VERLAG DER TECHNISCHEN UNIVERSITÄT GRAZ

Handbook of the Lower Ionosphere

Martin Friedrich

Martin Friedrich

Handbook of the Lower Ionosphere

Verlag der Technischen Universität Graz

Cover Photo: Launch of the rocket 41.094 on October 11th, 2011, at 21:15 UT from Andøya carrying a CHAMPS payload.
Photo by Kolbjørn Dahle, Andøya Space Center.

Textsetting: Martin Friedrich

Cover Design & Layout Editor: Nino Bijelić

Printed by Medienfabrik Graz (www.mfg.at)

www.ub.tugraz.at/Verlag

Print:
ISBN: 978-3-85125-485-3

E-Book:
ISBN: 978-3-85125- 486-0
DOI: 10.3217/ 978-3-85125-485-3

About the Author

Martin Friedrich was born in Graz in 1945. In 1963 he enrolled at Graz University of Technology in the field of electrical engineering graduating with a Dipl.-Ing. in 1969; in the same year he was employed at the then newly established Department of Communications and Wave Propagation. He obtained the PhD in 1974 and the habilitation (professorial qualification) in 1985. Since October 2010 he is formally retired, but still participates in several sounding rockets.

Abstract

This Handbook is not so much a classical textbook, than a guideline for further research in mesospheric physics with emphasis on experimental methods. But theoreticians may also profit from it by being made aware of problems existing in current theoretical models of the lower ionosphere.

Zusammenfassung

Dieses Handbook ist weniger ein Lehrbuch im üblichen Sinn, als vielmehr gedacht als Anregung für weitere Forschung, wobei das Hauptgewicht auf experimentellen Untersuchungen liegt. Aber auch für theoretische Studien kann es von Nutzen sein, da mancherlei Schwächen der Modellierung der Ionosphäre erkennbar werden.

CONTENT

Foreword

This write-up is a very personal/biased summary of the Institute's activities in upper atmospheric research. It all began in April 1969 when Prof. W. Riedler "acceded" to the chair of the then newly established Institute of Communications and Wave Propagation. As a legacy of his scientific activity at the (then) Kiruna Geophysical Observatory up to that date, he was invited to participate with an instrument in a sounding rocket to be launched later that same year. In retrospect this was a challenge and a unique opportunity to demonstrate our ability to provide space-qualified hardware under deadline pressure. Fortunately, the instrument, a three-receiver Faraday experiment combined with a slant-range receiver, performed as expected. As a result, we were invited to contribute with this instrument - later also with others hardware - in many more sounding rockets. The first of these rocket payloads were all integrated at the Norwegian Defence Research Establishment (today known by its Norwegian acronym FFI = forsvarets forskningsinstitutt), later invitations to contribute with plasma diagnostic instruments came from institutions in Germany, Sweden and the United States. The main experiment for which we gained a reputation, and eventually almost monopoly status, is the Faraday experiment, a radio wave instrument, the one best suited to measure electron densities in the lower ionosphere. The original idea was not ours, but our perseverance in continuing this particular measurement method led to a quasi-monopoly; the predecessors - and initial competitors - were notably the University of Illinois at Urbana-Champaign, the (erstwhile) Radio and Space Research Station in Ditton Park near Slough (now the Rutherford-Appleton Laboratory, Chilton), the Danish Space Research Institute, and the Uppsala Ionospheric Observatory. These institutions discontinued this type of rocket borne measurement, not so much because knowledge of the electron density was no longer a worthwhile topic, but primarily because the driving persons behind it had retired.

As mentioned above, the main altitude region where the Faraday experiment is at its best, is the lower ionosphere. Consequently, the main research emphasis is on the mesospheric ionosphere (aka the *D*-region). Another good reason for restricting our activities to these lower altitudes is that up to, say 150 km the time constants are short enough that steady-state can generally be assumed for most processes. Moreover, the world above 100 or 150 km is covered by decades of ionosonde measurements and many years of satellite data. To compete with this mass of data from a plethora of scientists and research institutions would have been an invidious task.

The initial trigger for this write-up was the stay of Dr. Lee-Anne McKinnell, a post doc fellow from South Africa at our institute (2003/4). Her stay led to the development of the Ionospheric Model of the Auroral Zone (IMAZ), using both data from EISCAT and from rockets from the auroral zone. She introduced us to the Neural Network approach for dealing with vast numbers of data with *a-priori* unknown behaviour (*cf.* Section 4.2.3.1). To track all stages of the development of IMAZ the various steps were documented, which eventually led to this review covering essentially all our ionospheric activities.

In order to be able to come to an end with the various chapters, there is an Outlook after each section listing things that can, or might be done to improve, should anyone want to continue these activities.

Martin Friedrich (editor),

Graz, October 2016

1. Introduction

The availability of satellites has deprived the ionosphere of most of its importance for communication; only in a very few cases does knowledge of its behaviour still matter for radio wave propagation. One such case is satellite navigation (*e.g.*, the Global Positioning System GPS or the future Galileo system) where the group delay caused by the ionosphere constitutes the single largest uncertainty. The absorption, which occurs in the lower ionosphere, is no longer a communications topic due to the use of much higher signal frequencies.

However, a good knowledge of the electron densities and other bulk parameters of the plasma in the mesosphere (*D*-region) is important for assessing the quality of theoretical models of the upper atmosphere. Notably the minor constituents NO and O, as well as temperature have such a significant impact on the plasma that their behaviour - or indeed their climatological models - allow for boundaries to be established for these parameters. Ultimately the results of empirical and theoretical models should converge demonstrating that all relevant processes are understood and accounted for. Or more provocatively expressed: If an atmospheric model can't properly predict plasma parameters, it is legitimate to question its ability to predict potentially more important outputs such as ozone or temperature. A plea for this aspect of ionospheric research was made by Friedrich (2004).

2. Waves in a Magneto-Plasma

The refractive index of plasma with a superimposed magnetic field has the following properties for radio waves:

a) complex, *i.e.* both retarding *and* absorbing,

b) dispersive, *i.e.* frequency dependent, and

c) dual, *i.e.* two characteristic waves (propagation modes) exist.

The theory, usually named in honour of Sir E.V. Appleton and D.R. Hartree, in a closed form describes the complex, dual refractive index $n = \mu + j\kappa$ as a function of (angular) signal frequency ω, electron density N_e, the magnetic field strength $\boldsymbol{B}$, its angle θ to the propagation vector, and the frequency of collisions ν between electrons and "stationary" particles (*i.e.* mostly neutrals). In this "classical" theory, the refractive index is given by:

$$n^2 = 1 - \frac{X(U-X)}{U(U-X) - \frac{Y_T^{\,2}}{2} \pm \sqrt{\frac{Y_T^{\,4}}{4} + Y_L^{\,2}(U-X)^2}} \qquad (2.1)$$

with the following notations:

$$X = \left(\frac{\omega_p}{\omega}\right)^2 = \left(\frac{f_p}{f}\right)^2 \quad \text{... (plasma frequency/signal frequency)}^2$$

$$Y = \frac{\omega_c}{\omega} = \frac{f_c}{f} \quad \text{... gyro frequency/signal frequency}$$

$$U = 1 - jZ$$

$$Z = \frac{\nu}{\omega} \quad \text{... collision frequency/(angular) signal frequency}$$

using:

$$\omega_p = \sqrt{\frac{N_e e^2}{\varepsilon_0 m}} \quad \text{...(angular) plasma frequency}$$

$$\omega_c = \frac{|e|}{m} B \quad \text{... (angular) gyro frequency of electrons}$$

$Y_T = Y\sin\theta$ ... magnetic field component transverse to the propagation direction

$Y_L = Y\cos\theta$ ... magnetic field component in propagation direction

Θ ... angle between the magnetic field $\boldsymbol{B}$ and the propagation direction

e ... electron charge

m_e ... electron mass

The above equation has two solutions for the complex refractive index, in absence of a magnetic field (or for large frequencies, *i.e.,* $f >> f_c$), the two solutions converge. Loss of energy (absorption) in this theory is implicitly accounted for by a "friction" term of the form

$$F = \nu v \qquad (2.2)$$

where ν is the collision frequency and v the velocity between electrons and the collision partners. For negligible v, $Z \to 0$ and $U \to 1$, and the imaginary component of the refractive index vanishes.

The "friction" term ν in Eq. (2.2) can be described as

$$\nu = \sigma n v \tag{2.3}$$

where σ is the collision cross section. Laboratory measurements by Phelps and Pack (1959) indicated proportionality (by a factor β) of the collision cross section with velocity v for N_2, the most common constituent of the Earth's lower atmosphere. Inserting this finding, *i.e.* $\sigma = \beta v$, in Eq. (2.3) one obtains

$$\nu = \beta n v^2 \tag{2.4}$$

According to the equation of state of an ideal gas, the number density n is related to pressure p and temperature T by

$$n = \frac{p}{kT} \tag{2.5}$$

k ... Boltzmann constant

The thermal velocity of particles in a gas is

$$v_{th} = \sqrt{\frac{2kT}{m}} \tag{2.6}$$

Because the electron rest mass m is orders of magnitude below that of any other constituent of the air, we may safely equate the velocity v in Eqs. (2.2) to (2.4) to the thermal velocity of electrons, *i.e.* we consider neutrals or ions to be stationary. By inserting Eqs. (2.5) and (2.6) into Eq. (2.4) we thus obtain the following relation for collision frequency

$$\nu = \beta \frac{p}{kT} \frac{2kT}{m} = \frac{2\beta}{m} p = Kp \tag{2.7}$$

where K, *i.e.* a proportionality factor between pressure p and collision frequency ν, replaces $2\beta/m$. This finding, together with an assumed Maxwellian velocity distribution, led to the development of the generalised magneto-ionic theory according to Sen and Wyller (1960). In that theory the terms U, X, Y, and Z are still used, but are defined by more complex relations.

Aggarwal and Setty (1980) compiled (laboratory) collision cross sections σ of various atmospheric gases. Using the ground level mixing ratios of air - which should be valid beyond 100 km - their data still suggest a linear relation between σ and thermal velocity, however with a finite value at zero velocity (or zero temperature, Figure 2.1). The relation of cross section σ and thermal velocity v_{th} is found to be:

$$\sigma = 6.713 \cdot 10^{-21} + 2.2250 \cdot 10^{-25} v_{th} \tag{2.8}$$

or expressed as a function of temperature

$$\sigma = 6.713 \cdot 10^{-21} + 1.238 \cdot 10^{-21} T^{0.5} \tag{2.9}$$

The magneto-ionic theory which incorporates this more detailed relation was developed in the MSc thesis by Spöcker (1984), published in a preliminary form by Friedrich *et al*. (1991) and is here termed ESW (= extended Sen and Wyller). One consequence is that now collision frequency is no longer proportional to pressure as in the Sen and Wyller theory, but also depends on temperature. We will illustrate the various degrees of complexity of the magneto-ionic theories by some sample calculations.

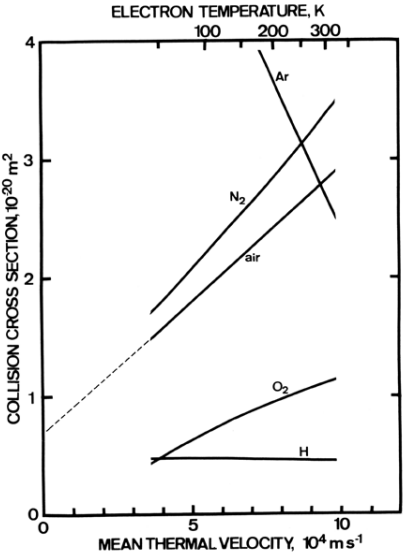


Figure 2.1 Laboratory collision cross section of the air's major constituents as a function of thermal velocity (Friedrich, 1985; data from Aggarwal and Setty, 1980). Note that the extrapolation of the cross section for air (dashed line) leads to a finite value a zero temperature.

In the paper by Sen and Wyller (1960) the difference between the "classical" Appleton-Hartree (AH) theory and the then new one (SW) was demonstrated by a dozen figures; we will here only show two of them, but with the ESW results added. Figure 2.2 displays the dispersion curve of the real and imaginary parts of the refractive indices (ordinary, o, and extraordinary, x, respectively). The calculation is carried out for strictly longitudinal propagation ($\theta = 0°$), a collision frequency ν half of the angular frequency of the signal ($\omega/2$) and a gyro frequency f_c of 1.4 MHz. As indicated above, in the ESW theory one also needs to specify the temperature; we here use 150 and 300 K to cover the extremes likely to be encountered in the lower ionosphere (mesosphere). The variation of the electron density is displayed by $X = (f_p/f)^2$, the squared ratio between plasma and signal frequencies. The ESW results are very close to the SW curves, and always slightly displaced in the direction of the AH results. Intuitively this makes sense, because the constant term in the collision cross section (Eqs. 2.8 and 2.9) corresponds to an inverse dependence on velocity as implicitly assumed in the AH theory. Varying the temperature only has a minor effect. Figure 2.3 shows the same conditions as before, but for transverse propagation ($\theta = 90°$).

Generally, the variation of the field strength E of an electro-magnetic wave experiencing a complex refractive index $n = \mu + j\kappa$ in a medium, can be described as a function of time t and distance z as:

$$E = E_0 e^{j\omega(t-\frac{n}{c}z)} = E_0 e^{-\kappa\frac{\omega}{c}z} e^{j\omega(t-\frac{\mu}{c}z)} \tag{2.10}$$

With the same components as used in Eq. (2.1) one can express the polarisation R as:

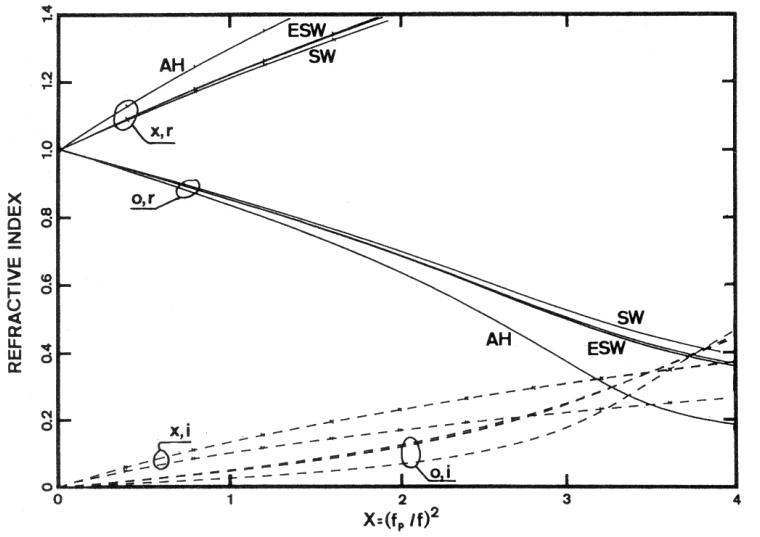


Figure 2.2 Real and imaginary parts of the refractive index (r, i) as a function of X. Indices o and x refer to the ordinary and extraordinary wave components, respectively. The conditions are the same as in Figure 2 of the publication by Sen and Wyller (1960), i.e. longitudinal propagation, 1.4 MHz gyro frequency and the collision frequency at half the angular signal frequency. AH, SW, and ESW refer to the results of the Appleton-Hartree, Sen and Wyller, and Extended Sen and Wyller theories, respectively (Friedrich et al., 1991).

$$R = \frac{-j}{Y_L}\left(\frac{\frac{1}{2}Y_T^2}{1-X-jZ} \mp \sqrt{\frac{\frac{1}{4}Y_T^4}{(1-X-jZ)^2} + Y_L^2}\right) \tag{2.11}$$

For strictly longitudinal propagation ($Y_T = 0$, $\theta = 0°$) R has two solutions $R_o = j$ and $R_x = -j$, *i.e.* the two characteristic waves are circularly polarised rotating in opposite directions. The other extreme situation is propagation transverse to $\boldsymbol{B}$, *i.e.* $Y_L = 0$, $\theta = 90°$. In that case the two polarisations are $R_o = \infty$ and $R_x = 0$, *i.e.* linearly polarised parallel and perpendicular to the magnetic field, respectively. In the general (intermediate) case $0° < \theta < 90°$ the characteristic waves will be elliptically polarised.

