The external magnetic sources over the polar caps - Feasible modelling \textit{versus} unrealistic expectations

Giovanni P. Gregori

Istituto di Fisica dell'Atmosfera (IF-A-CNR), Roma, Italy; and World Laboratory, Lausanne, Switzerland

Abstract

No average status can be defined for the sources of the external magnetic field over the polar caps, no index can provide with any such model, the observation of no single quantity or parameter can give the ultimate solution. Rather, every case history has to be considered independently. It is possible, however, to approach the problem from an interdisciplinary viewpoint, and to attempt to make an instant modelling of the electric currents that flow in the ionosphere and magnetosphere above the area of aeromagnetic prospecting. A few relevant previous such examples are discussed (such as Akasofu's inference on magnetospheric substorms derived by means of polar auroras, or the presently unfashionable Svalgaard vortex by means of the observed geomagnetic field and dealing with the pattern of the electric field within the magnetosphere, or the Sun-aligned auroral arcs inside the oval that are monitored by satellite, or, perhaps, the luminosity curve of polar auroras). It appears likely that some substantial achievement will be attained altogether with the progress in the understanding of the general pattern of a few typical recurrent configurations over the polar caps, in terms of a multidisciplinary input from different observations, either by ground-based observatories, or by space platforms.

Key words  magnetospheric model – plasma cavity – Hamilton’s principle – substorm – auroral oval – Svalgaard vortex – Θ aurora – luminosity curve – electric current – electric field – plasma drift – trapped particles

1. Introduction. Problem definition

It is here shown how a concrete modelling of the external magnetic sources over the polar caps can be carried out on a physical ground, although only provided that one gets rid of a few fundamental drawbacks, that are a more or less unconscious bias of the present standard way of dealing with the magnetosphere. Such an achievement, however, which is basically more a matter of philosophy rather than of formal improvement of previous models, can be concisely highlighted by some simple logical analogies, that are not here aimed at being colloquial, rather at focusing on some essential logical pertinent aspects.

Carrying out an aeromagnetic survey over a polar cap is almost like measuring the fading \textit{versus} time of the twilight-glow from your garden, while your neighbour continuously shoots a large amount of fireworks: for your aeromagnetic survey the neighbour is the solar wind. You can only attempt to search for some regularity in space and time of the disturbance from the fireworks, in order to subtract their unknown effects from your records, while you obviously will never afford to get rid of the entire perturbation unless you know the detailed schedule of your not very polite neighbour. In the case of the magnetosphere of the Earth, we know that there are quiet and storm times, and also some
generally intermediate conditions (e.g., Chapman and Bartels, 1940). Such a morphological characterisation was derived from simple visual analysis of standard magnetograms, according to the typical philosophy developed prior to the computer age, by which it was assumed that the status of such a complex physical system can be characterised just by one degree of freedom (d.o.f.) alone, i.e. by a geomagnetic index. One additional relevant morphological distinction, that was derived basically from auroral phenomena, led to the definition of magnetospheric substorm, a feature that however cannot be recognized, in terms of other observational parameters, as clearly as for auroras (Akasofu, 1968, 1977). Some additional detail is given in terms of the ionospheric dynamo, by which, upon consideration of the dynamics of the upper atmosphere (including tides), one obtains (e.g., Chapman and Bartels, 1940; Matsushita and Campbell, 1967; Akasofu and Chapman, 1972) a description of the diurnal variation ($S$ and $L$ fields), a phenomenon that, however, is very effective at low latitudes, and of little practical help over the polar caps.

If we want to get rid in some way of the magnetospheric perturbations that affect the aeromagnetic surveys over the polar caps, we must make some concrete step forward in their morphological understanding. It appears to be an excessive self-limitation using only one d.o.f. for depicting the magnetosphere: today, in the PC age, one can easily deal with multidimensional indices. On the other hand, the practical problem is not searching for more complicated indices, but envisaging some quantitative model for the perturbation field that has to be subtracted from our records. This is the target of the present paper.

Let us consider the case of a snow avalanche that rolls down the slope of a mountain. We want to model it, and we implicitly want to give a detailed description of what snow sample, within the valley, is being involved, or not, in increasing the size of the avalanche. Such a highly detailed description, however, is an obvious approximation, in the fact that there are several physical parameters that control phenomena, and that cannot be fed into the model (such as the compactness of the snow, its thermal history, i.e. whether after a snowfall the air temperature and wind caused a partial icing of its surface or not, the soil and air temperature, the possible partial melt at different depths within the snow, the tilt of the slope, the size and speed of an eventual triggering object, the wind blowing during the avalanche fall, etc.). However, apart all such indeterminacies, we know for sure that the snow mass must go down the slope, due to a variational principle. We cannot know whether any one given small amount of snow located at some specific site of the valley is going to be involved, or not, in the avalanche precipitation. However, we know that the potential energy of the avalanche within the gravity field must decrease, and transform into kinetic and friction energy.

