Abstract

The Sun and Earth are intimately related. Few decades back, it was assumed that the relationship was only through the incidence of solar visible and infrared radiation on the surface of the Earth. However, it was soon realized that many powerful solar radiations reached the top of the terrestrial atmosphere but got absorbed in the upper part of the atmosphere, causing significant changes in the terrestrial environment. In this review, various processes are described, first on the Sun where various solar structures evolve, later in the interplanetary space due to escaping solar wind, and further in the interaction of the solar wind with the Earth's magnetic field, containing it in the magnetosphere and entry through the neutral point in the magnetotail. Resulting phenomena like auroras, ring current etc. are described. Present status of solar and interplanetary environments and their terrestrial effects is briefly outlined.

1. Introduction

Sun-Earth relation has been a fascinating topic ever since humanity habilitated the Earth. The role of solar energy in sustaining agricultural activities and the water resources through cloud cover changes was noted long ago and human beings are ever grateful to the Sun for his bounty. Since prehistoric times, many cultures have regarded the Sun as a deity. However, until recent decades, the contribution of Sun was assumed to be only in heat and light, which everybody could feel easily. That the Sun might be emitting something more was suspected when people noticed that on hill stations, there were more sunburns. Soon, balloons with instruments were released to sound the upper atmosphere and, lo and behold, a plethora of other radiations were noticed in the solar inputs. Most of these, though in percentages smaller than the visible radiation, were highly energetic (ultraviolet UV, extreme ultraviolet EUV and even solar X-rays and gamma rays) and could cause substantial changes in the atmospheric structure. Air molecules were broken up into atoms,
and further, atoms were broken up into ions, so that the upper layers of the terrestrial atmosphere (F region) were almost completely ionized. At lower altitudes, the solar UV intensities decreased and so did the ionization levels so that in the E layer, there were enough neutrals to collide with ions and reduce their mobility. Since electrons were still free to move, electric currents could exist in the E layer and could cause external changes in the geomagnetic field, a bulk of which is otherwise of internal origin with no short-term changes. In what follows, the solar phenomena and their terrestrial effects as they evolved in time, are described (for more details, see reviews Kane, 1976, 1999).

2. The Sun

The Indian astronomer Aryabhatta (born in 476 A.D.) studied astronomy at the University of Nalanda (Clark, 1930; Kay, 1981; Sen and Shukla, 1985) and made significant contributions to the field of astronomy. He propounded the Heliocentric theory of gravitation, thus predating Copernicus by almost one thousand years. Aryabhatta's Magnum Opus, the *Aryabhattiya* was translated into Latin in the 13th century. Through this translation, European mathematicians got to know methods for calculating the areas of triangles, volumes of spheres as well as square and cube root. The lack of a telescope hindered further advancement of ancient Indian astronomy. Aryabhatta was the first one to have propounded the theory that the earth was a sphere, in the 5th century. Another Indian astronomer, Brahmagupta, estimated in the 7th century that the circumference of the earth was 5000 yojanas. A yojana is around 7.2 kms. Calculating on this basis, the estimate of 36,000 kms as the earth's circumference comes quite close to the actual circumference known today (40,000 kms). Old Sanskrit verse of that epoch say, "there are suns in all directions", and "the night sky is full of suns", indicating that in ancient times Indian astronomers had arrived at the important discovery that the stars visible at night are similar to the Sun visible during day time. In other words, it was recognized that the sun is also a star, looking big because of its nearness. This understanding is demonstrated in another verse which says, "when one sun sinks below the horizon, a thousand suns take its place". This apart, many Indian astronomers had formulated ideas about gravity and gravitation. Brahmagupta, in the 7th century had said about gravity that "Bodies fall towards the earth as it is in the nature of the earth to attract bodies, just as it is in the nature of water to
flow”. About a hundred years before Brahmagupta, another astronomer, Varahamihira had claimed for the first time perhaps that there should be a force which might be keeping bodies stuck to the earth, and also keeping heavenly bodies in their determined places. Thus the concept of the existence of some tractive force that governs the falling of objects to the earth and their remaining stationary after having once fallen; as also determining the positions which heavenly bodies occupy, was recognized.

In the west, Greek astronomers and philosophers (450-350 B.C.) recognized Sun’s sphericity and immensity and explained eclipses. Aristotle (384-322 B.C.) portrayed Sun as a distant sphere but the Earth was considered at the center of the universe. Aristarchus (270s B.C.) was the first one to propose a heliocentric system, the Earth (not the heavens) rotating daily and circling the Sun, but his ideas were discarded by succeeding astronomers. Within the Earth-centered framework, Apollonius (220s B.C.) proposed eccentric circles and deferent circles with epicycles, while Hipparchus (140s B.C.) used these for modeling the motions of the Sun and the Moon to predict whether solar eclipses would be total, partial, or unobservable at a given location. Few centuries later, Ptolemy (140s A.D.) synthesized earlier work and estimated the distances and sizes of the Sun and the Moon (crudely), all in the Earth-centered framework, and his ideas prevailed with minor variations for several centuries till finally, Copernicus (1473-1543) formulated the concept of sun-centered planetary system. During the seventeenth and eighteenth centuries, natural philosophers and astronomers recognized that the Sun was the nearest star and estimated its distance, size, mass, rate of rotation, and direction of motion through space within ~10 percent of today’s values (Zirin, 1988; Hufbauer, 1991, main sources of the material presented in the present brief review). Also, over more than a millennium before the seventeenth century, observers had noticed sunspots. However, detailed information came only after the invention of the telescope. Some astronomers interpreted sunspots as small planets passing in front of the solar disk. Galileo pleaded that the spots were features of a rotating spherical Sun but was puzzled that they appeared only within 30° of the solar equator. Soon, the Cartesian cosmology was invoked where an infinitude of solar systems existed, scattered through unbounded space. Further details about the Sun were provided, besides many others, by Newton (1642-1727, Sun’s mass and density) and Hershel (1738-1822, solar infrared radiation), though some of their ideas were speculative and erroneous,
for example that sunspots gave glimpses through the Sun’s luminous atmosphere to its
dark, solid body.