The ionosphere's refractive properties change an HF wave transmitted from the ground to the flying payload as function of electron density and collision frequency between transmitter and receiver. Differentiation with respect to altitude yields electron density and collision frequency. One can make use of one or more of the components of the refractive index n, *i.e.* the real part(s) μ, the difference between the real parts, the imaginary part(s) κ and the difference of the imaginary parts, or combinations of the above.

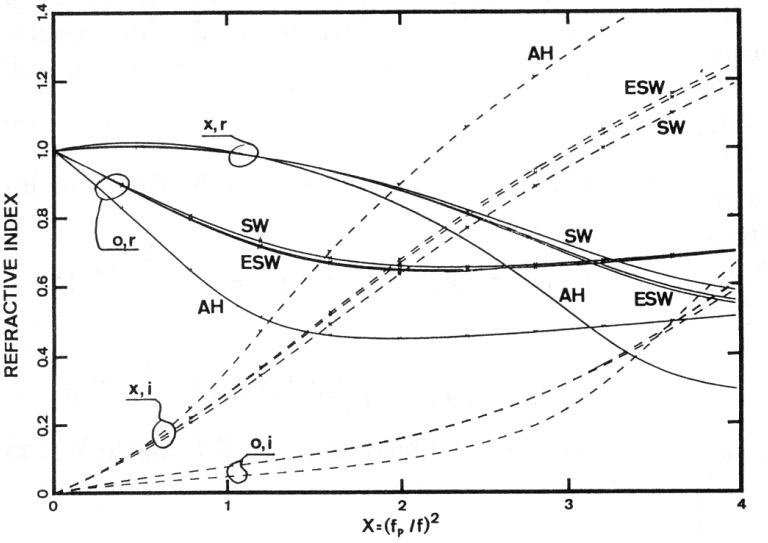


Figure 2.3 As Figure 2.2, but for transverse propagation (Friedrich et al., 1991).

2.1 Using the Real Part

2.1.1 Seddon Experiment

Since the real part μ of the refractive index is a function of electron density, the measurement of the phase velocity $v = c/\mu$ can be used to derive electron densities. The first rocket borne radio wave propagation experiments aimed at measuring electron densities made use of this effect. The simplest way to determine the phase velocity is to measure the phase difference between an electro-magnetic wave (largely) unaffected by the ionosphere (at a high frequency) and a signal of coherent lower frequency expected to be influenced by the total electron content (TEC). In the earliest such experiments two harmonically related frequencies were used (*e.g.*, 24.53 and 73.60 MHz), whereof for the higher one $\mu = 1$ was assumed, *i.e.* to be unaffected by the ionosphere. Initially the transmitters were aboard the rockets and the receivers on the ground (Seddon, 1953). This configuration poses potential problems (*a*) transmitters are heavier than receivers, and (*b*) in the mesosphere the residual air can lead to arcing due to the large electric fields near the radiating antennas, and (*c*) the radiated power can interfere with other instruments in the payload. Today the reverse set-up is more popular since telemetry capacity is readily available (Mayer *et al*., 1999). Sounding rockets are generally spin stabilised, hence simple dipoles will rotate with the payload. In order to eliminate dips in the received signal, one can transmit circularly polarised waves from the ground, or use circularly polarised receiving antennas on the ground in the case where the transmitter is on the payload.

2.1.2 Faraday Rotation

Any linearly polarised wave can be considered as the super-positioning of two circularly polarised waves rotating in opposite directions. If the right and left circularly polarised waves are the two characteristic waves o and x, they will have different real parts of the refractive

index. At any point along the propagation path, the two partial waves will add and the resulting wave will in general be elliptically polarised with the major axis rotated with respect to the original polarisation. This rotation of the polarisation plane is called ***Faraday rotation*** and its rate dF per unit length ds is

$$dF = \frac{\omega}{2c}(\mu_0 - \mu_x)ds \tag{2.12}$$

i.e. we make a differential Doppler measurement with only one frequency, but between the two modes of propagation. Figure 2.4 shows the simulated Faraday rotation at 10^7 electrons m^{-3} for a sounding frequency of 2.29 MHz, a magnetic field strength of 47000 nT and an angle of $\theta =$ 27° between the magnetic field vector ***B*** and the propagation vector as a function of collision frequency or - implicitly - altitude (the values are the ones used by Bennett *et al*., 1972, and are characteristic for launches from South Uist, Scotland). The figure also includes the results of the AH and the ESW theory. For low collision frequencies ($<10^5$ s^{-1}, or >90 km) Faraday rotation is proportional to electron density, for larger collision frequencies the proportionality factor decreases and - in the case shown - even becomes negative for $\nu > 10^7$ s^{-1}. The phase reversal (dF = 0) occurs when $\mu_o = \mu_x$; according to the (quasi-longitudinal) Appleton-Hartree magneto-ionic theory, this critical collision frequency ν_c is (Bauer, 1972):

$$\nu_c = \sqrt{\omega^2 - {\omega_c}^2 \cos^2\theta} \tag{2.13}$$

The above equation shows that a frequency below the gyro frequency ω_c does not experience a reversal of the Faraday rotation. Since the refractive index is dispersive (roughly proportional to f^{-2}), the choice of the sounding frequency is obviously important. Figure 2.5 shows a calculation carried out in deg km^{-1} of altitude assuming strict vertical propagation as a function of signal frequency. At frequencies a little above the local electron gyro frequency (in this case 1.15 MHz), the rotation sense is counter-clockwise (left), below that frequency to the right. Clearly, in the case shown either 2 MHz or a frequency below 1 MHz would yield the most pronounced Faraday rotation (best sensitivity). A more general presentation of the frequency

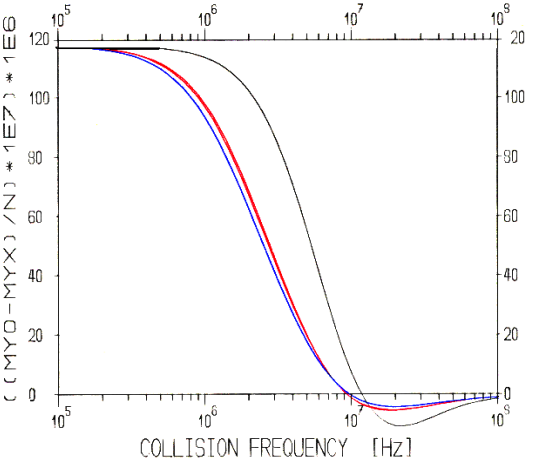


*Figure 2.4 Faraday rotation per electron for 2.29 MHz, a magnetic field of 47,000 nT, an angle Θ of 27° between the magnetic field and the propagation vector as a function of collision frequency. The three curves are for AH (**black**), SW (**blue**), and ESW (**red**). (Friedrich et al., 1991).*

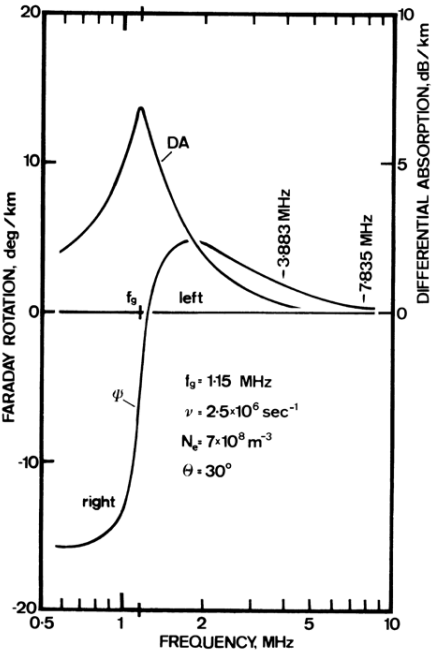


Figure 2.5 Faraday rotation as a function of sounding frequency simulated for conditions of a daytime launch in southern Spain at 75 km (Friedrich et al., 1977).

dependence of the Faraday rotation is given in Figures 2.6 and 2.7. Cases of two extreme electron density profiles are displayed, very large densities (daytime, rocket flight F34, 2.5 dB of riometer absorption), and a night flight (F31, 0.37 dB). Shown are lines of constant incremental Faraday rotation (*i.e.*, deg km^{-1} of altitude) as a function of altitude and frequency. Experience has shown that 0.5 deg km^{-1} is a realistic detection limit, the other limitation is detectable modulation of the signal from which phase can be derived (typically >1 dB). The absolute signal is a function of integral absorption between the beginning of the ionosphere and the height under investigation; values -20 and -30 dB are indicated (see also below). Generally, the sense of rotation changes sign at, or slightly above the gyro frequency f_c and is a weak function of electron density, but predominantly controlled by the collision frequency. A sounding frequency below f_c does not change its sense of rotation and is furthermore more sensitive than a higher frequency. In reality discrete sounding frequencies are used for each of which we can establish minimum altitudes based on the (assumed) detection limit of 0.5 deg km^{-1}. Figure 2.8 shows these altitude limits as a function of electron density for various sounding frequencies between 1.300 and 15.011 MHz. Clearly all frequencies $>f_c$ require infinitely large minimum electron densities at the height (collision frequencies) where the rotation sense changes direction. At larger altitudes (> 80 km), the threshold electron densities of the different frequencies become approximately proportional to f^{-2}. The frequency 1.3 MHz (*i.e.* below f_c) shows the lowest threshold and - more importantly - no singularity.

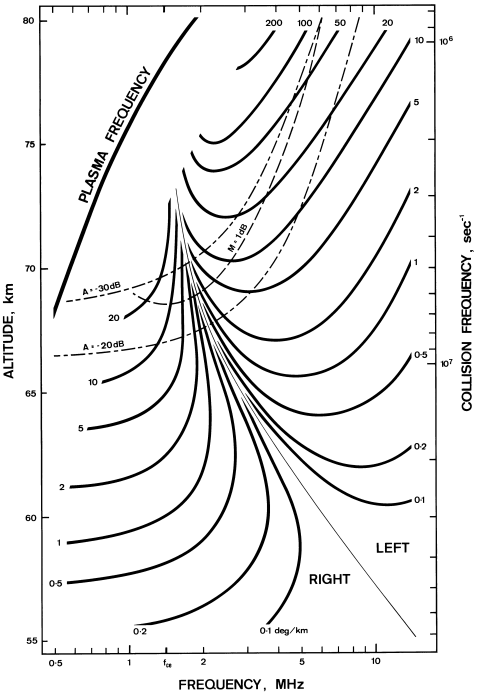


Figure 2.6 Incremental Faraday rotation for conditions at the Andøya Rocket Range and very large electron densities (in this figure f_g is the gyro frequency f_c; Jacobsen and Friedrich, 1979). A frequency below f_c will yield useful data about 4 km lower than any frequency above f_c.

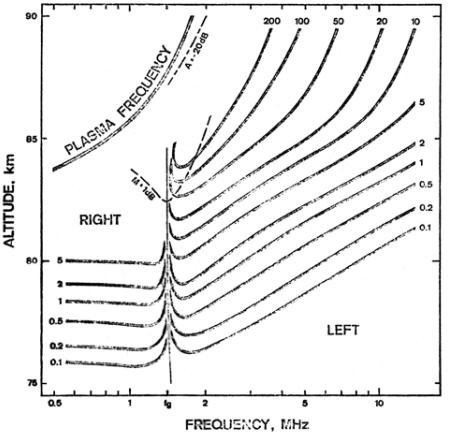


Figure 2.7 As Figure 2.6, but for low (night time) electron density profile (gyro frequency f_c is f_g in this figure; Jacobsen and Friedrich, 1974). Here the gain of using a frequency below f_c is only about 1 km.

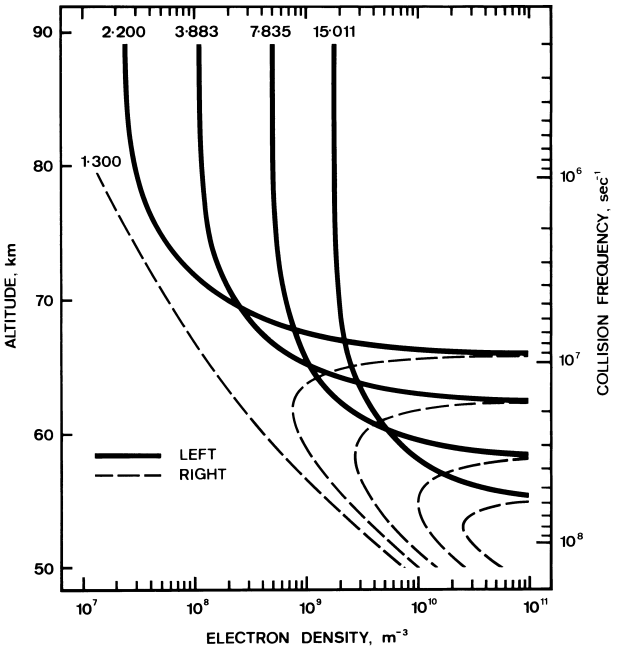


Figure 2.8 Electron densities required to produce 0.5 deg km^{-1} of incremental Faraday rotation as a function of collision frequency (the altitude scale is for guidance only). The lines represent the approximate detection limits for various common sounding frequencies (Jacobsen and Friedrich, 1979).