Such a situation is much like the case of the magnetosphere. When it was modelled, the example of the analogous problem of designing an airplane within a wind tunnel was followed: air blows in and out from it, leaving the model airplane in a steady condition, in which every molecule that leaves the tunnel is simultaneously replaced by a newly arriving one. Such an approximation is basically always satisfied. In rational mechanics this is called Eulerian viewpoint, in contrast with the Lagrangian viewpoint by which every single particle, or air molecule, has its own individuality, expressed in terms of its specific canonical coordinates. The model airplane computed in terms of such a Eulerian viewpoint results quite satisfactory. In practical reality, however, an airplane suffers from some small turbulence, by which e.g., even with a clear sky, the crew seldom warns the passengers to fasten their seat belts. The theory of fluid dynamics, suitably re-adapted for including e.m. interactions, leads to MHD (Magneto-Hydro-Dynamics). Hence, the model of the magnetosphere was developed by analogy with the airplane case, although in terms of MHD instead of fluid dynamics. The result was that, since the magnetic field lines (m.f.l.s) cannot be broken, because it must be

\[
\text{div } B = 0 \quad \text{curl } H = j \quad (1.1a,b)
\]

the m.f.l.s originated within the Earth are compressed into a limited space, while outside it the solar wind blows transporting its own magnetic
field. The resulting magnetosphere is called drop-like (fig. 1), and its boundary was named magnetopause. It can be formally modelled by considering the geometrical locus where the kinetic energy density of the solar wind (or its kinetic pressure, which is the same) equals the energy density of the magnetic field of the Earth. Outside the magnetopause, the kinetic energy density is overwhelming and the solar wind transports the frozen-in m.f.l.s. Inside the magnetopause, the magnetic energy density prevails, and the particles are trapped in the radiation belts.

Compared with the afore-mentioned airplane example, an unprecedented fact occurred: the solar wind contains some inhomogeneities, let us call them tout court «turbulence», by which, either occasionally or permanently, there are gaps within it, that are called plasma cavities, the effect of which is comparatively much more critical, compared to the airplane example. The most striking ultimate observational consequence was the unexpected discovery of the neutral sheet.

Some notes are needed for the reader who is not acquainted with the history of magnetospheric physics. Early in the space age, in the late ’50’s, the discovery of radiation (or Van Allen) belts was, maybe, the most impressive result. Their existence, however, had been foreseen early in our century by the great Norwegian mathematician Carl Størmer, who also constructed expressive 3D models of such orbits (see his account in Størmer, 1955). The next great discovery, early in the ’60’s, was the magnetopause by Ness’ school from Goddard, and shortly afterwards they also realised that the magnetosphere did not behave like a drop (see below), rather it had a completely unexpected feature, the neutral sheet. The fact that it was unexpected is testified by the incapability, for decades, of providing any physical explanation for it, other than in terms of some physically non-clear and ad hoc assumptions (see below), such as reconnection, friction across the magnetopause, motion of the m.f.l.s internal to the

Fig. 1. A «drop» magnetopause (here shown out of scale) results, on the front side i.e. on the left of the figure where the Sun is located, from the balance between the kinetic pressure of the solar wind and the inner magnetic pressure. When such a balance is computed far downstream, the resulting magnetopause is found to open more and more, having a cross-section with a plane though the Sun-Earth line that looks approximately like a parabola. When the thermal pressure within the solar wind is taken into account, its role, which is negligible on the front, becomes more relevant the farther one goes downstream, until it overwhelms all other pressures, thus making the magnetopause close much like a drop. After Gregori (1990). Figure and captions after Gregori (1998a).
magnetopause being transported by the motion of the m.f.l.s external to it, etc. All such items are better clarified in the following.

As far as the solar wind is concerned, one full specific discipline was soon developed, and its deviations with respect to an ideal perfectly steady flow were systematically investigated. The first anomaly being considered recalled Chapman and Ferraro (1931a,b, 1932a,b, 1933), who supposed an empty interplanetary space, occasionally experiencing some huge eruption of solar plasma that impinges over the Earth. The magnetic field of the Earth excavates some kind of a drop-cavity inside such incoming solar plasma. This is their classical theory of magnetic storms, that mutatis mutandis (*) still holds nowadays, although referring only to the sudden commencement of a magnetic storm. Akasofu (1998) briefly highlights the history of the interpretation of magnetic storms, and he also stresses the several apparent contradictions and still debated aspects. The opposite anomaly is attained whenever there occurs a gap within the solar wind (compared to some average flow, that is conventionally assumed to represent the reference ideal steady flow). In such a case, the conventional word used by specialists is plasma cavity, for expressing the fact that within some suitable volume inside interplanetary space there is some box with some smaller amount of plasma compared to the steady reference condition, or even empty space.

