We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

6,600
Open access books available

177,000
International authors and editors

195M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Chapter

Coupling between Geomagnetic Field and Earth’s Climate System

Natalya Kilifarska, Volodymyr Bakhmutov and Galyna Melnyk

Abstract

The idea about synchronized variations of geomagnetic field and climate appears in the middle of the twentieth century. Among others, one of the main reasons for its unpopularity is the missing mechanism of coupling between magnetic and non-magnetic media. This chapter offers such a mechanism, consisting of a chain of relations transmitting the geomagnetic spatial-temporal variations down to the planetary surface. The first element of this chain is energetic particles propagating in Earth’s atmosphere, whose density and depth of penetration are modulated by geomagnetic field. Thus, the non-dipolar geomagnetic irregularities are projected on the ionization layer in the lower atmosphere (known as Regener-Pfotzer maximum). This unevenly distributed ionization, in certain conditions (i.e. dry atmosphere), acts as a secondary source of ozone near the tropopause. Ozone at this level is of special importance due to its influence on the tropopause temperature and humidity, and consequently on the planetary radiation balance. Hence, the geomagnetic spatial and temporal variations are imprinted down to the surface, impacting the climate system and its regional structures. The chapter provides synthesized information about geomagnetic field variability, particles’ propagation in Earth’s atmosphere, ion-molecular reactions initiating ozone formation in the lower stratosphere, as well as evidence for its covariance with some atmospheric variables.

Keywords: geomagnetic variations, geomagnetic focusing of charged particles, lower stratospheric ozone, regionality of climate changes

1. Introduction

The co-variability of paleomagnetic and paleoclimate time series has been found in many sedimentary records, e.g. [1] and references therein. Most of the reversals of geomagnetic field polarity and magnetic poles’ excursions seem to appear in periods of cold climate [1, 2]. Other authors, however, announced that climatic cooling fairly well corresponds to episodes with a stronger geomagnetic field [3–5]. This controversy, together with objective difficulties for disentangling paleomagnetic from paleoclimate data – due to the high variability and climate dependence of marine sedimentation rates – determines the skepticism of the greater part of the scientific community regarding possible links between geomagnetic field and climate.
On the other hand, time series based on contemporary instrumental measurements do not contain the ambiguity of paleo-data records. Based on the magnetic and climate measurements collected from the beginning of 1900 up to 2010, this chapter provides not only more evidence for existing coupling between geomagnetic field and climate system, but also offers a physically rational explanation and results supporting its validity.

2. Spatial-temporal variability of geomagnetic field at different time scales

The Earth’s magnetic field interacts with all planetary shells – the core, mantle, and crust of the solid Earth, as well as with the atmosphere, hydrosphere, and biosphere. It comprises information about both the state of near-earth space and the internal structure of our planet. The Earth’s magnetic field is continuously changing in space and time. The sources of its variations are located inside and outside of the planet. The amplitude and periodicity of geomagnetic variations are very different, which affects the methods used for data acquisition.

2.1 Long-term variations related to the heterogeneity at the core-mantle boundary

The longest periods of reoccurrence have geomagnetic reversals, followed by geomagnetic excursions. Geomagnetic reversals define the exchange of positions of the North and South magnetic poles. For the last million years, the geomagnetic field has changed its polarity four times. The last one happened about 780,000 years ago. During the inversion of geomagnetic polarity, the magnetic field’s strength drops dramatically, leading to a severe weakening of the planetary magnetic shielding, which protects living organisms from harmful cosmic radiation. This is the argument of some scientists to suggest that episodes of mass extinctions of terrestrial biota could be attributed to geomagnetic reversal [6, 7]. No systematic pattern was found in the occurrence of inversions and they are treated as a random process.

The palaeomagnetic records reveals also the existence of shorter periods (with a duration of several thousand years) when the field has departed from its near-axial configuration. Such short-term events are called geomagnetic excursions. The excursions are usually defined as a deviation of the virtual geomagnetic pole equatorward of 45° latitude, or as a short-term change in the direction of a geomagnetic field, whose amplitude is at least three times greater than the secular variations for a given period of time. Excursions are short-term impulse fluctuations, which are mostly replaced by smoother secular variations in geomagnetic field intensity. The nature of geomagnetic polarity reversals and excursions is not fully understood. Their general characteristics suggest that they could be considered manifestations of various processes within the Earth’s liquid core.

Secular variations are another long-term variability of geomagnetic field ranging from decades to several thousands of years. It is generally accepted that the geomagnetic secular variations are associated with changes at the core-mantle boundary. They are studied using all available methods – paleomagnetic, archeomagnetic, and direct observations. Over the period of instrumental observations (approximately 120 years), secular variations are grouped in 4 intervals: 60–70, 30–40, 18–25, 10–11 years. Variations with a period of about 60 years have the greatest amplitude.
In the first approximation, the magnetic field is interpolated as the field of a magnetic dipole. However, the empirical models (incorporating all available measurements of field intensity), as well as satellite measurements, reveal the existence of a non-dipolar component in the real geomagnetic field. The irregularities in the spatial distribution of geomagnetic field intensity are well visible in Figure 1, based on the 13th generation of the International Geomagnetic References Field model (IGRF) [8]. The two-wave distribution of field intensity in the Northern Hemisphere and a single-wave in the Southern Hemisphere are well visible in Figure 1.