2.2 Using the Imaginary Part

The measurements of the phase as a function of total electron content (TEC) requires a known reference, such as a very high (*i.e.* unaffected) frequency. By contrast when using the absorption (imaginary part of the refractive index), the signal strength relative to free-space attenuation can be used to obtain total electron content. According to Eq. (2.10) the absorption can be expressed by the term $\exp(-\kappa\frac{\omega}{c}z)$ = $\exp(kz)$. For the longitudinal case (Y_T = 0) we obtain

$$\kappa = \frac{1}{2\mu}\frac{ZX}{(1\pm Y_L)^2+Z^2} = \frac{\nu\omega_p{}^2}{2\omega\mu(\omega\pm\omega_L)^2+\nu^2} = \frac{1}{\mu}\frac{e^2}{2\varepsilon_o m\omega}\frac{N_e\nu}{(\omega\pm\omega_c)^2+\nu^2} \tag{2.14}$$

or the absorption coefficient k:

$$k_L = \frac{1}{\mu}\frac{e^2}{2\varepsilon_o mc}\frac{N_e\nu}{(\omega\pm\omega_c)^2+\nu^2} \tag{2.15}$$

The analogous relation for transverse propagation is:

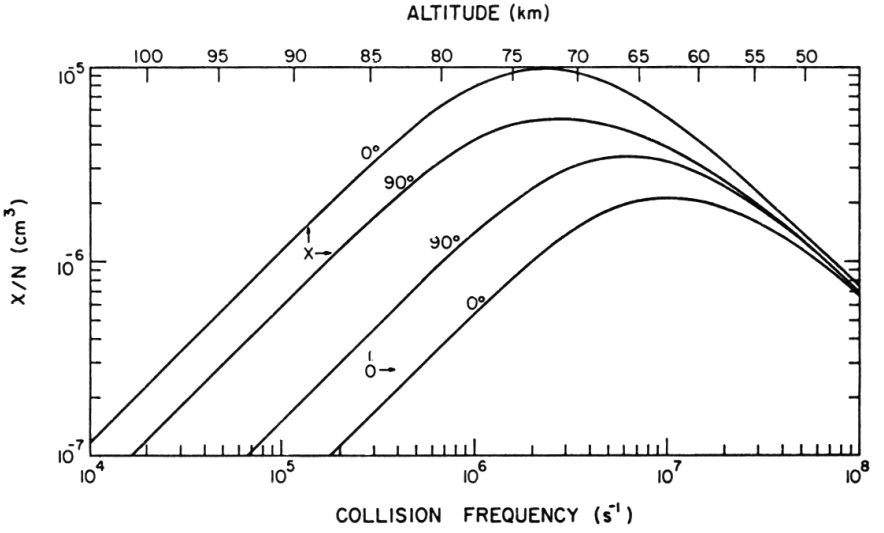


Figure 2.9 Specific absorption (by 1 electron per cm^3) of a 2.2 MHz signal in the presence of a gyro frequency of 1.4 MHz as a function of collision frequency. The components (ordinary and extraordinary) are shown for both longitudinal and transverse propagation (Θ = 0° and 90°; Smith, 1986).

$$k_T = \frac{1}{\mu}\frac{e^2}{2\varepsilon_o mc}\frac{N_e \nu}{\omega^2 + \nu^2} \tag{2.16}$$

hence for signal frequencies significantly larger than both the gyro and the collision frequency the absorption coefficient is proportional to the product of electron density N_e and collision frequency ν (or in consequence pressure p, *cf.* Eq. (2.7)). Figure 2.9 shows the specific absorption k per km (χ in this figure) and electron per cm^{-3}, as a function of collision frequency (or altitude). Shown are two curves each for the ordinary and extraordinary wave and for 0° and 90° angle of the propagation direction and the magnetic field $\boldsymbol{B}$. A sounding frequency of 2.2 MHz and a gyro frequency of 1.4 MHz are assumed. The full Sen and Wyller equations are employed, whereas the Appleton-Hartree theory always predicts proportionality to collision frequency, even for large ν (below 80 km; Smith, 1986). Clearly, the extraordinary wave is absorbed more than the ordinary wave component. For large frequencies (relative to f_c and ν) absorption is proportional to electron density N_e times pressure p, hence this product has a maximum generally found between 75 and 90 km. Using absorption to derive N_e above or below that height region is therefore not successful, irrespective of the sounding frequency. Cases of calculated incremental absorption using actual electron density profiles are shown elsewhere in another context (*cf.* Figure 4.31) and demonstrate the altitude restriction where absorption data are of potential use.

Given a known collision frequency profile (*i.e.* a known pressure profile), one can record the attenuation of an electromagnetic wave radiated from the ground to the rocket payload in flight. Because the refractive index has two values (and hence relations to electron density), one will generally intentionally either transmit the ordinary or the extraordinary wave. However, simply measuring the amplitude has several potential pitfalls: Any antenna, such as the one radiating from the ground, has an antenna pattern. Since the rocket is never launched exactly vertically, it will move out of the main beam of the ground antenna and thus the received signal

decreases unrelated to absorption in a physical sense; the same also applies to the receiving antenna aboard the rocket. The impedance of a short antenna (see Section 2.1.2) is capacitive. This capacitance forms part of the tuned input circuit of the receiver. For large electron densities with plasma frequencies approaching the signal frequency, the capacitance changes drastically and de-tunes the receiver; again this can not be disentangled from real absorption. But also a simple shift of the receiver's input/output characteristic (*cf.* Figure 2.23), *e.g.,* caused by the acceleration at launch, may lead to distorted results.

2.2.1 Differential Absorption

A way of largely bypassing most of these problems is to record the *difference* between the absorption of the ordinary and extraordinary modes using the same frequency, the same receiver, and the same antenna. Jespersen *et al*. (1964) intentionally transmitted alternatingly left and right hand circularly polarised waves (o- and x-mode, Figure 2.10). At the equator where strict transverse propagation prevails (assuming vertical transmission to the payload and a horizontal magnetic field) the two characteristic modes are waves linearly polarised in geomagnetic North-South and East-West directions, respectively. The latter arrangement is still used today for the determination of electron densities in the vicinity of the geomagnetic equator (Fig. 2.11; Friedrich and Torkar, 1995c; Goldberg *et al*., 1997; Friedrich *et al*., 1997).

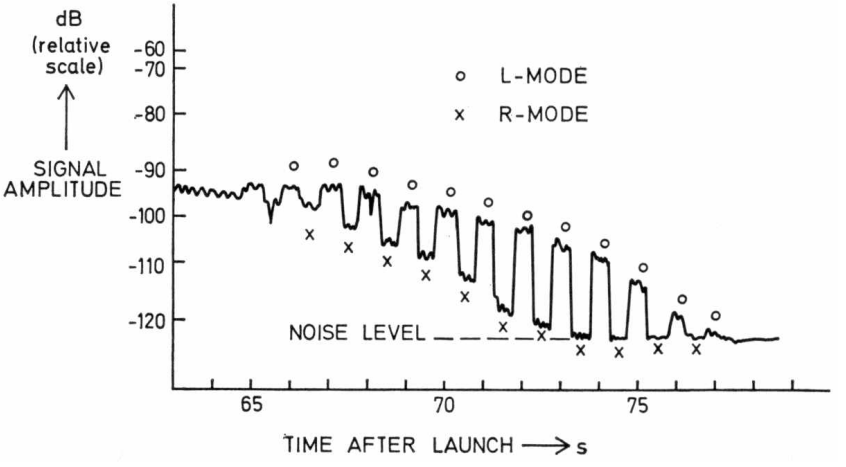


Figure 2.10 Signal strength recorded at 2.2 (?) MHz as a function of flight time. The periods of transmission of the ordinary and extraordinary wave are indicated (Jespersen et al., 1964, a rocket launch from Andøya).

For conditions where the propagation has a component along the magnetic field ($Y_L \neq 0$), a simpler method can be used to obtain differential absorption. Any linearly polarised wave can be considered decomposed into two circularly polarised waves rotating in opposite direction; the two circularly polarised waves experience the two characteristic refractive indices, o and x. The different real parts produce a rotation of the orientation where the sum of the two partial waves maximises (Faraday rotation, see above), whereas as a consequence of the different imaginary parts the two partial waves will have different amplitudes. Hence, they do not fully cancel out when their field strengths have opposite signs; the resulting polarisation therefore becomes elliptical with the major axis representing the sum of the o and x-modes, and the minor

axis their difference. Analogous to Eq. (2.12), the differential absorption *DA* per unit distance d*s* is

$$dDA = 20\frac{\omega}{c}(\kappa_o - \kappa_x)\log_{10} e\, ds \tag{2.17}$$

The differential absorption sensitivity for a given frequency as a function of collision frequency is displayed in Figure 2.12. In analogy to Figure 2.4 and for the same conditions, results according to the three theories (AH, SW, and ESW) are shown per electron. An example of the frequency dependence of differential absorption is depicted as one of the curves in Figure 2.5, a more general presentation is given in Figure 2.13. In this figure, contours of constant differential absorption are calculated for the same conditions as in Figure 2.6, *i.e.* for a case of large electron densities. More general, the detection limits can be calculated for various discrete sounding frequencies as a function of collision frequency (altitude); Figure 2.14 shows these detection limits, both for differential absorption (DA), and for the absorption of the ordinary wave (AOW). The figure demonstrates that at very low altitudes AOW is more sensitive than DA.

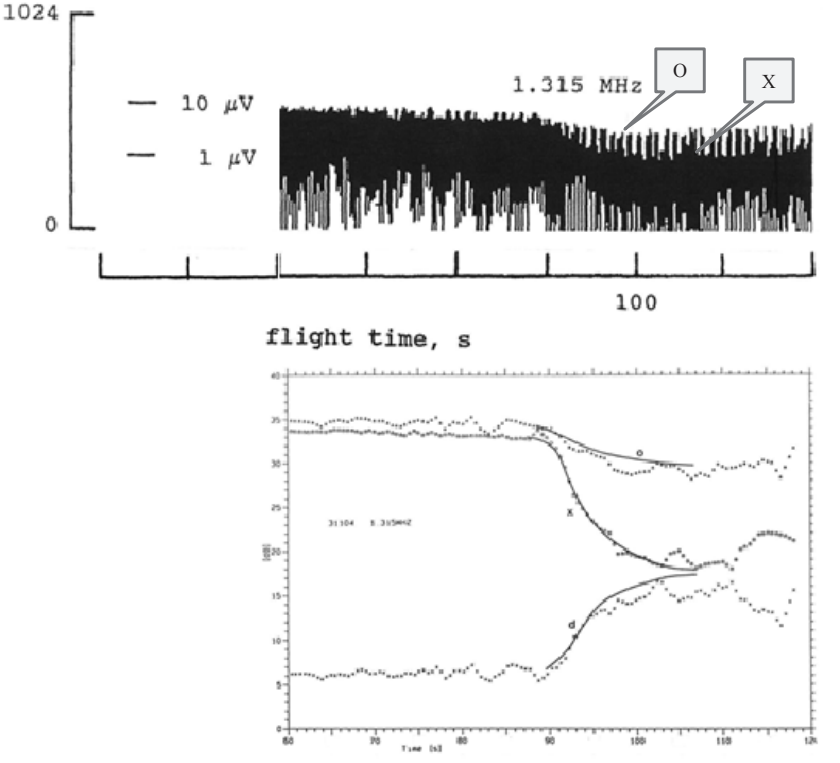


Figure 2.11 Signal strength recorded at 1.315 MHz as a function of flight time of a launch from Alcântara, Brazil at the equator (after Friedrich and Torkar, 1995c). The lines marked o, x and d are the simulated signals of the ordinary, extra-ordinary and differential absorption using the final electron density profile.

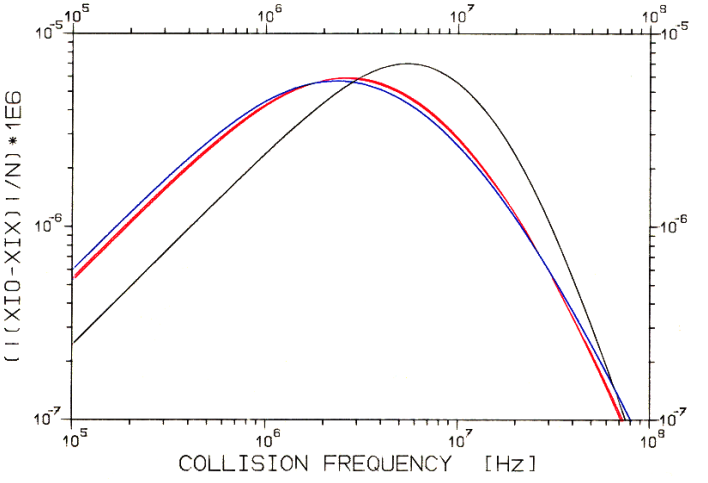


Figure 2.12 Differential absorption for the same conditions as in Figure 2.4 according to various magneto-ionic theories (Friedrich et al., 1991).

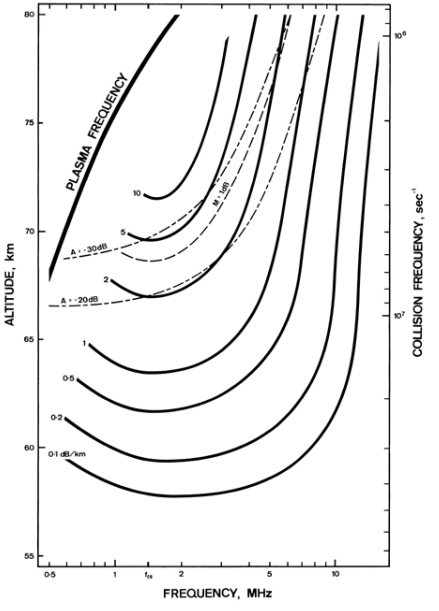


Figure 2.13 Incremental differential absorption for the same conditions as in Figure 2.6 (large electron densities; in this figure the gyro frequency is f_g; Jacobsen and Friedrich, 1979).

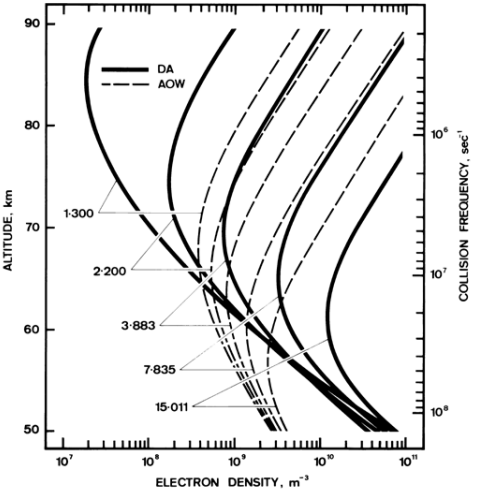


Figure 2.14 Electron densities required to produce 0.5 dB km^{-1} of incremental absorption, i.e. the detection limit, for various sounding frequencies (DA = differential absorption, AOW = absorption of the ordinary wave; Jacobsen and Friedrich, 1979).

2.2.2 Absorption of the Ordinary Wave

As demonstrated in Figure 2.14, at the very lowest altitudes the absorption of the ordinary wave is more sensitive than differential absorption. However, sounding rockets usually pass these low altitudes before antenna deployment. A more realistic application of the AOW data is at higher altitudes when the received signal is predominantly due the less absorbed ordinary extraordinary wave, *i.e.* when the x-mode is 20 dB or more below the o-mode (*cf*. Figure 2.18).

