9	fair		Kerguelen	0.5	poor	0.2
Azores	1.1	fair		Louisville	0.9	poor	3.0
Baja	0.3	poor		Macdonald	3.3	fair	3.9
Bermuda	1.1	good	1.5	Marqueses	3.3	fair	4.6
Bouvet	0.4	fair		Martin	0.5	poor	0.8
Bowie	0.3	poor	0.8	Meteor	0.5	poor	0.4
Canary	1.0	fair		Pitcam	3.3	fair	1.7
Cape Verde	1.6	good	0.5	Réunion	1.9	good	0.9
Caroline	1.6	poor		St Helena	0.5	poor	0.3
Crozet	0.5	good		Samoa	1.6	poor	
Discovery	0.5	poor	0.4	San Felix	1.6	poor	2.3
Easter	3.3	fair		Tahiti	3.3	fair	5.8
Fernando	0.5	poor	0.9	Tasman, Central	0.9	poor	
Galapagos	1.0	fair		Tasman, East	0.9	poor	
Great Meteor	0.5	poor	0.4	Tristan	1.7	fair	0.5
Hawaii	8.7	good	6.2	Yellowstone	1.5	fair	
Hoggar	0.9	fair	0.4	Sum	54.9		
Iceland	1.4	good					
Juan de Fuca	0.3	fair					

* Sleep (1990). [44]
† Davies [1988]. [45]

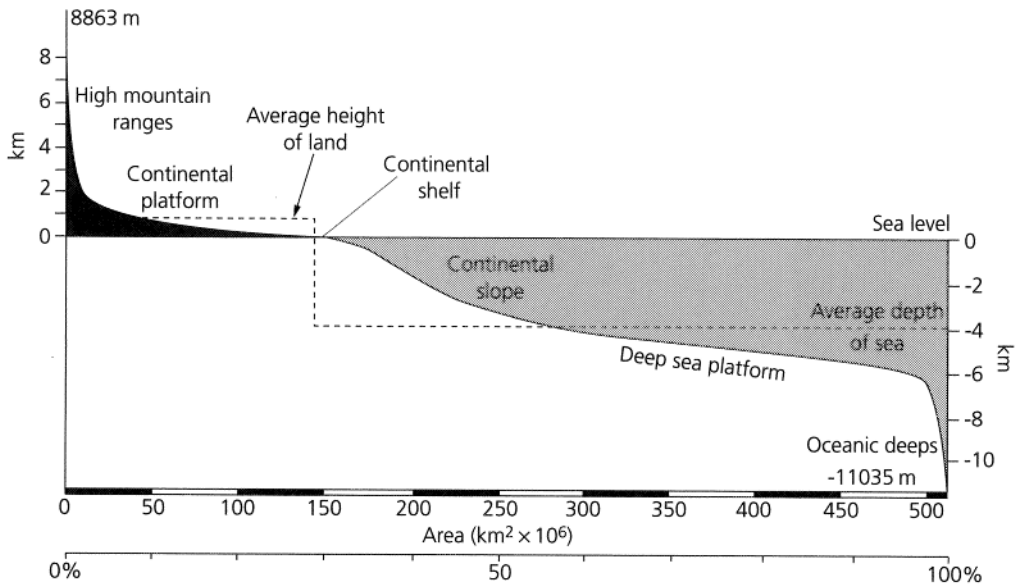


Fig. 1.37 The global hypsometric curve shows the elevation of the Earth's surface expressed as cumulative area (in millions of cubic kilometres) and as a percentage of the total surface area above a given height. The present-day land surface accounts for 29% of the Earth's surface area, but the break in slope on the hypsometric curve corresponding to the edge of the continental shelf gives a figure of 37% for the true spatial extent of the continents.

be situated over a mantle plume head. The present high elevation of southern Africa has been attributed to (i) the presence of underplated igneous bodies following plume activity in the Mesozoic, or (ii) the presence of an underlying present-day or recently active hotspot.