This led to a searching for some empirical correction of the formal MHD theory, by which it was supposed that, more or less sporadically or permanently, some m.f.l.s from the interplanetary environment result apparently reconnected across the magnetopause with m.f.l.s of the Earth. This is manifestly in disagreement with Maxwell's laws (1.1a,b). However, this is simply a suitable artifice, with no intrinsic physical implication, other than being a convenient way of realistically dealing with a model of the magnetosphere in terms of MHD, and by taking into suitable account some new empirical evidence related to some unexpected physical effect, originated by the fact that the flow of the solar wind is not perfectly steady. The incapability of understanding the real physical significance of such a concept is clearly testified by the following statements by two most authoritative scientists. Hultqvist (1990) states:

«... the reconnection process, which is still not well understood physically, has become a tool rather widely used by theoretical astrophysicists».

Akasofu (1998) states:

«... At a time, it was said that one could not be a magnetospheric physicist unless one tried to explain auroral substorms in terms of magnetic reconnection. ... In this paradigm, it was said that magnetic reconnection must occur, since the theory is so trustworthy and that all observations will eventually be understood in terms of it. ... However, in a powerful paradigm, an observation contradictory to the accepted theory gets little attention and often gets ridiculed. ... A powerful paradigm will delay the progress of its field. A high degree of agreement in a paradigm will suppress alternatives, so that researchers are lost when their paradigm is eventually found to be inoperative. During the period of a powerful paradigm, the progress of its field is actually retarded and sometimes it regresses. Much research time is lost as well. Indeed, we have lost about 30 years by pursuing the hypothesis of magnetic reconnection by believing that it is the only theory to explain substorms». 

Another concept, formerly introduced by analogy arguments, deals with the so-called convection within the magnetosphere. «Convection» of any given physical system is a concept having different logical implications, and either one of at least three logical origins. i) In some circumstances it relies on the physical assumption that the mass of the system must be conserved. Hence, whenever there is some flow that transports matter from some part of the system to another, there must be some corresponding return flow somewhere else. ii) In other circumstances the definition is mainly mathematical, rather than physical. For example,
refer to the authoritative statement by Elsasser (1966) dealing with the speculated convection within the Earth mantle:

«... in the geophysical literature, convection is sometimes associated with the concept of ‘cells’, that is, with comparatively simple, closed systems of circulation. Such cells, however, are primarily mathematical tools (they are orthogonal vector fields in the language of the mathematician) and the actual motions are likely to be much more complicated .....».

iii) In either case, convection is a concept needed by the human mind that must deal with simple concepts, and a convection cell is one such example, in the fact that some conspicuous amount of matter of the system ought to perform at least one complete round trip during some given non-null time interval. This implies that the system remains in a status of regular evolution in order for at least once such an entire round trip to be completed. In contrast, whenever the system experiences some temporal changes of its prime governing processes, it eventually begins to develop one such convection cell, without having however the time needed for its completion. In such a case, the concept of cell ought to be considered more as a mathematical description of the status of the system during some short time interval, rather than as a physical entity that can be fully exploited by natural reality. That is, convection can be a convenient mathematical tool for giving an expressive description of the system in some intuitively simple terms and during some given, and eventually comparatively short, time interval.

Within the magnetosphere, around the neutral sheet, conspicuous discontinuous earthward flows of particles (i.e. in the plasma sheet) were often observed. Hence, several scientists who carried out such observations started thinking that, perhaps, this was some kind of compensation for the particles that were supposed to be transported downstream just inside the magnetopause. Therefore, they began talking about convection within the magnetosphere, and such a «simple» concept for the human mind was appealing and rapidly became fashionable, much as had already happened for reconnection. However, it appears difficult to assess what definition among the three afore-mentioned ones they appealed to. In fact, it was not mass conservation. It was not an orthogonal function in a Hilbert space. Neither did it appear to be the beginning of some actual convection cell, the structure of which continuously changes in time, because the system experiences violent instantaneous variations by the impinging solar wind. Differently stated, if one wants to talk about convection within the magnetosphere, he can even do it correctly, provided that he specifies what he means by such a statement. To the author’s knowledge, however, such a specification was basically never clearly given within the literature (except, perhaps, within some conspicuous attempts at working out some formal numerical modelling). In general, such a concept was, rather, mentioned in terms of some intuitive feeling. In contrast, there are transient and more or less occasional situations within the magnetosphere that imply some closed patterns of plasma flow, that can be observed in terms of the projection of their respective effects over the polar caps (i.e. the so-called DP1 and DP2 features, that were quite a remarkable finding, see e.g., Nishida (1978) and references therein). But this is different from a steady large-scale convection within the magnetosphere.