2.3 Raw Data, Ideal *vs*. Real

A linearly polarised antenna rotating with the payload will scan the polarisation pattern (ellipse) twice in each spin period. The angle of the ellipse's major axis relative to the orientation prior to entering the ionosphere provides the major information on the electron content (Faraday rotation). Hence a reliable spatial reference to which to relate the momentary receiver output is of paramount importance. There are several possibilities to obtain such a reference: (*a*) a gyro (mechanical or optical) is an expensive device, (*b*) sun sensors provide the attitude to the Sun (obviously only in day flights), and (*c*) a star sensor yields this information at night. However the simplest device capable of providing attitude information, both in day and night, is a magnetometer. Ordinary magnetic aspect sensors are of the flux gate type, *i.e*. instruments that detect static fields superimposed on a nearly saturated iron core. Naturally such magnetic aspect sensors do not need doors in the payload skin or booms in order to establish the orientation to the Earth's geomagnetic field (which is assumed to be stable and known with sufficient accuracy). A predictable problem with this type of aspect sensor occurs at very high latitudes. Under these conditions the horizontal component - or more precisely the component transverse to the rocket's spin axis - is only small, in fact a coning payload may actually nutate about the Earth's magnetic field line. Another problem of a magnetic aspect sensor mounted

within a metallic payload cylinder occurs due to eddy currents. Figure 2.15 shows the scenario of a rotating metallic cylinder and the resulting disturbed magnetic field inside it. The error angle φ is (Ulf Fahleson, private communication, 1973):

$$\tan\varphi = \frac{\mu_o \omega_s R \rho t}{2} \tag{2.18}$$

with

μ_o ... free space permeability
ω_s ... angular spin frequency of the payload
R ... radius of the payload
t ... thickness of the cylinder wall
σ ... conductivity

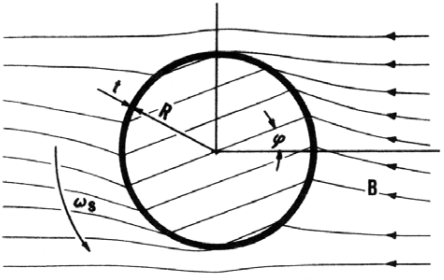


Figure 2.15 Distortion of an external magnetic field by a rotating metallic cylinder (Friedrich and Jacobsen, 1975).

The rocket payload F32 was a rare case where the magnetometer was in a section of the payload, which was exposed after clamshell ejection. Figure 2.16 shows the phase between the output of the receiver at 3.883 MHz and the magnetometer. Before 45 s flight time (below 50 km) the phase varies sinusoidally; this is not due to the ionosphere (Faraday rotation), but rather a signature of the payload's coning motion with a period of about 8 s. At 45 s, there is a clear discontinuity of the phase of about 20° coincident with the clamshell ejection, *i.e.* the exposure of the magnetometer and consequently a vanishing of the eddy currents. Inserting the values of that particular flight into Eq. (2.18) yields 28° which compares acceptably with the observed phase discontinuity. The distorted magnetic field direction inside the payload cylinder is constant and does not introduce a measurement error as long as the rocket's spin frequency is constant (no de-spin). A more serious problem is the phase error introduced by the coning motion. This can effectively be overcome by using the same antenna also for a receiver, which is primarily used a spatial reference. The erstwhile group in Slough (Bennett *et al*., 1972) used dedicated receivers at 138 MHz solely for the purpose of providing a rotational reference. We apply the cheaper method of simply using the signal of the Faraday receiver of the highest frequency as a reference; Figure 2.17 demonstrates that this procedure not only eliminates the apparent phase variation due to the payload coning, but also the magnetometer's phase discontinuity at clam shell ejection. Once all lower frequencies are absorbed, the reference for the remaining highest sounding frequency must be one of the dedicated aspect sensors. Usually, however, at such a height the real rate of Faraday rotation is large enough that the error introduced by the coning becomes negligible. On the downside it must be borne in mind that the phase is determined between two noisy signals, whereas a dedicated spin sensor (magnetometer, gyro, *etc*.) is noise-free, as a result the differential Faraday rotation will be somewhat more noisy; furthermore, the rate of differential Faraday rotation (deg km^{-1}) is less than that of the lower frequency *vs*. an aspect sensor. A Sun sensor as an aspect detector

naturally does not suffer a rotation offset, but essentially also shows phase variations due to coning. It furthermore requires an opening in the skin, and at night has to be replaced by a much more complex star sensor.

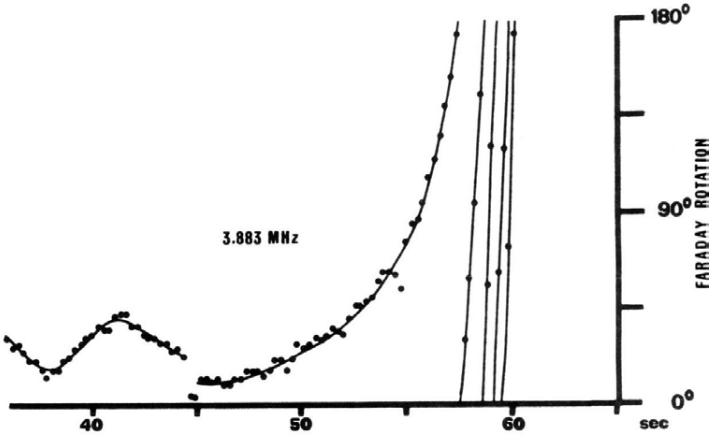


Figure 2.16 Phase of the 3.883 MHz receiver relative to the magnetic aspect sensor as a function of flight time (rocket F32; Friedrich and Jacobsen, 1975). Note the phase discontinuity at 45 s coincident with the exposure of the magnetic aspect sensor.

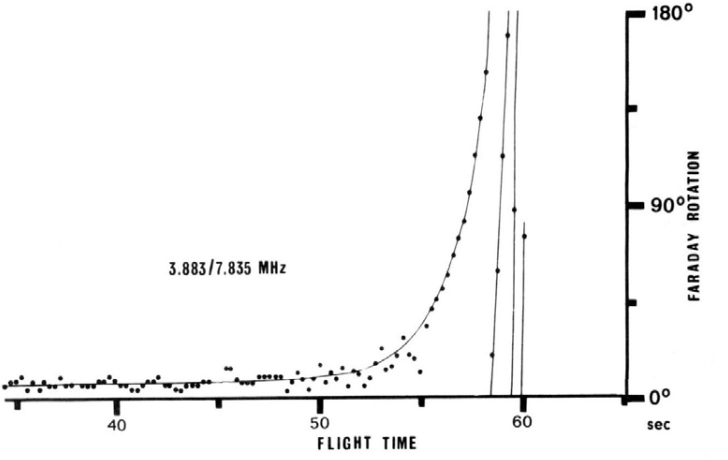


Figure 2.17 Differential rotation between 3.883 and 7.835 MHz. Both the coning signature and the phase discontinuity have disappeared (cf. the previous figure; flight F32; Friedrich and Jacobsen, 1975).

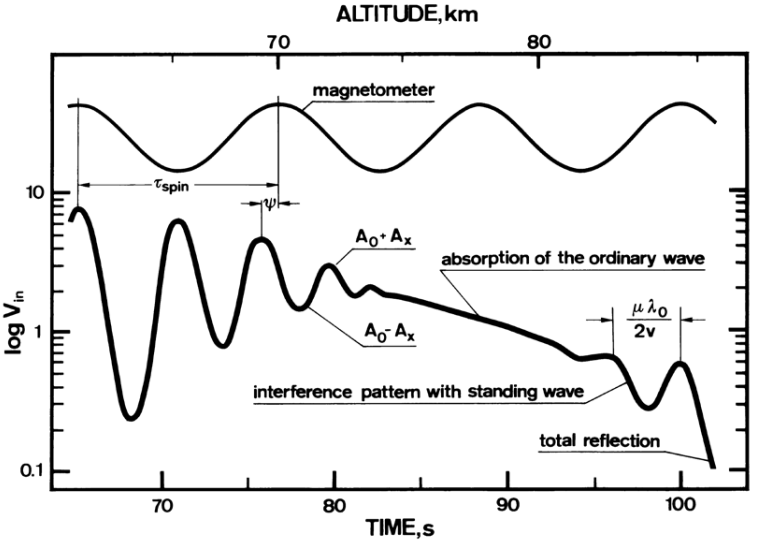


Figure 2.18 Idealized variation of the output of a Faraday receiver (Friedrich and Torkar, 1978).

An HF frequency is radiated *via* a linearly polarised antenna from the ground to the rocket payload, and is received at the payload with a likewise linearly polarised antenna. Since the receiving antenna spins with the rocket, it scans the polarisation pattern twice per spin period or typically every 200 m of altitude (assuming 4 rps spin and an apogee of 100 km). Figure 2.18 shows such an idealised variation of a receiver output. The information is contained in the phase shift ψ of the receiver's output signal *vs.* a reference (here a magnetometer is assumed), in the ratio between maximum and minimum (sum of ordinary and extraordinary component, and difference between the two), the strength of the ordinary component once the extraordinary component has virtually disappeared, and in the abrupt disappearance (total reflection where the signal frequency equals the plasma frequency). In addition, in this figure a standing wave pattern is indicated caused by the interference between upward propagating and reflected wave; this pattern can cause misinterpretation if it displays a similar frequency as the pattern caused by the rocket spin. Figure 2.19 shows actual Faraday receiver outputs from a night time flight (rocket WADIS-2, March 5th, 2015, 01:44 UT, Andøya). Reflection of the transmitted wave occurs when the refractive index *n* vanishes (see Eq. 2.1). However, already before *n* disappears, one encounters deviations of the wave propagation from free space geometry. As first demonstrated by Haselgrove (1957) a radio wave injected into a layered ionosphere at an angle from the vertical no longer propagates along the line of sight. As a result, the simple geometric assumptions on the propagation path no longer hold. In practice, one can see discrepancies starting several kilometres below the reflection height for a particular wave frequency (assuming the electron density to increases with altitude). Figure 2.20 shows another Faraday receiver output from the flight of rocket ECOMA-1 (September 6th 2006, 22:17 UT) with such standing wave pattern of the 1.300 MHz signal a few seconds before the respective reflection, most pronounced at 1:30 flight time.

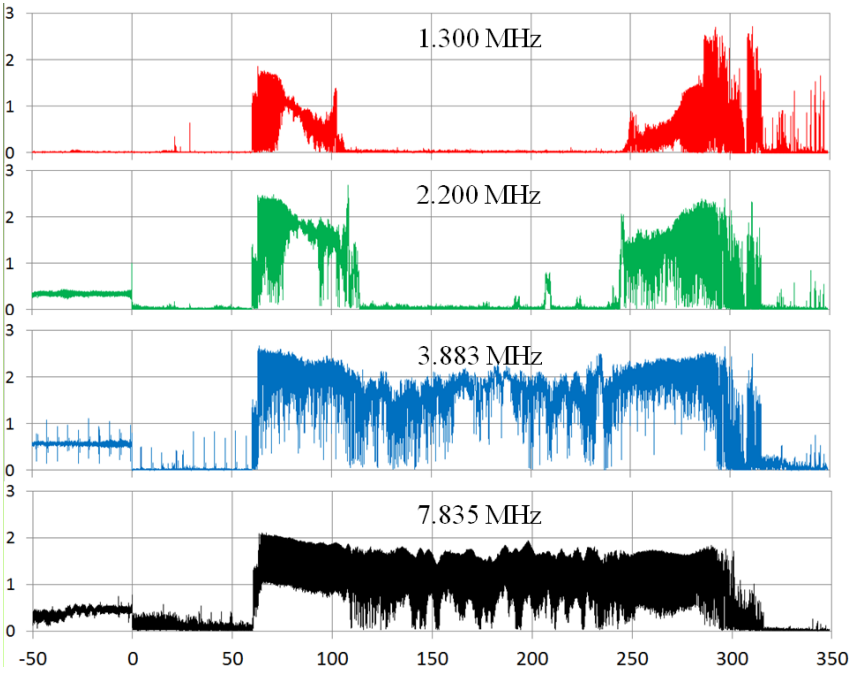


Figure 2.19 Signals of the four receivers aboard the rocket WADIS-2 as a function of flight time. The observed signal before take-off is noise inserted via the umbilical. The antennas were exposed at 55 s and fully deployed 2 seconds later.

The magnet-ionic theory predicts total reflection of the ordinary wave when the signal frequency f equals the plasma frequency (*i.e.*, $f = f_p = \frac{\sqrt{N_e e^2/\varepsilon_0 m}}{2\pi}$). The extraordinary wave, however, is reflected at $f = f_p + 0.5\, f_c$ (f_c = gyro frequency). The signal of the 1.300 MHz receiver depicted in Figure 2.20 is rare example where the reflections of the o- and x-modes are distinguishably different (flight ECOMA-1, September 8th, 2006, full night,

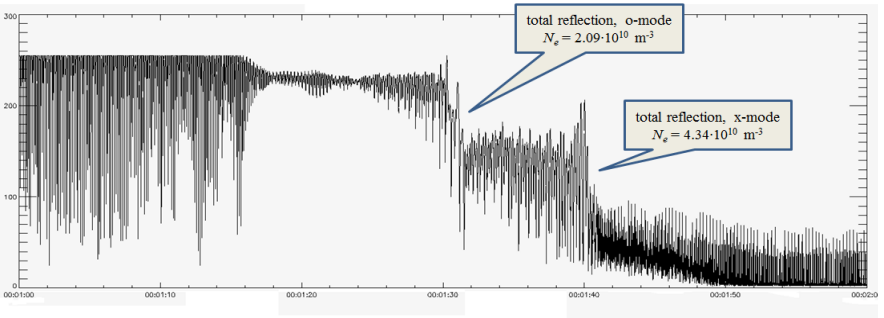


Figure 2.20 Signal of the 1.300 MHz receiver of the rocket ECOMA-1 as a function of flight time. Note that in this (rare) case total reflexion of the two partial waves occur at distinctly different times.

Andøya); usually at its reflection height the extraordinary wave is absorbed to below the receiver's threshold.

2.4 Hardware

One attraction of the wave propagation experiment described in this chapter is its relatively simple and robust design. There are, however, a few things to be considered, some of which are today only listed for the sake of completeness.

2.4.1 Receivers

The ionospheric magneto-plasma exhibits dual refractive and complex properties of the refractive index for electromagnetic waves. A linearly polarised wave is transmitted from the ground to the flying rocket payload. By measuring the received signal (strength, polarisation) one can infer the total electron content between ground and payload, and - by differentiation with respect to altitude - the electron density. For reasons of flight stability, a sounding rocket usually has a spin between 3 and 12 Hz mainly depending on the payload size, but also on its instrumentation. If both the antenna on the ground and aboard the payload are linearly polarised the receiver will therefore observe a modulation with twice the spin rate (*i.e.*, 6 to 24 Hz). Another more important factor determining the required minimum bandwidth of the receivers is the Doppler shift by the fast-flying rocket. Velocities of 1 km s^{-1} are typical for sounding rockets dedicated to the study the of the *D*- and *E*-regions, although 2 km s^{-1} should be taken into account. The highest frequency employed for diagnostic purposes is 7.835 MHz (on occasions higher frequencies have been used in the past, but retrospectively did not yield much additional information or provided a larger measurement range than 7.835 MHz). The Doppler shift for the highest frequency and a velocity of 2 km s^{-1} is about 50 Hz, the IF filter with the narrowest bandwidth commercially available is 300 Hz, which is therefore safely wide enough for all cases. To reduce noise and unwanted signals the bandwidth should be kept to a minimum and wider bandwidth is therefore not desirable anyway.

		from			
		1.300 MHz	2.200 MHz	3.883 MHz	7.835 MHz
to	1.300 MHz	X	90 dB	-	-
	2.200 MHz	73 dB	X	70 dB	-
	3.883 MHz	89 dB	-	X	125
	7.835 MHz	118 dB	-	-	X

Table 2.1 Crosstalk between receivers at IF level (455 kHz).