Moreover, the temporal evolution of geomagnetic field also differs in different regions over the world. The greatest amplitude of changes is observed in the Western Hemisphere, in the regions of the Canadian (Figure 2a) and South Atlantic (Figure 2c) world anomalies. In the Eastern Hemisphere amplitudes of these changes are smaller (Figure 2b and d). The spatial structure of these irregularities is well visible in the maps of geomagnetic secular variations (Figure 3), which are calculated by the formula: 

\[ F_{sv} = \frac{(F_{t_2} - F_{t_1})}{(t_2 - t_1)}, \]

where \( F_{t_1} \) and \( F_{t_2} \) are field intensity in two moments in time, and \( \frac{(t_2 - t_1)}{C_0} \) is the length of the period in years.

Figure 3 illustrates fairly well that focuses of the strongest secular variations evolve with time, in their strength and position over the globe. All the features of the spatio-temporal structure of the geomagnetic field, the problems of its observations and modeling, are described in great details in [9].

2.2 Heterogeneous interplanetary environment and its imprint on the geomagnetic field’s short-term variability

Short-term changes in geomagnetic field (from seconds to days) are caused by the external sources – i.e. the current systems in the magnetosphere and ionosphere. In the absence of solar-terrestrial disturbances, the Earth’s magnetic field shows regular daily variations with small amplitude (~tens of nT), which are primarily composed of 24, 12, 8, and 6-hour spectral components [10–12]. These variations are known as

Figure 1.
Spatial structure of the modulus of the total vector of the geomagnetic field intensity, calculated for 2021 by the IGRF-13 model. (https://www.ngdc.noaa.gov/geomag/calculators/magcalc.shtml#igrfgrid).
solar quiet (Sq) variations. Today it is well understood that Sq variations are induced by the electric currents existing in the ionospheric dynamo region (between 90 and 150 km), where the neutral wind drives an electromotive force through the ionospheric wind dynamo mechanism [13, 14]. The Sq variations are sensitive to the sunspot numbers [10]. For example, the midlatitude Sq currents’ intensity is approximately twice higher in solar maximum than in solar minimum conditions [15, 16].

In addition, the geomagnetic perturbations at the planetary surface also have lunar spectral components. The stronger one is the semidiurnal lunar variation with a period of 12 hours in lunar time or $\frac{1}{C24}12.42$ hours in solar time. The typical amplitude of lunar variation is much smaller – approximately one-tenth of the Sq variation [17]. Geomagnetic lunar variability is a consequence of atmospheric lunar tides, inducing ionospheric currents in the ionospheric dynamo region, which are furthermore projected on the ground [18, 19].

The maximum amplitude of quiet Sq and lunar variations has a maximum during the daytime hours, and when the moon is in opposition. These are smooth periodic variations with intensities reaching 200nT, increasing from the equator to the poles [20].

The quiet conditions, however, are frequently disturbed by active processes on the Sun (e.g. solar flares, coronal mass ejection, coronal holes, etc.). The ejected solar mass and magnetic fields propagate in the interplanetary magnetic field (IMF) as a shock wave, which distorts significantly geomagnetic field when it splashes on the Earth. Only $\sim$1% of energy carried by the solar wind is transferred to the Earth’s magnetic field because the reconnection between interplanetary and geomagnetic fields depends on their directions. It is well established that the southward direction of the interplanetary magnetic field favors its reconnection with the Earth’s magnetic field.
The energy transferred to the magnetosphere in such periods abrupt dramatically by one-two orders of magnitude, reaching power of $\geq 10^{11}$ W [21]. These periods are known as geomagnetic storms or substorms.

The dominant interplanetary phenomena causing intense magnetic storms (with an equatorial Dst index lower than $-100$ nT) depends on the solar cycle. Around the solar maximum, the interplanetary medium is dominated by fast coronal mass ejections (CMEs). Two interplanetary structures are important for the development of storms, involving intense southward IMFs: the sheath region just behind the forward shock, and the CME ejecta itself. Whereas the initial phase of a storm (manifesting itself as a sudden impulse in geomagnetic field) is caused by the increase of plasma pressure at, and behind the shock, the storm’s main phase is due to southward IMFs. The storm recovery begins when the IMF turns less southward, with delays of $\approx 1–2$ hours, and has typically a decay time of 10 hours [22].

Magnetic clouds are large-scale interplanetary formations, caused by coronal mass ejection on the Sun, in which the magnetic field strength, propagating speed, and plasma concentration are higher than in the surrounding flows [23]. The vertical Bz component of IMF slowly changes from negative to positive sign in SN clouds, and vice versa in NS clouds. The interaction of the Earth’s magnetosphere with magnetic clouds, as a rule, is accompanied by intense geomagnetic disturbances [24, 25].
According to some estimates, the geoeffectiveness of magnetic clouds to disturb Earth’s magnetic field is 77% [25, 26].