Another warning is concerned with the definition of the average model of anything. For example, the average air temperature, or rain precipitation, etc. are badly representative of the climate of a site (in fact, you can have large deviations, such as either drought or heavy rainfall, hot summer and cold winters, etc.). When higher altitudes are considered, dealing with the middle and upper atmospheres, the percent deviations from average condition get even larger, and are further increased when entering into the magnetosphere and/or the solar wind. Therefore, one should be quite concerned, prior to dealing with any average model of anything: an average model of the magnetosphere normally is no actual physical status; it is, rather, some curious mathematical average among several physical states, every one of which is governed by Maxwell’s equations, etc., unlike their average, that per se has no need to satisfy any physical law, or constraint (e.g., Williams et al., 1992).
2. Modelling the magnetosphere

The description of the magnetosphere can be accomplished in terms of different observational quantities, such as, e.g., optical emissions (i.e. polar auroras), particle fluxes (i.e. radiation belts; James Van Allen’s school), field lines either of the magnetic field $B$ (Norman Ness’ school), or of the electric current density $j$, or of the electric field $E$, or of the plasma drift $v$, or e.m. waves such as pulsations or else, or energy densities, etc. A more extensive account is given elsewhere (Gregori, 1998a). The previous state of the art is highlighted e.g., by Akasofu and

---

**Fig. 2a,b.** The solar wind is here symbolically represented (without loss of generality) only by one loop of electric current $j_1$, and the source of the Earth magnetic field only by one loop $j_2$. Within a «drop» magnetosphere (a) $j_1$ flows all outside the magnetopause and it links no flux $\Phi(B_2)$ of the magnetic field $B_2$ generated by $j_2$. However, whenever some physical cause makes $j_1$ link as much $\Phi(B_2)$ as possible, such as occurs in (b), Hamilton’s principle states that, by this and only by this, stable equilibrium can be attained. That is, both cases of (a) and (b) can be considered physically possible and meaningful states of equilibrium, although (a) is unstable, while (b) is stable. After Gregori (1990). Figure and captions after Gregori (1998a).
Fig. 3. The earthward termination of the neutral sheet occurs wherever the kinetic energy density (or pressure) of the charged particles, that produce the $j$ electric current density (here shown by arrows and that are responsible for the «stretching» of the tail), balances with the magnetic energy density (or pressure) of the Earth magnetic field. That is, this is basically the same computation as for the magnetopause, with the sole difference that the magnetopause is a 3D geometrical locus, while this figure is 2D. Close to the Earth the former $j$ that flows within the neutral sheet is split, and becomes aligned with the geo-m.f.l.s (i.e. $j$ thus originates the «Birkeland currents»). The circuit finally closes across the ionosphere, either within the auroral oval or elsewhere. In the case of the Earth, the neutral sheet does not extend up to the front of the magnetosphere. In the case of Jupiter, however, when a suitable scaling of the involved pressures is taken into account, it can be shown that the neutral sheet affords to «go around» the planet (i.e. it extends even on the noon side); this is the explanation of Jupiter's «magneto-disk» in terms of Hamilton's principle, i.e. without appealing to centrifugal force (Gregori, 1998a). The discussion of planetary magnetospheres is to be given elsewhere. After Gregori (1990). Figure and captions after Gregori (1998a).
Chapman (1972), Nishida (1978), or Kamide (1988), or Williams et al. (1992), or Brekke (1997). The physical system and its phenomena are, however, a unique entity, that ought to be considered altogether. Neither should we attempt to provide with an excessively detailed description of what happens to every particle of the system, or at every specific site inside its geometrical space. Much as in the avalanche example, we should rather be concerned with some relevant large-scale features. This can be accomplished by considering the classical Hamilton variational principle. For applying it, let us first consider that any \( j \) must always flow within a closed loop, because

\[
\text{div } j = 0. \tag{2.1}
\]

By Hamilton’s principle, every such \( j \)-loop must link as much \( B \)-flux as possible. This implies that a drop magnetosphere (fig. 1) is highly unstable, in the fact that an eventual perturbation rapidly (at light speed) triggers a transformation according to the principle idea sketched in fig. 2a,b. To the author’s knowledge, this is the only physical explanation for the existence