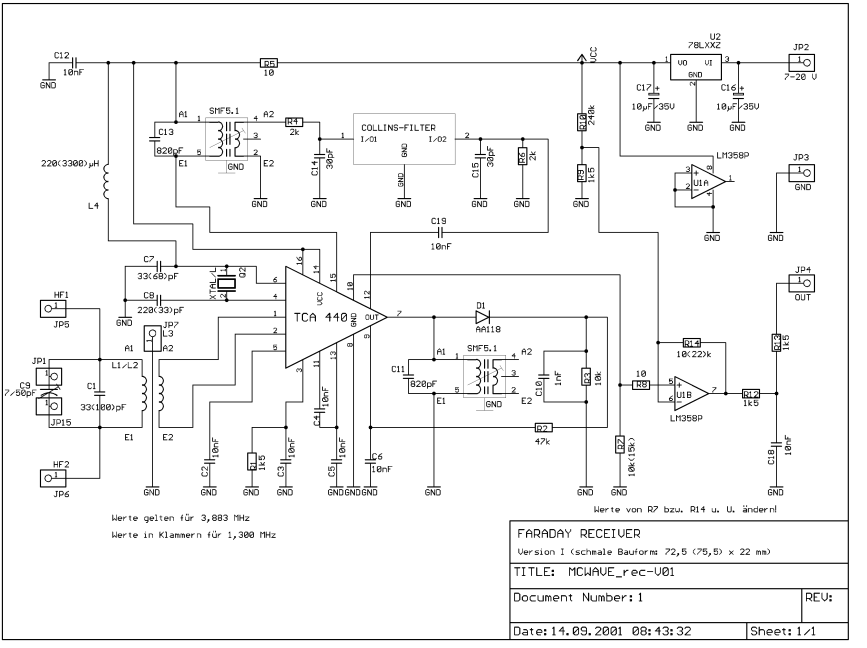


Figure 2.21 Faraday receiver schematic used since the McWave payloads (2002/2003).

For good height coverage, one will generally use more than one frequency (up to four). The receivers are usually fed from the same antenna, and in early years it was feared that using the same intermediate frequency (IF) for all receivers might lead to cross talk between them. In those Faraday receivers we therefore used a different IF for each receiver (*e.g.,* 0.7, 1.0 and 1.5 MHz for 1.300, 2.200 and 3.883 MHz input frequencies, respectively; Friedrich, 1972). Since 1973 we use commercial filters (ceramic or mechanical) at the standard IF of 455 kHz.

The signals at IF level are de-coupled by at least 70 dB and also the poorer image frequency rejection by the low IF poses no problem with antennas that are at most 50 cm long (Table 2.1). Similarly, since 1998 we have abandoned the bandpass (two tuned circuits) before the mixer stage, and have not experienced problems that one could attribute to interference at the image frequency (rejection 48 dB for a receiver at 8.855 MHz, and better for lower frequencies). Figure 2.24 shows the selection curve of a current receiver with a mechanical filter. Because of the poorly predictable receiver input voltages (see Ch. 2.4.2), we use a logarithmic input-output characteristic with a 2 to 3 decade input range. Figure 2.21 shows the schematic diagram of the current design. Although only a positive supply voltage ≥12 V is used, the output can practically reach zero for small inputs. The current consumption is about 8 mA. Figure 2.22 depicts the most recent design of only 72 by 22 mm which predominantly uses SMD (= surface mounted devices) components and a mechanical IF filter. Figure 2.23 shows the calibration curves (input *vs*. output) of the receivers of one of the McWave payloads flown in 2003.

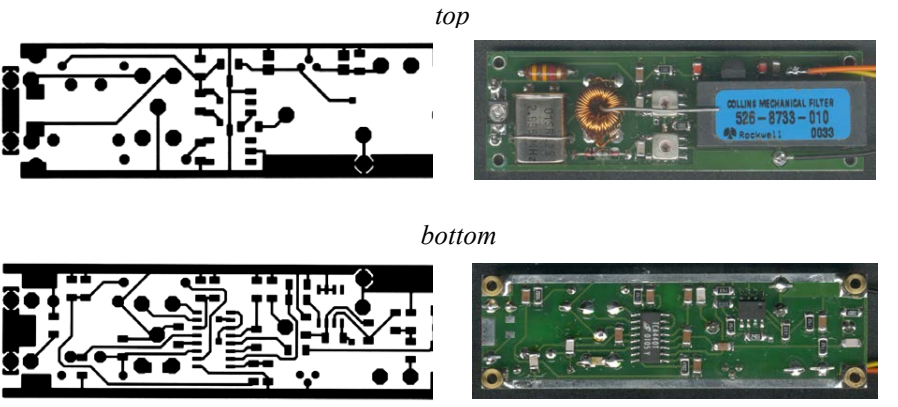


Figure 2.22 Faraday receiver first used in the McWave payloads (2002/2003), and ever since. The temporary wire through the toroid is used for inserting a signal to establish the calibration curve and has an approximate impedance of 50 Ω to ground. The patch at the left hand side visible at the bottom is for direct attachment (soldering) of a tape measure antenna. In the populated receiver shown on the right this part of the PC board was not needed and was therefore cut off.

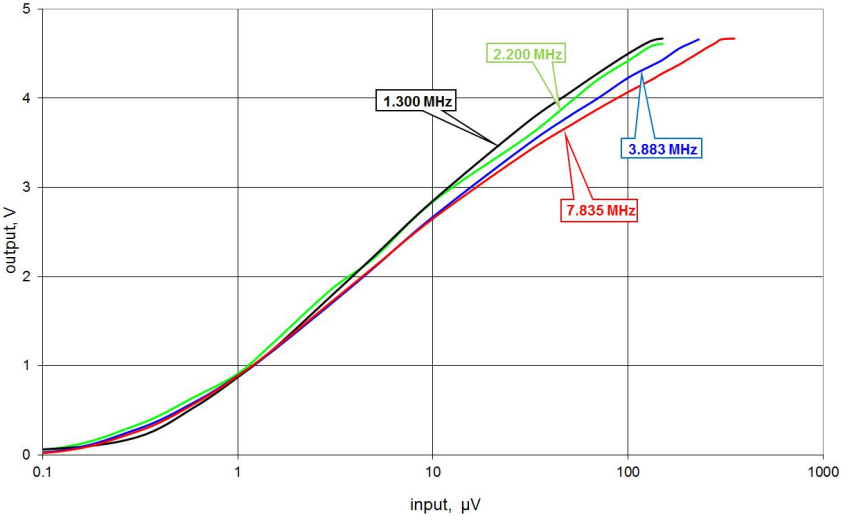


Figure 2.23 Calibration curves of the receivers flown on one of the McWave payloads (41.030, January 24, 2003, Esrange).

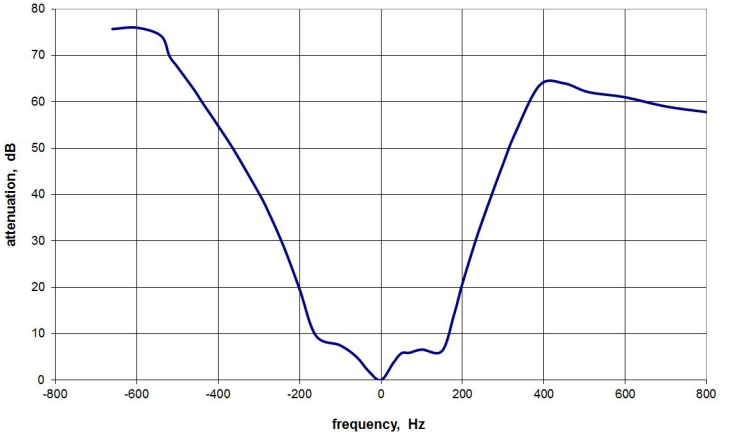


Figure 2.24 Overall selection curve of a current Faraday receiver with a mechanical 300 Hz IF filter (centre frequency 8.855 MHz).

2.4.2. Antennas, Matching and Link Budget

The antennas aboard the rocket payload must be linearly polarised in order to detect the polarisation of the ellipse to deduce its aspect (Faraday rotation) and the axis ratio (differential absorption). Some groups have employed ferrite rods (notably the University of Illinois at Urbana-Champaign), which require a non-conductive payload section which is mechanically not desirable. Most other groups use(d) electrical antennas, which are necessarily short compared to the wavelength λ. For such a short dipole antenna, the textbook gives the following approximation of its impedance (Heilmann, 1970; or Johnson, 1993):

$$Z_A = R_{As} + jX_{As} = R_A + \frac{\ell}{j\omega C_{As}} = 80\left(\frac{\pi\ell}{\lambda}\right)^2 + j\frac{120}{\pi}\left(\frac{\lambda}{\ell}\right)\left[ln\left(\frac{\ell}{a}\right) - 1\right] \qquad (2.19)$$

where
Z_A ... complex impedance,
R_{As} ... real part of the antenna impedance (series),
X_{As} ... imaginary part of the antenna impedance (series),
$2\times\ell$ ... length of the antenna (tip-to-tip), and
a ... radius of the antenna rod.

Equivalently one can express the imaginary part of the antenna impedance X_{As} as a (frequency independent, serial) capacitance C_{As}:

$$C_{As} = \frac{\ell}{120c\left[ln\left(\frac{\ell}{a}\right)-1\right]} \qquad (2.20)$$

where c is the speed of light.

In reality, the dipole halves are generally not thin cylinders (wires), but can have various shapes and sizes. We will attempt to empirically establish Z_A, including its real part. Inserting actual frequencies and using measurements in the above equations yields real parts of the order of mΩ, whereas the imaginary part is of the order of several kΩ; such an arrangement is therefore essentially capacitive. Directly measuring the impedance's real part with a commercial vector impedance meter (Hewlett Packard 4815A) is impossible, given that the instrument has full-scale ranges from 10 Ω to 100 kΩ for $|Z|$, and a phase scale of ±90°. We use the following arrangement to bypass the problem of measuring an almost perfectly capacitive impedance.

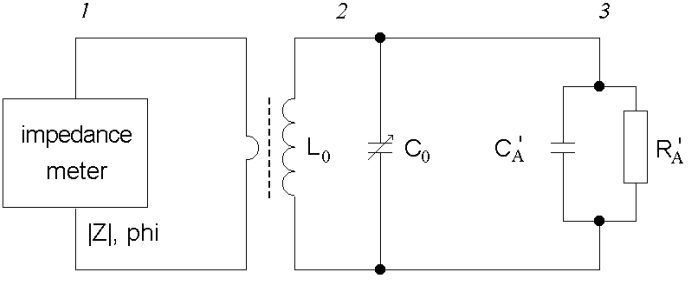
For almost purely capacitive short antennas $C_{As} \approx C_{Ap}$ (hence also $X_{Ap} \approx X_{As} \approx X_A$) and R_{As} of the series equivalent circuit can be converted to R_{Ap} (and *vice versa*) in a parallel circuit by the following relation:

$$R_{Ap} = \frac{X_A^{\,2}}{R_{As}} \tag{2.21}$$

In the equivalent parallel circuit, the impedance of a short antenna is a small capacitance C_A parallel to a large resistor R_{Ap}. By choosing a large transformation ratio one can readily bring R_{Ap} to a value within the range of the vector impedance meter. With the above arrangement, one can determine R_{Ap} and C_A by two methods:

a) When attaching the antenna both the resonance frequency f and Q of the tuned circuit (L_o, C_o) will change, namely to lower Q and lower frequency. Provided both C_o and L_o were accurately established, one can deduce the complex antenna impedance from the change (lowering) of the resonance frequency (*cf.* Preiß, 2004) and the increase of the bandwidth (lowering of Q).

b) Alternatively one can substitute the real and imaginary parts of the antenna impedance by



1 ... vector-impedance meter
2 ... tuned circuit with low impedance tap to impedance meter
3 ... antenna shown in parallel equivalent circuit

Figure 2.25 Arrangement to determine the real parts of the various antenna antennas using full-scale mock-ups.

discrete components (resistors and capacitors) until the same Q and f is obtained as with the antenna attached (Friedrich, 1974).

For measuring convenience, we chose a combination of the two methods and proceed as follows: (*a*) we determine the antenna capacitance from the shift of the resonance frequency, and (*b*) we then re-tune C_o for resonance, and measure the (transformed) lowered impedance

with the vector meter. In the following the measured impedances of the antenna designs actually flown in the past are listed and compared to theory.

1) The series of MIDAS payloads had diameters of 7" and hinged welding rods as antennas (2 mm diameter, 26 cm on each side, Figure 2.26); this simple geometry is the one best suited to test the validity of the theoretically expected impedance. For the calculation of the theoretical value according to Eq. (2.22) we ignore the diameter of the payload such that ℓ in the equation is twice the length of the spikes. The arrangement yields a capacitance in fair agreement with the empirical value, although - in order to obtain the measured capacitance - one either has to increase the physical diameter of the rods (2 mm) to an effective diameter of 4 mm or the physical length (52 cm) to an effective one of about 63 cm. After some manipulation of Eq. (2.19), using, Eq. (2.21), one obtains for the real part in the parallel equivalent circuit

$$R_{Ap} = \frac{45c^4}{\pi^4 \ell^4}\left[ln\left(\frac{\ell}{a}\right) - 1\right]^2 f^{-4} \qquad (2.22)$$

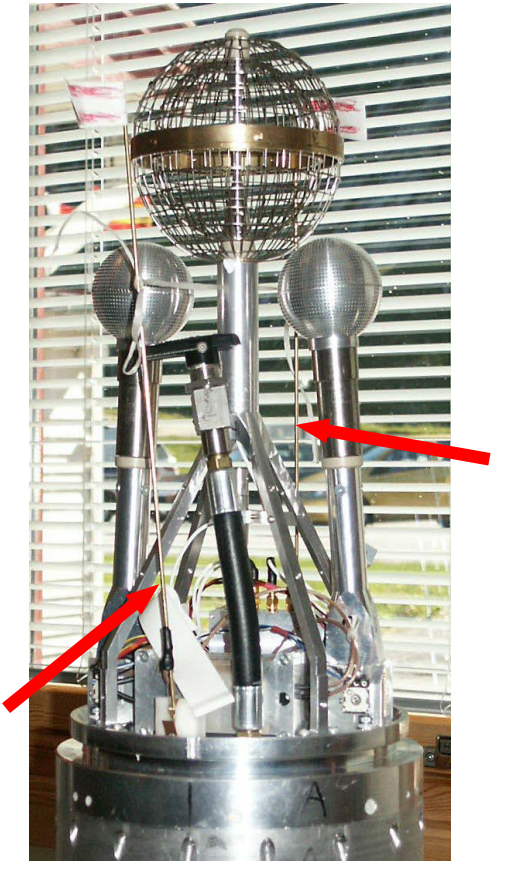


Figure 2.26 Hinged antenna of a MIDAS payload in stowed position (Type 1). The central ion probe uses a gridded collector to allow free airflow (see later, Ch. 3.2).

