Chapter 2
The General Structure of the Ionosphere

2.1 Introduction

The ionosphere of any planet is defined as that portion of the atmosphere where free electrons and ions of thermal energy exist under the control of the gravity and magnetic field of the planet. There are a number of definitions for the Earth's ionosphere. The main interest of this book is electromagnetic wave propagation though ionospheric plasma as relevant to telecommunications, and so the IEEE Std 211-1997 definition is adopted. The ionosphere of the Earth is therefore taken to be that portion of the upper atmosphere where ions and electrons of thermal energy are present in quantities sufficient to affect the propagation of radio waves.

Most information about the terrestrial ionosphere originated from ionosonde observations at mid-latitudes, and it is reasonable to use this as a basis for describing the behaviour of the ionospheric regions. The physical processes underway in the mid-latitude ionosphere could be assumed to extend over a wide range of latitudes, although significant differences are found at high and low-latitudes due to the orientation of the geomagnetic field. These will be taken into account when relevant for ionospheric prediction and forecasting.

Chapter 2 describes the formation of the Earth's ionosphere and the principal causes for its ionization due to solar–Earth processes and interactions. The main solar phenomena and the Earth space environment including solar wind, the magnetosphere, and the geomagnetic field are briefly presented, with specific emphasis on the importance of the solar and geomagnetic indices for predicting and forecasting the effects on the terrestrial ionosphere.

Then, before discussing the main ionization processes that induce the formation of the ionosphere, there is a short introduction to the neutral atmosphere, its composition, the important parameter of the scale height and the different regions determined by the temperature profile in relation to altitude.

The general structure of the ionosphere and its regular morphology is then described, together with the main mid-latitude ionospheric irregularities.

B. Zolesi and L. R. Cander, *Ionospheric Prediction and Forecasting*,
Springer Geophysics, DOI: 10.1007/978-3-642-38430-1_2,

2.2 Main Sources of Ionization of the Earth's Upper Atmosphere

2.2.1 Sun and Solar Interactions

2.2.1.1 The Sun's Interior

The Sun is the closest star to the Earth, at a distance varying from 147,100,000 to 149,157,000 km with an absolute magnitude of 4.8, and is classed as a yellow dwarf in the principal sequence of the Hertzsprung-Russell (HR) diagram. The Sun is the most important source of terrestrial upper atmosphere ionization and is illustrated in Fig. 2.1 in a photograph from NASA.

The other main characteristics of the Sun are listed in Table 2.1.

Like similar stars the Sun is composed of about 90 % hydrogen and about 10 % helium. Its activity, i.e. the production of energy, is due to a continuous transformation by nuclear fusion of hydrogen into helium inside the solar nucleus where the temperature is 15 million degrees K.

These transformations can be schematized into three steps as shown in Fig. 2.2. In the first step, Fig. 2.2a, there are two hydrogen nuclei H^1, or only protons, which combine to form a *nucleus* of deuterium D^2 (1 proton and 1 neutron) plus one positron e^+ and a neutrino υ, according the following equation:

$$H^1 + H^1 = D^2 + e^+ + \upsilon \tag{2.2.1}$$

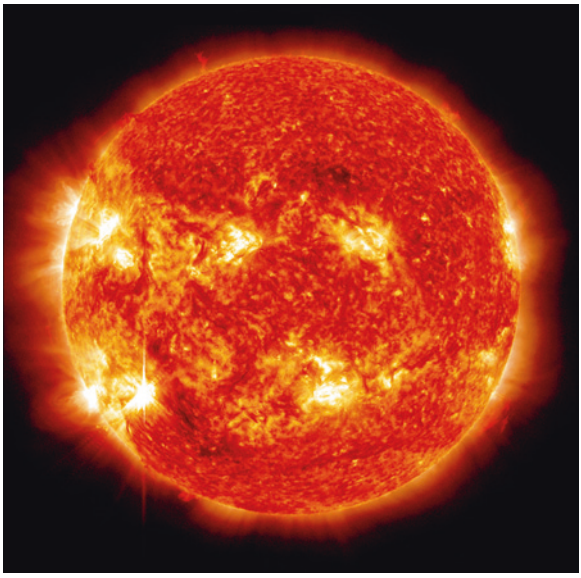

Fig. 2.1 The image of the Sun captured by NASA's solar dynamics observatory (SDO) with a solar flare on 12 November 2012 peaking at 9:04 p.m. EST. (*Note* that EST is 5 h behind coordinated universal time (UTC). (Credit: NASA/SDO at www.nasa.gov/mission_pages/sunearth/news)

Table 2.1 Main characteristics of the Sun

Solar radius	695,990 km
Solar mass	1.989×10^{30} kg
Solar luminosity (energy output of the Sun)	3.846×10^{33} erg/s
Surface temperature	5770 K
Surface density	2.07×10^{-16} kg/m^3
Surface composition	70 % H, 28 % He, 2 % (C, N, O, etc.)
Central temperature	15,600,000 K
Central density	150×10^{-9} kg/m^3
Solar age	4.57×10^9 years

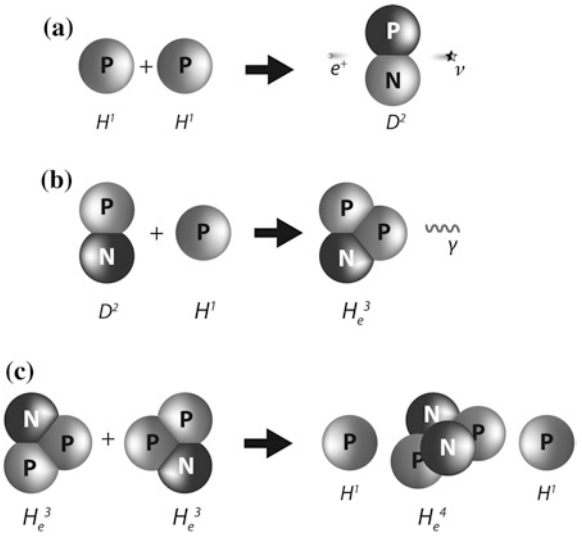

Fig. 2.2 **a** First step: two hydrogen *nuclei* H^1 combine to form a nucleus of deuterium D^2 plus one positron e^+ and a neutrino **υ. b** Second step: an atom of deuterium D^2 combines with a proton H^1 to form a helium *nucleus* H_e^3 plus electromagnetic radiation γ. **c** Third step: the fusion of two H_e^3 *nuclei* into one H_e^4 *nucleus* plus two hydrogen atoms H^1 with consequent release of energy

Next, the combination of deuterium D^2 with a proton H^1 forms a helium *nucleus* H_e^3 (2 protons plus 1 neutron) plus electromagnetic radiation γ (Fig. 2.2b):

$$D^2 + H^1 = H_e^3 + \gamma \tag{2.2.2}$$

Finally, two reactions as described above produce two H_e^3 *nuclei*, which can then fuse into one H_e^4 *nucleus* (2 protons plus 2 neutrons) plus two hydrogen *nuclei* H^1 (1 proton and 1 proton) with consequent release energy (Fig. 2.2c):

$$H_e^3 + H_e^3 = H_e^4 + H^1 + H^1 \tag{2.2.3}$$

The final result is the transformation of 4 protons and 2 neutrons into one atom of helium H_e^4, the difference between the initial mass and the final mass is transformed, according to Einstein's law, into kinetic energy.

The energy is radiated from the *nucleus* of 160,000 km in radius as γ rays to a distance of 450,000 km from the centre, losing energy and increasing in wavelength to become first X-rays then UV (ultraviolet) rays, and finally visible light.

This region of the solar interior where the energy produced by the inner layers is continuously absorbed and emitted is known as the radiative zone. Above this there is another zone known as the convective zone, where energy is transmitted by the transport of matter. Figures 2.3 and 2.5 provide schematic diagrams of the principal features of the Sun.

2.2.1.2 The Solar Atmosphere

The Sun is a gaseous body so it is not easy to define a surface and the borders of its atmosphere. However, it is this external part of the Sun that plays the most important role in the ionization of the Earth's atmosphere, making it possible to predict the resulting influences on ionospheric radio propagation.

What appears from the Earth to be the surface of the Sun, a yellow disk, is the photosphere. As seen in the photo of Fig. 2.1 and in the schematic diagram of Fig. 2.3, this is a very thin layer of about 500–700 km in thickness, and a very small component in the radius of the Sun. Small convection cells are active in this layer and appear on the surface as granulation, or like grains of rice, with diameters of approximately 700–1,000 km and lifetimes of a few minutes. Similar to bubbles from a boiling liquid, the energy is radiated from the photosphere into space. A typical example of this granulation is shown in the photo in Fig. 2.4.

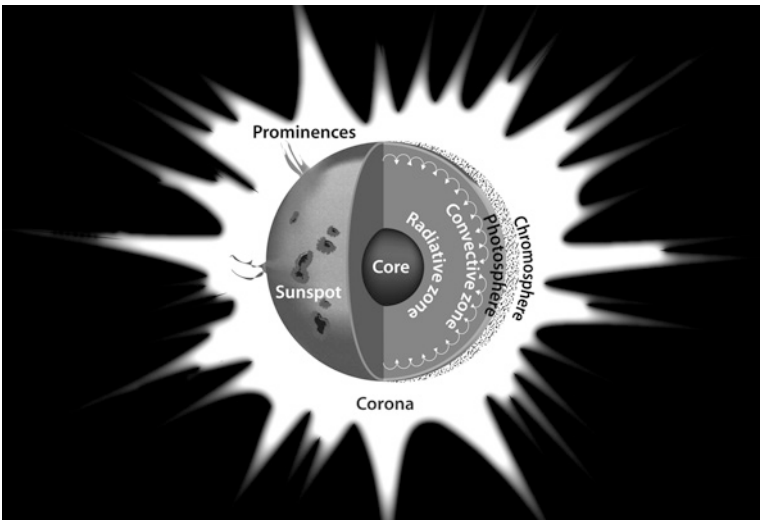

Fig. 2.3 The Sun: schematic diagram of the internal and external features

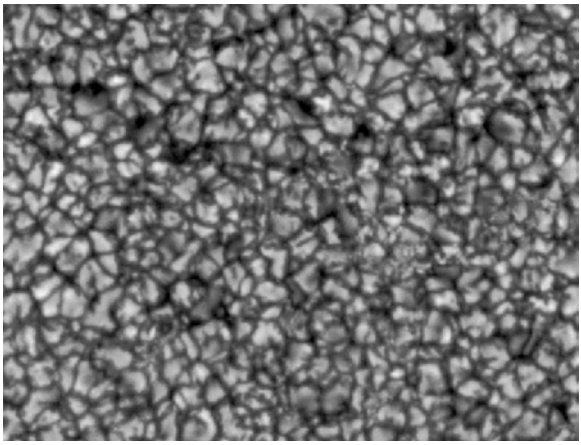

Fig. 2.4 A magnification of the photosphere reveals the small active convection cells that appear on the surface as granulation or like grains of rice

The effective temperature of the photosphere established by black body physics is about 6,000 K. Most of the energy radiated by the Sun in the range from UV to IR (infrared) rays comes from the photosphere.

2.2.1.3 Sunspots and the Solar Cycle

The most important phenomenon occurring in the photosphere, in relation to all other solar phenomena, and for the overall influence and interaction with the terrestrial ionosphere, are certain small regions on the solar surface that appear darker on the solar disk because their temperature is about 3,000 K. These are known as sunspots and are transient phenomena lasting from a few days to some months, of variable dimensions in some rare cases observable with the naked eye, as for example in January 1989. Large sunspots have in fact been observed without optical instruments since antiquity by Chinese astronomers, Greek philosophers like Theophrastus (the pupil and successor of Aristotle also known as the first author of a History of Physics) and on the eve of the modern age in 1610 by Galileo Galilei, who first observed the Sun with a telescope. At the beginning of the 17th century people were surprised to discover that the Sun was not a perfect body but had spots or *maculae*, and observation of the spots and their movement through time proved the differential rotation of the star with a period of about 27 days, faster in the equatorial region at 24 days, and slower in the polar regions at over 30 days. The cause of cooler zones within the sunspot was explained by the existence of strong magnetic fields that obstruct the transfer of hot gases from the lower layers to the photosphere (Fig. 2.5).