**Fig. 4.** A model magnetosphere (for an «away» interplanetary sector), based on a 3D mapping of the electric current density \( j \), envisages the existence of three classes of \( j \) circuits, that are here schematically indicated (out of scale) as \( J_1, J_2, \) and \( J_3 \), respectively. A \( j \) loop of the class \( J_1 \) envelopes the northern lobe of the tail like a coil, or solenoid; \( J_1 \) flows downstream, counter-clockwise as seen from the Sun. Analogously, \( J_2 \), envelops the southern lobe like a solenoid; \( J_2 \) flows sunward, clockwise as seen from the Sun. The \( J_2 \) circuits look rather like closed orbits of particles, which are «trapped» and flow over the surface either of the neutral sheet or of the magnetopause. The \( J_3 \) circuit shown here is a simplified and symbolic version, its actual structure being depicted in fig. 5. In the case of a «toward» interplanetary sector, this model has to be changed accordingly, resulting indicatively in a model generated from the present figure, simply by symmetry with respect to the plane of the neutral sheet. After Gregori (1990). Figure and captions after Gregori (1998a).
of the neutral sheet. Its earthward termination is depicted in fig. 3, deriving from the fact that, close to the Earth, the pressure by \( j \) inside the neutral sheet is eventually contrasted by the pressure of the magnetic field of the Earth; hence, particles can no longer penetrate, and must spiral along m.f.l.s down over the ionosphere. In this way, the ionosphere becomes an essential segment of the electric circuits of the magnetosphere. In the final analysis, within the neutral sheet there occurs, in 2D, the same balance between kinetic and magnetic energy densities or pressures, that is responsible, in 3D, for the formation of the magnetopause.

Given one (well known and classical) model magnetosphere in terms of \( B \), it is possible by means of (1.1b) to provide a model magnetosphere in terms of \( j \) (fig. 4), that envisages three kinds of \( j \)-loops. Two of them envelop, like two coils, the two lobes of the tail. A third family of closed \( j \)-loops results into some new kind of trapped radiation (fig. 5). Moreover, when we
know the $B$ and $j$ model magnetospheres, it is also possible to derive the $E$ and the $v$ models. For details the interested reader ought to refer to Gregori (1998a). We cannot forget $E$, that can be originated by different mechanisms as follows: $SW$ (by which the $E$, that is embedded within the solar wind and transported by it, is projected along electrically-equipotential m.f.l.s into the magnetosphere); $IO$ (by which the ionosphere is the unique resistor within the entire system of $j$-loops; hence, some Ohmnic, or other, potential drop is generated across it, and project-
ed along the electrically-equipotential m.f.l.s into the outer magnetosphere); TP (by which the plasma inflow from the tail generates $E$ much like an MHD generator of electric potential); $P$ (by which front injection of plasma generates $E$ much like an MHD potential generator); and $S_q$ (it is the afore-mentioned ionospheric dynamo). The resulting pattern is synthesized in figs. 6a,b and 7a,b. Evidently, it is nonsense even to attempt to search for some average model for $E$ (this nonsense was a comparatively late finding by experimenters). In fact, even its basic geom-
Fig. 7a. Figure after Gregori (1998a). Electric field generated inside the magnetosphere by either the $IO$, or $TP$, or $FI$ mechanism (see text). Arrow symbols are as in fig. 6a,b. After Gregori (1991). Cross-section of the tail far downstream as seen from the Earth (out of scale). Dotted circles (not in the neutral sheet) denote $B$ pointing toward the reader, crossed circles denote $B$ pointing away from the reader. Dotted circles (within the neutral sheet) denote $v$ pointing toward the reader. It is $E$ (dawn to dusk) along the dashed lines (the outer one of which is here drawn as the magnetopause itself, but this is only a simple way of drawing the figure, just an assumption for clarity and simplicity purposes, as there appears to be no physical reason for such a coincidence). Such mechanisms result to imply a permanent earthward $v$, i.e. there is an earthward flow of particles within the neutral sheet (see fig. 7b).

Geometry dramatically changes, depending on the different situations that are eventually encountered, following the relevant plasma drifts that affect the entire system, by which either one of the aforementioned possible mechanisms temporarily overwhelms the others. Much like a zoologist, who cannot make an average between a fish and a bird, we must distinguish between different «animals» in our «zoo», if we want to be able to get rid in some way of the disturbances to our aeromagnetic surveys.

