Chapter 2. Physical Processes

The previous chapter has considered the structure of near-Earth space and the interplay of that structure with the solar wind. In this chapter some of the physical processes operative in near-Earth space are discussed, mainly with the intention of placing in context the study of geomagnetic variations occurring with a timescale of a few minutes. Pulsational phenomena on shorter time scales are not considered in detail in this study and the physics behind them is thus also not discussed here.

a. Physical State of Near-Earth Matter

Above an altitude of roughly 90 km the tenuous gasses which fill near-Earth space have a significant degree of ionization which generally increases with height. The collision frequency is low, which does not favor recombination. On the other hand, there are abundant sources of energy for ionization, including solar photons and high-energy particles of magnetospheric, solar, and cosmic origin. At ionospheric height solar photons dominate in maintaining ionization on the dayside, but are not available to do so on the nightside. The auroral zones, where ionization is maintained by particle precipitation, have an ionospheric conductivity roughly equivalent to that on the dayside while other regions of the nightside have lower conductivity due to the lower ionization. The actual day/night ionization variation outside the auroral zone is roughly a factor of 100.

The most important aspect of the presence of ions is that the medium becomes conducting and subject to electromagnetic forces. The dynamics of such an ionized gas, or plasma, are more complex than those of a neutral gas. In lower regions of the ionosphere where there is still an appreciable fraction of neutral molecules, their interaction with ionized constituents may be important.

b. Magnetohydrodynamic Representation

Considerable simplification of the dynamical equations can be achieved if suitable averaged quantities can be defined and fluid-like equations used to determine their behaviour. The dynamical relations between the bulk or averaged quantities may be derived by averaging of the equations governing single particle motion which were discussed in the previous chapter. Perhaps the most widely used such simplification is magnetohydrodynamics (MHD). In this approximation, a set of dynamical equations is developed which govern macroscopic behaviour. Quantities may occur in these equations which are derived from a particle treatment. One such quantity is the conductivity, as described below. It allows a relation to be established between the macroscopic electric current and the electric and magnetic fields, but must itself be calculated from a more general theory than MHD.

c. Conductivity, Electric Field, Electric Current

The effect of an electric field applied to a conducting medium is to cause electric currents to flow in that medium. In an isotropic medium with mobile electrons, their acceleration opposite to the direction of the applied field is the primary source of current. The limiting factor on the magnitude of the current is collisions, whose frequency may be denoted ν. In the noncollisional regime of the magnetosphere these considerations are not applicable. However, in the denser ionosphere collisions are important and the effect of electric fields may be considered in a similar manner to that in, say, a metallic conductor. Due to the effects of the magnetic induction field through the Lorentz force, the medium is nonisotropic in its response to the applied electric field. The magnetic effect enters the conductivity expressions which follow in the form of the gyrofrequency $\omega_{cs} = \frac{q_s B}{m_s}$, where the subscript s denotes the species, either i for ions or e for electrons. This electric current due to the applied electric field **E** may be written as $\boldsymbol{J} = \sigma \cdot \boldsymbol{E}$ where σ is the conductivity tensor [see, for example, Parks, 1991, p. 283] with structure, in a system aligned with the magnetic field,

$$\sigma = \begin{bmatrix} \sigma_1 & -\sigma_2 & 0 \\ \sigma_2 & \sigma_1 & 0 \\ 0 & 0 & \sigma_0 \end{bmatrix}$$

where

$$\sigma_1 = \frac{nq}{B}\left[\frac{\omega_{ci}\nu}{(\nu^2 + \omega_{ci}^2)} - \frac{\omega_{ce}\nu}{(\nu^2 + \omega_{ce}^2)}\right]$$

is the Pedersen conductivity, effective in the direction of that component of the electric field which is perpendicular to the magnetic field ,

$$\sigma_2 = \frac{nq}{B}\left[\frac{-\omega_{ci}^2}{(\nu^2 + \omega_{ci}^2)} - \frac{\omega_{ce}^2}{(\nu^2 + \omega_{ce}^2)}\right]$$

is the Hall conductivity reflecting the magnetic field's deviation of accelerating particles (after a collision) and is effective in the direction perpendicular to both **B** and **E**, and