During solar minimum, high-speed streams from coronal holes dominate the interplanetary medium activity. The high-density, low-speed streams (associated with the heliospheric current sheet plasma) impinging upon the Earth’s magnetosphere cause positive Dst values in the initial phase of the storm. In the absence of shocks, sudden impulses are infrequent in periods of low solar activity. The interaction between fast stream (emanated from coronal holes) and the slow heliospheric current sheet plasma leads to the formation of a compression region with a high magnetic gradient, called Corotating Interaction Region (CIR). The main phase of magnetic storms generated by the CIR is typically weaker and highly irregular. The recovery of geomagnetic storms that happened in periods of inactive Sun is also quite different – lasting from many days to weeks. The southward magnetic field component of Alfvén waves, existing in the high-speed plasma stream, causes intermittent reconnection and substorm activity, as well as sporadic injections of plasma sheet energy into the outer portion of the ring current, prolonging its final decay to quiet day values [22].

For certain classes of magnetic storms, the interaction of CIR with the Earth’s magnetosphere is more efficient than CME [27]. On the other hand, comparisons of the geoeffectiveness of various interplanetary structures, such as shock waves, magnetic clouds, IMF sectors boundaries, and CIR, showed that 33% of CIR are accompanied by moderate or intense storms. This means that every third phenomenon of the observed CIR at the Earth is geoeffective [28].

It is statistically confirmed that geoeffective disturbances can be caused by a whole spectrum of various phenomena on the Sun: flares (especially with the release of high-energy protons); the sudden disappearance of filaments, followed by a transition to a coronal mass ejection; high-speed streams of solar wind; Earth passage through the IMF sectors’ boundaries, etc. However, the features of the magnetic storms are primarily determined by changes in IMF and solar wind parameters [29–32].

The influence of geomagnetic storms on the lower atmospheric variables is studied by many authors. The storm imprint on the near-surface pressure and temperature has been reported by [33–34], on circulation by [35–40], on total ozone density by [41], etc. The latter authors have compared geomagnetic storm manifestation in upper, middle, and lower atmosphere, emphasizing on differences in the atmospheric response to geomagnetic storms. Their main conclusions are summarized as follow: (i) unlike the prevailing latitudinal dependence of storm impact on the upper-middle atmosphere, the tropospheric effects manifest itself with a well pronounced regionality; (ii) the weak seasonal dependence of the storm effect in the upper middle atmosphere is altered by a strong seasonal dependence of detected tropospheric response; and (iii) the geomagnetic effect in the upper middle atmosphere are caused primarily by energetic particles, while the origin of tropospheric effect is still not well understood [41].

All these effects are due to the short-term geomagnetic disturbances, initiated by the external influence – i.e. solar variability and inhomogeneity of interplanetary medium. Although important, these fluctuations of Earth’s magnetic field are short-lasting and their impact on the climate system is negligible. Oppositely, this publication is focused on the long-term variations of geomagnetic field on interdecadal and multidecadal time scales (initiated at the core-mantle boundary) and their relation to climate variability with its regional specifics.
3. Geomagnetic modulation of charged particles in Earth’s atmosphere

3.1 Van Allen radiation belts

Important structures in Earth’s magnetosphere are its radiation belts, which consist of relativistic electron and proton populations, trapped in the Earth’s magnetic field. The Earth has two such belts and some others may be temporarily created. The outer radiation belt, occupying the space between 3 and 10 Earth’s radii (Re), consists mainly of electrons with energies 0.1–10 MeV (million electron volts). The outer belt’s particles have both – solar (mainly helium ions) and atmospheric origin. The protons of the outer belt, however, possess much lower energies than those of the inner belt. The most energetic particles of the outer belt are electrons, achieving energies of several hundred MeV.

An electrons population is found also in the outer edge of the inner radiation belt at a distance 1.5÷3 Re [42]. The outer and inner electrons’ belts are separated by the slot region, where the interactions with the electromagnetic waves called “whistlers” are the main reason for the lower density of the electron population. The electrons’ loss in the slot region is due to the pitch angle scattering (related to the impact of whistlers) – facilitating their escape from the geomagnetic trap [43]. The inner belt electron population is periodically refreshed by the transport of electrons from the outer radiation belt [44]. Moreover, it has been recently recognized that the decay of thermal energy neutrons, produced by cosmic rays striking the upper atmosphere, contributes to energetic electrons in the inner belt and acts as the dominant source of energetic electrons at the inner edge of the inner belt [45].

The inner radiation belt, occupying the near Earth space between 0.2 and 2 Re, is largely populated, however, by energetic protons with energies exceeding 30 MeV. According to the current understandings, these protons originate from: (i) the decay of neutrons – produced within the interaction between galactic cosmic rays and atmospheric atoms and molecules [46], and (ii) solar protons – injected into the interplanetary space during the solar flares and coronal mass ejections [47–49]. The solar energetic protons are the primary source of particles for the inner belt, which energy is beneath ~100 MeV [49] and sometimes produces a long-lived proton belt – distinct from the inner radiation belt [47, 50].

3.2 Particles’ focusing in a heterogeneous magnetic field

Particles trapped within the geomagnetic field are urged by the Lorentz force (1) to move along the magnetic field lines on spiral trajectories (the result of a combined circular and a field align motions), continuously bouncing between the Northern and the Southern Hemispheres.

$$m \frac{dv}{dt} = q(E + (v \times B)); \quad v = \frac{dr}{dt}$$

where: $B(r,t)$ is external magnetic field – function of the spatial dimensions and time, $r$ and $v$ are respectively particle’s radius vector and velocity; “m” is particle’s mass and “q” – its charge.