The sunspot number has been routinely calculated since 1849 at the Zurich observatory according to the empirical formula defined by its director Rudolf Wolf:

$$R = k(10\,g + s) \tag{2.2.4}$$

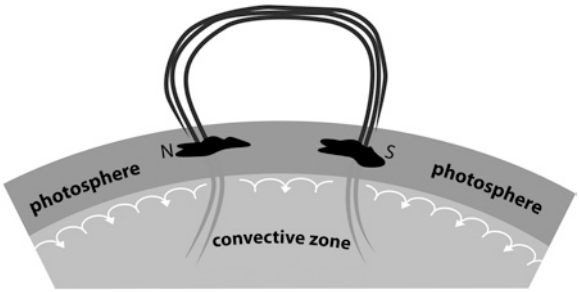

Fig. 2.5 Strong magnetic fields extend out from the sunspots, which are darker, cooler zones of the photosphere

where k is a correction factor approximately near to 1 taking into account the equipment characteristics, g is the number of groups of sunspot, s is the number of isolated sunspots, and the resulting R is known as the Wolf number. Monthly averages of this number, from 10 to more than 100, and up to 200, reveal a clear 11 year cycle. Although systematic observation only started in the 19th century, it was nevertheless possible to reconstruct all the cycles back to the beginning of the 16th century. Figure 2.6 shows the monthly average of solar sunspot numbers over the last three hundred and 10 years. It is interesting to note that during the period from 1645 to 1715 (the Maunder minimum) as well as at the beginning of the 19th century, the sunspot number remained very low and this behaviour was thought to have some implications for the coincident climate phenomena known as the mini ice-ages. In addition to the main 11 year cycle (not always precisely 11 years but varying from 8 to 14), another 22 year cycle has been observed, taking into account the reversal of the dipolar magnetic field.

The prediction of the behaviour of future solar cycles, naturally only as a smoothed trend, is important in order to predict interactions with the Earth's atmosphere and all the phenomena relevant to space meteorology.

Long-term prediction of solar activity, some years or even some solar cycles in advance, is extremely difficult and at present scientific literature proposes different methods that offer less than adequate results. These methods estimate the extent of the next cycle maximum by considering the length of the previous cycle, the level of activity at sunspot minimum, and the range of the previous cycle. Furthermore, geomagnetic precursors are also considered, for example the variation in the Earth's magnetic field at and before sunspot minimum. The best results are obviously obtained when predicting the remaining part of a current solar cycle 3 years after the sunspot minimum, or by using a method that utilizes an average of different techniques.

2.2.1.4 The Chromosphere, the Corona, and Other Solar Phenomena Including Solar Flares and Coronal Mass Ejections

Just above the photosphere there is a region, about 2,000–2,500 km thick, clearly visible during a total solar eclipse by the moon and reddish in colour: the

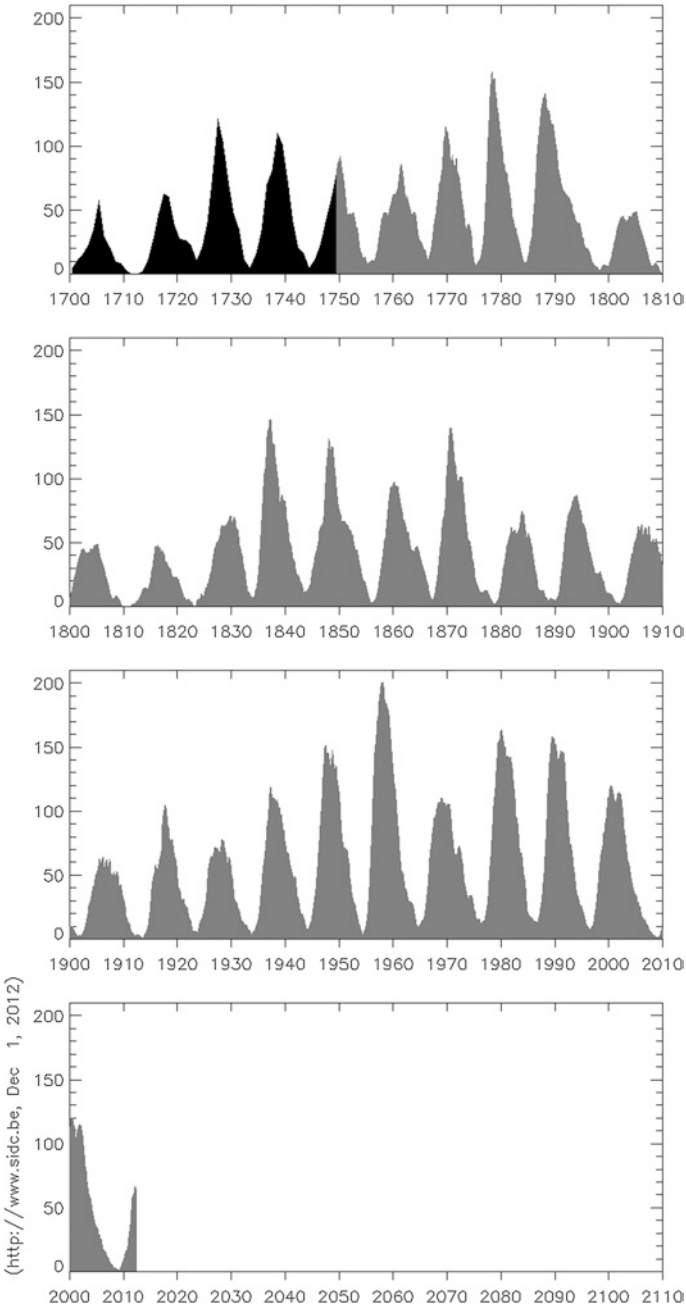

Fig. 2.6 The yearly and monthly sunspot numbers, from 1700 up to 2012 (http://www.sidc.be)

chromosphere. This region is characterized by progressively increasing temperature from 6,000 to 20,000 K. Within the chromosphere different kinds of plasma eruptions are observed, although in reality they represent different aspects of the same phenomenon. Prominences are emissions of cooler and denser clouds of plasma along the magnetic field lines, visible above a darker zone on the edge of the solar disk known as the limb and lasting for more than one solar rotation. Seen over the solar disk these cooler clouds appear darker and are called Filaments, while *Spiculae* are instead small transient eruptions of material from the surface into the corona region (Fig. 2.3).

Above the chromosphere the corona is known as the outer atmosphere of the Sun, visible during a solar eclipse as white light extending over many solar *radii*. The plasma of the corona is extremely low density but with a temperature of millions of degrees K. Most of the light of the corona comes from the light diffused by the photosphere.

Solar flares are large explosions occurring above the surface of the Sun, close to active regions characterized by the presence of sunspots. They are located along the neutral lines between opposite directions of the magnetic field and produce exceptionally large emissions of electromagnetic γ and X-rays, very hot (millions of degrees K) plasma material composed of protons and electrons (Fig. 2.1).

Flares are often associated with Coronal Mass Ejections (CMEs), enormous emissions of plasma in the form of giant bubbles of protons and electrons coming from the corona and flowing along the solar magnetic field. These can last several hours and interact with the solar wind and the Earth's magnetosphere.

All the phenomena occurring in the solar atmosphere result in variations in emissions of electromagnetic energy, and variations in the energy of the particles dispersed into the interplanetary medium. They are directly dependent on the periodic activity of the Sun, as clearly expressed by the presence of sunspots. However, low numbers of sunspots does not necessarily indicate low solar activity or the absence of other solar features.

2.2.2 The Solar Wind, the Geomagnetic Field, and the Magnetosphere

2.2.2.1 The Solar Wind

At the end of the 1950s, just before the launch of the first satellites, Sydney Chapman tried to explain the high temperatures measured in the upper atmosphere as a result of heating by hot plasma from the solar corona. The corona was thought to extend throughout the solar system immersing all the planets. As noted in the previous paragraph, the external portion of the Sun's atmosphere, the solar corona, is composed of very hot plasma, from 1 to over 2 million degrees K, a gas with extremely high kinetic energy with the electrons free from the positive *nuclei*.

Starting from this hypothesis Eugene N. Parker proposed a complete explanation for the phenomenon, also taking into account observations of a continuous flux of particles arriving from the Sun. Near the Sun the corona can be approximated as a static atmosphere but with increasing distance from the star the internal pressure of the gas becomes greater than the weight of the upper plasma such as to generate a dynamic flux of particles of increasingly higher velocity. Like a wind these particles flow across the entire solar system, reaching a velocity of hundreds of km/sec. This is called the solar wind and is illustrated in Fig. 2.7.

Together with the continuous flux of particles from the Sun, it is important to note the behaviour of the Solar Magnetic Field, or rather the Interplanetary Magnetic Field (IMF). In the case of a stationary Sun the magnetic field and the solar wind would expand symmetrically along the radius. In reality, solar rotation results in a circular field being superimposed over the radial magnetic field: the result is a spiral form IMF of an intensity decreasing with the square of the distance (see Fig. 2.8). The solar wind is composed of plasma, which being a conductor

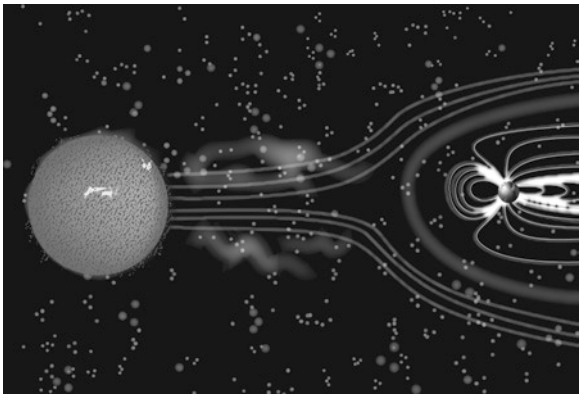

Fig. 2.7 A pictorial representation of the solar wind

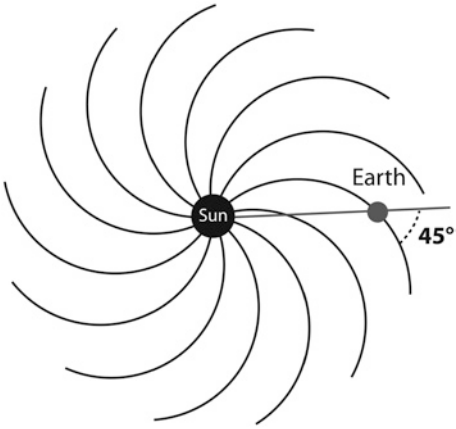

Fig. 2.8 Spiral form of the IMF, the field lines meet the Earth's orbit at an angle of 45° as confirmed by observations

"freezes" the solar magnetic field, dragging it from the region in which it originated into outer space. This effect is similar to that of flowing water ejected from a garden sprinkler when the individual water drops maintain the radial direction of rotation.

Seen from above on the solar equatorial plane, the IMF appears divided into different magnetic sectors in which the polarity is observed from the Earth as intermittently positive or negative, as shown in Fig. 2.9.

This complex behaviour can be explained by taking into account that the open lines of the solar magnetic field, seen on a plane orthogonal to the equatorial plane, lie parallel and are separated by a neutral current sheet. An observer on the Earth observes opposite polarity over several days when the Earth, along its orbit, is above or below the neutral sheet shaped like a ballerina's tutu (Kelley 1989).

The IMF is a vector of weak intensity varying from 1 to 37 nT near the Earth, with *Bx* and *By* components on the plane of the ecliptic, and *Bz* component perpendicular to it.