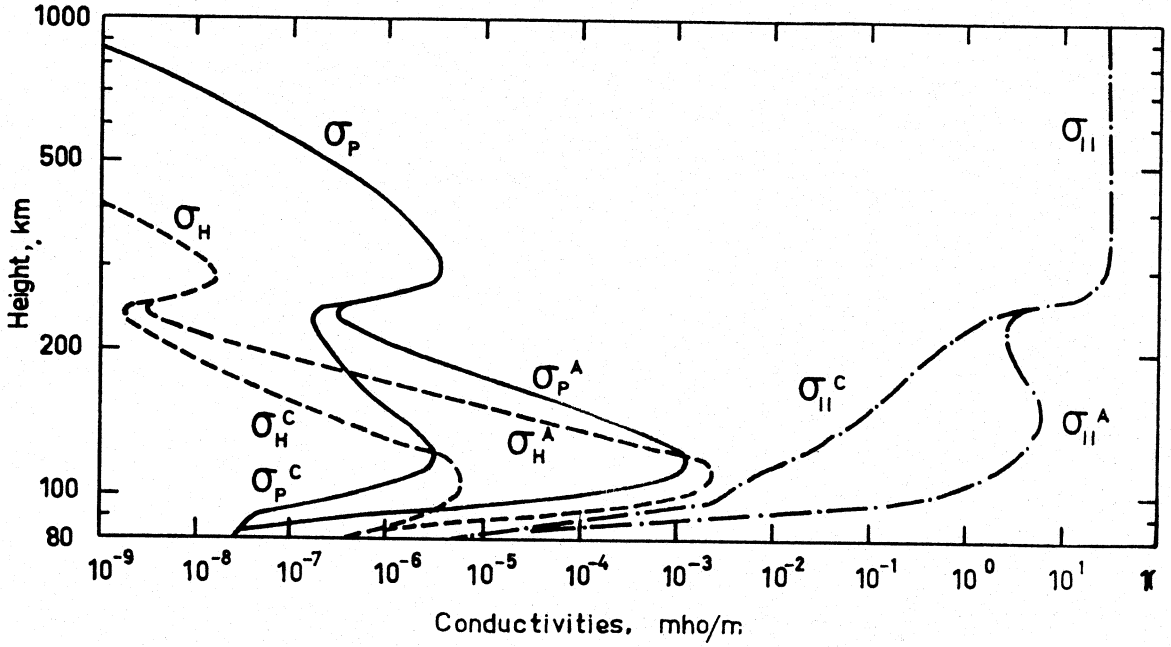
$$\sigma_0 = \frac{nq}{B}\left[\frac{\omega_{ci}}{\nu} - \frac{\omega_{ce}}{\nu}\right]$$

is the parallel conductivity, unaffected by **B**. Boström [1964] gives forms which are more correct in that the differing collision frequencies between the species (including neutrals) are taken into consideration. However, he notes that at all heights where there is significant conductivity, the Hall and Pedersen conductivities are determined

solely by the properties of the ions, resulting in slightly simpler forms for the above expressions. The parallel conductivity has no functional dependence on B since it is cancelled from the denominator by appearing in the numerator in the gyrofrequencies. It essentially depends only on quantities characterizing the electrons, which have great mobility parallel to the magnetic field.

These expressions may be evaluated by substitution of appropriate quantities in and outside auroral arcs, and are shown graphically in Figure 2.1. The conductivity structure has two peaks due to the interplay of number density and collision frequency. The conductivity within an arc is greatly enhanced due to the increase in number density there. This conductivity is strongly peaked at roughly 110 km, which is the height of ionospheric currents used in modelling in this study. Due to the strongly peaked conductivity distribution, the approximation is made in modelling that the currents flow in a thin sheet at 110 km height. Kisabeth [1972] studied the effect of slightly changing this height and concluded that the effects of varying it between 100 and 145 km were not significant, and further that a thin sheet approximation was adequate and that more precise modelling with integration over a layered current distribution was not warranted.