In the nineteenth century, with better instrumentation and eclipse observations, striking
features of the Sun (‘corona’, ‘prominences’, the colourful region ‘chromosphere’ between
corona and the ‘photosphere’) were observed, and the study of the Sun’s structure and
behavior was named as “solar physics”. Simultaneously, sustained sunspot monitoring by
Schwabe (1843) revealed a ~10-year cycle in sunspot numbers, which Wolf (1876) later
corrected to an 11-year cycle. Sabine (1852) revealed that the sunspot maxima and minima
coincided in time with maxima and minima of geomagnetic variations, establishing an
important Sun-Earth link. Carrington (1858) reported that after a minimum, sunspots
appeared on both sides of the solar equator in zones between 20° and 40° latitude, and as the
cycle progressed, the spot zones contracted towards the equator, eventually disappearing
there at the next minimum. Also, spots near the equator traversed the solar disk more
rapidly than those toward either pole (differential rotation, Carrington 1859a), and on one
occasion, short-lived intensely bright and white patches were seen above a sunspot group
(first observation of a ‘solar flare’, Carrington 1859b).

The nineteenth century and the early part of the twentieth century were intervals of the
consolidation of solar physics, first by coming out of the earlier domination of pure
astronomy, and later by introducing ideas of physics, notably spectroscopy and
thermodynamics, developing new instruments, establishing new observatories, and
arranging meetings and symposia where solar physicists could compare notes and develop
new ideas. By 1910, there were good grounds to consider that the Sun’s atmosphere
consisted of terrestrial elements heated to the gaseous state (Fraunhofer lines interpreted by
Kirchhoff, 1859), the photosphere had a temperature of about 6000 K and radiated about
4x10^{24} cal/sec, the angular velocity was greater at the solar equator than its poles, there was
an 11-year cycle in sunspot numbers, chromospheric activity and coronal shape varied
along with the sunspot cycle, and sunspots were the seat of strong magnetic fields. A
substantial contribution came from George Hale who developed spectroheliography, started
the Astrophysical Journal, established the Yerkes Observatory and Mount Wilson Solar
Observatory, organized the International Union for Cooperation in Solar Research, installed
the first tower telescope, and came up with convincing evidence of high magnetic fields in sunspots (detailed references in Hufbauer, 1991), and a 22-year cycle of solar magnetic field variations. Hale regarded the Sun as key to the study of stars.

In the next three decades (1910-1940), solar physicists from various observatories around the globe were keeping the Sun under constant surveillance, publishing the results in the *Quarterly Bulletin of Solar Activity*, and using the results to examine Sun’s influence on radio transmissions and geomagnetism. Also, new means of monitoring the Sun were developed such as, coronographs, solar cinematography, the monochromatic filter. Equally impressive was the contribution to solar physics of new theoretical tools and results based on recent physical research. The Sun’s internal constitution was examined using Eddington’s theory of radiative equilibrium, Russel’s work on the Sun’s composition, and Bethe’s identification of thermonuclear reactions. It was concluded that there was an overwhelming abundance of hydrogen in the Sun (and probably all stars). The internal temperature of the Sun was estimated to be about 15 million K, mainly caused by a chain of nuclear reactions (carbon cycle), which began with a proton tunneling into a carbon nucleus and culminated, after the tunneling of three more protons one by one into the nucleus, with the division of the resultant nucleus into a carbon nucleus and a helium nucleus. The energy came from the conversion of mass (m) into kinetic and radiant energy by the Einstein formulation \( E=mc^2 \). The outward flow of radiation created a pressure that counterbalanced the superincumbent matter’s immense weight. Only near and below the surface where the temperature and density were much lower, a ‘convection zone’ supplemented radiation as an important means by which Sun’s heat continued its outward journey. A major puzzle was that there were several coronal emission lines which could be attributed only to emissions from atoms stripped of many of their electrons (e.g., coronal green line 5303 angstroms arose from thirteen-times ionized iron atoms) and such large scale stripping needed coronal temperatures of about 2 million K. What was the origin of this abnormal coronal heating?