2.2.2.2 The Geomagnetic Field

It is important now to schematically describe the geomagnetic field, not only because of its influence over the terrestrial ionosphere, but also to understand how the arrival of the charged particles comprising the solar wind can modify the lines of this field in the region of space around the Earth, generating what is known as the magnetosphere. As an initial approximation the geomagnetic field can be described as the field generated by a simple dipole with its axis slightly offset from the rotational axis. The lines of the field extend outwards from the South Pole and converge inwards towards the North Pole, as was postulated in historical times by William Gilbert in his *De Magnete*, in 1600. This field is the product of different sources which include the dynamics of the metallic fluids within the Earth's core, according to the dynamo theory; residual magnetism in the Earth's

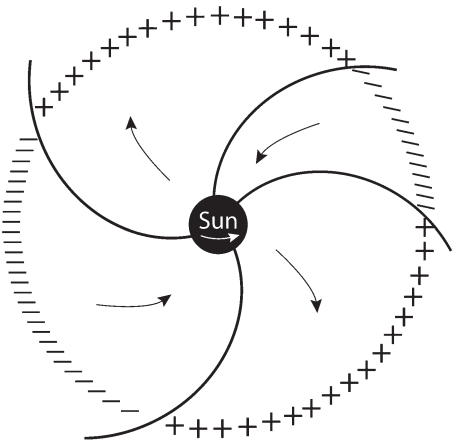

Fig. 2.9 Different magnetic sectors in which the polarity of the IMF is observed from the Earth as intermittently positive or negative (from Wilcox and Ness 1967)

crust; currents flowing in the ionosphere; and finally the influence of circum terrestrial space through interactions with charged particles in the upper atmosphere. An idealized image of the Earth's magnetic field is shown in Fig. 2.10.

The intensity of the magnetic field is expressed as a vector of magnetic induction **B** and it is measured in Tesla, although it is more practical to use nanoTesla (nT), which are equal to 10^{-9} T and also referred to as gamma and equal to 10^{-5} gauss, as traditionally used in the CGS system.

The Earth's magnetic field is subject to different categories of variation. There are very slow variations, like the slow secular variation of the magnetic field since the inversion of polarity which occurred around a hundred thousand years ago. There are much faster diurnal variations with a regular trend through the day, and the perturbations induced by solar phenomena or by interaction with the upper atmosphere and magnetosphere.

The vector **F** of the geomagnetic field is described within the reference system as shown in Fig. 2.11 by the so-called seven magnetic elements, 5 of intensity and 2 angular, when H is the horizontal component, Z the vertical component, I the inclination defined as the angle of the vector **F** with the horizontal plane, and D the declination as the angle between the **H** and the geographical meridian:

$$F = \sqrt{X^2 + Y^2 + Z^2} \qquad H = \sqrt{X^2 + Y^2} \tag{2.2.5}$$

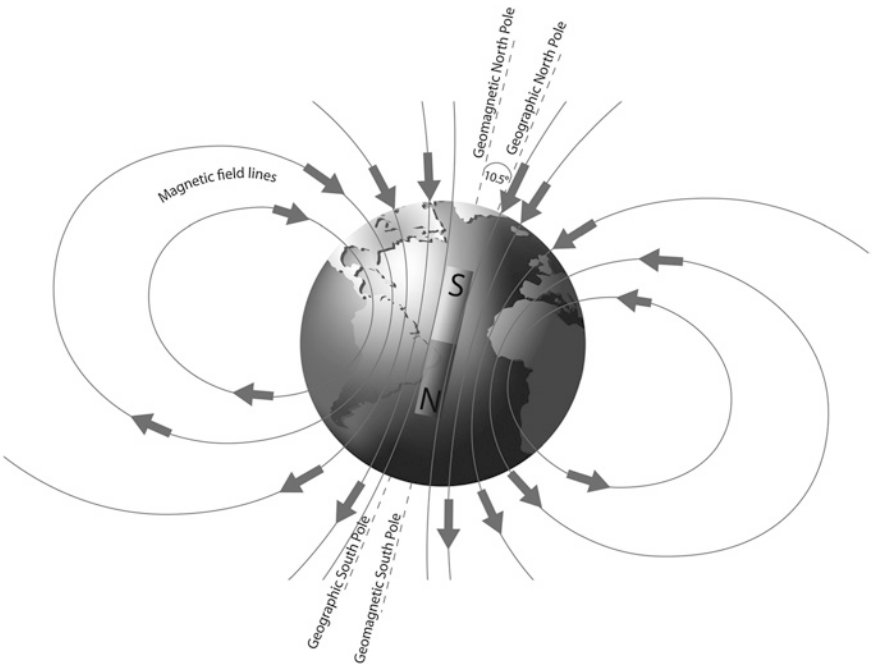

Fig. 2.10 Idealized view of the Earth's magnetic field as a dipole with a tilt of 10.5° relative to the rotational axis

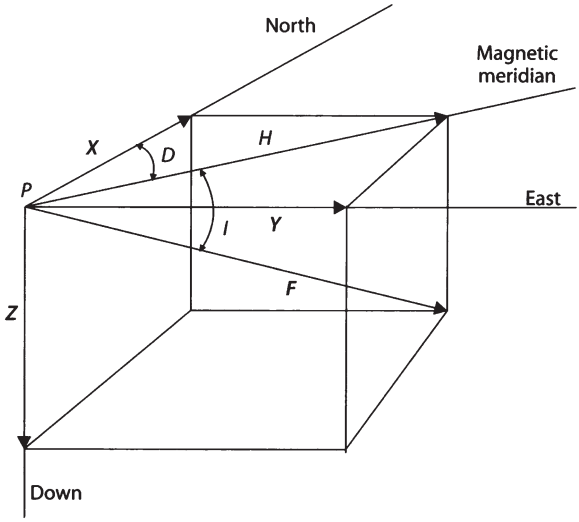

Fig. 2.11 Elements of the Earth's magnetic field

$$H = F\cos I \quad Z = F\sin I \quad Z = H\tan I \quad X = H\cos D \quad Y = H\sin D \tag{2.2.6}$$

To analytically represent the geomagnetic field over the entire Earth, it is better to use a reference system with the origin in the centre of the Earth.

A magnetic potential V, generated by a dipole at point P with polar coordinates r and ϑ, is:

$$V = \mu_0 \frac{\mathbf{M}\cdot\mathbf{r}}{4\pi r^3} = \mu_0 M \frac{\cos\vartheta}{4\pi r^2} \tag{2.2.7}$$

when $\mathbf{M}$ is the moment of dipole. The intensity F can be obtained using its gradient or with the equivalent *nabla* operator ∇:

$$\mathbf{F} = -\text{grad}\ V \quad \text{or} \quad \mathbf{F} = -\nabla V \tag{2.2.8}$$

See Appendix A for a review of the mathematical operators. The two components along the radius F_r and transversal F_t are obtained by:

$$F_r = -\frac{\delta V}{\delta r} \quad \text{and} \quad F_t = -\frac{\delta V}{r\delta\vartheta} \tag{2.2.9}$$

$$F_r = 2\mu_0 M \frac{\cos\vartheta}{4\pi r^3} \quad \text{and} \quad F_t = \mu_0 M \frac{\sin\vartheta}{4\pi\, r^3} \tag{2.2.10}$$

To analytically represent the geomagnetic field on the Earth's surface it is useful to apply the analytic techniques of spherical harmonics first used by Karl F. Gauss in 1838. The idea is to start from the general physical law also explained by

the Maxwell equations, when for magnetic induction **B** the following equivalence applies:

$$\mathbf{B} = -\text{grad } V \tag{2.2.11}$$

Considering that in free space, where there are no currents, the Maxwell equation is:

$$\text{div } \mathbf{B} = \text{div}(-\text{grad } V) = 0 \tag{2.2.12}$$

$$\text{div grad } V = \text{div } (-\nabla V) = \nabla^2 V = 0 \tag{2.2.13}$$

where ∇^2 is the *Laplacian* operator:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{2.2.14}$$

Applying the *Laplacian* operator to the equation of the geomagnetic potential V in spherical coordinates:

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial V}{\partial r}\right) + \frac{1}{r^2 \sin\vartheta}\frac{\partial}{d\vartheta}\left(\sin\vartheta\frac{\partial V}{\partial \vartheta}\right) + \frac{1}{r^2 \sin^2\vartheta}\frac{\partial^2 V}{\partial \lambda^2} = 0. \tag{2.2.15}$$

Its general solution may be represented in a simplified form by:

$$V(r, \vartheta, \lambda) = a\sum_{n=0}^{\infty}\sum_{m=0}^{n}\left(\frac{a}{r}\right)^{n+1} P_n^m(\cos\vartheta)\left[g_n^m \cos(m\lambda) + h_n^m \sin(m\lambda)\right] \tag{2.2.16}$$

where a is the radius of the Earth's surface, ϑ is the co-latitude, λ is the east longitude, $P_n^m(\cos\vartheta)$ are the spherical harmonic Smith functions, and the coefficients g_n^m and h_n^m are the Gauss coefficients. This equation is important because a similar mathematical approach has been used to represent ionospheric characteristics over a sphere in order to obtain global ionospheric maps, as described in Chap. 6. See Fig. 2.12 for an example of a global map of the total magnetic field F in nT for the year 2005.

2.2.2.3 The Magnetosphere and the Plasmasphere

The lines of the geomagnetic field shown in Fig. 2.10 and generated by a simple magnetic dipole can approximate geomagnetic behaviour close to the Earth's surface, but they cannot be extended into space. The exploration of the first satellites during the 1960s revealed a completely different situation at distances of several Earth radii from the planet.

The charged particles of the solar wind flowing from the Sun interact with the lines of the geomagnetic field and change their shape. The lines are compressed on the side facing the Sun and extended on the opposite side in a complex way, generating something resembling the tail of a comet (Fig. 2.13).

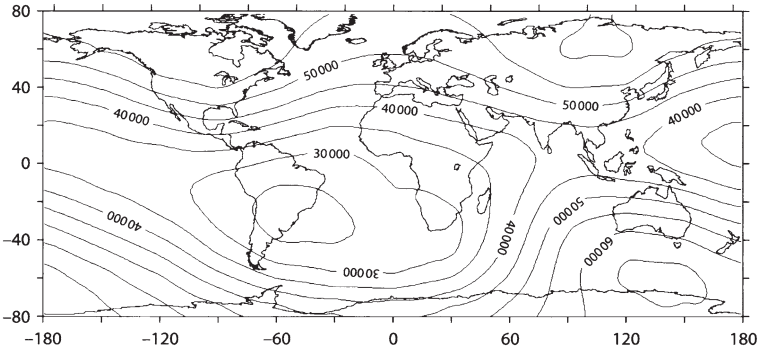

Fig. 2.12 The Earth's total magnetic field *F* with isolines in nT, year 2005 (from Lanza and Meloni 2006)

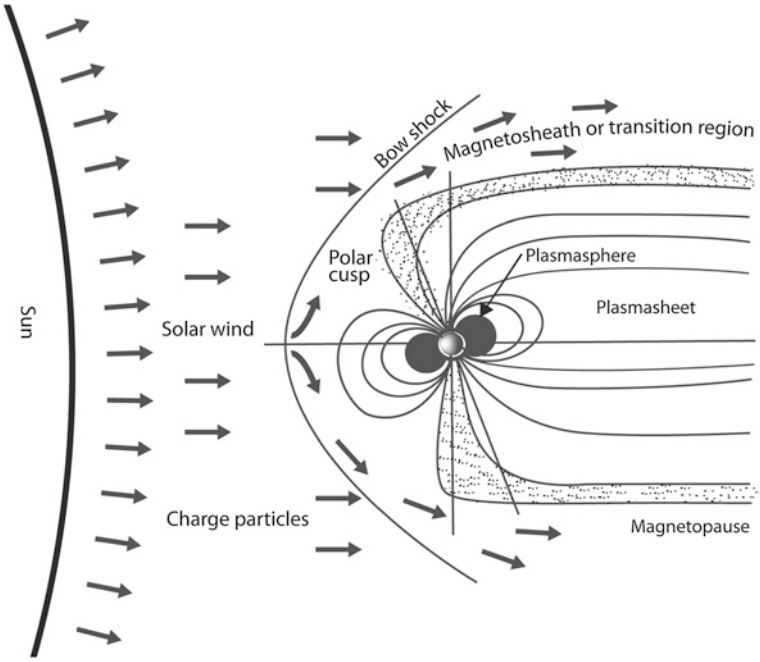

Fig. 2.13 Diagram of the magnetosphere produced by the interaction of the particles of the solar wind with the geomagnetic field

In the classical and simplified schema of the magnetosphere, the charged particles composing the solar wind start to interact with the lines of the geomagnetic field on the dayside direction producing a shock wave, known-as a bow shock for its shape, which could be considered as the border of the circum terrestrial region. Most of the particles deflected by the bow shock flow along a transition region

called the magnetic sheath while a fraction of the particles penetrate into the so-called magnetosphere, the region bounded by the magnetopause. Within this region the charged particles are influenced by the Earth's geomagnetic and electric fields.