Figure 2.1 Conductivities within and outside an auroral arc. From Boström [1964]. Subscripts H, P, and || denote Hall, Pedersen, and parallel conductivities, respectively. Superscript A indicates results within an arc, C outside. Units of mho/m shown are equivalent to SI units of S/m (siemens per meter).



Since the approximation of a thin conducting sheet is applicable, it is convenient to define height-integrated conductivities Σ and current densities $\mathbf{j}$. The height-integrated conductivity is more properly called the conductance and has SI unit S (siemens). In the case that a two-dimensional Cartesian coordinate system is defined locally having the x

axis positive eastward and the y axis positive northward, Ohm's law in the ionosphere may be written [Hughes *et al.*, 1979] as

$$j_x = \Sigma_P E_x + \Sigma_H E_y$$
$$j_y = -\Sigma_H E_x + \Sigma_P E_y.$$

Formulation in a more general coordinate system [Lyons, 1992] but with the desire to separate effects of Hall and Pedersen conductivity, would result in the expressions

$$\mathbf{j}_P = \Sigma_P \mathbf{E}$$

and

$$\mathbf{j}_H = -\Sigma_H \mathbf{E} \times \frac{\mathbf{B}}{B}$$

which are equivalent to the above with $\mathbf{j} = \mathbf{j}_P + \mathbf{j}_H$. $\mathbf{j}_P$ and $\mathbf{j}_H$ are referred to as (height-integrated) Pedersen and Hall current densities, respectively, with units Am^{-1}.

Indirect observations of the conductivity may be made if the electric current and electric field can be observed simultaneously. This is one motivation for the present approach to measuring currents, since the electric field may be observed remotely by radar or *in situ* by satellites [Kamide, 1988, Section 2.3]. The ratio of Hall to Pedersen conductivities may itself be of interest and can be measured from satellite data under certain assumptions. For example, Sulzbacher *et al.* [1982] used magnetic measurements from the ground and from the TRIAD satellite, along with the assumption that the ionospheric electric field was exactly southward, to deduce that ratio, finding a value which varied from 4 at the northern boundary of the electrojet to about 0.8 near the southern border. In many existing routines for inversion of ground data the conductance plays a crucial role (see Chapter 4). In the method developed in this study, the conductance has no direct role, but could be derived from the modelling results if supplemented by measured electric field.

d. Induction Effects

Faraday's law describes the induction of an electric field as the result of a time-varying magnetic (induction) field: $\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$. In space physics, there are two important low-frequency results of this effect in conductors. Within the conductor, currents flow, and the entire set of relevant equations gives rise to wave solutions [Kraus and Carver, 1973, Section 15-10]. These Alfvén waves are similar to the familiar electromagnetic waves but arise through the interplay of electric currents with the plasma (through the $\mathbf{J} \times \mathbf{B}$ force) rather than displacement currents. They are involved with higher frequency phenomena, such as pulsations [Samson, 1991], than are considered here and will not be discussed further. The second induction effect is important, in practice, in regions outside the plasma. This is induction of currents in the Earth or other conductors which may be near a point of observation.

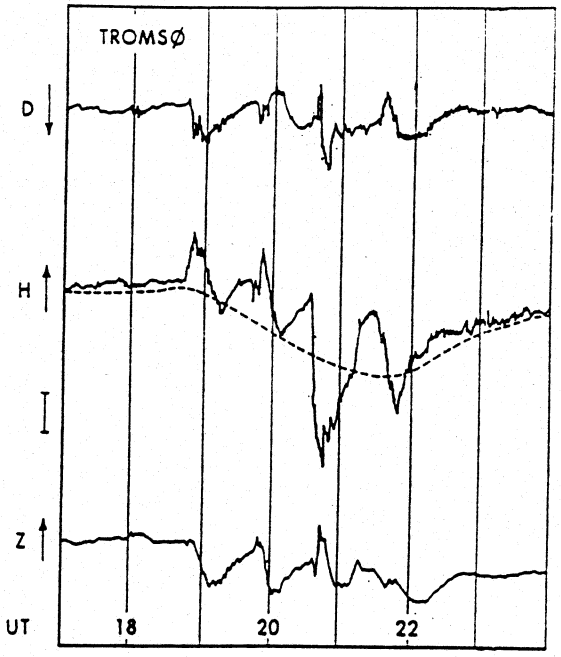
In practice, the induction of currents in the Earth is important due to the proximity of these currents to ground magnetic observatories. In cases of nonuniform ground conductivity, there may be no effective way to incorporate the induction effect, and a nearby station would be considered 'anomalous'. Such cases must be treated individually, and are, in the chapters presenting modelling results. Induction in a uniform Earth also occurs, and is incorporated into the modelling routine as described in Chapter 4.