During World War II, observational and interpretive solar physics programs were disrupted, but some scientists were able to convince their governments that solar observations were useful for forecasting ionospheric shortwave transmission characteristics.
and had military value. Thus, some additional facilities for research could be established which came handy in the decade following the war. British and American scientists had detected solar radio emission (start of solar radio astronomy) and German V-2 rockets could send instruments above the atmosphere and observe solar extreme ultraviolet and X-ray radiations, which do not reach the Earth’s surface due to absorption in the atmosphere. A magnetograph was devised for studying the solar magnetic field outside sunspots. An International Geophysical Year (IGY, 1957-1958) was planned and more than 20000 scientists at more than 2000 locations obtained synoptic observations of dozens of solar and terrestrial phenomena. Following the work of Tousey and his group (details in Tousey 1986) in obtaining spectrograms in rocket flights, Friedman (1981) developed electronic methods of detection, which were sensitive and rapid and the measurements could be relayed to the ground during flight via the rocket’s telemetry system. His study of solar flares by means of rockets conclusively demonstrated that solar X-ray bursts caused shortwave fadeouts. Also, since the nuclear reactions powering the Sun involve emissions of neutrinos, some scientists thought of detecting these particles (which were theoretically expected to be traveling outward from the Sun’s core at the speed of light) by their ability to transmute the nuclei of chlorine 37 into argon 37. In the Homestake Mine of South Dakota, Davis put his detectors in 1967, but the rate turned out to be very low, far less than theoretically expected, perhaps requiring a major revision in stellar or neutrino theory (Bahcall, 1969).

The most important contribution of the IGY to solar physics was an indirect one, the launching of the soviet satellite Sputnik on October 4, 1957. Partly political, the launching of Sputnik stirred up considerable scientific activity in 1957-1958, with scientific instruments placed in orbit by the Soviet Union on Sputnik 1, 2, and 3 and by USA on Explorer 1, 3, and 4, Vanguard 1, and Pioneer 1. US government made operational in October 1958 a civil agency named National Aeronautics and Space Administration (NASA), which announced soon that it was ready to help other nations put instruments, or even complete scientific satellites, into orbit. This encouraged space research in many countries and by 1975, scientists in more than twenty nations had joined United States and Soviet Union in studying natural phenomena with the aid of spacecraft. Meanwhile, American and Soviet programs were getting more and more sophisticated, with larger
spacecrafts going on longer journeys and communicating more rapidly and reliably with ground stations, besides promoting greater support to scientists using traditional approaches and cultivating closer ties with them. For observations of solar electromagnetic radiations, many satellites have been launched (OSO 1, 3, 4, 6; SKYLAB; AEROS A, B; AE-C,E; SOLRAD 11; PROGNOZ 7-10; SMM; San Marco 5; PHOBOS 1,2; YOHKOH; CORONAS 1, INTERBALL 1; ELECTRO; SOHO; TIMED; ISS; GOES, details in Tobiska et al., 2000) and many of the ground observatories had improved instrumentation (e.g., Kitt Peaks Mcmath Solar telescope). Tobiska et al. (2000 and updates) have produced an empirical solar irradiance model, which can be used by workers as input for studying terrestrial effects. A journal Solar Physics established itself as a thriving international forum for the solar physics community. In recent years, funds have also gone for nonsolar astrophysical missions such as International Ultraviolet Explorer, Einstein High Energy Astronomical Observatory, Infrared Astronomical Satellite, Hubble Space Telescope and Chandra. The Challenger disaster of January 1986 proved a big setback for spacecraft launching programs as well as ground-based solar observing programs. However, a part of this loss has been compensated by improving the precision, versatility and reliability of the payloads and acquiring a new generation of high-resolution solar telescopes and upgrading auxiliary instrumentation as new technologies become available.

An important development since World War II was solar radio astronomy, mainly due to the lead of a group in Sydney, Australia, where Pawsey (1946) and his team showed that particular spot groups were associated with enhanced solar radio emission, but there was a steady background component also which indicated a coronal temperature of 600,000-1,200,000 K. Since then, many radiospectrographs have been operative in different parts of the world, supplying very useful information for studies of solar phenomena like solar flares. Synoptic radio observations of the Sun started in Canada in November, 1946, when Covington used a 4-ft reflector from a Type IIIC Gun Laying radar system to start recording the solar flux at 3-cm wavelength. In 1947, Covington's landmark measurements developed into a regular observing program at 10.7 cm. The daily flux was measured in Ottawa at Algonquin Radio Observatory until June 1991, when the program was moved to Penticton at Dominion Astrophysical Observatory, three time zones west, where it continues today. There has been some controversy about whether the microwave flux from
active regions is due to thermal bremsstrahlung or gyro-resonance emission, which would then indicate whether the source of emission was the optically-thin corona or optically-thick sunspots. Spatially-resolved spectra in the cm-lambda range indicate that gyro-resonance emission (with peaked spectra) usually dominates the flux from active regions, while thermal bremsstrahlung emission (flat or rising spectra) comes from plage regions displaced from spots.