There is another region below the lower portion of the magnetosphere called the plasmasphere, which is shaped like a torus around the Earth. The plasmasphere is composed of relatively cool plasma of hydrogen ions and can be considered the upper or outermost portion of the ionosphere, lying above 1,000 km in altitude. This transition region is very influential during measurements of total electron content by satellites, and in all issues affecting Earth-space radio propagation.

The ogive shaped geomagnetic tail extends for over 200 terrestrial radii. The plasma in this portion of the magnetosphere has a different composition: there is an external plasma mantle of very low density and an internal central plasma sheet. The complex behaviour of the particles in the distant regions of the tail are not relevant to the present discussion. It is important to note the two polar cusps at the points where the magnetosphere is in contact with the Earth. Through these two zones there is precipitation of charged particles from the solar wind into the upper atmosphere of the so-called auroral regions.

Finally, the presence of two radiation layers trapped by the fields of the magnetosphere and discovered during space missions is also very important. These are known as the Van Allen Belts. The first layer of radiation is found along the magnetic equator at a distance between 1,000 km to 3 Earth *radii* from the surface of the planet and is composed of protons from the solar wind and from the lower ionosphere. The second layer is at a distance of between 4 and 6 Earth radii and is composed mainly of electrons, again originating from the solar wind. These particles are subject to a rotary movement around the lines of the geomagnetic field as shown in Fig. 2.14, oscillating between North and South. In addition to this oscillation, the electrons and protons are characterized by another movement in opposite directions longitudinally, generating a current called the ring current. On the Earth the effects of this current can be observed and measured as variations in the geomagnetic field.

2.2.3 Solar and Geomagnetic Indices

Prediction of solar and geomagnetic activity is essential for forecasting conditions in the near-Earth space environment in general and for ionospheric modelling and propagation processes in particular. This is an important subject in solar-terrestrial studies and significantly depends on the types of solar events and effects described in Sects. 2.2.1 and 2.2.2. For example, solar flares have an immediate and dramatic effect on ionospheric propagation and therefore it is important to provide a useful service for predicting the occurrence of large solar flares. When forecasting geomagnetic activity it is vital to monitor the solar wind upstream of the Earth. It is well established that the critical frequencies of the ionospheric layers depend in a systematic way on measurable quantities related to solar radiation. At the same

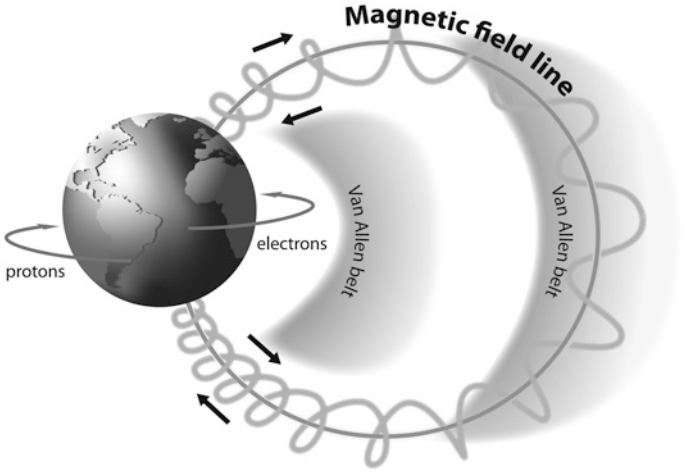

Fig. 2.14 Electrons and protons in the two layers of radiation known as the Van Allen belts, shaded regions in the figure, oscillate from North to South rotating around the lines of the magnetic field. They simultaneously move longitudinally, generating a current called the annular current

time the structure and dynamics of the ionosphere are profoundly dependent on measurable quantities related to geomagnetic activity.

As described in Sect. 2.2.1, changes in solar activity are cyclic, with one cycle of about 11 years, usually referred to as the 11-year sunspot cycle, another quasi-cycle of about a year, and irregular fluctuations with periods of less than a month. Solar indices such as R, Ri, R_{12}, Φ and Φ_{12} are measurable quantities of solar activity or a specific solar radiation. The most widely quoted average sunspot number is the Zurich number (Rz) which was replaced after January 1981 with the International Sunspot Number (Ri). Figure 2.15 shows the variation of the monthly and smoothed values of the sunspot number Ri for the last five solar cycles including the current 24th cycle.

The 12-month running mean sunspot number R_{12}, which considerably reduces complicated rapidly-varying components but does not obscure the slowly varying component, is generally used to study the main component of the solar cycle. The definition of R_{12} is:

$$R_{12}(m) = \frac{1}{12}\left[\frac{1}{2}R_{m-6} + \sum_{k=m-5}^{k=m+5} R_k + \frac{1}{2}R_{m+6}\right] \tag{2.2.17}$$

where R_k is the mean of the daily sunspot numbers for a single month k, and R_{12} is the smoothed index for the month represented by $k = m$. From Eq. 2.2.17 it is clear that the most recent available R_{12} value is centred on a month at least 6 months earlier than the present time. Therefore, R_{12} cannot be used to forecast short-term variations in solar activity.

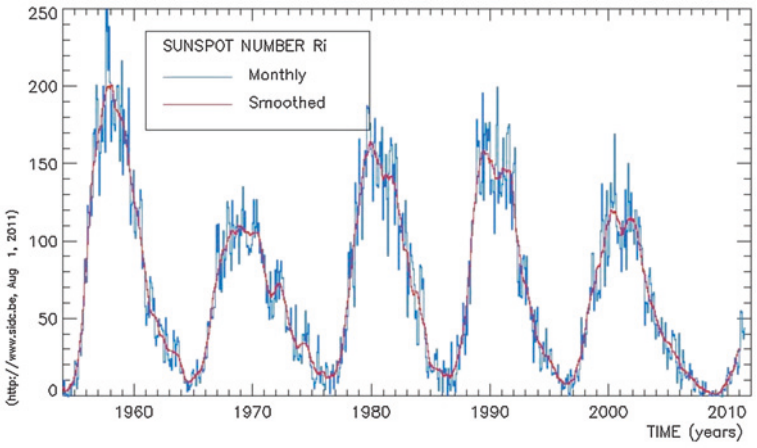

Fig. 2.15 The monthly and smoothed values of the sunspot number *Ri* from July 1959 to July 2011

Another useful indicator of the level of solar activity is the solar radio noise flux Φ of 10.7 cm wavelength (frequency 2.8 GHz) which can be regarded as a reference index for dates up to one month. Extended consistent observations of this solar radio noise flux, also known as $\Phi 10.7$ flux and expressed in units of 10^{-22} Wm^{-2} Hz^{-1}, have been conducted by Canadian, Japanese and other laboratories. The relationship between R_{12} and Φ_{12} was found to be:

$$\Phi_{12} = 63.7 + 0.728\, R_{12} + 8.9\, x\, 10^{-4} R_{12}^2 \tag{2.2.18}$$

The R_{12} and Φ_{12} indices have generally been adopted by the International Telecommunication Union for long-term and monthly median ionospheric predictions. Measured and predicted values for R and Φ and their 12-month running mean values (R_{12} and Φ_{12}) are published by different organizations. An example of the results from the Sunspot Index Data Centre (SIDC) in Brussels (http://www.sidc.be/) is shown in Table 2.2.

Geomagnetic indices have been used for a number of years as a suitable technique to represent the complex set of processes associated with the geomagnetic activity discussed in Sect. 2.2.2. Several indices have been applied in solar-terrestrial physics to describe geomagnetic field fluctuations. The most widely-used index is the planetary index 3 h *Kp*. It is derived from the *K* index, which is obtained by observing the *H*, *D*, and *Z* components for eight daily intervals of 3 h (00–03, etc.) at selected geomagnetic observatories. The planetary 3 h *Kp* index is usually expressed in one-third units by adding the signs −, 0, $^+$ to the numbers 0–9, providing a 28-step scale from 0_0 to 9_0. The daily ΣKp is the sum of the eight 3 h indices for a universal day. The *Kp* index is extremely useful for evaluating geomagnetically disturbed days (D-days) and the quietest days (Q-days). The selection

Table 2.2 SIDC news of daily and mean sunspot numbers for April, May, and June 2012 (see http://www.sidc.be/)

SIDC Definitive international and hemispheric sunspot numbers for 2012									
	April			May			June		
Date	Ri	Rn	Rs	Ri	Rn	Rs	Ri	Rn	Rs
1	42	19	23	69	10	59	85	30	55
2	51	24	27	68	11	57	106	49	57
3	46	26	20	69	15	54	116	61	55
4	41	27	14	62	15	47	104	68	36
5	34	34	0	63	25	38	106	72	34
6	34	34	0	62	33	29	107	67	40
7	19	19	0	57	35	22	89	62	27
8	16	16	0	61	39	22	75	51	24
9	9	9	0	65	40	25	74	42	32
10	10	10	0	69	41	28	85	42	43
11	19	11	8	79	51	28	90	34	56
12	37	16	21	72	51	21	78	16	62
13	35	27	8	75	51	24	79	13	66
14	33	33	0	85	51	34	86	12	74
15	58	37	21	87	50	37	82	11	71
16	39	23	16	98	54	44	74	11	63
17	51	32	19	79	57	22	62	11	51
18	72	26	46	73	58	15	40	8	32
19	95	35	60	78	56	22	36	8	28
20	108	35	73	83	61	22	21	0	21
21	104	31	73	79	58	21	11	0	11
22	94	31	63	57	57	0	11	11	0
23	89	28	61	55	46	9	11	11	0
24	86	38	48	62	48	14	14	7	7
25	81	41	40	67	45	22	12	0	12
26	75	35	40	57	30	27	25	8	17
27	68	36	32	59	29	30	49	10	39
28	72	30	42	78	29	49	61	15	46
29	71	21	50	56	17	39	72	27	45
30	67	15	52	52	8	44	73	25	48
31				64	14	50			
Mean:	55.2	26.6	28.6	69.0	38.2	30.8	64.5	26.1	38.4

is based on three criteria related to the *Kp* values for each day and the order numbers are selected as the ten quietest and the five most disturbed days of the month.

Another index often used for monitoring geomagnetic activity is the *ap* index. It is a linear index derived from each 3 h planetary index *Kp* in such a way that at 50° dipole latitude the *ap* index approximately represents the maximum disturbance range of the largest of the three components *H*, *D*, and *Z* in units of 2.0 nT. The *ap* index ranges from 0 to 400. The sum of eight 3 h *ap* values gives the daily geomagnetic planetary *Ap* index for a universal day. *Kp* and

Ap values are published regularly by the GFZ Helmholtz Centre, Potsdam at http://www-app3.gfz-potsdam.de/kp_index/index.html.

A particular class of fluctuations in the geomagnetic field, with decreases of several hundred nT in the H component, are known as geomagnetic, or simply magnetic, storms. Storm variation D can be defined as $D = Dst + DS$, where DS is due to auroral electrojet activity and Dst is due to the ring current in the magnetosphere. Currently, the storm-time variation index Dst is calculated using the H component from four observatories: uniformly distributed in longitude and remote from the effects of the auroral and equatorial electrojets: Honolulu, San Juan, Kakioka, and Hermanus. Although the continuous hourly values of the Dst index primarily give information on geomagnetic storm variations, they are also useful for characterizing persistent quiet-time geomagnetic variations.

Geomagnetic field fluctuations due to the currents in the auroral-zone ionosphere are represented by the AE index. The AE index is calculated from the H components at geomagnetic observatories situated on auroral or sub-auroral latitudes and uniformly spaced in longitude. The maximum positive (upper) amplitude of the H component is called the AU index and the maximum negative (lower) amplitude is called the AL index. The index $AE = AU - AL$, and so represents the difference between the upper and lower limits of the magnetic fluctuations in ΔH at any time. Dst and AE values are published regularly by the Data Analysis Center for Geomagnetism and Space Magnetism Faculty of Science, Kyoto at http://wdc.kugi.kyoto-u.ac.jp.