It has been noted in the previous chapter that induction localized in regions of the magnetosphere may be responsible for particle energization. By Faraday's law, temporal variation of the magnetic field leads to spatial variations in the electric field. If the geometry is appropriate and temporal variation rapid enough, large electric fields and thus forces on charged particles may result. The role of induction is intimately tied, because of the involvement of a time derivative, to the time scale of the event under study. We now examine the relevant time scales for substorms.

e. Substorm Time Scale

The gross physical effects of substorms which are studied here take place on a scale of minutes to hours as shown in Figure 2.2. The short period (high frequency) cutoff is dictated by the desire not to study pulsational phenomena (although they may be of great importance) and by the availability of data with one minute temporal spacing.

Figure 2.2 Prototypical relation of polar magnetic substorms to geomagnetic bay. Major deviations in the H component are spaced by about one hour. Scale bar at left indicates 53 nT in H, 48 nT in D, 72 nT in Z. From Rostoker [1969].

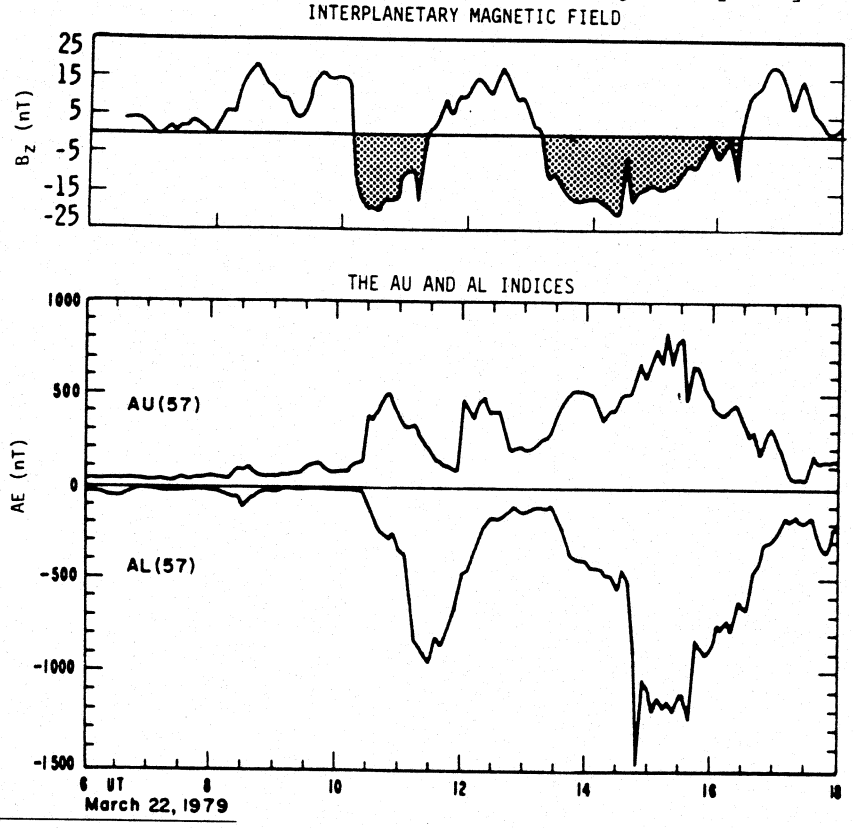


Rostoker [1969] pointed out the existence of two basic time scales in the gross development of auroral current systems. At the time there was controversy about whether the overall layout of such currents was as a one-cell or two-cell equivalent current system. Rostoker identified the one-cell system as being responsible for polar magnetic substorms, with a time scale typically in the tens of minutes. Longer period geomagnetic bay variations were attributed to the two-cell system. Figure 2.2 shows the prototypical case of a sequence of substorms with associated longer-period geomagnetic bay. These data are from Tromsø at corrected geomagnetic latitude of 66.3° N. H perturbations at stations further north (70° to 75°) were found to be consistently negative indicating near-overhead passage of the electrojet associated with the current wedge. Weak H signatures at 19 UT and 20 UT along with negative Z at Tromsø confirm this. Stronger negative H signatures near 21 UT and 22 UT, along with weak and transitional Z, indicate that these disturbances occurred nearer to Tromsø, that is further south than the previous activity. The dashed line indicates the long-period geomagnetic bay associated with this activity. The activity of the two systems is now recognized to be possible simultaneously, with the two-cell system often referred to as the 'driven' system, and the one-cell system as the 'substorm current wedge'. The latter was introduced in Chapter 1 and some of its observational aspects discussed there. As the three-dimensional nature of the auroral zone currents is now firmly established (see introduction to Chapter 5), reference is not usually now made to the cell nature of the equivalent current systems. However, the observation that there are two time scales involved in substorm activity remains valid and is taken to reflect differences in the physical mechanisms or the spatial scale of the regions involved in causing the currents of the two systems.