Hypsometry is strongly influenced by dynamic topography (Section 1.5.4). The continental hypsometry also determines the extent to which continents are flooded by certain absolute sea level changes. Consequently, an analysis of the extent of continental flooding in the geological past may reveal past variations in continental hypsometry. The simplest pattern to emerge is of two main cycles through the Phanerozoic (the last 540 My), with elevated sea levels particularly in the Ordovician and Cretaceous, and low sea levels particularly in the Cambrian and Permo-Triassic. Some authors link this broad pattern to the aggregation (inflation and exposure) and dispersal (continental flooding) of supercontinents. However, separating the effects of absolute (eustatic) sea level change from those of variations in isostatic or dynamic topography is a difficult task. Since most of the observations on continental flooding were made on stable continental interiors which have not suffered a great deal of tectonics leading to lithospheric thickness changes, the topographic contribution is most likely due to dynamic topography. By comparing different continents and deciphering the common eustatic signal, it can be seen, for example, that Africa has undergone broad uplift in the Tertiary compared to other continents. This may explain its present anomalous hypsometry.

Further reading

P.A. Allen & J.R. Allen (1990) *Basin Analysis: Principles & Applications.* Blackwell Scientific Publications, Oxford.

E.K. Berner & R.A. Berner (1987) *The Global Water Cycle: Geochemistry and Environment.* Prentice Hall, Englewood Cliffs, NJ.

D.M. Moore (1982) *Green Planet: the Story of Plant Life on Earth.* Cambridge University Press, Cambridge.

J.M. Wallace & P.V. Hobbs (1977) *Atmospheric Science: An Introductory Survey.* Academic Press, San Diego, CA.

References

1 E.K. Berner & R.A. Berner (1987) *The Global Water Cycle: Geochemistry and Environment.* Prentice Hall, Englewood Cliffs, NJ, Table 2.3, p. 20.