2.3 General Atmosphere

2.3.1 General Description of the Atmospheric Regions, Composition, and Temperature

The atmosphere is the gaseous envelope surrounding the planet Earth and comprising a mixture of gases. The major components of the atmosphere are summarized in Table 2.3. With the exception of oxygen, the more abundant components are chemically inactive and therefore maintain the same volumetric concentration up to a height of 100–110 km. Inactive components, like carbon dioxide, play an important role in the energy balance, while active components like ozone, despite its very low concentration, play an important role in the absorption of UV radiation and X-rays.

An important property of the atmosphere is pressure p, equal to 1.03 Pascal at sea level, and decreasing together with density ρ at increasing altitude. The pressure p at an altitude h is defined per unit of surface area as the weight of all the air above:

$$p = \int_0^h g\rho dh \tag{2.3.1}$$

where g is the acceleration of gravity.

Table 2.3 Composition of the atmosphere and concentration of the major components

Inactive elements		
Nitrogen	N_2	78.11 %
Oxygen	O_2	20.95 %
Argon	Ar	0.93 %
Neon	Ne	18.18×10^{-4} %
Helium	He	5.24×10^{-4} %
Active elements		
Water	H_2O	From 0 to 7 %
Carbon dioxide	CO_2	From 0.01 to 0.1 %
Ozone	O_3	From 0 to 0.00001 %

Another important parameter characterizing the neutral atmosphere is the scale height H as will be seen in the equations modelling ionization.

The atmosphere is taken to be a fluid in hydrostatic equilibrium composed of a mixture of gases, in which the law of ideal gases is still valid with acceptable approximation:

$$p = nKT \tag{2.3.2}$$

with K the Boltzmann constant (1.38×10^{-23} J/K), n the number of molecules per unit of volume, and T the absolute temperature.

Then, assuming that atmospheric pressure can vary only with height h, the upward force due to the gradient is equal to the weight of the atmosphere above:

$$\frac{dp}{dh} = -g\rho \tag{2.3.3}$$

and assuming that the density ρ is also $= n\ m$, with m the mean molecular mass, gives:

$$\frac{dp}{dh} = -g\, n\, m \tag{2.3.4}$$

Then taking n from Eq. 2.3.2:

$$\frac{dp}{p} = -\frac{gm}{KT}dh \tag{2.3.5}$$

By integrating this equation:

$$p = p_0 \exp\left(-\int_0^h \frac{gm}{KT}dh\right) = p_0 \exp\left(-\int_0^h \frac{dh}{H}\right) \tag{2.3.6}$$

When $H = KT/gm$ is defined as the scale height and p_0 is the pressure at a height $h = 0$.

To understand the physical meaning of this parameter some approximations are introduced. It is assumed that up to 100 km, m can be considered constant so H should vary only in proportion to T and in the case also of uniform temperature:

$$p = p_0 \, e^{-h/H} \tag{2.3.7}$$

So H results as the difference in altitude corresponding to a variation in pressure p equal to e^{-1} i.e. about 0.37 %, in other words the difference in altitude when the atmosphere has a pressure or a density ρ equal to 1/e of the reference level. At sea level, H is equal to 8 km and remains almost constant, as assumed above, up to an altitude of 100 km. Above 100 km the approximation defined above is no longer valid. In fact, in addition to T varying with altitude, m is also not constant but varies according to the molecular mass of the different elements, each of which is distributed according to its own scale height. This means that components with lower molecular mass, like hydrogen or helium, have a greater scale height and are more numerous at higher altitude. Moreover, the scale height represents a measure of the density and pressure gradient, so a lower scale height represents a greater gradient of pressure or density.

The most important parameter used to describe the vertical structure and regions of the atmosphere is the behaviour of temperature with altitude, because this trend reflects the numerous physical and chemical processes in play. As will be discussed later, electron density behaviour can be used to describe the vertical structure of the ionosphere (Fig. 2.16).

From the Earth's surface up to 10 km above the poles and up to 15 km above the equatorial regions the average temperature decrease is about 6.5 ° per km up to

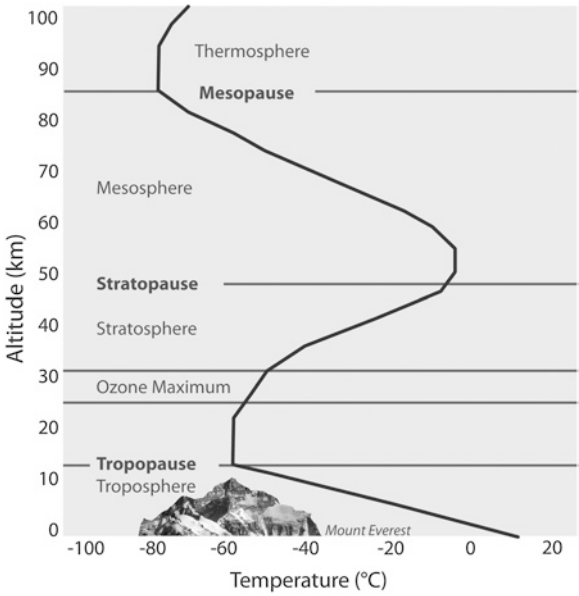

Fig. 2.16 The temperature of the atmosphere versus altitude and the regions characterized by the inversion of this physical parameter

a level known as the tropopause. Up to this level is the section of the atmosphere that supports life, where most weather phenomena occur, and where the physical process is mainly thermodynamic. The reason for this decrease in temperature is the vertical convective movement of the air: the gas in contact with the warm surface of the Earth expands and cools as it moves upwards in an adiabatic process. Above the troposphere and tropopause the temperature begins to increase again from −70 to 0 K, up to an altitude of 50 km limited by the stratopause. This region is called the stratosphere and the increase in temperature is due to the existence of a layer of ozone, a tri-atomic molecule of oxygen that absorbs UV radiation from the Sun generating new bi-atomic and mono-atomic molecules of oxygen in the process illustrated in Fig. 2.17.

Unlike the troposphere, the temperature inversion blocks vertical movements. Above a maximum altitude of about 50 km the temperature begins to decrease again in a transition region called the mesosphere, continuing up to about 90 km where the temperature reaches about −70 to −90 K. The upper limit of the mesosphere is the mesopause, above which lies the thermosphere where the temperature increases again due to the absorption of UV rays by oxygen and nitrogen, reaching a temperature of 1,500 K at an altitude of 200–300 km. Matter at this height is extremely low density and consequently the temperature parameter is meaningless and it is better to consider the mean quadratic velocity of the particles. In this portion of the upper atmosphere the molecular oxygen and nitrogen are not evenly mixed, instead being stratified according to their molecular mass with a particular scale height. In the thermosphere and upper region of the exosphere the strong UV radiation and X-rays produces an intense ionization so these two regions overlap with the ionosphere.

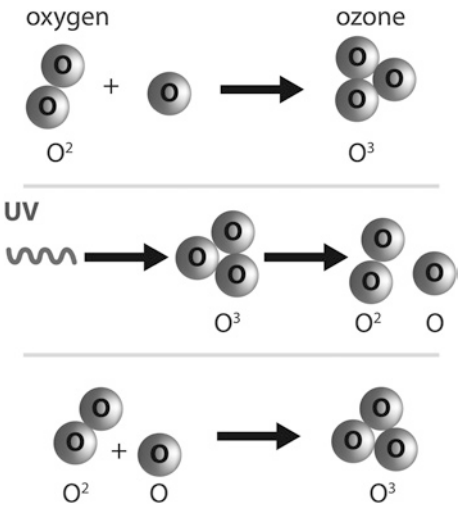

Fig. 2.17 The stratosphere contains mono-atomic, bi-atomic, and tri-atomic molecules of oxygen. The absorption of UV radiation by an ozone molecule produces an oxygen atom plus a molecule of bi-atomic oxygen. Subsequently the oxygen atom is recaptured by a bi-atomic molecule of oxygen reforming an ozone molecule

2.3.2 Formation of the Earth's Ionosphere

Above the stratosphere, which means beyond an altitude of 50 km, the atmosphere is characterized by a high density of free electrons and free ions, mostly produced by the energetic photo-ionization of UV and X-rays arriving from the Sun, and to a minor extent over high-latitudes by corpuscular ionization. This section of the atmosphere, including the layers referred to as the mesosphere, thermosphere, and exosphere, is defined as the ionosphere.

A high or relevant density of free electrons and free ions is not a clearly defining characteristic because electrons and ions are present at every altitude in the lower and upper atmosphere. Therefore, a more practical definition, originating from the first application of long distance radio communications, is that part of the atmosphere in which the density of ionization is sufficient to deflect radio waves in the 2–30 MHz range.

However, while it is easy to define the lower limit of the ionosphere opinions differ on how to establish its upper limit. A reasonable compromise would be up to 1,000 km, well above any electron density peak and above the limits where the ionosphere is identified as a transition region with the plasmasphere.

2.3.2.1 Photo-Ionization

A simplified photo-ionization process is shown in Fig. 2.18, in which a UV radiation strips electrons from neutral atoms which absorb the energy of a photon provided by the quantity hc/λ, where h is the Plank constant 6.62×10^{-34} Js, c the velocity of light, and λ the length of the incident electromagnetic wave, as represented by the equation:

$$a + \frac{hc}{\lambda} \rightarrow a^{+} + e^{-} \tag{2.3.8}$$

Here a represents one generic neutral atom, a^{+} represents a positive ion, and e^{-} the free electron, so that the quantity hc/λ must be $> Li$ in J, the energy necessary for ionization and related to the specific atomic component i.

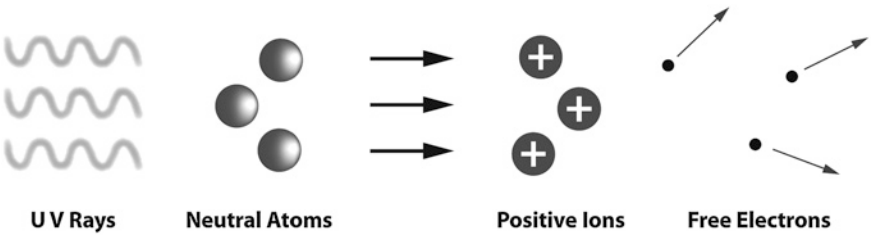

Fig. 2.18 A simplified scheme of photo-ionization: electrons are detached from their neutral atoms by the energy of UV rays from the Sun, producing ions and free electrons

The amount of energy required for the ionization of the different components of the atmosphere is well known and tabulated. This energy is expressed in electron-volt (eV) or simply with the length λ expressed in nm (nanometres) or Angstrom ($\mathring{A} = 10^{-10}$ m) of the associated electromagnetic wave. Furthermore, the production of ion electron pairs is also conditioned by the probability of collision or interaction between a photon and a neutral atom. This is proportional not only to the intensity of radiation I, which means the number of photons, but also to the absorption cross section σ_f of the ionization which depends on the frequency f for a specific chemical component i. The production q_i is given by:

$$q_i = \sigma_f n I \tag{2.3.9}$$

where n is density of the neutral particles of the element i.