In ensuing years there has been much discussion of substorm time scales and possible physical implications in terms of overall magnetospheric processes. Discussion is provided by Rostoker [1994, 1995], who points out further that use of certain magnetic indices (in particular AE: see section 3.a.1.b) can obscure the substorm time scales by combining effects from the driven and wedge systems. Further, relating ground magnetic data to satellite or other data can be difficult if use of the index causes timing discrepancies. If activation of the driven system is rapid enough it may be confused with an expansive phase onset if only an index is used for onset timing. In fact, driven system activity may involve near steady-state dissipation of energy in near-Earth regions while there is 'storage' of energy in the tail as discussed in the first chapter. The expansive phase onset would then correspond to 'unloading' of stored tail energy in a not completely understood manner, triggered by an unknown mechanism. In this sense the driven system activity reflects an aspect of the substorm process which may be most readily distinguished during the 'growth' phase.

The triggering mechanism for substorm expansive phases remains in dispute and indeed may be nonunique as are many effects in the domain of chaotic dynamics. Figure 2.3 illustrates the relationship, for the CDAW 6 March 22 1979 substorms[1], between IMF B_z and AU and AL indices computed using 57 stations rather than the normal 12. The use of many stations should in principle allow better timing of expansive phase onsets. In this case driven system activity is seen from 1030 UT on as manifested by a decrease in the AL index and an increase in AU. Shortly after 1100 there appears to be an onset as shown by a steep drop in AL and a period during which AU decreases. Again at approximately 1330 UT this pattern repeats on a larger scale. The general relationship of substorms to southward turning of the IMF is clear; in both cases the driven system activity commences, after a short delay, in synchrony with a southward turning. The expansive phase onsets, however, appear to be more closely related to northward changes in the IMF. This is particularly clear in the case of the very sharp and strong onset at 1450 UT, which follows an impulsive decrease in the magnitude of southward GSM B_z which appears to mark the start of a fairly gradual northward turning. Extension of this prototypical case to general behaviour appears to be unwarranted: northward turnings may trigger some onsets but do not appear to trigger all. This question will be considered in detail in what follows.

Figure 2.3 Comparison of IMF B_z and Auroral Electrojet indices. The relationship of IMF southward B_z to activity is clear, and northward variations in the IMF appear to trigger expansive phase activity. From Kamide and Baumjohann [1985].