3. Solar wind

Even before 1950s, many scientists suspected that the Sun might be sending corpusscular matter into space with speeds much lesser than that of light (Kipenheuer, 1953). Strong aurorae and some geomagnetic disturbances seem to recur in 27 days (solar rotation period) and were stronger at the terrestrial poles, indicating some sort of chanelling of charged particles by magnetic field. Even in quiet periods, there seemed to be some solar inputs. Biermann (1951, and later papers) precipitated this thinking by suggesting that the comet tails always pointed away from the Sun because of impinging solar corpuscular material (100-1000 ions and electrons at speeds of 500-1000 km/sec) all the time, and solar radiation pressure was grossly inadequate to produce this effect. Not many people agreed with Biermann. The famous solar-terrestrial physicist Sydney Chapman (1957) claimed that the Sun had a static atmosphere (just like the Earth) but so large that its corona could engulf and affect the Earth. However, Parker (1958, 1959, and later papers) at the University of Chicago examined this issue and found that the Sun was not only not static but highly dynamic and could be emitting solar efflux of the same order as indicated by Biermann. Parker termed the efflux as “solar wind”, attributed it to the expansion resulting from coronal temperatures of ~2 million K over an extended region around the Sun, and expected that the expanding gas would draw magnetic field lines out of the corona far into
the solar system and, because of the solar rotation, the resulting interplanetary field would have a spiral pattern in the Sun’s equatorial plane. As a mechanism for the high temperatures of the corona, Parker suggested that hydromagnetic (Alfven) waves propagating upward from the photosphere dissipated their energy in the coronal plasma by magnetically accelerating the fastest protons encountered there to still higher velocities.

Parker’s theory got a very lukewarm reception. Chapman (1959) was not certain that ejection was continual over the whole Sun. Chamberlain (1960) criticized Parker’s idea as arbitrary and proposed an alternative model which needed a much lower velocity for the expanding plasma - a solar breeze. Only Thomas Gold (1959) seems to have appreciated the theory and made his own conjectures about the behavior of plasma and magnetic fields in the solar system. Parker realized that a confirmation of his theory could come only through experimental observations by satellites. The first such evidence of a solar wind came from the Russian group of Gringauz et al. (1960) who reported observations from Lunik 2 (launched in September 1959), indicating a flux of high-speed ions of \( \sim 2 \times 10^8 \) ions/cm\(^2\)/s, but their speed and direction was not known. This was followed by the American experiment on Explorer 10 (launched in March 1961) which rose above the Earth’s night side in a highly eccentric orbit with an apogee of 240,000 km and probably never reached the undisturbed interplanetary medium, but reported a flux of \( \sim 4 \times 10^8 \) ions/cm\(^2\)/s (double that of Lunik 2) with speeds of 120-660 km/s (generally away from the Sun), yielding plasma densities of \( \sim 6-20 \) protons/cm\(^3\) (Bridge et al., 1962). More convincing results were obtained by Mariner 2 (launched in August 1962), where 104 days’ worth data indicated a continuous plasma flow but with peaks of activity and quiet periods. The velocities ranged between 400 and 700 km/sec but occasionally exceeded 1250 km/sec (Neugebauer and Snyder, 1962), always away from the Sun, and the spectra showed two maxima indicating the presence of helium nuclei as well as protons. Soon after, Snyder et al. (1963) reported that the solar wind contained high-speed corpuscular streams with velocities of 600-700 km/sec, recurrent at a 27-day (solar rotation) period and correlated with the recurrent peaks in geomagnetic activity, yielding a linear relationship between solar wind velocity and geomagnetic index (clear Sun-Earth link). These streams did not seem to arise in a hydrodynamic expansion of a homogeneous solar corona but came instead from long-lived local regions in the corona which were abnormal in some respect.
In subsequent years, several satellites have been used to study the interplanetary plasma parameters of solar wind.

An important discovery during this period was that of a magnetic sector structure in the solar wind, where the interplanetary space in the equatorial plane around the Sun seemed to be divided into alternate sectors of magnetic field directions ‘away from the Sun’ and ‘towards the Sun’ (Ness and Wilcox, 1965). Wilcox (1968) suggested that the distribution of large, long-lived unipolar regions in the equatorial latitudes in the photosphere mapped out in the interplanetary field as sectors. His student Schatten developed a ‘source surface model’ wherein the photosphere’s complex field was smoothed out as a sectorlike pattern at a thin region (source surface) in the corona about 0.6 solar radii above the photosphere, and the pattern was carried out in interplanetary space by the solar wind (Schatten et al., 1969). However, the sector structure has a more plausible explanation in terms of a ‘two hemisphere model’. The magnetic fields which originate in the northern hemisphere of the Sun point in one direction (inward or outward) while fields originating in the southern hemisphere point in the opposite direction. The boundary between the two magnetic hemispheres consists of a thin neutral sheet, in which the magnetic directions are not consistent. The neutral sheet is slightly warped, so that it does not lie quite flat in the plane of the Earth's orbit. As the Sun rotates, the sheet also turns, so that the Earth is alternately on one side of the warped region and the other. As this happens, satellites near the Earth observe the change in the direction of the interplanetary magnetic field as the sector boundaries pass the Earth.

Several decades ago, Bartels (1932) noted the 27-day recurrence tendency in geomagnetic storms but did not find any associated striking features on the Sun, and hypothesized invisible M regions as the possible sources. Ness and Wilcox (1965) identified the M regions to unipolar magnetic regions. The recurring geomagnetic storms are found to coincide with streams that are much faster than the normal solar wind. By comparing the arrival times of these high-velocity streams with pictures of the Sun's corona taken by Skylab X-ray telescopes on known dates, the high-speed streams were traced to parts of the corona which emit no X-rays, the so-called coronal holes. The temperatures and densities of coronal holes are much lower than those of other parts of the corona. Investigations
show that in the holes, the magnetic field has no loops, but extends directly out into the solar wind. We do not yet know how and why coronal holes form, but we do know that they are a major source of the solar wind. Two apparently permanent coronal holes exist at the north and south poles of the Sun, and it may be that much of the solar wind that leaves the Sun originates in these polar coronal holes. Solar wind velocity variation is not completely parallel to that of sunspots, neither is the variation of geomagnetic activity.