2.3.2.2 Processes of Electron Loss: Recombination

The reverse phenomenon of photo-ionization is recombination, when the free electrons recombine with the positive ions to produce neutral atoms again. Recombination can occur in two ways. In the first, defined as radiative recombination, the electrons e^- interact directly with the positive ions a^+ producing a neutral atom a or molecule plus excess energy L:

$$a^+ + e^- \rightarrow a + L \tag{2.3.10}$$

The second and more frequent is called dissociative recombination and involves first an interaction between positive ions a^+ originating from the photo-ionization and existing neutral bi atomic molecules b_2, like oxygen or nitrogen, according the following equation:

$$a^+ + b_2 \rightarrow ba^+ + b \tag{2.3.11}$$

Substituting in the molecule b_2 a neutral atom b with positive ion a^+; and then by the combination of free electrons with the positive molecules ba^+ producing two neutral atoms:

$$ba^+ + e^- \rightarrow b + a \tag{2.3.12}$$

The probability of occurrence of the first process is proportional to the density of electrons or negative ions N^- and the density of the positive ions N_i^+ according to a recombination factor given by:

$$l_r = \gamma_i N^- N_i^+ \tag{2.3.13}$$

where γ_i is a coefficient dependent on the specific ion component. This process is peculiar to the higher parts of the ionosphere while, less frequently, in the lower ionosphere the loss of electrons is due to attachment of electrons to neutral atoms, which produces negative ions a^- plus excess energy L:

$$e^- + a \rightarrow a^- + L \tag{2.3.14}$$

Again the probability of occurrence is proportional to the electron density N and to the density of the specific element n, and also depends on a coefficient factor δ_i determined by the capability of the element to attract the electrons:

$$l_c = \delta_i N^- n \tag{2.3.15}$$

So finally the equation of continuity or the equation of electronic equilibrium can be written as:

$$\frac{dN}{dt} = q - l + d \tag{2.3.16}$$

where dN/dt represents variation in time and in unit volume of electron density, q is the rate of electron production, l the rate of electron loss, and d a factor that takes into account losses due to diffusion, neutral winds, and vertical electromagnetic drift. Substituting Eqs. 2.3.9, 2.3.13, and 2.3.15 into Eq. 2.3.16, for a specific element i:

$$\frac{dN}{dt} = \sigma_f nI - \gamma_i N^- N_i^+ - \delta_i nN^- + d \tag{2.3.17}$$

Then for all the components i:

$$\frac{dN}{dt} = \sigma_f nI - \alpha N^- N^+ - \beta nN^- + d \tag{2.3.18}$$

where α and β are the effective coefficients of recombination and attachment respectively.

2.3.2.3 The Chapman Model

A simple theory of photo-ionization was formulated by Sydney Chapman in 1931 and is able to describe the behaviour of an ionospheric layer and its variations during the day. The Chapman model was of seminal importance for the many models describing the behaviour of the ionosphere that followed, even in recent times. The mathematical model known as the Chapman function is based on a list of simple assumptions:

1. the atmosphere is composed by only one chemical element exponentially distributed;
2. the atmosphere is plane stratified and not subject to turbulence, diffusion or horizontal variations;
3. the temperature is constant so the scale height $H = KT/gm$ is constant too;
4. ionization is caused only by photo-ionization through absorption of solar monochromatic radiation, and electron loss only by recombination.

According to the above assumptions the Eq. 2.3.18 can be simplified taking into account that $N^- = N^+$ because it is caused by the production of pairs of electron ions, and the contribution of βN^- is not considered:

$$\frac{dN}{dt} = \sigma_f nI - \alpha N^2 \tag{2.3.19}$$

Bearing in mind the Eq. 2.3.7 it could be assumed that the number of particles also follows the same law:

$$n = n_0\, e^{-h/H} \tag{2.3.20}$$

which becomes $n = n_0\, e^{-z}$ if it is assumed that $z = (h - ho)\,/H$ as a new system of coordinates.

In this system (Fig. 2.19) the decrease in radiation intensity I along the path $ds = \sec\chi\, dh = \sec\chi\, H dz$ is given by:

$$dI = -\varepsilon I n H \sec\chi\, dz = -\varepsilon I n_0 H \sec\chi\, e^{-z} dz \tag{2.3.21}$$

when ε is a numerical coefficient of absorption and χ is the solar zenith angle. It is important to note that the solar zenith angle χ, the angle between the perpendicular to a point on the plane of the Earth's surface and the line of direction of the Sun, can be calculated at any point on the planet for any time of day using the relation:

$$\cos\chi = \sin\varphi \sin\delta + \cos\varphi \cos\delta \cos\omega \tag{2.3.22}$$

when φ is the geographic latitude of the point, δ the declination of the Sun on a given day of the year, and ω the hourly angle of the Sun. This angle represents true solar time, i.e. the angle between the plane containing the point on the Earth and the Sun and the plane containing the local meridian ($\omega = 0$). Then, integration of the Eq. 2.3.21 between $z = 0$ ando $z = \infty$ gives:

$$I = I_\infty e^{-\varepsilon\, n_0 H \sec\chi \exp(-z)} \tag{2.3.23}$$

where I_∞ may be considered as the intensity of radiation out of the atmosphere or before any absorption. Then considering that $q = \sigma_f\, nI$:

$$q = \sigma_f\, n_0 e^{-z} I_\infty e^{-\varepsilon\, n_0 H \sec\chi \exp(-z)} = \sigma_f\, n_0 I_\infty e^{(-z-\varepsilon\, n_0 H \sec\chi \exp(-z))} \tag{2.3.24}$$

This function shows how q depends on the solar zenith angle χ and the height z. In fact, q is maximum when the solar zenith angle is minimum or $\chi = 0$, i.e. $sec\ \chi = 1$:

$$q = \sigma_f\, n_0 I_\infty e^{(-z-\varepsilon\, n_0 H \exp(-z))} \tag{2.3.25}$$

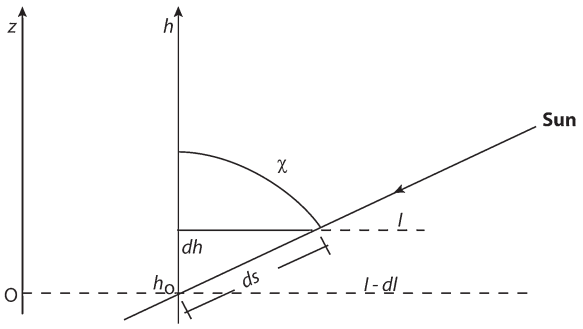

Fig. 2.19 The decrease in radiation intensity I along the path dS (from Dominici 1971)

Then to establish at which height z_M q reaches maximum it is necessary first to calculate the derivate *dq/dz,* with $\chi = 0$, then imposing the derivate of this function = 0. So:

$$\frac{dq}{dz} = \sigma_f n_0 I_\infty \left[\left(-1 + \varepsilon n_0 H e^{-z} \right) e^{(-z - \varepsilon n_0 H \exp(-z))} \right] = 0 \tag{2.3.26}$$

This is true when:

$$e^{z_M} = \varepsilon n_0 H \tag{2.3.27}$$

If now we consider the height $z = 0$ in the new system of coordinates as the height where q is maximum, true for $\chi = 0$ and now $z_M = 0$, this means that $\varepsilon n_0 H = e^0 = 1$. At this height:

$$q_0 = \sigma_f n_0 I_\infty e^{-1} \tag{2.3.28}$$

so the function in Eq. 2.3.24 could be rewritten for a generic height as

$$q = \sigma_f n_0 I_\infty e^{(-z - \varepsilon n_0 H \sec \chi \exp(-z))} = q_0 e^{(1 - z - \sec \chi \exp(-z))} \tag{2.3.29}$$

Then imposing the derivate of this function = 0:

$$\frac{dq}{dz} = q_0 \left(-1 + \sec \chi \cdot e^{-z} \right) e^{(1 - z - \sec \chi \exp(-z))} = 0 \tag{2.3.30}$$

which is true if $\left(-1 + \sec \chi \cdot e^{-z} \right) = 0$. This means that:

$$z_M(\chi) = -\ln \cos \chi \tag{2.3.31}$$

Substituting this equation in Eq. 2.3.29 also gives:

$$q_M(\chi) = q_0 \cos \chi \tag{2.3.32}$$

the maximum production rate of ionization $q_M(\chi)$ and its height $z_M(\chi)$ as a function of the solar zenith angle. The consequences of these functions are clearly seen in Fig. 2.20 where it is shown how the quantity q/q_0 can vary according to altitude z and varying solar zenith angle χ: the maximum rate of ionization q_M decreases when the Sun goes down to the horizon or moves to its highest position.

2.3.3 General Electron Density Profile

As described in the previous section the principal cause of the high density of free electrons and positively charged ions and molecules is photo-ionization from UV radiation arriving from space and in particular from the Sun, ignoring here other causes like energetic particles, again originating from the Sun, or radioactive elements from natural sources or fall-out pollution. Bearing in mind that the density of neutral atoms decreases with altitude and that ionization becomes weaker because of absorption, the two opposite phenomena should logically result in a layer containing a maximum concentration of free electrons.

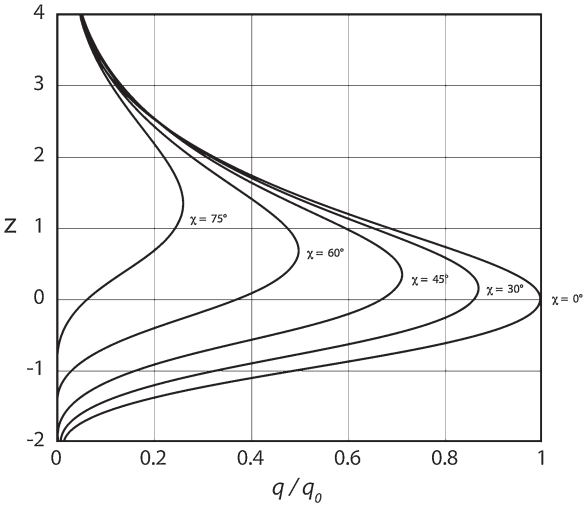

Fig. 2.20 The quantity q/q_0 versus altitude z and for different solar zenith angles χ: the maximum rate of ionization decreases when the Sun goes down to the horizon or moves to its highest position

Since the upper neutral atmosphere is composed of different chemical elements which interact differently with UV radiation, more than one electron density maximum is found. Therefore, in addition to the temperature of the neutral atmosphere, electron density behaviour *versus* altitude is another useful parameter for describing different regions of the ionosphere, with the relative maxima and minima used to identify the ionospheric layers (Fig. 2.21).

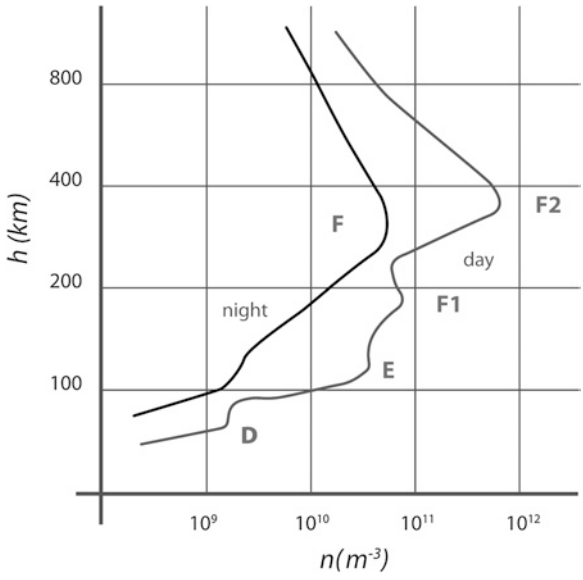

Fig. 2.21 Diurnal and nocturnal electron density profiles versus altitude

Historically, the letter E was used for the first layer observed by experimental measurements by Edward V. Appleton because of the electric field E reflected by the ionospheric layers, and subsequently the layer above was called F and the layer below D.

The two generic electron density profiles of the figure, one nocturnal and one diurnal, show that during the day all the ionospheric regions can be observed by routine ionospheric soundings and in some seasonal contexts even the splitting of the F region into F1 and F2 layers. Conversely, the nocturnal profile shows the very weak ionizations in the D and E regions. At those altitudes the high density of neutral and ionized particles induces rapid recombination of free electrons and positive molecules, while at higher altitudes of the F region the recombination time is sufficiently long to permit photo-ionization to start again at dawn.

The situation in the polar regions is much more complex because the night can last for several months and a very weak ionosphere is still observed due to marked phenomena of diffusion and transport. Nocturnal profiles logically exhibit a lower density than diurnal profiles at all altitudes.

2.4 Ionospheric Regions

The structure and dynamics of the Earth's ionosphere are subject both to large spatial and frequent temporal variations, which can be periodic as well as irregular. The changes that occur in the ionosphere are different at different altitudes because of varying relative ionization, loss, and transport phenomena as described in Sect. 2.3. Consequently, the terrestrial ionosphere is divided into regular and sporadic regions according to the thermal and chemical properties of the neutral gas and ionized components.