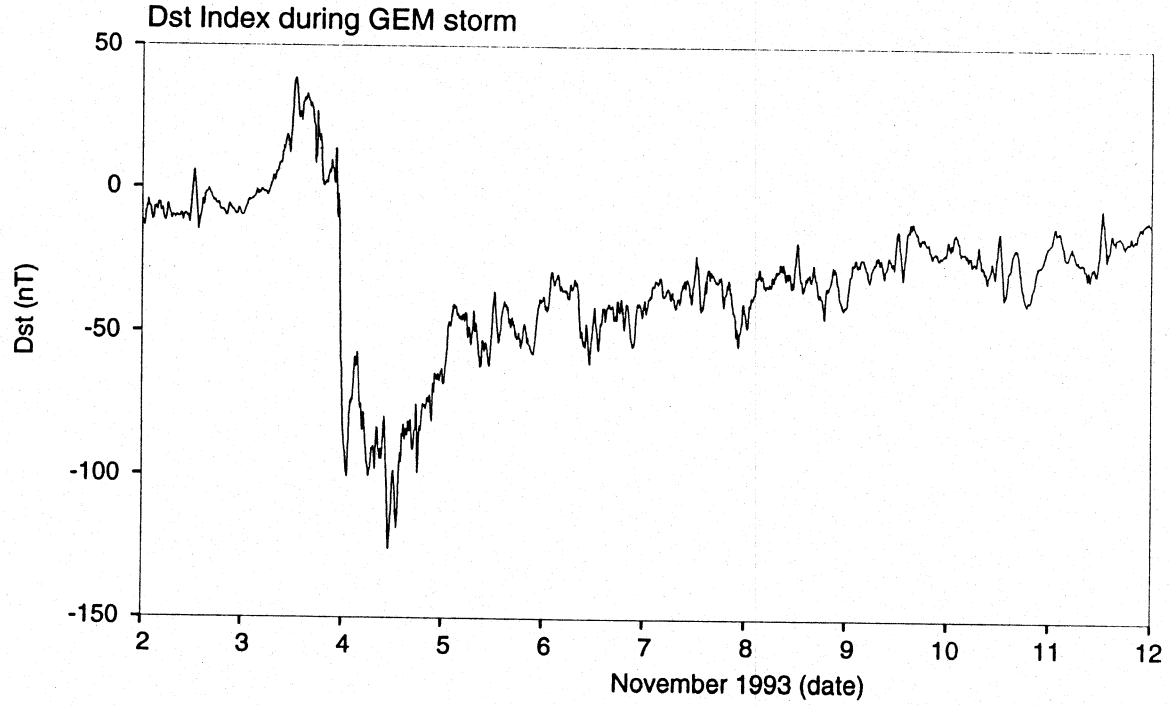


[1] This substorm sequence is often regarded as prototypical and has been described extensively, notably in a special section of the Journal of Geophysical Research (Volume 90, Number A2, February 1, 1985, pages 1175-1374).

In a prefatory article about the CDAW 6A substorm, McPherron and Manka [1985] emphasize that the major question related to substorms is the relative importance of energy dissipation directly driven by the solar wind and that driven by release of energy stored in the geomagnetic tail. This question remains largely unanswered. It suggests that attempts to model magnetic activity should include current systems analogous to the driven system and to the substorm current wedge, so that the behaviour of each system may be examined separately.

Another relevant substorm-related timescale is readily observed in the horizontal component at mid to low latitudes. The D_{st} index, derived from low-latitude but not equatorial stations, is generally depressed during magnetic storms for periods of up to several days and this has been shown to be due to the growth of the Earth's ring current [Nishida, 1978, Chapter 4]. A similar but shorter-lived phenomenon occurs during substorms, possibly giving rise to 'partial' or 'asymmetric' ring currents not extending fully around the Earth [Nishida, 1978, p. 127]. Certainly cases are known in which particles, after a substorm injection, display 'echoes' [Lopez and Baker, 1993], so that it is not necessary to consider that injections form only partial ring currents, but rather they contribute to the global ring current [Liu and Rostoker, 1995]. Figure 2.4 illustrates the behaviour of the D_{st} index during an active period, that of the GEM study in early November 1993. One particular onset during this period is treated in detail in Chapter 10.

Figure 2.4 Behaviour of D_{st} Index during a period of storm activity. Initial phase with positive values is followed by the main phase with negative values. Data are from November 1993 and an event from this period is discussed in Chapter 10.