Besides the helical movement of particles along geomagnetic field lines, they also perform the additional movement in a direction perpendicular to the magnetic field.
lines – known as *magnetic drift*. This type of motion is determined by the non-uniformity of Earth’s magnetic field in the direction perpendicular to B, and by the magnetic field curvature (2)

\[
v_{\text{drift}} = \frac{m}{q} \cdot \mathbf{B} \cdot \left( v_{\parallel} B \times \nabla B - \frac{v_{\perp} \rho \times B}{\rho^2} \right)
\]

where \( \mathbf{B} \) is the magnetic vector, \( \rho \) – the radius of the geomagnetic lines curvature, \( v_{\parallel} \) and \( v_{\perp} \) are projections of particle’s velocities parallel and perpendicular to geomagnetic field line; \( q \) and \( m \) are respectively particle’s charge and mass. The first term in the brackets corresponds to the magnetic gradient perpendicular to the field lines, while the second term – to their curvature.

Formula (2) shows also that particles’ drift across the magnetic field lines depends on their charge \( q \), and consequently leads to a charge separation, which in turn generates electric field \( E \) along the drift direction. The combined effect of \( E \) and \( B \) fields induces an \( E \times B/B^2 \) drift of particles, which displaces positive ions and negative electrons in the same direction – perpendicular simultaneously to \( B \) and to \( E \). These charged particles are then “lost” in the ambient atmosphere, where they release their energy, producing showers of secondary particles.

In a dipolar geomagnetic field (with its cross-latitudinal magnetic gradient) the protons are drifting westward, while electrons – eastward. The real geomagnetic field has, however, a non-dipole component creating additionally a cross-longitudinal gradient. In this case, the protons (entering the denser atmosphere from the west) are shifted sought-westward in regions with a positive cross-longitudinal gradient and sought-eastward – in regions with a negative gradient (refer to Eq. (1)). Consequently, the overall westward drift (forced by the magnetic curvature and cross-latitude gradient) is reduced by the eastward component – exerted in regions with a negative azimuthal magnetic gradient. Furthermore, the electric field (induced by the charge separation of impending particles) is significantly reduced in these regions. Finally, the number of particles expelled outside the magnetic trap (due to the \( E \times B/B^2 \) electric drift) is much less. More precisely, only a few of them have a “chance” to be lost in the atmosphere in said regions.

Oppositely, in regions with positive azimuthal geomagnetic gradients, the southward drift component changes slightly in the direction, but not the amplitude of the westward drift, impelled by the magnetic curvature and latitudinal gradient. Consequently, in these regions, the induced electric field – resulted from the charge separation of arriving particles – is much stronger. It will intensively expel the charged particles outside the magnetic trap through the imposed \( E \times B/B^2 \) drift. Furthermore, these particles interact with the atmospheric molecules creating secondary electrons, ions, and nuclear products, giving rise to the ionization of the lower atmosphere.

### 3.3 Hemispherical asymmetry of geomagnetic non-dipolar field and its influence on particle precipitation in Earth’s atmosphere

The confinement of any particle in the gradient magnetic field \( B \) depends on the ratio between the maximum field strength \( B_{\text{max}} \) in the polar regions (where the backward reflection of trapped particles occurs) and the equatorial magnetic field strength \( B_0 \), i.e.
\[
\sin(\alpha) = \sin(\alpha_0) \cdot \sqrt{\frac{B_{\text{max}}}{B_0}}
\] (3)

where the angle \(\alpha_0\) between velocity vector of arriving particle and corresponding magnetic line in the equatorial region, is known as an equatorial pitch angle, and \(\alpha\) is the continuously changing pitch angle, when particle is moving along the magnetic field line. Thus \(\alpha\) increases with particles’ movement toward the pole, due to the reduction of filed aligned component of particles velocity, and increase of its velocity in a direction perpendicular to geomagnetic field line (refer to formula (4) and Figure 4), are decreases when particle is moving toward the equator.

\[
\alpha = 2\pi \left( \frac{v_B}{v_\perp} \right) \cdot r_B, \quad \text{where } r_B = \frac{m}{qB} v_\perp
\] (4)

Any particle is assumed trapped by the magnetic field, when the angle \(\alpha\) becomes greater than \(\frac{\pi}{2}\), because at this point – known as a magnetic mirror – the particle reverses its direction of movement, remaining confined by the magnetic field line. Formula (3) shows that particles approaching Earth’s magnetosphere at very small angles could not exceed the pith angle \(\frac{\pi}{2}\), and when enter the mirror point these particles are “lost” in the atmosphere. The minimum value of angle \(\alpha_{\text{om}}\) (for which the maximum magnetic field is still able to reflect particles) is called loss cone. If a particle arrives at angle lower than the solid angle defined by \(\alpha_{\text{om}}\), it will be lost in the ambient atmosphere on its motion along the magnetic field line. Formula (3) shows also that the efficiency of magnetic mirror to reflect charged particles does not depend

---

**Figure 4.**
*Orientation of the particle’s velocity vector, with respect to the equatorial magnetic field \(B_0\), and changing particles pitch angle \(\alpha\) (from \(\alpha_0\) at the equator, to 90 degrees at magnetic mirror point).*
neither on the particles' speed, nor on their charge and mass (in the guiding center approximation, known also as adiabatic approximation).