2 W.D. Sellers (1965) *Physical Climatology.* University of Chicago Press, Chicago.

3 E.F. Harrison, P. Minnis, B.R. Barkstom & G.G.

Gibson (1993) Radiation budget at the top of the atmosphere. In: *Atlas of Satellite Observations Related to Global Change* (eds R.J. Gurney, J.L. Foster & C.L. Parkinson). Cambridge University Press, Cambridge, pp. 19–38.
4 M. Chahine (1992) The hydrological cycle and its influence on climate. *Nature* **359**, 373–80.
5 R.C. Ward (1975) *Principles of Hydrology*, 2nd edn. McGraw-Hill, Maidenhead.
6 R. Keller (1984) The world's fresh water: yesterday, today and tomorrow. *Applied Geography and Development*, **24**, 7–23.
7 P.J. Webster (1994) The role of hydrological processes in ocean–atmosphere interaction. *Reviews of Geophysics* **32**, 427–76.
8 S.I. Rasool & C. de Bergh (1970) The runaway greenhouse and the accumulation of CO_2 in the Venus atmosphere. *Nature* **266**, 1037–9.
9 A. Gill (1982) *Atmosphere–Ocean Dynamics*. International Geophysical Series **30**, San Diego, CA.
10 R. Atlas, R.N. Hoffman & S.C. Bloom (1993) Surface wind velocity over the oceans. In: *Atlas of Satellite Observations Related to Global Change* (eds R.J. Gurney, J.L. Foster & C.L. Parkinson). Cambridge University Press, Cambridge, pp. 129–40.
11 G.A. Meehl (1982) Characteristics of surface current flow from a global ocean current data set. *Journal of Physical Oceanography* **12**, 538–55.
12 H. Stommel (1958) The abyssal circulation. *Deep Sea Research* **5**, 80–2.
13 H. Heinrich (1988) Origin and consequences of cyclic ice rafting in the northeast Atlantic Ocean during the past 130 000 years. *Quaternary Research* **29**, 142–52.
14 A. Miller *et al.* (1983), *Elements of Meteorology*, 4th edn. Charles E. Merrill.
15 E.G. Njoku & O.B. Brown (1993) Sea surface temperature. In: *Atlas of Satellite Observations Related to Global Change* (eds R.J. Gurney, J.L. Foster & C.L. Parkinson). Cambridge University Press, Cambridge, pp. 237–50.
16 S.C. Porter & A. Zhisheng (1995) Correlation between climate events in the North Atlantic and China during the last glaciation. *Nature* **375**, 305–8.
17 J. Tricart & A. Cailleux (1972) *Introduction to Climatic Geomorphology* (translated by C.J. Kiewiet de Jonge). Longman, London.
18 M. Newson (1994) *Hydrology and the River Environment*. Oxford University Press, Oxford.
19 H.H. Lamb (1972) *Climate: Present, Past and Future. Vol. 1: Fundamentals and Climate Now*. Methuen, London.
20 R.G. Barry (1970) A framework for climatological research with particular reference to scale concepts. *Transactions of the Institute of British Geographers* **49**, 61–70.
21 I.A. Shiklomanov (1993) World fresh water resources. In: *Water in Crisis* (ed. P.H. Gleick). Oxford University Press, Oxford, pp. 13–24.
22 A. Baumgartner & E. Reichel (1975) *The World Water Balance*. Elsevier, Amsterdam.
23 R.J. Gibbs (1970) Mechanisms controlling world water chemistry. *Science* **170**, 1088–90.
24 I.D. White, D.N. Mottershead & S.J. Harrison (1992) *Environmental Systems*, 2nd edn. Chapman & Hall, London.
25 D.E. Reichle (1970) *Analysis of Temperate Forest Ecosystems*. Springer-Verlag, Heidelberg.
26 C.R. McClain, G. Feldman & W. Esaias (1993) Oceanic biological productivity. In: *Atlas of Satellite Observations Related to Global Change* (eds R.J. Gurney, J.L. Foster & C.L. Parkinson), Cambridge University Press, Cambridge, pp. 251–63.
27 M.D. Brasier (1995) Fossil indicators of nutrient levels. 1: Eutrophication and climate change. In: *Marine Palaeoenvironmental Analysis from Fossils* (eds D.W.J. Bosence & P.A. Allison), Geological Society Special Publication 83. Blackwell Science, Oxford, pp. 113–32.
28 J.T. Houghton, G.J. Jenkins & J.J. Ephraums (eds) (1990) *Climate Change: The IPCC Scientific Assessment*. Cambridge University Press, Cambridge.
29 D.L. Turcotte & G. Schubert (1982) *Geodynamics*. Wiley, New York.
30 G.J. Masek, B.L. Isacks, E.J. Fielding & J. Browaeys (1994) Rift flank uplift in Tibet: evidence for a viscous lower crust. *Tectonics* **13**, 659–67.
31 P. Molnar, P.C. England & J. Martinod (1993) Mantle dynamics, uplift of the Tibetan plateau, and the Indian monsoon. *Reviews of Geophysics* **31**, 357–96.
32 P. Bird (1979) Continental delamination and the Colorado plateau. *Journal of Geophysical Research* **84**, 7561–71.
33 D.P. McKenzie (1978) Some remarks on the development of sedimentary basins. *Earth & Planetary Science Letters* **40**, 25–32.
34 M. Gurnis (1991) Continental flooding and mantle-lithosphere dynamics. In *Glacial Isostasy, Sea Level and Mantle Rheology* (ed. R. Sabadini *et al.*). Kluwer Academic Publishers, Dordrecht, p. 446.
35 A. Cazenave, A. Souriau & K. Dominh (1989) Global coupling of the Earth surface topography with hotspots, geoid and mantle heterogeneity. *Nature* **340**, 54–7.
36 W.J. Morgan (1981) Hotspot tracks and the opening of the Atlantic and Indian oceans. In: *The Sea* (ed. C. Emiliani). Wiley, New York, vol. 7, pp. 443–87.
37 R.W. Griffiths & I.H. Campbell (1990) Interaction of mantle plume heads with the Earth's surface and onset of small scale convection. *Journal of Geophysical Research* **96**, 18 295–310.
38 R.W. Griffiths, M. Gurnis & G. Eitelberg (1989) Holographic measurements of surface topography in

laboratory models of surface hotspots. *Geophysical Journal* **96**, 477–95.
39 N.H. Sleep (1990) Monteregian hotspot track: a long-lived mantle plume. *Journal of Geophysical Research* **95**, 21 983–90.
40 K.G. Cox (1989) The role of mantle plumes in the development of continental drainage patterns. *Nature* **342**, 873–7.
41 R.I. Hill (1991) Starting plumes and continental break-up. *Earth & Planetary Science Letters* **104**, 398–416.
42 R.W. Hartley & P.A. Allen (1993) Interior cratonic basins of Africa: relation to continental break-up and role of mantle convection. *Basin Research* **6**, 9–114.
43 C.J. Ebinger *et al.* (1989) Effective elastic plate thickness beneath the East African and Afar plateaus and dynamic compensation of the uplifts. *Journal of Geophysical Research* **94**, 2883–2901.
44 N.H. Sleep (1990) Hotspots and mantle plumes: some phenomenology. *Journal of Geophysical Research* **95**, 6715–36.
45 G.F. Davies (1988) Ocean bathymetry and mantle convection, 1, large-scale flow and hotspots. *Journal of Geophysical Research*, **93**, 10 467–80.