2.4.1 Regular Ionospheric Regions

Modern experimental and theoretical investigations divide the ionosphere into three regions: D, E, and F. Under certain solar-terrestrial conditions these regions are split into four main layers D, E, F1, and F2 as shown in Figs. 2.21 and 2.22. In principle, the lower ionosphere (up to 100 km) is a zone in which photochemical processes are the main influence on its formation and ionization balance. The boundary of the middle ionosphere (100–170 km) marks the limit of ionization-recombination processes together with thermal and dynamic processes. The upper ionosphere is the F region, characterized by the transfer of charged particles in plasma by ambipolar diffusion, thermospheric winds, and ionosphere-magnetosphere interactions. The real heights of the ionospheric layers vary with solar zenith angle time, time of day, seasons, solar cycles, and solar activity.

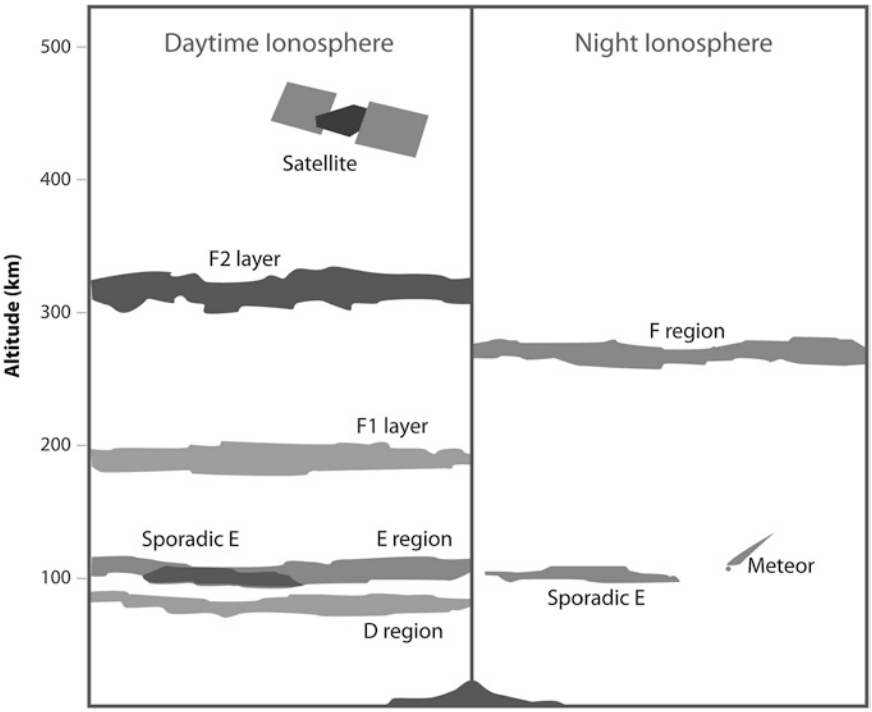

Fig. 2.22 Day and night structure of the terrestrial ionosphere

2.4.1.1 D Region

The D region of the ionosphere lies at an approximate altitude of 50–90 km, with the section from 50 to 70 km sometimes referred to as the C region and produced mainly by cosmic rays. The neutral component of the D region consists mostly of N_2, O_2, Ar, CO_2, He, and a highly variable quantity of O_3 and H_2O. The ionized component of the D region consists mostly of NO^+ (nitric oxide) as the major positive charge carrier, while electrons, O_2^- and possibly other negative ions are the negative charge carriers (Fig. 2.23). Although the ionizing agents of the D layer are still the object of intense study, solar L-alpha (121.6 nm) radiation appears to be the most important for ionizing NO, solar X-rays (<0.8 nm) for ionizing N_2, O_2, and Ar, solar ultraviolet (UV) radiation ($\lambda < 111.8$ nm) for ionizing unstable O_2, and galactic cosmic rays for ionizing all atmospheric constituents (Fig. 2.24). Rocket and satellite observations suggest that solar cosmic rays, solar protons of 1–100 MeV, and possibly solar electrons of energy >10 keV are the likely contributors to the ionization of the D region.

As Fig. 2.23 shows, D region electron densities are typically around 10^8–10^9 el/m^3 depending on height. They are subject to typical diurnal, seasonal, and solar-cycle variations. Since the electron collision frequency in the D region is as high

as $2 \times 10^6\ s^{-1}$, radio wave absorption is quite significant and is very important for the calculation of the Lower Usable Frequency (*LUF*) in radio links as will be discussed in Chap. 9.

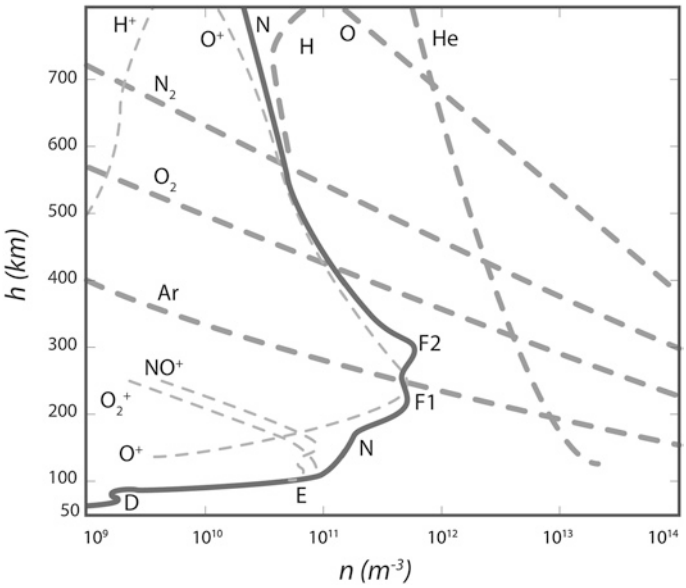

Fig. 2.23 Principal neutral and ionic constituents of the ionosphere (N is for free electrons) (from Dominici 1998)

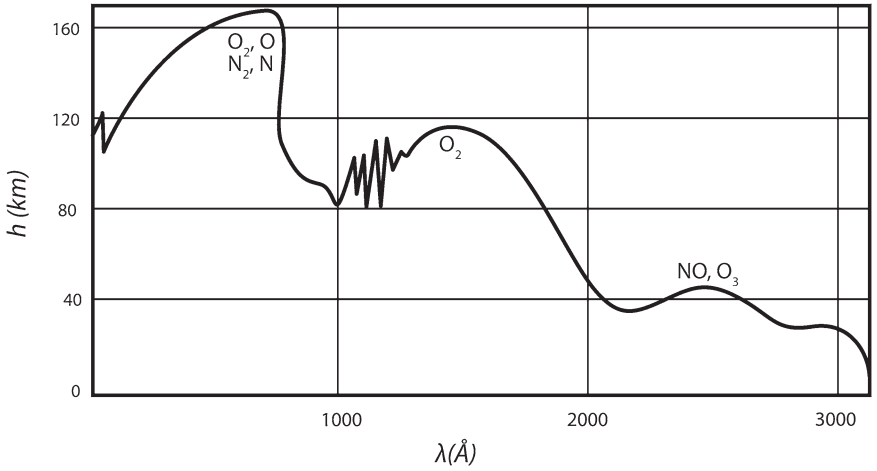

Fig. 2.24 Atmospheric absorption of solar electromagnetic radiation (from Dominici 1971)

2.4.1.2 E Region

The ionosphere from an altitude of 90–140 km is called the E region and contains the E layer, sometimes an E2 layer, and a sporadic Es layer (Fig. 2.22). Since the sporadic Es layer is considered to be an independent phenomenon from the normal E region, it is described in Sect. 2.4.2. Ionization in the E layer is mainly caused by X-rays in the 8–10.4 nm range and UV radiation from 80 nm to L-beta (102.6 nm), producing the principal NO^+, $O_2{}^+$ and secondary O^+, $N_2{}^+$ ion components (Figs. 2.23 and 2.24). Together with positive ions, the E region also contains electrons with an average E layer electron density of ~10^{11} el/m^3 at around ~110 km (Fig. 2.23). The E layer exhibits the Chapman model behaviour with daily maximum at local noon, seasonal maximum in summer, and solar cycle dependence. The E layer does not disappear completely at night and remains weakly ionized. When an E2 layer occurs it is between the normal E and F1 layers at an altitude of about 150 km. Even though the cause of E2 layer formation is still being studied, it seems that the E2 layer is controlled purely by solar ultraviolet radiation.

The electric field and electron drift velocities in the E layer in general, and the low-latitude E layer irregularities arising from the equatorial electrojet in particular, are the key E region features and have been studied in numerous radar experiments.

2.4.1.3 F Region

The F region is a major segment of the terrestrial ionosphere and the most important from the point of view of radio communications and navigation systems. It lies between 140 and 600 km in altitude, occasionally with altitudes extending to the upper limits of the ionosphere. As a result of the complex physical mechanisms involved in its formation, solar radiation causes the F region to split into two separate layers, called F1 and F2. The F2 layer is always present during the night and so these layers merge a few hours after sunset, reshaping the F region.

2.4.1.4 F1 Layer

The F1 layer of the ionosphere lies at an altitude of approximately 140–210 km (Fig. 2.23). The ionized component of the F1 layer consist mostly of NO^+ and $O_2{}^+$ as primary, and O^+, N^+ as secondary positive charge carriers (Fig. 2.23). The main source of ionization is extreme ultraviolet (EUV) solar radiation in the wavelengths $\lambda \approx 58.4$ and 30.4 nm (Fig. 2.24). The height of F1 varies with solar activity, season, and geomagnetic activity. Rocket experiments have revealed a maximum density height h_mF1 from 160 to 180 km, while ionograms from ionosonde measurements indicate ≈200 km.

Like the E layer, the F1 layer is a Chapman layer with maximum electron density reaching about 2×10^{11} el/m^3 at midday (Fig. 2.23). It exhibits dependence on solar zenith angle χ and sunspot number R. Accordingly, the F1 layer is more pronounced in summer than in winter, always disappears during the night and sometimes in winter even during the day. Under those conditions the F1 and F2 layers form a single portion of the ionospheric F region.

2.4.1.5 F2 Layer

The F2 layer is present 24 h a day under all solar-terrestrial conditions, making it the most important layer of the ionosphere (Figs. 2.21, 2.22, and 2.23). The F2 layer has a different nature from the photochemically dominated lower layers and the diffusion dominated topside ionosphere. The main characteristics of the F2 layer are its high variability, on timescales ranging from the 11 years of a solar cycle and even longer, to a few seconds during strong interactions with the plasmasphere above (at altitudes >1,000 km) depending on solar-terrestrial conditions. It is reasonably obvious that replenishment from the plasmasphere plays a significant role in maintaining the F region during the night.

In the F2 layer the dominant charge carriers are O^+ ions with secondary H^+ and He^+ (Fig. 2.23). The main electron–ion source in the F region is the process of photo-ionization by EUV radiation in the wavelength range from 5 to 102.7 nm. The F2 layer has a maximum electron density between 10^{10} and 8×10^{12} el/m^{-3} (Fig. 2.23). Because the F2 layer is a non Chapman layer, there is no solar zenith angle dependence of the diurnal and latitudinal variations in electron density. However, the Sun does affect the electron density of the F2 layers causing a rapid increase after sunrise, with maximum values occurring at any time during the day.

Moreover, the complexities of F2 layer behaviour can be understood in terms of: (1) chemical changes, because the electron density equilibrium depends strongly on the [O]/[N_2] ratio and also on the [O]/[O_2] ratio; (2) diurnal heating and cooling, since the height distribution of the F_2 layer is governed by the plasma scale height, $H_p = K\,(T_i + T_e)\,/m_i\,g$ (see Eq. 2.3.6); (3) winds in neutral air at heights around 300 km that range between tens and hundreds of ms^{-1}, at mid-latitudes directed towards the poles by day and towards the equator at night; and (4) electric fields, since the electrical structure and dynamics of the upper atmosphere generate electrodynamic drifts in the F region that can significantly modify the vertical distribution of ionization. However, because of the dip angle of the magnetic field, the effects of electric fields are much more important for the equatorial ionosphere while the effects of neutral winds are much more important at mid-latitudes.