The geomagnetic field near the poles is stronger in the Southern Hemisphere, compared to those in the Northern Hemisphere. Consequently, in the case of isotropic particles' flux arriving at magnetopause – almost every third particle will be confined in the Southern Hemisphere, while in the Northern Hemisphere less than \( \frac{1}{4} \) of all arriving particles are trapped, because of its larger loss cone \([51]\). This means that some of particles confined in the Southern Hemisphere could not be held by the weaker geomagnetic field in the Northern Hemisphere. The expected result is – more particles precipitating in the Northern Hemisphere.

3.4 Regener-Pfotzer maximum and its influence on the lower stratospheric chemistry

Energetic particles penetrating deeper in the atmosphere create showers of secondary particles, produced from their interaction with atmospheric molecules – the deeper the penetration is, the wider the showers are. In the lower stratosphere, the number of secondary products dramatically increases, becoming maximal at a certain level. This level is known as a Regener-Pfotzer maximum. Beneath it, the concentration of secondary ions and electrons decreases again.

The longitudinal geomagnetic gradient and hemispherical asymmetry of geomagnetic field determine the uneven distribution of geomagnetically trapped particles' precipitation over the globe (refer to Subsections 3.2 and 3.3). Existence of such an effect is illustrated in \([52]\).

3.4.1 Ozone formation in the lower stratosphere

For almost a century – since the creation of the theory about ozone production in the upper atmosphere by Sydney Chapman \([53]\) – the single source of stratospheric ozone is believed to be the photo-dissociation of molecular oxygen by solar ultraviolet radiation. Recently it has been shown that in the dry lowermost stratosphere the lower-energy electrons in the Regener-Pfotzer maximum initiate ion-molecular reactions producing ozone \([54]\).

The mean energy of electrons in the Regener-Pfotzer max (\(~35\) eV \([55]\)) is not sufficient to break the molecular bounds of the major atmospheric constituents. It is, however enough to ionize the molecular oxygen (Reaction (5)). The oxygen cation interacts furthermore with neutral oxygen molecule, producing a tetra-oxygen ion \(O_4^+\) \([56, 57]\), (see Reaction (6)).

Being very unstable, this oxygen complex rapidly dissociates into two different channels \([57]\). The first channel (7) produces \(O_3^+\) and \(O\), while the second one restores the \(O_2^+\) ions (8). The weakly bonded \(O_3^+\) molecule easily dissociates or exchanges its charge with \(O_2\), yielding a neutral ozone. Most efficient, however, appears to be the dissociative recombination of ozone cation \(O_3^+\) to three oxygen atoms, occurring in 94% of all cases \([58]\), in prevailing conditions typical for the lower stratosphere (i.e. ground state ozone cations and lower energetic electrons).

\[
O_2 + e^- \rightarrow O_2^+ + 2e^- + 12.07 \text{ eV} \quad (5)
\]

\[
O_2^+ + O_2 + M \rightarrow O_4^+ + M + 3.5 \text{ eV} \quad (6)
\]
As a result, the dissociation of one $O_4^+$ molecule leads to the formation of four new $O_3$ molecules (reactions (7) and (9)), while reaction (8) and continuous ionization of $O_2$ by the atmospheric lower-energy electrons support a steady production of $O_4^+$ (more detailed analysis could be found in [51, 53]. Thus, the reactions (6)–(9) form an autocatalytic cycle for continuous $O_3$ production in the lower stratosphere. An absolutely necessary condition for the activation of autocatalytic ozone production is a dry atmosphere. Otherwise, water clusters of $O_2^+$ are formed instead of $O_4^+$ [59]. The maximum efficiency of this ozone-producing cycle should be expected near the level of the highest secondary ionization produced by GCRs, i.e. near the Regener–Pfotzer maximum.

### 3.4.2 Evidence for particles’ influence on the lower stratospheric $O_3$ density

At middle and high latitudes, the Regener–Pfotzer maximum is placed well above the tropopause [60], which provides the necessary conditions for activation of the autocatalytic cycle of ozone production – i.e. a dry atmosphere and plenty of low energy electrons. As discussed in Sections 3.2 and 3.3, and shown in [52], the ionization in the Regener–Pfotzer maximum is unevenly distributed over the globe. Remind that an increased particles’ flux is expected in regions of geomagnetic field strengthening. Consequently, if the autocatalytic production of ozone is significant, the longitudinal variations of the Regener–Pfotzer maximum ionization should be projected on the ozone profile.

Figure 5 presents a comparison between ozone profiles in regions with increasing and decreasing geomagnetic field, during solar minimum in 2009. Note that the $O_3$...
values beneath the peak ozone density are higher in regions with increasing geomagnetic field (i.e. the longitudinal sector 90–50°W in the Western Hemisphere and 120–140°E – in the Eastern one), relative to corresponding O$_3$ values in regions with a geomagnetic field weakening in the sectors: 140–110°W and 30–50°E.