and with the dimensions of the MIDAS antennas one obtains $R_{Ap} = 17.03\, f^{-4}$ (in MΩ and MHz), in reasonable agreement with the measurement at lower frequencies, but with very different frequency dependence. Similarly, the serial real part of the measured impedance *increases* with frequency, whereas according to theory (Eq. 2.22) it should *decrease.* This is explicable by the fact that primarily the ohmic loss (including that of the toroid) is determined and the real part of the antenna impedance is but a negligible fraction thereof. The measured capacitance inevitably includes inseparable stray capacitances which do not constitute part of the antennas impedance; hence in order to emulate the calculated impedance (capacitance) either the length of the antenna or its effective diameter has to be increased unrealistically.

	capacitance, pF	***R_{Ap}, (f in MHz), MΩ***	***R_{As}, (f in MHz), Ω***	***equiv. diameter, mm***	***equiv. length, cm***
measured	3.9	$14.00\ f^{-1.3833}$	$119\ f^{-0.6167}$	30.0	80
theoretical	1.6	$17.03\ f^{-4.000}$	$0.00237\ f^{2.000}$	2	52

2) Measuring tape antennas were flown on many payloads with 14" diameter payloads (since 2002; Figure 2.27). The dimensions of each dipole half are 1.4 by 48 cm and the arms are separated by the diameter of the deck plate (30 cm). Measurements with a mock-up were made

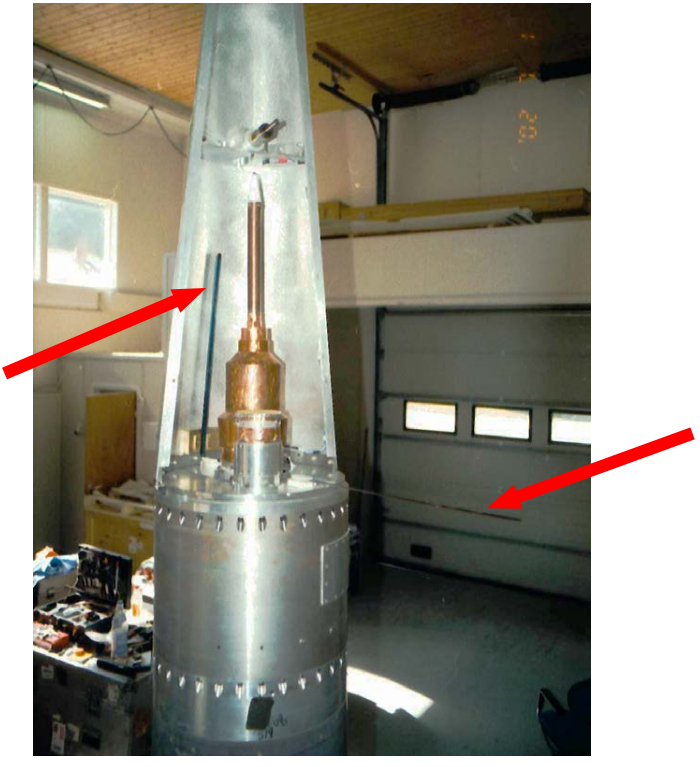


Figure 2.27 Type 2 antenna of a McWave payload (41.030 to 41.033). The left half of the dipole is in stowed position "leaning" against the clam shell, whereas the right half is in measuring position.

length, cm	*capacitance, pF*	R_{Ap}, *(f in MHz), MΩ*	R_{As}, *(f in MHz), Ω*	*equivalent dia., mm*
2×48	5.4	$10.60\ f^{-1.5214}$	$81.95\ f^{-0.4786}$	29.9
2×30	4.3	$17.90\ f^{-1.5364}$	$73.09\ f^{-0.4636}$	31.8
2×15	3.2	$25.40\ f^{-1.5750}$	$91.46\ f^{-0.4250}$	30.0

not only for the actual length of 48 cm as flown, but also for shorter tapes. One can define an equivalent diameter for these antennas maintaining the physical length by varying the effective diameter for agreement between measurement and theory.

3) MiniMIDAS, the hitherto smallest payload, flown by a Viper-III rocket, had an outside diameter of only 2". The antennas were tape measure strips slit in half (7 mm wide) and only 8 cm on either side (Fig. 2.28). Here the equivalent diameter of the dipole arms turns out to be 18 mm.

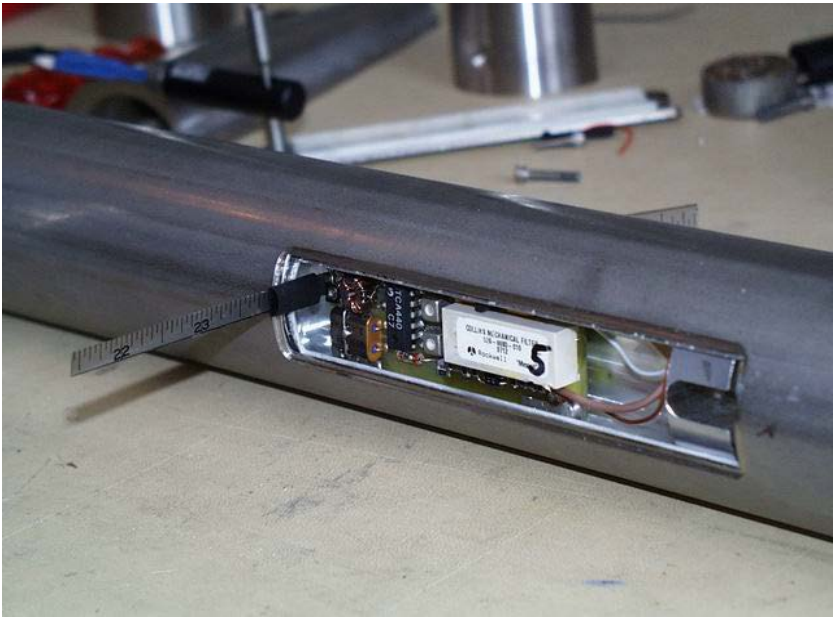


Figure 2.28 Antenna of a MiniMIDAS payload in flight position (Type 3). Note that the antenna tapes are directly soldered to the receiver PC board.

capacitance, pF	R_{Ap}, *(f in MHz), MΩ*	R_{As}, *(f in MHz), Ω*	*equivalent dia., mm*
1.9	$26.44\ f^{-1.4791}$	$265.5\ f^{-0.5209}$	18.3

4) In the early days of sounding rocketry antennas for the Faraday experiment were an extra pair of 250 MHz telemetry antennas. The simplest design of this type are λ/4 spikes under 45° to the axis of the 6¼" Apache or Cajun payloads (Fig. 2.29).

Figure 2.29 Antennas of a payload of an Apache sounding rocket (Type 4, rocket F44, January 1976, El Arenosillo, Spain).

capacitance, pF	***R_{Ap}, (f in MHz), MΩ***	***R_{As}, (f in MHz), Ω***
8.4	$0.63\ f^{-0.8780}$	$569.8\ f^{-1.1220}$

5) Payloads aboard Tomahawk rockets (9" diameter) have used doubly bent antennas (Fig. 2.30). The real part of their impedance showed the lowest value, but also the smallest dependence on frequency of all antenna geometries investigated.

capacitance, pF	***R_{Ap}, (f in MHz), MΩ***	***R_{As}, (f in MHz), Ω***
7.8	$0.56\ f^{-0.7923}$	$748.6\ f^{-1.2072}$

6) Orions were flown in the NLTE campaign from Esrange, Sweden, in 1998 (Fig. 2.31). The antennas were originally quarter wave dipoles for 250 MHz offset from the payload skin. For the present purpose, they were mounted fully isolated.

capacitance, pF	R_{Ap}, *(f in MHz), MΩ*	R_{As}, *(f in MHz), Ω*
17.4	$0.77\ f^{-1.0906}$	$341.4\ f^{-0.9094}$

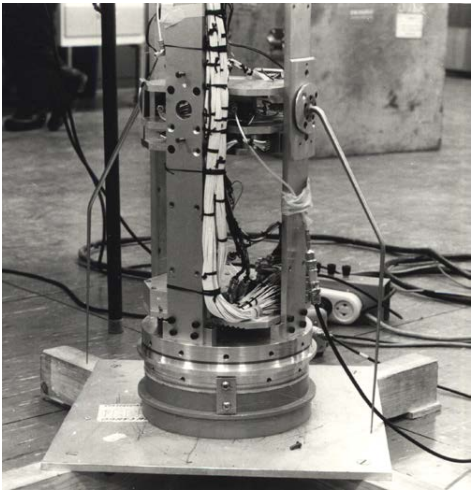


Figure 2.30 Antennas of a payload for a Tomahawk sounding rocket (Type 5, F32, Andøya, Norway).

Figure 2.31 Antennas of an NLTE payload for an Orion sounding rocket (Type 6, March 1998, Esrange, Sweden).

All antenna arrangements investigated showed frequency dependencies considerably smaller than the exponent 4 anticipated from theory (Eqs. 2.19 to 2.21).

Figure 2.32 shows the measured real parts R_{As} of the serial equivalent circuit of various antenna mock-ups *vs*. frequency. Clearly, the frequency dependencies of all types of antennas can be approximated by power laws. The curves appear in two groups with different impedance and frequency dependencies, namely types 1 to 3, which are linear antennas, and types 4 to 6, which are under an angle or bent. Within the accuracy of the measurements the capacitance (responsible for the imaginary part) turns out to be frequency independent as expected from the relation 2.20. The parallel R_{Ap} depicted in the diagram can readily be converted to serial R_{As} for comparison with theory (Eq. 2.21). The values thus derived are orders of magnitude larger than what theory predicts because the measurement includes ohmic losses. Since these values are still significantly lower than X_{As}, we may ignore the antenna's real impedance altogether and treat it as a source with a purely imaginary (capacitive) impedance.

The electric field E due to an isotropically radiating source with power P at a distance r is:

$$E = \sqrt{\frac{Z_o P}{4r^2\pi}} \tag{2.23}$$

$Z_o = 1/\varepsilon_o c \approx 120\pi \approx 376.7\ \Omega$ (free space impedance)

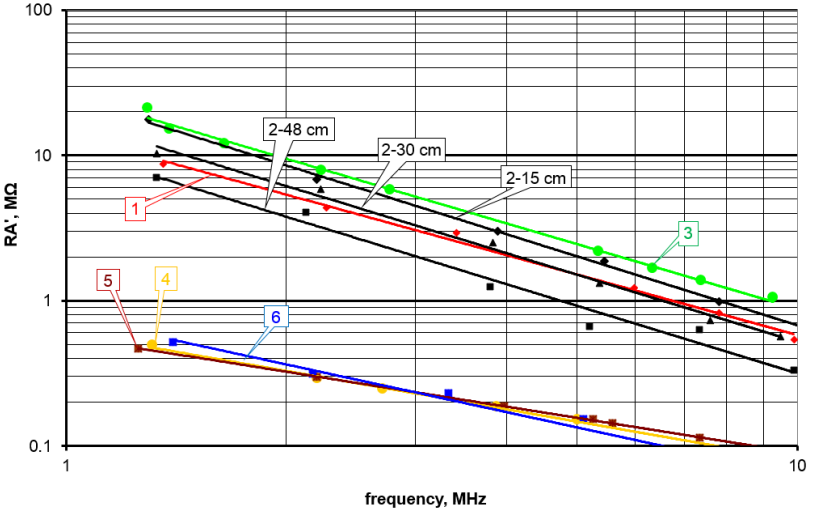


Figure 2.32 Measured real part of the (parallel) antenna impedance R_{Ap} as a function of frequency for various antenna designs (Types 1 to 6; Preiß, 2004).

The ground antennas radiating to the sounding rocket are half-wave dipoles, generally *not* suspended at the optimum height of a quarter wave length. To test the link budget, we ignore the influence of the ground (with its poorly known properties) as a reflector; in this case the gain G_{TX} is 1.64 (or 2.15 dBi). With a typical radiated power of 500 W we can thus expect a field strength in 70 km altitude (*i.e.* at the beginning of the measurements and below the absorbing ionosphere) of 2.24 mV m^{-1}.

A test of the validity of the link budget is best performed with the Type 1 antenna (welding rods, 2 mm dia., 2×26 cm) since this configuration is the one described theoretically by Eq. (2.19). The voltage induced is the electric field E expressed by Eq. (2.23) times the length (tip-to-tip, 52 cm) of the antenna times ½ (Hertzian dipole). The impedance Z_{As} of this antenna is almost purely capacitive, namely $Z_{As} \approx 1/\omega C$, where $C = 3.9$ pF as measured for this particular arrangement. This impedance is matched by the tuned circuit at the input of the receiver by adjustment for resonance, effectively by an inductive impedance of the same magnitude as represented by the 3.9 pF. The reference is the 1-turn calibration input to the receiver which, for this test, is approximated by an impedance of 50 Ω. The voltage thus "seen" by the input tuned circuit is transformed by $1/w$ to the reference/calibration input. The resulting formula relating electric field strength to the voltage $V_{50\Omega}$ at the reference input is:

$$V_{50\Omega} = El\frac{50w}{Z_{As}+50w^2} \qquad (2.24)$$

From the above equation one can derive the optimum winding ratio $w_{opt} = \sqrt{Z_{As}/50\Omega}$ in the tables below w_{opt} is rounded to the nearest integer.

frequency, MHz	*winding ratio, w*	$V_{50\Omega}$, *µV*	*winding ratio,* w_{opt}	$V_{50\Omega(opt)}$, *µV*	*matching loss, dB*
1.300	55	9.3	25	11.6	-1.9
2.200	40	10.9	19	15.1	-2.8
3.883	23	8.9	15	20.1	-7.1
7.835	9	25.2	10	28.5	-1.1

The only other antennas that can potentially be compared to theory are Types 2 and 3 (measuring tape), the results for the four different lengths are listed below.

	frequency, MHz	*winding ratio, w*	$V_{50\Omega}$, *µV*	*winding ratio,* w_{opt}	$V_{50\Omega(opt)}$, *µV*	*matching loss, dB*	$V_{50\Omega(measured)}$, *µV*
2×48 cm	1.300	55	35	21	51.2	-3.30	86
	2.200	40	48	16	68.4	-3.08	29
	3.883	23	75	12	90.6	-1.64	15
	7.835	9	129	9	129.0	0.00	31

In the above table the actual winding ratios are a little different from what is used in the other tables. Some of the voltages observed at the reference input are even larger than what one would expect for optimum winding ratio (1.300 MHz). This may be attributable to the uncertain gain of the ground antenna, the significant deficit, on the other hand, can be explained by a detuning of the respective receiver (3.883 MHz) during launch or launch operations.

	frequency, MHz	*winding ratio, w*	$V_{50\Omega}$, *µV*	*winding ratio,* w_{opt}	$V_{50\Omega(opt)}$, *µV*	*matching loss, dB*
2×30 cm	1.300	50	13.4	24	28.0	-1.3
	2.200	45	14.9	18	37.3	-8.0
	3.883	27	19.7	14	48.0	-7.7
	7.835	17	39.5	10	67.2	-4.6

	frequency, MHz	*winding ratio, w*	$V_{50\Omega}$, *μV*	*winding ratio,* w_{opt}	$V_{50\Omega(opt)}$, *μV*	*matching loss, dB*
2×15 cm	1.300	50	5.2	28	6.1	-1.4
	2.200	45	6.1	21	7.9	-2.3
	3.883	27	9.2	16	10.5	-1.2
	7.835	17	19.8	11	30.5	-3.8

	frequency, MHz	*winding ratio, w*	$V_{50\Omega}$, *μV*	*winding ratio,* w_{opt}	$V_{50\Omega(opt)}$, *μV*	*matching loss, dB*
2×8 cm	1.300	50	2.3	36	2.5	-0.7
	3.883	27	6.6	21	8.5	-2.2

All of the above back-of-the-envelope calculations are for coarse guidance only; not least the assumption of 50 Ω at the calibration input is a simplification. Actual values range from 30 to 80 Ω. A margin of 10 dB to account for a mismatch should under all circumstance be foreseen, although the received signals are sometimes larger by about a factor of two. Also, the winding ratios *w* used here may be (have been) different depending on the permeability of the toroid cores that were actually used.