In the F2 layer, the rate of electron loss depends on the concentration of molecular ions, particularly O_2 and N_2, whereas the production rate depends on the concentration of atomic oxygen. The photoelectrons produced by ionization processes are hotter than the neutral atoms from which they were formed. This excess energy is gradually shared with the positive ions, but transfer to the neutral component

is less efficient. Consequently the temperature of the plasma exceeds that of the neutral air. By day the electrons are considerably hotter than the ions ($T_e > T_i$), but by night the temperatures are more similar. When the plasma is hot it moves up to higher altitudes, where it is effectively stored because of the slower rate of recombination with height. When the plasma cools, at night, it moves down again and so helps to maintain the F region against loss by recombination at lower altitudes.

In 1997 evidence for another layer, named the F3 layer, in the equatorial F region ionosphere was found. Since the presence of such a layer could affect HF communication and surveillance systems, experimental and theoretical investigations were conducted. These suggest that a combined effect of electrodynamic drift and neutral wind during morning to noon periods can drift an F2 layer upwards to form an F3 layer, while the usual photochemical and dynamic effects maintain a normal F2 layer at lower altitudes.

2.4.2 Sporadic Ionospheric Layer Es

Although a sporadic E layer occurs at altitudes from 90 to 140 km (the E region), it is usually considered independent of the normal E layer of the ionosphere (Fig. 2.22). The main reason for this view is that Es has very diverse spatial (it can be spread over a wide area or confined to a small area), diurnal (it can appear at any time during the day or night), and seasonal patterns at high, mid, and low-latitudes generated by different physical mechanisms. An Es in low and mid-latitudes occurs mostly during the day and prevalently during the summer months. At high-latitudes, an Es is more likely to occur at night and is frequently associated with aurora. Most importantly, sporadic E layers can have an electron density similar to the F region. However, its random time of occurrence and presence at any particular place makes Es layer prediction very difficult.

There have been many theories for the sporadic E layer based on a large quantity of observational data but the most widely accepted theory still appears to be the wind-shear theory. According to this theory a reversal of the east–west neutral wind combines with the Earth's magnetic field to produce upward and downward motions of heavy ions, thereby compressing them into a layer coincident with the height of wind reversal. Some more recent work indicates that the electric field may play a part in auroral regions where it is very large and the magnetic field is nearly vertical.

2.5 Ionospheric Irregularities

Ionospheric irregularities generally include conditions that cannot be accurately described by standard ionospheric models, and events that do not follow conventional patterns predicted on the basis of their physical causes. As illustrated in

Sect. 2.4.1, the F region is highly variable and exhibits, among other irregularities, two very well-known phenomena recognized as travelling disturbances and spread-F.

2.5.1 Travelling Ionospheric Disturbances

Travelling ionospheric disturbances (TIDs) play an important role in the dynamics of the thermosphere and ionosphere. For example, the existence of atmospheric gravity waves (AGWs) in the Earth's upper atmosphere, originating from the weather system of the stratosphere and troposphere, was discovered by Hines (1960) from ground-based observations of travelling ionospheric disturbances. TIDs are irregularities of the F region expressed as wave-like oscillations of the contours of constant electron density, descending slowly through time (Fig. 2.25). TID occurrence and direction of travel show significant diurnal, seasonal, and sunspot cycle variations. They are classified into two types: large-scale TIDs and medium-scale TIDs.

Large-scale travelling ionospheric disturbances (LS TIDs) have a period of 1–3 h and a horizontal wavelength of 1,000–4,000 km, moving faster than 300 m/s. They are thought to be a manifestation of atmospheric gravity waves excited by sources in polar regions of the northern and southern hemispheres. Consequently, the study of LS TIDs could provide important information on auroral processes under

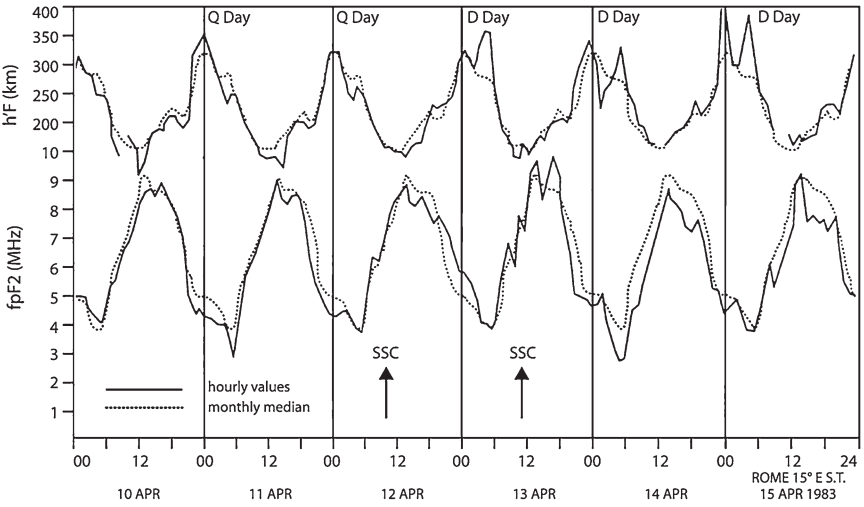

Fig. 2.25 Variations in some *fpF2* and *h'F* values showing local appearance of TIDs during geomagnetically quiet and disturbed days in April 1983 at the Rome (41.9° N, 12.5° E) ionosonde station

quiet and disturbed geomagnetic conditions. It is known that surges in the auroral electrojet generate trains of large-scale TIDs in both hemispheres, and these trains subsequently propagate towards the equator, causing substantial increases in the height of the ionosphere.

Medium-scale travelling ionospheric disturbances (MS TIDs) have shorter periods of 10 min to 1 h and a wavelength of up to 300 km, moving more slowly (50–300 m/s). MS TIDs are related to meteorological phenomena like neutral winds or the solar terminator, which produce AGWs expressed as TIDs at ionospheric heights.

2.5.2 Spread-F

Usually the F region reflections of radio waves are clear echoes indicating well-defined ionospheric layers as described in Sect. 2.4. Spread-F occurs when the ionospheric F region becomes diffused and irregularities scatter the radio waves (Fig. 2.26). The size of F region irregularities, which occur in patches, varies from 20 km to more than 100 km. The solar cycle and seasonal variations, as well as local time dependence, are well established characteristic of spread-F occurrence. In addition, the probability of spread-F occurring is strongly dependent on latitude since spread-F occurrence rates and characteristics at low-latitudes differ significantly from high-latitudes.

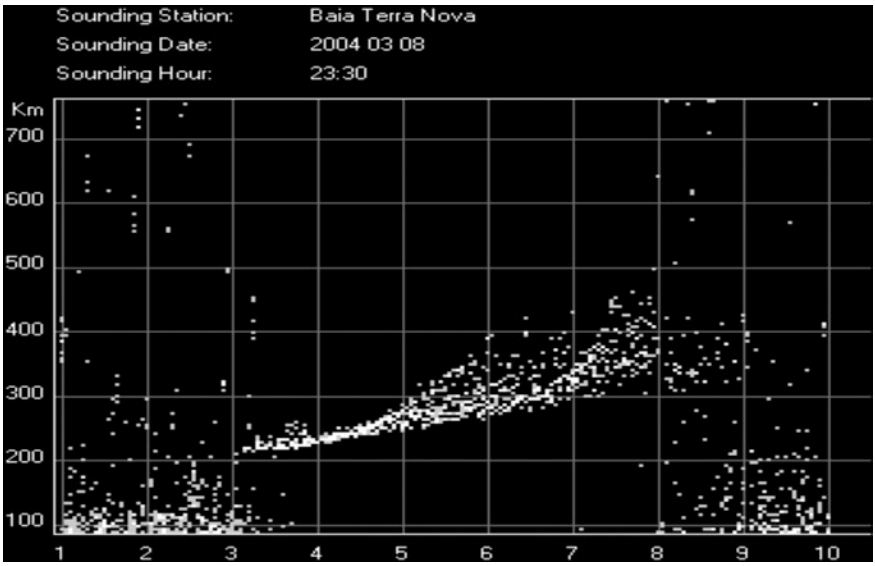

Fig. 2.26 A typical spread-F ionogram recorded at an Antarctica Baia Terra Nova (74.7° S, 164.1° E) ionosonde station at 2330 UT on 8 March 2004

At low-latitudes, spread-F occurs mostly during the night and around the equinoxes. At mid-latitudes, spread-F is more likely to occur at night and in winter. At latitudes greater than about $\pm 40^\circ$ geomagnetic latitude, spread-F is a night-time phenomenon, mostly around the equinoxes. Near the magnetic poles, spread-F is frequently observed both during the day and night. At high-latitudes spread-F is correlated with geomagnetic activity. On the contrary, at the equator there is no geomagnetic control over spread-F.

Suggested Readings

Akasofu S.-I, Chapman S (1972) Solar Terrestrial Physics. Oxford University Press, Oxford

Balan N, Bailey GJ, Abdu MA, Oyama KI, Richards PG, MacDougall J, Batista IS (1997) Equatorial plasma fountain and its effects over three locations: evidence for an additional layer, the F3 layer. J Geophys Res 101:2047–2056

Boschi E, Visconti G, Rizzi V, Zolesi B (1991) Il Buco dell'ozono: indizio del cambiamento globale. Editor: Il Cigno Galileo Galilei

Bowman GG (1964) Spread-F in the ionosphere and the neutral particle density of the upper atmosphere. Nature 201:564–566

Bowman GG (1990) A review of some recent work on mid-latitude spread-F occurrence as detected by ionosondes. J Geomag Geoelectr 42:109–138

Brekke A (1997) Physics of the Upper Atmosphere. Wiley published in association with Praxis Publishing

Campbell WH (1997) Introduction to Geomagnetic Fields. Cambridge University Press, Cambridge

Davies K (1990) Ionospheric Radio. IEE electromagnetic waves series 31, Peter Peregrinus Ltd, London

Dominici P (1971) Radio Propagazione Ionospherica. Supplement to vol XXIV of Ann Geofis

Dominici P (1998) My first fifty years in ionospheric research. Ann Geofis 41:857–883

Dominici P, Cander LjR, Zolesi B (1997) On the origin of medium-period ionospheric waves and their possible modelling: a short review. Ann Geofis XL:1171–1178

Hargreaves JK (1992) The Solar-Terrestrial Environment. Cambridge Atmospheric and Space Science Series 5, Cambridge University Press, Cambridge

Hines CO (1960) Internal gravity waves at ionospheric heights. Can J Phys 38:1441–1481

IEEE Std 211-1997, IEEE standard definition of terms for radio wave propagation. The Institute of Electrical and Electronics Engineers, New York (1998)

ITU-R Choice of indices for long-term ionospheric predictions, Recommendation ITU-R P. 371-8, International Telecommunication Union, Geneva (2001)

Kelley CM (1989) The Earth's Ionosphere. Academic Press Inc, California

Lanza R, Meloni A (2006) The Earth's Magnetism. Springer, Heidelberg

McNamara LF (1991) The Ionosphere: Communications, Surveillance, and Direction Finding. Krieger Publishing Company, Malabar

Sugiura M (1964) Hourly values of equatorial Dst for the IGY. Ann Int Geophys Year 35:9, Pergamon Press, Oxford

Villante U (2001) Al di là delle nuvole, la fisica delle relazioni Sole-Terra. Bollati Boringhieri, Saggi Scienze

Whitehead JD (1989) Recent work on mid-latitude and equatorial sporadic-E. J Atmos Terr Phys 51:401–424

Wilcox J, Ness NF (1967) Solar source of the interplanetary sector structure. Solar Phys 1:437–445

Additional Web Sites

http://www.astronomynotes.com/
http://solarscience.msfc.nasa.gov/
http://www.windows2universe.org/
http://www.daviddarling.info/encyclopedia/ETEmain.html

http://www.springer.com/978-3-642-38429-5

Ionospheric Prediction and Forecasting

Zolesi, B.; Cander, L.R.

2014, XII, 240 p. 138 illus., 34 illus. in color., Hardcover
ISBN: 978-3-642-38429-5