The Chapman and Ferraro (ibid.) theory for the sudden commencement of magnetic storms was already mentioned, being the consequence of an excess amount (compared to conventional steady flow) of solar wind impinging on the magnetosphere. Instead, one brief mention is needed to explain what happens on the occasion of a plasma cavity within the impinging solar wind. Relation (1.1b) cannot work on the front side of the magnetosphere, due to a lack of particle-supply for sustaining the steady flow that is strictly required by the Eulerian viewpoint. Hence, either (1.1b) does not hold (which is impossible), or the m.f.l. topology must change (that is, the system shifts to a reconnected configuration). In fact, the new physical system is completely different from the previous one: it contains far fewer particles, due to the plasma cavity. Such a lack of particles propagates downstream, intuitively (although improperly) almost like an air bubble within the water pipes that, in

182
the magnetosphere, can be likened to the j-loops depicted in figs. 4 and/or 5. Concerning the neutral sheet, it cannot remain stretched as before, because (1.1b) is no longer locally supported by the incoming solar plasma. Hence, there is a transformation of magnetic energy into kinetic energy of particles, and this originates the well known earthward flow within the plasma sheet, etc. The ultimate result is that the tail is repeatedly squeezed much like a toothpaste tube. In
this way, a geomagnetic storm results almost like a damped oscillation, every oscillation being a substorm. The ultimate fate is that, in a few days, the magnetosphere has re-adapted its configuration to the new parameters of the solar wind. Such an entire model could even be verified by an artificial trigger of a small substorm. The details are given in Gregori (1998a) and will not be repeated here.

Let us just note that, according to such a general interpretation, several pieces of observational evidence apparently have been for a long time (and still are) misinterpreted in the literature, as a support to the concepts of reconnection and/or convection (refer e.g., to Akasofu and Chapman, 1972; Nishida, 1978; Kamide, 1988; Williams et al., 1992; Brekke, 1997; and references therein). In the ultimate analysis, such concepts played an ad hoc role, very much like the epicycles in Ptolemy’s astronomy, that were not wrong per se as they only implied a simple change of reference frame; but, they were ad hoc assumptions for explaining the observed apparent motion of the planets within the previous logical frame, although they finally completely failed when Galileo observed the phases of Venus and of Mercury and the Jupiter satellites. Concerning the magnetosphere, there is no mystery, no new fundamental law, no unexplained feature (refer to the aforementioned authoritative statements by Akasofu and by Hultqvist): everything simply derives from the Hamilton-Jacobi variational approach, which is an alternative concept, it is classic, well known since the last century, and eventually much more pregnant and effective, for dealing with physical systems, compared to the microphysical approach typical of Newton’s principles and of Maxwell’s laws (that are used either from the Lagrangian or the Eulerian view-point).

3. Modelling external magnetic field sources over the polar cap

No mathematical model shall ever provide more physical information than what is intrinsically contained within its observational input. For example, it is customary in geophysics to compute numerical models of some given physical system, on the basis of a few assumed known fundamental physical laws, and of some very restricted set of available observations, that are fed into the model by means of its boundary conditions. The output of such modelling provides plenty of details, referring both to every minor corner of the physical volume of the system, and to some high time resolution. Such a deterministic approach is eventually suited and successful for representing some average status of the system, while it is completely unviable for depicting physical situations that rapidly change versus time. Such a last and difficult case is the concern of magnetospheric physics, where we must, rather, model every single specific situation from a much more pragmatic viewpoint, by which we can only deal with models that have a number of d.o.f.s that is never larger than the number of d.o.f.s of the actual available observational input. The distinction between such two basic viewpoints (i.e. deterministic versus pragmatic) is discussed by Gregori (1998b). Therefore, no average model, characterised either by one, or even by a few, indices, will ever afford any credible solution for our problem. We must deal with instant situations, and monitor how the different classes of geometrical or topological features and structures of j-loops, B, E, v, etc. within the polar cap magnetosphere do change in time. The detail of every instant situation is strictly related to the diagnostic tools that we are capable of monitoring, and of giving as input to our model. Therefore, the problem is shifted on what specific observations have to be monitored in order to fit them into our computation.

Figures 6a,b are characterized by a v or j pattern over the polar caps, that for a long time has been known as the Svalgaard vortex, a discovery that originally raised great enthusiasm, but later became unfashionable when it was realized that it was not a permanent feature. According to the present paper, the heuristic value of such a vortex is substantially reevaluated: it works fairly well, provided that one considers that when the configuration of figs. 7a,b occurs, no such vortex exists, and searching for it is even nonsense (as in zoology searching for wings on a worm).

In contrast, figs. 7a,b recall about the Sun-aligned auroral arcs inside the auroral oval (Θ
The external magnetic sources over the polar caps - Feasible modelling versus unrealistic expectations

aurora; see e.g., Kamide, 1988; Williams et al., 1992). Differently stated, whenever such Sun-aligned arcs are observed, no Svalgaard vortex should exist, and vice versa (this observational fact ought to be checked).

A huge cluster of different probes and satellites are presently crowding outer space, providing direct or indirect information on the local \( B \), \( E \), \( j \), \( v \), etc. The networks of the ground based geophysical observatories provide a space and time coverage that is sometimes insufficient, but that nevertheless can give important information. The ultimate target should be to fit such an entire database into a unique all-inclusive model that depicts, to some extent, the geometrical, physical, quantitative details of the entire system, including the information needed for cleaning our aeromagnetic records.