4. Magnetosphere and geomagnetic storms

Geomagnetic field is basically dipolar and should fall off with distance \( r \) as \( r^{-3} \) in the equatorial region. However, the solar wind exerts a pressure and the field is compressed on the sunward side. Early satellite measurements showed that the geomagnetic field was confined to what is known as the ‘magnetosphere’, snub-nosed like a bullet on the sunward side up to about 10 earth radii, and stretched far back to several tens of earth radii in the magnetotail, in quiet time solar wind. When solar flares occur, shocks are produced which propagate in interplanetary space with high solar wind pressures. If the Earth encounters these, the sunward boundary of the magnetosphere may be compressed to even up to 7 earth radii, but solar wind cannot penetrate the magnetosphere easily and is mostly diverted to the tail side. It was noticed that only on certain occasions, the solar wind penetrated the magnetosphere from the tail side and the necessary condition seemed to be a negative Bz component of the magnetic field in the shocks. The reason for this remained a mystery till Dungey (1961) gave an explanation. As the geomagnetic dipole field is stretched in the magnetotail, a neutral sheet is formed, with geomagnetic field away from the Earth above the neutral sheet and toward the Earth below the neutral sheet. At the end, in a small region far away from the Earth, the field is still north-south. If the field in the interplanetary shock has a component (negative Bz) which can neutralize the geomagnetic field, a neutral point is formed and solar wind gets an entry into the magnetosphere. Low energy particles spiral around the stretched geomagnetic field lines and impinge on the terrestrial atmosphere in the polar regions, causing enhanced aurora. Higher energy particles rush towards the Earth but are diverted around the Earth in circular orbits in the equatorial plane and cause large geomagnetic field reductions which recoup slowly when the Earth comes out of the shock region and solar wind input stops. Thus, for geomagnetic storms to occur, two conditions
are necessary, firstly, the Earth should enter a shock region, and secondly, the shock should have a magnetic field component (negative Bz), which can neutralize geomagnetic field in a small region in the magnetotail and create a neutral point which will facilitate solar wind entry into the magnetosphere. If the shock is not produced by a solar flare but is produced by a (fast) stream- (slow) stream interaction, same conditions are still applicable. If the Earth encounters a shock, the first effect would be a compression of the geomagnetic field (an increase called SSC, storm sudden commencement), and if there is a negative Bz in the shock, solar wind would enter the magnetosphere, auroras would be caused by low energy particles, a ring current would be formed in the equatorial plane by high energy particles, and there will a sharp decrease in the geomagnetic field lasting for a few hours (main phase) followed by a slow recovery when the ring current particles disappear by collision etc. in due course.

5. Ionospheric effects

When solar flares occur, ionospheric number densities may increase causing SFE (solar flare effects) but these are short-lived. Major effects occur when low energy particles precipitate in the auroral regions, an ‘auroral electrojet’ is formed, joule heating occurs, and ionization travels towards lower latitudes. This occurs preferentially along the geomagnetic field lines, which are not parallel to the ground but are rising, and firstly, the ionospheric heights increase, and secondly, the particles enter in regions of different, altitude-dependent, loss processes. Thus, ionospheric storm effects can be very complicated, depending considerably upon the local time when the geomagnetic storm commenced (e.g., Kane 1973 and many other later papers). Hence, predictions can be hazardous. However, for long-term changes, both foF2 and thermospheric temperatures increase parallel with the sunspot activity.

On short-term scales, effects of solar flares on magnetospheric radiation belt particles at 1000-6000 km can bring down a torrent of particles affecting the ionosphere and brightening auroras. High energy solar protons from CME can enter directly into the Earth’s atmosphere, warm the outer layers of the polar atmosphere (above 50 km) by
several degrees, and create NOx compounds which can deplete ozone and cause lesser UV absorption and hence, cooling of the atmosphere.

6. **Mesospheric and stratospheric effects.**

For the mesospheric region, Clemesha et al. (1997) reported long-term and solar cycle changes in the atmospheric sodium layer, while Jacobi (1998) reported on the solar cycle dependence of winds and planetary waves in the mesopause region. In general, the connection in the mesopause region is weaker. Short-term effects are small or erratic. In the stratosphere, there is a naturally formed ozone layer and it has a small (a few percent) solar cycle effect, which may get reflected in the filtered ultraviolet in certain wavelength bands. Of particular interest is the effect on UVB, which is harmful to human skin. The changes in UVB due to solar cycle changes of ozone are rather small. A greater hazard is due to the depletion of ozone by man-made chlorofluorocarbons, as this may increase the UVB doses considerably and cause skin cancers.