The injection of particles during substorms was discussed in Chapter 1. Such injections can contribute to the ring current as the energized electrons drift eastward and the accompanying ions drift westward, their differential motion causing a current. The initial energization of the bulk of ring current particles may be due not solely to injections, but also partly to non-adiabatic processes occurring during compression of the magnetosphere at the time of storm onset. In any case, D_{st} may be seen in the figure to decline rapidly after an initial phase due to magnetospheric compression. The continued (for hours or days) low values of D_{st} are diagnostic of a storm: the decay of the ring current appears to be a slow process in the essentially collisionless plasma of the near magnetosphere. Impulsive decreases seen in D_{st} may be due to later injections associated with substorms during the storm, which revitalize the ring current. From a modelling point of view, the worldwide depression of the horizontal component must be included in the forward model if mid or low latitude stations are included. However, the generally small D_{st} values during a typical substorm episode are easily overwhelmed by noise at auroral zone stations. Attention must be paid to incorporation of lower latitude stations if a D_{st}-like parameter is to be extracted from the modelling. In global modelling, a D_{st}-like parameter can be determined, but in regional (auroral zone) modelling it should be supplied as an input parameter or may be neglected. These issues are further discussed in Chapter 5[2].

An important consideration in discussion of substorm time scales is the applicable physics, briefly alluded to above. Although Gauss' law $\nabla \cdot \mathbf{B} = 0$ is universally true and has no implications for time changes of the magnetic induction field **B**, Ampère's law in the general case [Slater and Frank, 1949, p. 85] is $\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} + \mathbf{J}$ where **H** is the magnetic intensity, **D** the electric displacement, and **J** the electric current density (A m^{-2}). In a plasma, bound charges are unimportant and the constitutive equation relating **H** and **B** is simply $\mathbf{B} = \mu_0 \mathbf{H}$, where μ_0 is the vacuum permeability ($4\pi \times 10^{-7}$ H/m). However, free charges do move in response to applied fields, giving rise to polarization currents and an effective dielectric constant (in fact dielectric tensor) ε so that the electric field and electric displacement are related by $\mathbf{D} = \varepsilon \mathbf{E}$. If time variations are fast enough, the polarization currents, even in the usual case of a near-neutral bulk plasma, may be large enough that the electric displacement term becomes large compared with the local current density. In this case the situation is analogous to that prevailing in free space, where displacement currents exist although there is no actual current. That situation allows electromagnetic radiation (waves) to occur. In the situation where polarization currents become important in a plasma, wave phenomena may similarly occur. Although an entire rich literature exists about plasma waves, and although they are of importance in studying the magnetosphere [Samson, 1991], from the point of view of modelling in this study it is desirable to use a simpler form of Ampère's law in which time variations may be neglected and in which wave phenomena are not encompassed. As discussed in Section 4.a.1, in this case there is a unique relationship between currents in all of space and the magnetic induction at any given

[2] Using low latitude perturbations, it is also possible to define an asymmetry parameter A_{sym} [McPherron, 1995], which might be of use in separating intensified tail current effects from those of the ring current. A forward modelling study to do this would likely be a feasible project.

point, given by the Biot-Savart integral. The boundary between magnetostatics and magnetohydrodynamics in the magnetosphere is generally taken to be at periods of tens of seconds. Since wave phenomena often are periodic, it is also the case that averaging over wave periods will effectively remove them from the data considered. For the most part, the work done here is done with one-minute averaged magnetic induction values. It will thus be considered that the Biot-Savart integral effectively relates the observed values to the currents (everywhere in space) which generated them.

When magnetic data with fast time sampling (i.e. at intervals of less than one minute) are discussed here, it will generally be in the context of the use of pulsations as signatures of characteristic events. The use of the damped low-frequency (5-15 mHz) Pi 2 pulsations as signatures of substorm onset is notable in this regard (again see Samson [1991], and also Harrold [1986]). Since data with one minute or larger sampling are used in the modelling, and since stations are generally needed in as many locations as possible, pulsation data, when available, were time averaged to remove pulsations and used along with other data.

We now proceed to a discussion of what types of data are in fact available, how they are obtained, and the coordinate systems which are relevant to their processing.