Akasofu’s substorm, Svalgaard’s vortex, and \( \Theta \) auroras, are all information concerned with the latitudinal/longitudinal distribution of phenomena, while they give no direct evidence dealing with the radial coordinate. However, some other evidence concerning this can be achieved by ground observations, in terms of the currently unexceptionable luminosity curve of polar auroras, \textit{i.e.} the relative distribution \textit{versus} height of the light intensity emitted by an aurora. Such an aspect that was historically investigated from the ground by Leiv Harang and Anders Omholt, depends on two concurrent factors: i) the energy spectrum of the impinging particles, that depends on the characteristics of every single event of particle precipitation, and different such events have a different statistical frequency, depending on solar or other activity; and ii) the density distribution of the atmosphere \textit{versus} height, that depends on a calorimetric effect, associated both with a several-year integral accounting for solar activity, and with the short time- and space-range that is concerned with the previous particle precipitation at every given location. Notwithstanding such drawbacks, including the difficulty (due to geometrical perspective) of measuring such luminosity curves from the ground, it seems possible to recognize a solar cycle dependence of such a phenomenon, although one should need a much better calibration, prior to implementing its results in a quantitative magnetospheric model. Refer to Gregori (1992) and references therein. Such observations, however, can now be much better carried out by limb scanning from satellite.

In general, two basic warnings ought to be borne in mind. The first deals with multidisciplinarity: monitoring only one physical parameter at a time, and attempting to envisage the underlying physics simply by means of formal data handling, mathematics, sophisticated numerical modelling, etc. for providing some average trend or structure, that is eventually characterized by either one or a few indices, that appear more or less correlated with geomagnetic indices or with sunspot number, etc. all this appears to be, for the purposes of the present paper, an ultimately hopeless effort. Neither should one hope to search for multivariate linear or non-linear regression analysis: in no case, will it conclude that a worm has wings. Every index, defined in any way, always deals with the morphological stage of knowledge, much like a medical doctor who uses a few indicators (such as body temperature, headache, heartbeats, etc.) to guess the health condition of his patient: to get rid of his problem he has to rely on physiology and pathology. A geophysicist must also understand the prime mechanisms that control his system, and he cannot rely only on some empirical index, even defined in terms of several variates. Several solar, ionospheric, and geomagnetic indices were proposed in the past (e.g., Perrone and de Franceschi, 1998), and two most useful ones are the \( D_s \) and the \( AE \) geomagnetic indices, because unlike the others they attempt at giving physical information, \textit{i.e.} a relative indication, respectively, of the ring current and of the current intensity flowing along the auroral oval, \textit{i.e.} along the auroral electrojet. However, Akasofu (1998) emphasises that «both the \( AE \) and \( D_s \) indices are products of the 1960s and their accuracy is very limited. A blind use of them could cause confusion. The present and future advanced study should devise better indices for the ring current and electrojets». For example, a few decades ago some discussion was concerned with the possibility of defining a polar cap index of auroral activity. Such an item, however, soon became obsolete when satellite images of polar caps were available that showed the instant position of auroras (in any
The historical controversy between the Birkeland-Alfvén \( j \)-system and the Chapman-Vestine \( j \)-system can be briefly summarised as follows. Birkeland (1908) and later on Alfvén (1939, 1940) formerly attempted to explain the geomagnetic storms and auroras in terms of a \( j \)-system involving currents aligned along m.f.l.s (called field-aligned or Birkeland currents), plus currents flowing within the ionosphere. In contrast, Chapman (1935) and Vestine and Chapman (1938) proposed a pattern composed only of currents flowing within the ionosphere. The latter proposal appeared to be preferred by most geophysicists, until the Alfvén school resumed it. The last great debate between the two schools occurred in September 1967, in Sandefjord (Norway) (Egeland and Hollet, 1968). Late in 1967, Naoshi Fukushima announced his classical argument in a short preprint (Boström, 1969, mentions it as a private communication), that later appeared in Fukushima (1969, 1972), and it is also referred to, by Berdichevskiy et al. (1972). Fukushima’s argument is now classic, and is concerned with a plane horizontal Earth, and with field aligned currents flowing along vertical lines. Vytenis M. Vasyliunas (according to Fukushima, 1972) generalised the argument to the case of a spherical Earth with field aligned currents along radial directions. The argument can be fully generalised to any geometrical configuration (Gregori, in preparation), by which it is concluded that all that can be inferred from ground-based observations is only that part of the magnetospheric disturbances that succeed in escaping the damping and shielding by the ionospheric currents. Differently stated, the ground-based observed effects are the difference between two current systems (i.e. the impinging one, minus that one induced into the ionosphere) that cannot be separated only by means of ground-based records.