An alternative to the (largely futile) attempt of matching the receivers' impedance to that of the antenna is to use a very high impedance RC pre-amplifier (*e.g.* FET source follower). Such antennas cover a wide frequency range and can provide 50 Ω output. Such a low impedance signal can then easily be fed to the receivers even over longer distances inside the payload. Furthermore, in such a scenario the receivers can be tuned (peaked for maximum) irrespective of the length of the cable and the actual antenna impedance.

2.4.3. On-Board Processing

The basic time (height) resolution with which the rocket borne wave propagation experiment (Faraday rotation, [differential] absorption) can provide data is tied to the rocket's roll rate. Per spin period the antenna rotating with the payload twice scans the polarisation ellipse and yields two maxima and two minima and their respective phases relative to a spatial reference; measuring (transmitting) the received signal between the maxima and minima is (theoretically) not needed. It is therefore tempting to search for the amplitude and phase of the maxima (minima) on-board the rocket and only transmit these values. The most widely used spatial reference was a transverse magnetometer, which has a sinusoidal variation with the spin frequency as output, whereas the output of a Faraday receiver has twice the spin frequency as the basic period. Figure 2.33 shows telemetry signals from a magnetometer and a Faraday receiver. The magnetometer signal is indeed sinusoidal, whereas the receiver signal is not only at twice that frequency, but shows broad maxima and pronounced minima; this distorted waveform is due to the logarithmic input-output characteristic of the receiver.

Whereas it is fairly easy to detect the phase of the magnetometer signal by *e.g.*, a Schmitt trigger, due to noise on the signal and the variable shape of the Faraday output one has to apply

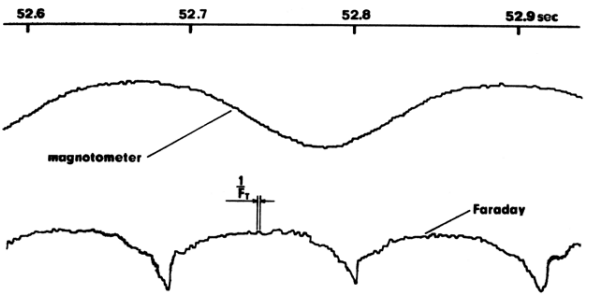


Figure 2.33 Telemetry signals of a transverse magnetometer (top) and a Faraday receiver (bottom; rocket F33, Andøya, May 31, 1972, 7.835 MHz, Torkar and Friedrich, 1976). $1/F_T$ indicates the time between two telemetry samples.

a more sophisticated procedure to obtain the phase of the receiver output. France and Williams (1976) used Schmitt triggers for the signals of both the spin sensor and the Faraday receiver. From the logical signals thus obtained, they digitally derived the phase of the receiver outputs to the spatial phase (in their case a dedicated reference receiver).

An electronic procedure to account for the distinctly non-sinusoidal shape of the Faraday receiver and which is able to detect the maxima and minima of the receiver output was originally developed within an MSc thesis by Nutz (1972), flown aboard the sounding rocket F33 from Andøya on May 31, 1972, and presented by Friedrich (1973). The essential novelty was a phase locked loop (PLL) used to obtain a logical signal in phase with the Faraday receiver output. A PLL is a square wave oscillator locked in frequency and phase to the incoming signal, in our case the output of the Faraday receiver with a signal of twice the spin rate (*ca*. 10 Hz). The square wave thus synchronised by the incoming signal is 90° out of phase to the incoming signal; however, the further away the synchronising frequency is from the natural (free-running) frequency of the PLL, the more likely lock can be lost and the phase between input and output deviates from 90°. Ideally, the edges of the synchronised square wave coincide with the maxima and minima of the synchronising signal.

Figure 2.34 shows the block diagram of the phase meter flown in 1972. For the (sinusoidal) magnetometer signal, a modified Schmitt trigger is used to obtain a square wave. The spin period is converted to a voltage, which in turn controls the natural (free running) frequency of the PLL *via* an opto-coupler. Thus, the phase between input and output of the PLL is kept close to 90°. The Faraday signal is fed to a follow-and-hold circuit, which is controlled by the restored Faraday signal, *i.e.* 90° out of phase. Figure 2.35 shows the principle of detecting the maxima and minima where separate capacitors follow the maxima and minima, respectively. The actual phase measurement (Faraday rotation) was accomplished by an RS flip-flop, which provided a pulse length proportional to the phase between magnetometer and Faraday receiver output; every 128th phase reading was interrupted for a transmission of the spin period.

An alternative approach to measure the phase between magnetometer and Faraday receiver signal was pursued by Reisinger (1977). First the square of the magnetometer signal was formed (with a four-quadrant multiplier) in order to double the spin frequency, *i.e*. the same basic frequency as the receiver output. Then the frequency-doubled magnetometer signal was delayed in an analogue delay line (bucket brigade). This delayed signal was compared in a four-quadrant multiplier (phase detector) with the Faraday signal. The delay (≈ Faraday Rotation)

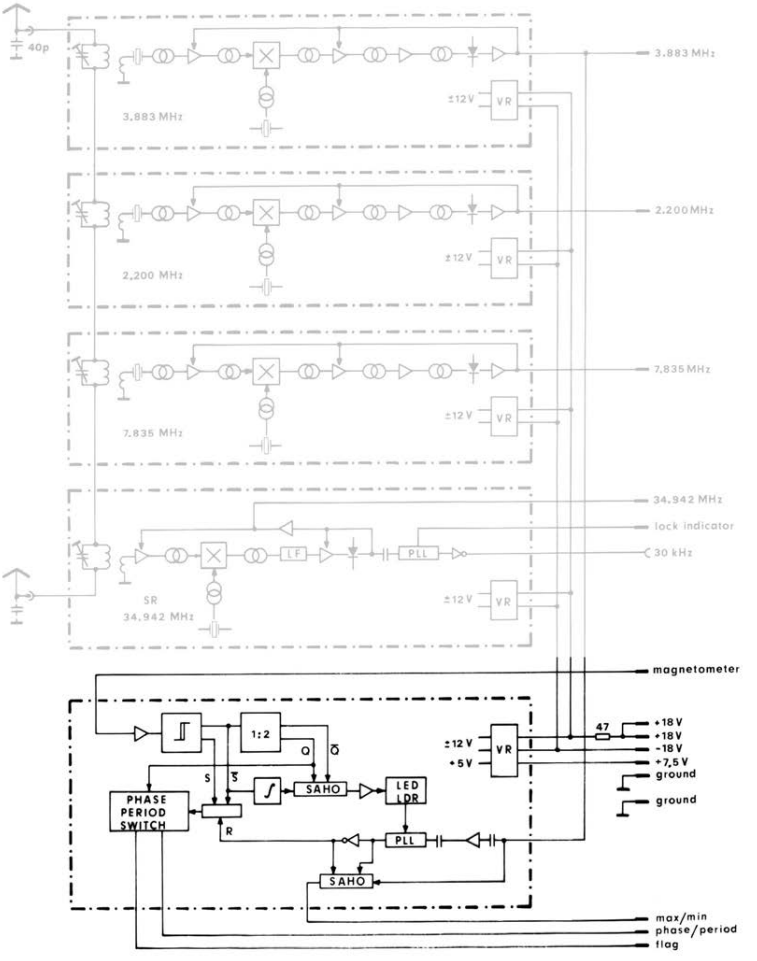


Figure 2.34 Block diagram of the wave propagation experiment aboard the payload F34. The phase meter and maximum/minimum detector are highlighted (Friedrich, 1974).

was controlled for maximum output of the phase detector. This instrument was built as a laboratory model only, but not actually flown on a sounding rocket.

Development of on-board data pre-processing is no longer pursued since generally sufficient telemetry capacity is available, and post-flight data processing is more flexible and avoids the risk of instrument failure.

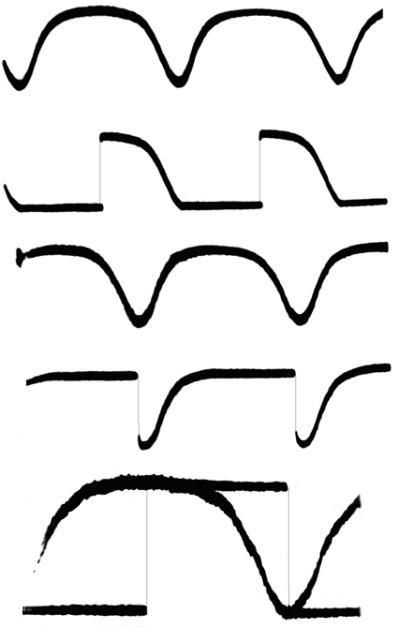


Figure 2.35 Maximum-minimum detection of a simulated Faraday signal (top trace). Second panel: capacitor tracking the minimum, third panel: capacitor tracking the maximum, bottom panel: output of the Faraday receiver together with the amplitude channel representing maxima and minima only (Friedrich, 1974).

2.5. Data Processing

The linearly polarised antenna aboard the rotating payload scans the polarisation pattern (in general an ellipse); the observed maxima $\bar{A}$ represent the sum of the ordinary and the extraordinary wave components (A_o, A_x), the minima $\underline{A}$ their differences. The orientation of the ellipse's major axis is the Faraday rotation, *i.e.* the main quantity from which one derives the total underneath electron content and - by differentiation with respect to altitude - the electron density. Figure 2.36 shows the variation of a receiver output. Ideally, the signal should be the projection of the two axes of the ellipse onto the receiving antenna of the form:

$$E(\varphi) = \sqrt{\bar{A}^2 \cos^2(\varphi - \psi) + \underline{A}^2 \sin^2(\varphi - \psi)} \tag{2.25}$$

where $E(\varphi)$ is the field strength received as a function of rotation angle $\varphi = \omega t$, $\bar{A} = A_o + A_x$, $\underline{A} = A_o - A_x$, ω is the payload's angular spin frequency, and ψ an arbitrary initial angle. Introducing differential absorption $\Delta = A_o/A_x$, Eq. (2.25) can be re-written as:

$$E(t) = \frac{A_o}{1-\Delta}\sqrt{1 + \frac{4\Delta}{(1-\Delta)^2}\cos^2(\omega t - t)} \tag{2.26}$$

or combining all angle-dependent terms the signal follows:

$$s(t) = \sqrt{1 + F\cos^2(\omega t - \psi)} \tag{2.27}$$

According to Eqs. (2.26) and (2.27) the signal is modulated by twice the spin frequency. Faraday rotation (*i.e.* the phase of the receiver output $s(t)$ *vs.* a spatial reference) and differential absorption (contained in F) can ideally be determined twice per rocket spin period. In reality, however, there are usually two "families" of shapes of $s(t)$. In the example shown in Figure 2.36 one can clearly see alternating high and low maxima. This example (payload CHAMPS, rocket 41.094) is a very rare case of an attitude controlled flight in which the angle between rocket axis and velocity vector was kept below 2.3°. Nonetheless the amplitude of the maxima alternates; in the case of a coning payload a plausible explanation can be put forward: The hitherto most compelling explanation is that the two halves of the antenna contribute differently to the input voltage of each receiver. The receivers are connected in series (*cf.* Fig. 2.34) and - considering stray capacitances - the contribution to a receiver directly connected to one antenna half (usually the ones for the highest and the lowest frequency, respectively) will be greater than the signal provided by the other antenna half which has to pass two or three receivers. The payload generally exhibits some coning, and the ground transmitter is located nearly, but never exactly along the nominal spin axis (*i.e.* at the launch site). In such a configuration even for little or no coning, the antenna arms will - one at a time - be in the shadow of the rocket in the course of a spin period. This explanation of the asymmetry is supported by the following observations:

a) the central receiver shows the least asymmetry (same stray capacitance to either antenna section, *i.e.* the same contribution of each of the antenna sections),

b) the asymmetries in the maxima of the two edge receivers occur in antiphase,

c) the behaviour reverses in phase with the payload coning, and

d) at times where one knows from simulation that only the (circularly polarised) ordinary wave exists, it is modulated with the single spin frequency (NB: the standing wave pattern between up going and reflected wave may produce a "modulation" with a frequency of the same order as the spin rate!).

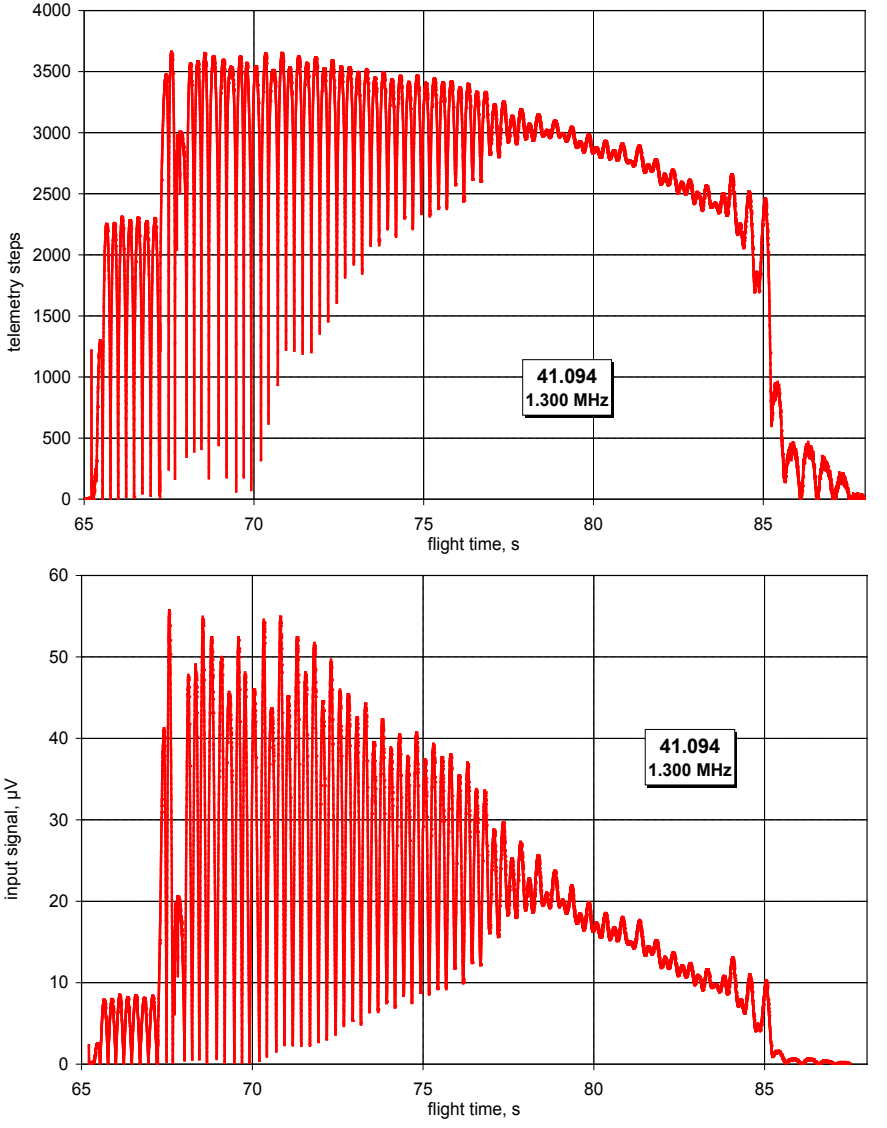


Figure 2.36 Typical signal of a Faraday receiver (1.300 MHz, flight of rocket payload 41.094, October 11th, 2011, Andøya). Top panel: telemetry raw data (receiver output), bottom panel: telemetry data converted to receiver input voltage. At 65 s the antenna is exposed (folded forward) and fully deployed at 67 s.