7. **Climatic changes**

The study of the effects of solar radiation on terrestrial climate has been very copious and has a long history (Pittock, 1978). However, the conclusions have been mostly uncertain and sometimes confusing. Attention has been paid to (a) Radiative forcing (Lean and Rind, 1999 and references therein), (b) Magnetospheric REP events causing ozone depletion (Lastovicka, 1991) and abrupt changes in atmospheric circulation (Bucha and Bucha Jr., 1998), (c) cosmic rays (controlled by solar magnetic field extension in the heliosphere and showing an 11-year cycle) affecting global cloud coverage (Tinsley, 2000), (d) Ionospheric ground electrical circuit variability by controlling cloud micro-physics (Baker, 1986), and through their strong connection with sudden commencement storms (Bochnicek et al, 1999). Recently, one more mechanism has been suggested, namely, when solar wind energy is deposited in the auroral electrojet during storms, atmospheric gravity waves are generated. If these are transmitted downward and get amplified by wind shears or seeding instabilities that generate gravity waves in the mid-latitude troposphere, cloud formation may occur and cause weather changes (Prikryl et al., 2003).
Solar contribution to climate is small and on short-term time scales, there are major earthly effects like those of greenhouse gases, volcanoes, sulfate aerosols, El Niños and probably many others (unrelated to solar activity) which are overwhelmingly larger than the solar effects, which can therefore be detected (if at all) by sophisticated statistical analyses. In rainfall series, an 11-year signal is often found for some locations, but these are not stationary and are not phase locked with sunspot activity (Lean and Rind, 1999). However, Reddy (2001) reported a 11-year cycle in the equatorial lower stratosphere, Alaskan climate, Indian summer monsoon, and Reddy and Karim (2003) presented evidence showing a modification of the solar cycle effect by phases (easterly or westerly) of the stratospheric wind QBO (see also Labitzke and van Loon, 1990).

Total Solar Irradiance (TSI) has been measured accurately during the last two decades and shows a small (~0.1%) variation over the sunspot cycle, and there is a great controversy whether such a small change can cause significant climatic changes. Whereas direct effects of visible solar radiation may be negligible, indirect effects through cosmic rays etc. which have a large solar cycle variation may be substantial. Also, the solar UV flux has a considerable solar cycle fluctuation and through photochemistry, may influence stratospheric ozone and therefore, stratospheric temperature. On a long-term time scale, the global warming seems to have increased from the late nineteenth century to around 1940, decreased up to the mid-1960s, and increased substantially thereafter. Lean and Rind (1999) have looked carefully at the historical record of the sun's varying activity levels, including direct observations of solar radiation over the last 20 years and indirect evidence of solar activity implied through the study of ice cores and tree rings (Eddy, 1976). Lean and collaborator David Rind made simulations with computer models of climate change in response to changes in solar radiation during the past 400 years. They then used the model results to compare with both pre-industrial and current climate change trends to determine the role of the sun in the heating on the Earth. The general conclusion of their study is that the sun may have played a dominant role in pre-industrial climate change (from 1600 to 1800, for example) but it has not played a significant part in long-term climate change during the past few decades. It is furthermore unlikely that the sun accounted for more than half, at most, of climate change from 1900 to 1970. Stott et al.
(2000) developed a computer model which indicated that whereas anthropogenic emissions alone could explain the rapid rise in temperatures in the past 30 years and that solar variation alone could have caused the warming observed during 1910-1940, a model including both these causes could explain only up to 60% of the variations of the entire century. Lawrence et al. (2000) have developed an “extremely crude model” of three interrelated equations to stimulate the flow patterns in the atmosphere at middle latitudes. It calculates the average speed of the westerly flow of winds as a function of latitude. The model exhibits chaotic behavior, where even small changes in the inputs to complex systems can cause large changes in the answers. The model's calculations produce correlations that appear during the early phases of a simulation, disappear later in the simulation, and then reappear as anticorrelations. This matches the past behavior of the solar cycle. Between 1860 and 1920, cooler temperatures occurred when sunspot numbers were large. From the 1920s to the 1960s, there was no clear correlation between sunspot numbers and temperature. But after 1960, increased sunspots correlated with higher temperatures. Finally, Lawrence's model allows the Northern and Southern Hemispheres to fluctuate independently of one another and to have different correlations to the solar variation, as has been observed. Incidentally, the comparisons made by Duhau (2003) show that the observed temperature decrease during 1920-1960 (when sunspot activity was still rising) can be reproduced if a correlation analysis includes a geomagnetic SSC index (product of the magnitude and the duration time of a storm sudden commencement, averaged over an year). Georgieva et al. (2003 and references therein) have noted that the correlation between the Earth’s surface temperature and sunspot activity in the 11-year solar cycle depends on the period studied and changes sign in consecutive secular Gleissberg cycles (~80 years), and this relationship depends upon the solar activity asymmetry, positive when the northern solar hemisphere is predominantly more active, and negative when the southern solar hemisphere is predominantly more active. The two hemispheres rotate differently and the interplanetary magnetic field at the Earth’s orbit is related to the differential rotation of the more active hemisphere. Also, the two hemispheres have different magnetic helicities, which are carried to the Earth by magnetic clouds preserving the helicity of the source region of their origin. The reaction of the terrestrial atmosphere to the arrival of the magnetic clouds depends on the helicity of the clouds, in addition to a stratospheric QBO phase effect. Incidentally, the N-S asymmetry
of solar activity seems to have a QBO of its own (Badalyan et al., 2003).