Summarising, never expect to find some ultimate, simple, or easy-to-handle, numerical model, derived from a deterministic viewpoint. Rather, always search, from a pragmatic viewpoint, for physical instant-models, based on a number of d.o.f.s that never exceeds the number of d.o.f.s of the actual available observational input. By this, search for envisaging quantitative models for every instantaneous status of the system.
4. Conclusions

The aforementioned rationale appears to be the only realistic way of getting rid of the tremendous difficulties in modelling the external sources of the magnetic field observed over the polar caps. There appears to be no easy way of solving the problem, no «magic» index or procedure. Neither does it make sense to search for one. In some way, only common sense, and an extended use of interdisciplinary information, can solve the problem. Every time-instant has to be considered as dealing with a case history of its own.

One should hesitate in relying on simple indices (geomagnetic, solar, or else), unless one clearly understands the substantial underlying physical approximations and drawbacks. Mathematics, conventions, and opinion science never provided any physical insight into any problem. Fashionable ideas (like e.g., the concepts of reconnection or of convection) are certainly expressive (and eventually heuristically useful) for depicting observations. They are effective tools, as far as the morphological stage of knowledge is concerned. They are brilliant intuitions by scientists who had to organise a huge amount of observations that, prior to their guess, appeared to everybody almost random. At the time, their brilliant idea was useful for setting our knowledge in order. But, whenever those same (or other) authors were pulled, by their enthusiasm for the success of such ideas, to consider them either as an intrinsic, essential, although eventually non-understood, aspect of natural reality, or some kind of a new aspect of some still not-clear implications of the ultimate laws of nature, etc., in such a case the fact of relying on such concepts eventually became misleading, much like relying on the epicycles that were very well suited for describing the morphological features of the motion of the planets in Ptolemy’s astronomy, but that had no physical basis.

In the ultimate analysis, the modelling of the external sources of the magnetic field in the polar caps shall successfully proceed altogether with the international interdisciplinary integration of the different networks of observations, either ground-based or space-borne. The present paper wants to warn about the correct logical approach. For the time being, the ideal situation is to perform several repeated airborne surveys, and store their original records in suitable data centres. Their contribution will be important for the progress of our understanding of polar cap phenomena. For example, the discovery of the auroral substorm by Akasofu is one of the most brilliant scientific contributions of our century, based on objective observational inference, and derived through a tremendous effort by direct visual synthesis of a huge amount of all-sky camera pictures collected during the IGY (1957-1959): this is still within the classical heritage of the old fashioned way of doing science with no help from computers or automatic devices, etc. He provided an expressive and quite effective description of what happens over the polar caps during a magnetic storm. He gave a planetary scale description of phenomena, not simply a local monitoring of a microphysical aspect. His conclusion appears, so to speak, much more akin to Hamilton-Jacobi’s theory than to Newton’s or Maxwell’s. But, from the quantitative viewpoint, his discovery gave no immediate contribution. We can rather state that, according to Akasofu’s rationale, we can observe polar auroras (e.g., from a space platform), and by this we can recognize the instant position of the auroral oval. Hence, on an instant basis we know whether the airplane performing the aeromagnetic survey is either inside or outside the oval, and the status of evolution of the magnetosphere, i.e. whether it is more or less strictly quiet, or whether a substorm has been triggered and is expected to develop during the next 2 or 3 h, etc. From this we have an idea of where we should locate the polar electrojet, in order to give an instant model of its associated perturbation. Prior to such Akasofu’s finding, we had been completely incapable of even attempting at computing such information. Mutatis mutandis, similar comments also apply to the Svalgaard vortex, although it had a less lucky acceptance by the international scientific community, or to Θ auroras (also monitored by satellite), or to the luminosity curve of auroras, etc.

The conclusion of the present paper is that, much like Akasofu’s rationale for polar auroras, suitably applied and supported by space plat-
forms, other observations, such as the Svalgaard vortex, or the Sun-aligned auroral arcs, or the luminosity curve of auroras, etc., suitably interpreted according to a correct rationale, can effectively and practically provide an ever improving, actual, quantitative model of the external sources of the magnetic field over the polar caps. Such quantitative models can be more or less detailed and/or sophisticated, depending on the observational input that is fed into them. However, it is always fundamental that we avoid any basic conceptual misconception. The epicycle hazard is shared with ancient Ptolemy’s astronomy even by the most advanced present science: we must be conscious of this, and we should get rid of it.

REFERENCES


The external magnetic sources over the polar caps - Feasible modelling versus unrealistic expectations


NISHIDA, A. (1978): Geomagnetic Diagnosis of the Magnetosphere (Springer-Verlag, New York, etc.).


189