The longitudinal variations in atmospheric ozone have been noticed long ago [61, 62]. The authors have suggested that this variability could be related to the planetary wave structure. However, the maximal amplitude of the stationary planetary waves is found at ~300 hPa [61], while the highest amplitude of O$_3$ longitudinal variations in ERA Interim reanalysis is placed near 150–70 hPa [51]. These and some other problems, e.g. [63, 64] suggest that other factor(s) (e.g. energetic particles) may have an important influence on the spatial and interannual variability of the extratropical near tropopause O$_3$.

In order to assess quantitatively the coupling between energetic particles precipitating in Earth’s atmosphere and lower stratospheric ozone, as well as its spatial distribution, we have performed a cross-correlation analysis in a grid with 10° increments in latitude and longitude. Ground-based measurement of galactic cosmic rays (GCR) by neutron monitors, has been used as an indication of energetic particles flux. The Moscow record of GCR has been expanded backward in time by the paleo constructed GCR intensity [65]. The 11-year periodicity of GCR has been removed by moving averaging procedure with 22-year running window. The winter values of ozone at 70 hPa have been taken from ERA twentieth century reanalysis, covering the period 1900–2010. Data have been preliminarily smoothed by 11-year running window.

The map of ozone-GCR correlation is presented in Figure 6 (colored shading). It is important to note that the map has been created from correlation coefficients, being preliminary weighted by the autocorrelation function of GCR with time lag corresponding to the delay of O$_3$ response to the GCR forcing. This procedure, which reduces correlation coefficients with longer time lags, allows a comparison of correlations with different time lags. The introduction of weighs for the lagged correlation coefficients is justified by the assumption that the effect of the applied forcing in a given moment of time decreases with moving away from this moment [66].

Figure 6.
Figure 6 shows that the ozone responds differently to particles’ impact at different regions over the world – not only by amplitude but even by sign. Thus, at high latitudes and in the Indo-Pacific region, ozone varies synchronously with GCR. On the other hand, at the Northern Hemisphere extratropics and near the southernmost edge of Latin America, both variables covariate in antiphase – meaning that in these regions ozone increases with time.

Such heterogeneity in ozone response to particles’ forcing could be attributed to the different origins of impacting particles. For example, the polar regions are vulnerable to the particles from interplanetary space, propagating along the open geomagnetic field lines. The long-term variations of these particles are modulated mainly by the interplanetary magnetic field in the heliosphere. The latitudes shielded by the closed geomagnetic field lines (i.e. the tropics and mid-latitudes) are accessible to very highly energetic particles (which are very few), and to the radiation trapped in the Van Allen radiation belts. The latter are subject to geomagnetic lensing (in the lowest part of their helical trajectories along the magnetic field lines) and asymmetrical precipitation in both hemispheres, due to the asymmetry of geomagnetic field (refer to Sections 3.2 and 3.3).

Figure 6 shows in addition the effective vertical cut-off rigidity of geomagnetic field (contours), with the values greater than 12 GV being colored in red. Note that the strongest GCR-O₃ correlation over the equatorial Indo-Pacific region fairly well coincides with the higher geomagnetic cut-off rigidity. Having in mind the centennial negative trend in GCR, the positive correlation coefficients indicate ozone depletion during the examined period (1900–2010). Consequently, the reduced ozone density could be attributed to the weaker particles’ fluxes accessing the said region.

On the other side, the negative GCR-ozone correlation in extratropics suggests enhancement of ozone density near 70 hPa. This result indicates that particles confined in the outer radiation belt are involved in ozone production in the lower stratosphere. Powered by the solar wind, the population of this radiation belt is highly variable [45], reflecting the changes in solar activity. The examined period is characterized by enhanced solar activity, which appears to be projected on the extratropical latitudes as enhanced ozone density at 70 hPa – due to the enhanced particles’ population in the outer radiation belt.

The positive GCR-O₃ correlation at polar latitudes suggests a centennial ozone depletion, which corresponds to the decreased flux of GCR, modulated itself by the stronger interplanetary magnetic field in the heliosphere during the twentieth century [51].

The centennial changes in ozone mixing ratio at 70 hPa, between the first decades of twenty-first and twentieth centuries, is presented in Figure 7. Note that ozone changes deduced from the correlation map in Figure 6 fairly well corresponds to the observed changes of ozone at 70 hPa.

4. Ozone as a mediator of geomagnetic field influence on climatic variables

The sensitivity of atmospheric temperature profiles and climate to the ozone density (particularly near the tropopause) has been noticed long ago [67–71], etc. The detected synchronization between the spatial and temporal variability of particles’ flux reaching the ground, and the lower stratospheric ozone, is a hint that ozone could serve as a mediator of the geomagnetic field-energetic particles’ influence on climatic
variables (i.e. temperature, pressure, etc.) [72]. The following section throws some more light on this problem.

4.1 Ozone imprints on climatic variables

4.1.1 Hemispherical and longitudinal asymmetries of ozone-temperature covariance

The potential synchronization between ozone at 70 hPa and near-surface temperature variability, within the period 1900–2010, is examined by the use of lagged cross-correlation analysis. The leading role of winter ozone in the ozone-temperature correlation, have been analyzed in a spatial grid with 10° steps in latitude and longitude. The time series of both variables are taken from the monthly values provided by the ERA twentieth century reanalysis. The correlation map presented in Figure 8 is created from the preliminary weighted correlation coefficients by the autocorrelation function of ozone, with lag corresponding to the time delay of temperature response – to account for the reduced weight of covariances being away from the moment of applied forcing.