In short, Sun-climate relationship is very complex by itself and meteorological changes by other effects such as of greenhouse gases, volcanoes, sulfate aerosols, El Niños and probably many others, which are overwhelming larger than the solar effects, can complicate matters still further. That is why no single effect is seen invariably and consistently and for the same reason, accurate predictions are not possible. Incidentally, some terrestrial phenomena apparently unrelated to solar activity may not be unrelated completely. Volcano activity displays no 11-year periodicity, but 21-year running averages seem to indicate that volcanic activity is generally lower in periods of prolonged maxima of solar activity (Strestik, 2003), and their spectra show similar periodicities (200-215 yr, 100-105 yr, 80-90 yr). If true, a connecting mechanism needs to be discovered. Similarly, a possible connection between El Niño events and solar activity reported by Landscheidt (2000) needs scrutiny.

8. Present status

8.1 Solar physics

The origin of all solar activity is in the convective zone (immediately below the photosphere), and helioseismology has provided considerable information about the interior structures and dynamics, from the global dynamo to small-scale flow associated with solar flares (Kosovichev, 2003). Solar oscillations have typical periods of 3-10 minutes with maximum power at about 5 minutes. These are excited near the surface. The f-modes (surface gravity waves) propagate in a thin layer just beneath the solar surface and are useful for measuring the solar seismic radius, while the p-modes (acoustic waves) propagate in the deep layers of the Sun and are finally reflected back. Two regions, the tachocline and the upper convective boundary layer are critical for understanding solar variability. Both regions have strong rotational shears and provide evidence for a 1.3-year periodicity but no indication of a 11-year periodicity. Sunspots as cool objects appear to be only 4-5 Mm (millionmeters) deep, but accumulate
significant heat in the deeper layers and form converging downfalls. Sunspots have a

tree-like magnetic structure. (Mechanisms of sunspot formation and stability are not yet understood). Large active regions are formed as a result of multiple flux emergence.

Flow maps show that in the subphotosphere, there are divergent supergranular flows
and strong converging flows in magnetic regions, but there is a remarkable multiple-

scale reorganization on the larger and global scales. There are zonal flows which
migrate to the equator (reason not yet fully understood) and meridional flows
(presently) from the equator to the poles. Bumba (2003) mentions that besides the 11-
year and 22-year cycles in solar activity, there exist several modes of cyclic variations
of lesser periodicities (QBO 2-3 years, 1.3 years, 150-160 days) and all of them seem
to be related to the regularities in the appearance and distribution of the magnetic flux
in the photosphere by its more or less spatial grouping through the local magnetic fields
in active longitudes, and temporal grouping in the formation and development of
complexes of activity. However, there is also a suggestion that these may be harmonics
or subharmonics of a certain basic period.

Another interesting observation relates to the multiple peaks (mostly two) in solar
activity at sunspot maximum. In cycle 23, there were two distinct peaks, one near July
2000 and another near February 2002 (separation about 20 months). The relative
magnitude of the first peak with respect to the second peak are different for different
solar indices. For sunspots, the second peak was lower than the first peak by ~4%, but
the second peak was higher than the first peak by ~3.5% for solar EUV (26-34 nm) and
higher by ~10% for 2800 MHz radio flux and for Lyman alpha (Kane, 2003). These
differences need explanation.

CMEs and solar flares are two important phenomena responsible for solar emissions
entering the interplanetary space. During the last few years, copious observations of
CMEs were possible. Using the SOHO/LASCO coronograph, Gopalswamy et al.
(2003) reported the results of a study of nearly 7000 CMEs which occurred during
1996-2002. The peaks of CMEs and sunspot number were almost two years apart and
the CME mean speeds doubled from sunspot minimum to maximum. High latitude
CMEs were intimately related to the solar polarity reversal during solar maximum,
and polarity reversal seemed to be an energetic process involving the release of large
amounts of energy. Both sunspot activity and high latitude CME activity were high at sunspot maximum. Maricic et al. (2003) have studied the initiation and development of two CMEs and both show clearly a three-part structure already at low heights during the initial gradual rise in the pre-eruptive phase. Many other details are given.

A major lacuna in solar physics is the failure of solar flare theories to account for the fact that the total power and the number of particles required to explain the emissions cannot be supplied by the active region. Simnett (2003) invokes a global view where an erupting magnetic structure plays the central role and the active region plays a minor role. Magnetic reconnection in the high corona gradually pumps up the erupting structure with mildly energetic particles, mainly protons. Finally the stability of the structure is destroyed, and it erupts, dumping the particles into the evolving active region, where they are reaccelerated to produce the high energy flare protons and other emissions. The energy and matter supplied by the erupting structure is sufficient to overcome the active region deficit.