The most impressive of the results shown in Figure 8 is the asymmetry of the temperature response to ozone variations. The positive $O_3 - T_{2m}$ correlation coefficients – over Eurasia and the extratropical Pacific Ocean, unlike the overall negative correlation, require their explanation. In addition, the analysis of the long-term variations of ozone and temperature at 60°N latitude, and at longitudinal zones, 140 and 70°W (corresponding to the regions with positive and negative GCR-ozone correlation) are presented in Figure 9. It is important to note that the short-term variations are preliminarily filtered by data smoothing through 11-year running average procedure.

Figure 9 clarifies that the lower temperature trend of Eastern Asia corresponds to the higher ozone density at 70 hPa. Oppositely, the stronger warming in south-eastern Canada corresponds to a lower ozone density at 70 hPa, with a negative centennial trend. Examination of the global picture of twentieth century warming (presented in Figure 10) reveals that the “hot spots” of contemporary global warming (i.e. north-eastern Canada and Greenland, and the Southern Ocean – southward of Africa) correspond to the regions of negatively correlated ozone and temperature (refer to Figure 8). In opposite, the regions with in-phase co-varying ozone and temperature are characterized by weaker warming.
Figure 8.
(top) correlation map of winter ozone at 70 hPa and air surface temperature, calculated over the period 1900–2010; (bottom) time lag in years of temperature response following ozone changes.

Figure 9.
(left) Time series of winter ozone at 70 hPa and 60°N latitude, obtained at Eastern (140°E longitude) and Western (70°W) longitude; (right) air surface temperature at the same latitude and longitudes.
Conclusively, the above results indicate that the strongest warming during the twentieth century is observed in regions with reduced density of the lower stratospheric ozone.

4.1.2 Climatic modes and lower stratospheric ozone density

Climate variability is not homogeneous in space and is usually described as a combination of some “preferred” spatial regimes, called modes. In meteorology and climatology, the term ‘mode’ is used to describe a spatial structure with at least two strongly connected centers of action [73]. The most famous of these spatial structures – known as climatic modes – affect weather and climate on different spatial and temporal scales. Most climatic modes are defined by statistical classifications of the observed variability of surface temperature, sea-level pressure, precipitations, etc. They could be a result of the action of fundamental physical processes such as the instability of the climatic mean flow, mesoscale interactions between the atmosphere and the ocean, etc. [74]. However, these statistical patterns may also be artifacts of nature, whereby they are not stable over long periods of time, or they may be statistical artifacts.

Although the spatial-temporal variations of climatic modes are extensively studied, the reasons for their occurrence and variability over time are not fully understood. Internal variations of the climate system are usually associated with the processes of energy exchange and redistribution between the planetary atmosphere and ocean. The huge heat capacity of the ocean is the reason for its inertia in response to short-time fluctuations of atmospheric variables, which transforms them into long-period variations of the ocean surface temperature. This understanding does explain the phase alteration, but it is not able to explain neither the various manifestations of climatic modes [75] nor their long-term changes.

Analysis of the spatial-temporal variability of GCR and ozone at 70 hPa reveals the important role of the latter in the formation of regional specificity of air surface temperature variability (refer to Subsection 4.1.1, or to [76]). Examination of the temporal synchronization between two of the most important climatic modes – North

Figure 10. Centennial changes of the air surface temperature between the first decades of twenty-first and twentieth centuries, derived from the ERA twentieth century reanalysis.
Atlantic Oscillation (NAO) and El Niño Southern Oscillation (ENSO) – confirms the existence of statistical relation in the regions of modes’ manifestation [77].

Figure 11 illustrates the projection of the long-term variations of ozone at 70 hPa on the NAO index (which describes the variability of the surface pressure between Azores and Iceland). The coupling between both variables has been estimated by the use of the lagged cross-correlation analysis between annual values of NAO index (smoothed by 5 points averaging) and winter ozone values at 70 hPa (smoothed by 11 points moving window). The stronger smoothing of ozone is due to its higher temporal variability. The leading factor (i.e. the “forcing”) in calculated ozone-NAO variability is ozone. As in the previous case, the correlation coefficients have been preliminarily weighted (according to different delay of NAO response) with the ozone’s autocorrelation function. The physical reasoning behind this weighting is that the memory of the climate system for the applied impact weakens with time. This suggests that the high correlation coefficients with a large delay are more or less random.

Figure 11 shows that the ozone’s impact on the NAO climatic pattern fairly well coincides with both centers of action (Azores and Iceland) determining the phase of NAO mode. Unlike the previous results (stressing the leading role of the northern [78] or the southern part of NAO spatial structure [79]), Figure 11 indicates that the variations of lower stratospheric ozone density can impact each center of action.

Figure 11.
(top) Cross-correlation maps of the winter lower stratospheric ozone and NAO index, calculated for the period 1900–2010; (bottom) time lag of NAO response in years.
(Azores or Icelandic), or simultaneously both of them – altering in such a way the phase of NAO mode [76].