8.2 Interplanetary disturbances and their magnetospheric response

Only 1-2% of the nearly 7000 CMEs studied by Gopalswamy et al. (2003) during 1996-2002 were geoeffective. Those resulting in SEP (Solar Energetic Particle) events need to drive a shock that accelerate particles, and hence, need to be fast and wide. The storm-causing CMEs need to be directed towards the Earth and must contain a southward component Bz magnetic field. Halo CMAs (those which appear to surround the occulting disc of a coronograph) and fast and wide CEMs are important from the point of view of space weather. Presently, considerable effort is put in studying the relationship between parameters of geomagnetic storms, e.g., Dst magnitudes, and parameters of halo CMEs, notably magnetic cloud speeds (e.g., Gonzalez et al., 2004). The principal interplanetary parameters controlling the magnetospheric response are the solar wind ram pressure and the IMF magnitude and direction. Feldstein et al. (2003) examined a two-stream solar wind interval (two interplanetary CME events) during May 1-7, 1998, modeled the magnetospheric response to these
events, and compared with satellite data. For the intense storm of May 4, they estimated the disturbance fields as $-208$ nT for the ring current contribution DR, $112$ nT for the Chapman-Ferraro magnetopause current system DCF, and $-161$ nT for the tail current system DT. They pointed out that these currents significantly modify the magnetospheric geometry and size and must be included for any accurate magnetic field representation during storm periods. An interesting test of large changes in magnetospheric geometry came when during May 10-12, 1999, solar wind almost disappeared. While the wind velocity maintained its normal value of $\sim 360$ km/s and the magnetic field was $\sim 6$ nT, the number density dropped below $1/cm^3$ and the dynamic pressure dropped below $0.1$ n Pa. The bow shock along the Earth-Sun line, normally at a distance of $\sim 10$ RE, reached a maximum value of $53$ RE (Youssef et al., 2003 and references therein).

8.3 Climate

Whereas as several mechanisms are suggested for solar effects on climate, all these seem to be mostly of academic value, as few indicate substantial effects like those in Labitzke and van Loon (1997) and Labitzke (2001). Ramaswamy et al. (2001) reviewed the stratospheric temperature data from various sources and found that the stratosphere has, in general, undergone considerable cooling over the past 3 decades, and the major radiative factor responsible for this is the depletion of stratospheric ozone, though some contribution from the increases of greenhouse gases is also expected. Superposed on this trend is a solar cycle variation of about 1K during a solar cycle. At the SORCE Science Meeting at Sonoma, Sultan Hameed (2003) suggested the following possibility. “Variations in global heating rates and circulation cause changes in the intensities and the morphologies of the atmospheric centers of action (such as the Aleutian Low and the Hawaiian High). In turn, these systems influence atmospheric and oceanic circulations over their respective domains. The centers of action therefore may be considered to act as bridges between variations on the global and regional scales, and could provide a viable link between small magnitude solar activity changes and large changes in local climate. It is known that solar activity induces changes in UV radiation and stratospheric ozone. The primary response of the atmosphere to this
direct forcing is in the zonal circulation in the stratosphere. This circulation change induces changes in the centers of action as stationary wave nodes. These changes may be small. However, the changes induced by a center of action in regional circulation and clouds feedback to the center of action. As a result, significant changes in regional climate are observed associated with the solar cycle”. In another presentation at Sonoma, David Rind (2003) said “We used various climate change experiments simulations from the GISS global climate/middle atmosphere model to investigate the impact stratospheric perturbations have on the troposphere with emphasis on solar forcing. Atmospheric radiation, advection, stability influences and wave-mean flow interactions allow the stratospheric changes to be felt at lower levels. Changes in stratospheric zonal winds can affect planetary wave propagation extending down into the troposphere, and hence the phase of the Arctic/North Atlantic oscillations. Changes in stratospheric temperatures can also affect temperatures in the upper troposphere, with a corresponding influence on tropospheric eddy energy generation and Hadley Cell intensity. Stratospheric radiative perturbations in general have a smaller impact on surface temperature than those of well-mixed or direct surface forcing, due to the cloud cover response. The magnitude of the tropospheric response is generally on the order of 0-10% of control run values, while some local/extreme effects can be higher”. Thus, various possibilities are under study.

For short-term time scales (hours to days), effects like those reported by Svensmark and Friis-Christensen (1997) relating variation of cosmic ray flux and global cloud coverage, are certainly detectable, but on long-term time scale, effects become obscure. On very long-term time scale, some relationships seem to be partly valid (Eddy, 1976; Lean and Rind, 1999). A major complication is because of effects of nonsolar natural phenomena like greenhouse effects, El Niños, volcano activity etc., which are often overwhelmingly large. Accurate measurements will probably establish solar effects beyond doubt, but the recent global temperature increases indicate that the effects of other nonsolar sources are on the increase and detecting solar effects will be increasingly difficult.

9. Conclusions
Solar effects are felt very strongly in the upper atmosphere but are reduced at lower altitudes. In particular, climatic changes in the troposphere are overwhelmingly due to nonsolar phenomena like greenhouse effects, volcanic eruptions, El Niños and solar signals, if any, are obscure in recent decades.

Acknowledgements

This work was partially supported by Fundação Nacional de Desenvolvimento Científico e Tecnológico (FNDCT), Brazil, under contract FINEP-537/CT.

References


Pawsey, J. L., 1946, Observations of million degree thermal radiation from the Sun at a wavelength of 1.5 meters, Nature, 158, 633-634.


Sabine, E. 1852, On periodical laws discoverable in the mean effects of the larger magnetic disturbances – No. 2, Phil. Trans. 142: 103-124.


Captions for Figures

Fig. 1: Solar spectrum above the terrestrial atmosphere.

Fig. 2: Major processes on the Sun during a typical solar flare (Piddington, 1969).
Fig. 3: Spiral and sector structure of the interplanetary magnetic field (Wilcox and Ness, 1965).
Fig. 4: Synoptic views of (a) flare-associated ‘driven’ shock, and (b) stream interface (Hundhausen, 1972).
Fig. 5: Magnetosphere.