Analysis of the time delay of NAO response to ozone changes shows that surface temperature near the Icelandic Low respond with a delay of 1–2 years. In the subtropical center of action, however, the atmospheric response is delayed approximately by a decade (see the bottom panel in Figure 11).

Figure 12.
Comparison of correlation maps of ozone at 70 hPa with GCRs (dark shading) and water vapor at 150 hPa (contours), for winter (a) and (c), and summer (b) and (d) panels.
4.2 Mechanism of ozone influence on climatic variables

Direct ozone influence on the surface temperature is quite small due to the mutually exclusive effect of stratospheric and tropospheric ozone in the planetary radiation balance [70]. Ozone's ability to absorb the incoming solar radiation (and to a lesser extent the longwave radiation emitted from the Earth), makes it a radiatively active gas. The covariance between the near tropopause ozone and temperature has been noticed long ago [80, 81]. However, the tropopause temperature determines the moist adiabatic lapse rate and accordingly the static stability of the upper troposphere [82, 83], which in turn alters the humidity near the tropopause [51]. For example, ozone depletion cools the near tropopause region making the upper troposphere more unstable [82, 83]. The upward propagation of the more humid air masses from the lower atmospheric levels moistens the upper troposphere, and strengthens the greenhouse warming of the planet. The satellite measurements show that water vapor at these levels ensures 90% of the greenhouse warming of the total atmospheric humidity [84]. Consequently, ozone variability in the lower stratosphere is projected on the planetary surface through the modulation of the strength of greenhouse warming.

**Figure 12**, which compares the lag-corrected correlation maps of ozone mixing ratio at 70 hPa with: (i) GCR, and (ii) humidity at 150 hPa, is a good illustration of our hypothesis validity. Note that the latitudinal band of antiphase correlation between GCRs and ozone (dark shading), and in phase correlation between ozone and water vapor (red contours), coincide impressively well. In the Northern Hemisphere, this coincidence persists round the year, although being slightly reduced in summer season (compare panels (a) and (b) in **Figure 12**). In the winter Southern Hemisphere, the area of synchronous variations of GCR, ozone, and humidity is narrower and practically disappears in summer (**Figure 12d**). The results presented in **Figure 12** are a good indication that ozone–humidity variations, which are projected down to Earth's surface by the strengthening or weakening of the greenhouse effect, are actually related to GCR variability.

5. Conclusions

Historical and contemporary changes in climate system put a lot of questions, the answers to which are difficult. This motivates scientists from different branches to look for various factors with a potential influence on the climate system. Geomagnetic field is one of the proposed factors, due to the rendered multiple evidence for spatially or temporary co-varying geomagnetic field and climate, at different time scales. In this chapter, we clarify that hypothesized geomagnetic influence on climate could be reasonably explained through the mediation of energetic particles, propagating in Earth's atmosphere, and their influence on the ozone density in the lower stratosphere.

More specifically, the non-dipolar part of geomagnetic field creates irregularities in the spatial distribution of lower atmospheric ionization in the Regener-Pfotzer maximum [51]. The bulk of low-energy electrons and dry lower stratosphere favors activation of autocatalytic ozone production at these altitudes. Thus geomagnetic irregularities are projected on the ozone density near the tropopause. Being a radiatively active gas, the ozone itself affects the temperature and humidity in the tropopause region, altering in such a way the greenhouse effect and consequently – the near-surface temperature.
This chapter provides evidence for the validity of this chain of sequences, which gives an adequate explanation of hemispherical and longitudinal asymmetry of the lower stratospheric ozone distribution, regionality of climate change, formation of regional climate patterns, known as climatic modes, etc.

Acknowledgements

The authors are grateful to the project teams of ERA 20C and ERA Interim reanalyses, providing gridded data for ozone temperature and pressure, as well as to the Climatic research unit of University of East Anglia – for the data of NAO index. We are thankful to the National Centre for Environmental Information (NOAA) for providing IGRF model.

Funding

This research was funded by National Science Fund of Bulgaria, Contracts KP-06-N34/1/30-09-2020, and DN 14/1 from 11.12.2017.

Author details

Natalya Kilifarska*, Volodymyr Bakhmutov and Galyna Melnyk

1 Climate, Atmosphere and Water Research Institute, Bulgarian Academy of Sciences, Sofia, Bulgaria

2 Institute of Geophysics by S.I. Subbotin name, National Academy of Sciences of Ukraine, Kyiv, Ukraine

*Address all correspondence to: natalya_kilifarska@yahoo.co.uk

© 2022 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
References


[30] Yaremenko LN, Melnik GV. Magnetic storms on 15–16 July 2000 and


[45] Li W, Hudson MK. Earth’s Van Allen radiation belts: From discovery to the Van Allen probes era. Journal of
2019;124:8319-8351. DOI: 2018JA025940


[59] Kilifarska NA. Hemispherical asymmetry of the lower stratospheric O₃ response to galactic cosmic rays forcing. ACS Earth and Space Chemistry. 2017;1:80-88. DOI: 10.1021/acsearthspacechem.6b00009

in Space Research. 2017;60:991-998. DOI: 10.1016/j.asr.2017.05.014