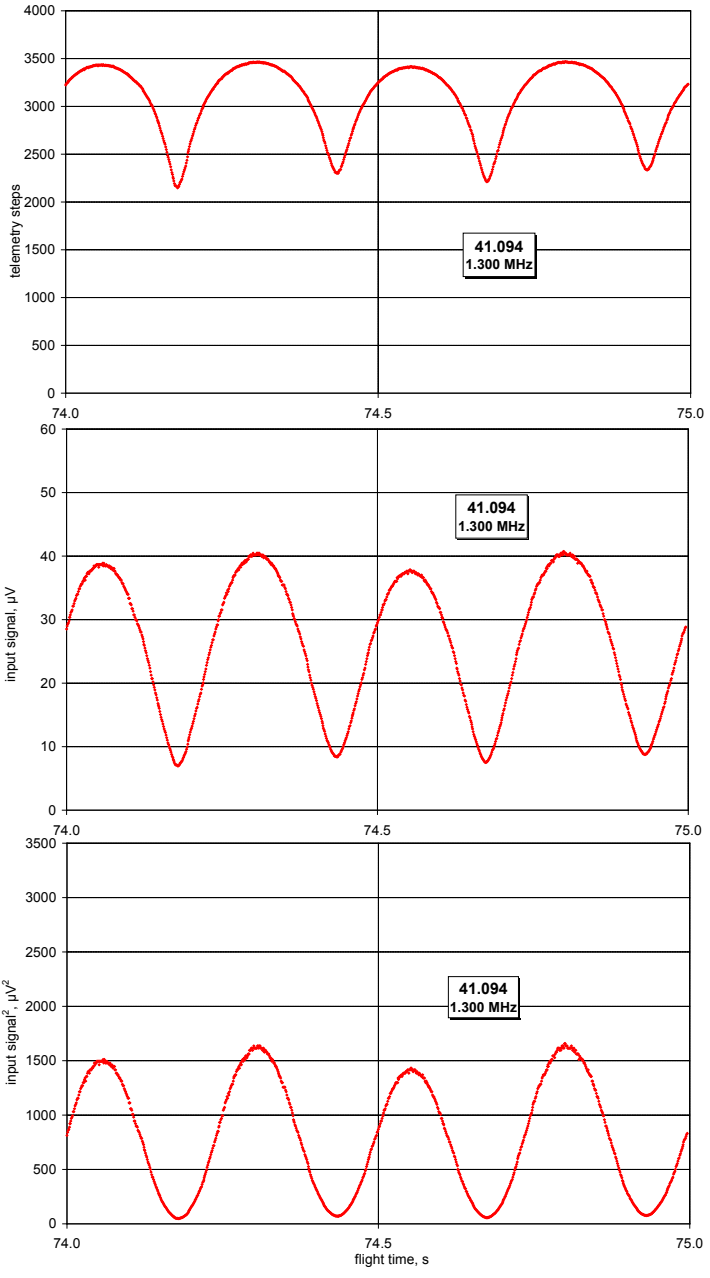


Figure 2.37 Blow-up of the 1.300 MHz signal of rocket 41.094 near 74 s. Top panel: receiver output (in telemetry steps), central panel: receiver input (in µV), and the squared receiver input (bottom panel) which (ideally) should be sinusoidal.

Procedures were developed to compensate for the single spin component in the receiver signals (Torkar and Friedrich, 1981). However, today we simply fit the receiver output by the expected curve (sine wave) over a full spin period in increments of half a spin period; we thus still obtain a time resolution of half a spin period.

The receiver's calibration curve (input *vs.* output, *cf.* Fig. 2.23) is applied to the telemetry data. According to Eq. (2.27) it is the square of the reconstructed input signal $s(t)$ which should have the form of a constant superposed by a sine function at twice the spin frequency, whereby the constant as well as the amplitude and phase of the sine function are unknowns and can be fitted using a least-mean-square (RMS) method (Torkar, 1975; Torkar and Friedrich, 1976). Data in saturation or below the receiver's noise floor must be removed before fitting, as including them in the fit would invalidate the calculated amplitude and possibly also compromise the result for the phase (thinking of asymmetries in the signal due to time lags in the receivers or due to eddy currents). The signal variation thus reconstructed yields maxima and minima even if they should have been in saturation or below the receiver's noise floor (in which cases the result would be an extrapolation from the data, associated with some increase of the error bars of the fitting procedure); an example of a receiver signal and its reconstruction is depicted in Figures 2.37 and 2.38.

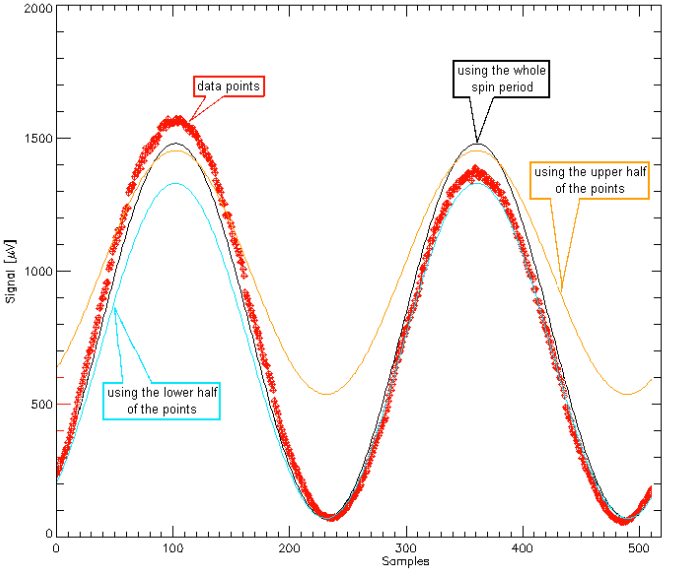


Figure 2.38 Receiver signal near 74 s flight time, converted to $(\mu V)^2$ together with three ways of assimilating it by sine functions. The samples are in approximately 1 ms increments (CHAMPS payload, rocket 41.094).

The fitting process should account for possible errors of the calibration. In practice it is worthwhile to visualise the quality of the sine function fit to the data. In case of systematic deviations as a function of phase it may be advisable to replace the fitting of all data together by separate fits in time intervals near the maxima and minima, respectively (Fig. 2.38). Due to the higher susceptibility of absorption data to the calibration compared to the phase, absorption data typically provide less reliable results than phase data.

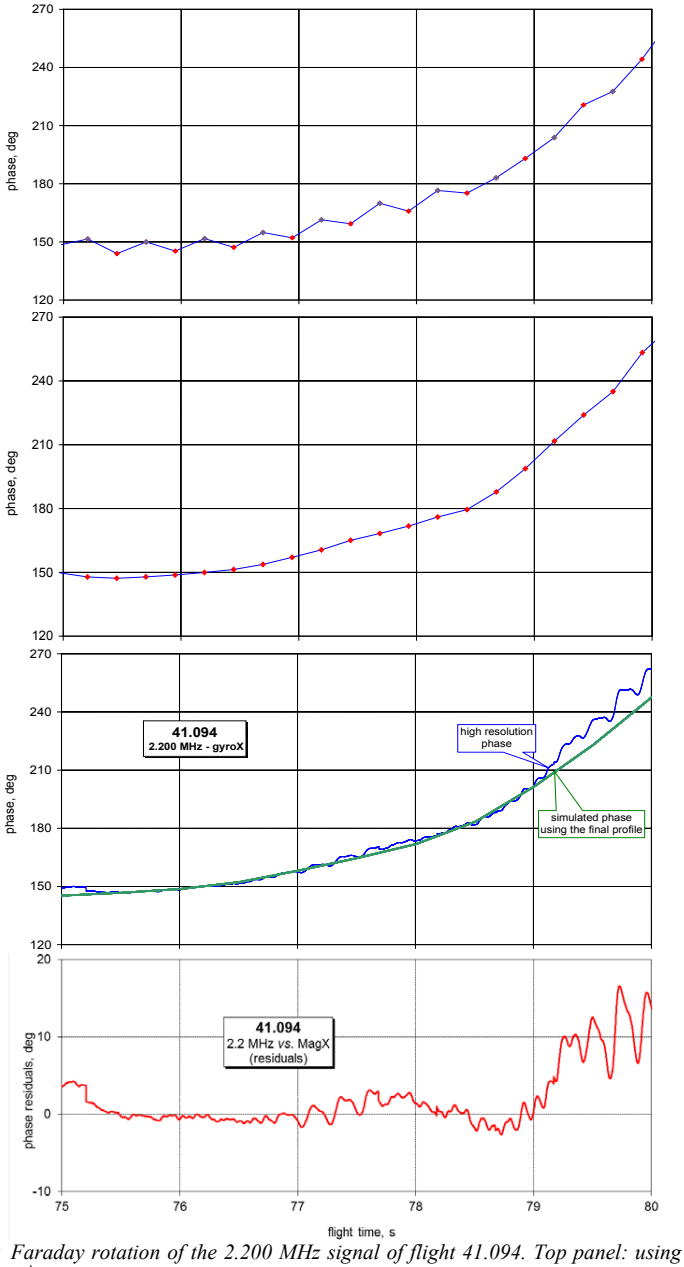


Figure 2.39 Faraday rotation of the 2.200 MHz signal of flight 41.094. Top panel: using every half spin period, 2nd panel: fitting the sine wave over a whole spin period (i.e. two Faraday periods). The time resolution is half a spin period for both cases. 3rd panel: high resolution phase over full spin periods (one telemetry frame increments) is plotted together with the simulated Faraday rotation using the final electron density profile ("Best Guess"). The bottom panel shows the residual phase (measured minus simulated).

The weight, w_i of each data point in an RMS fit should be taken as the inverse square of the error bars σ_i: $w_i = 1/\sigma_i^2$. Although some contributions to the total error bar are constant, in which case no weighting would be necessary, the logarithmic response function of the receiver introduces components of the error bars which are proportional to the signal (think of

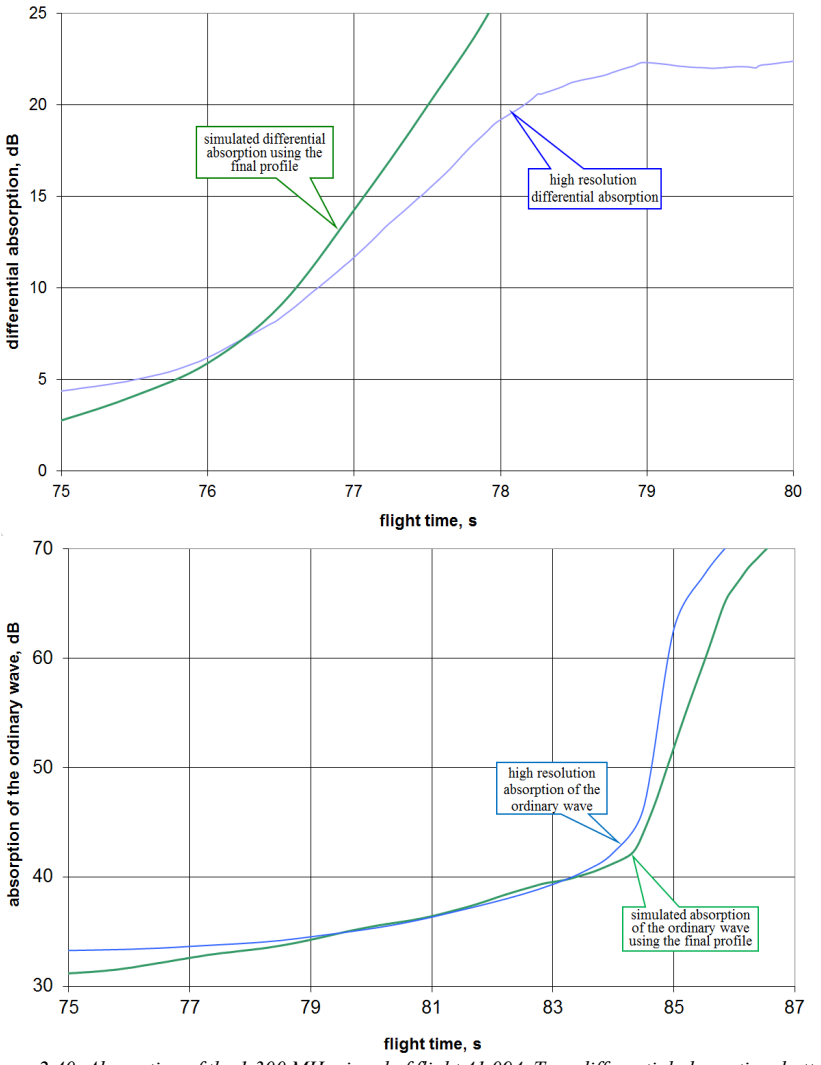


Figure 2.40 Absorption of the 1.300 MHz signal of flight 41.094. Top: differential absorption, bottom: absorption of the ordinary wave. The data are due to the high resolution processing (single frame increments). The simulated raw data are obtained using the final electron density profile ("Best Guess" in Fig. 2.41). In order to reduce the discrepancy between data and simulation one could increase the collision frequency. However, the necessary increase by almost a factor of 2 would imply a similar - and unrealistic - increase of the neutral pressure.

constant deviations from a signal plotted on a logarithmic scale). Assuming that these errors dominate, one should apply weights according to the signal s squared $w_i = 1/s_i^2$. Again, in practice different error sources may show up, so that the optimum weighting function should be established empirically, considering $w_i = 1/s_i^2$, $w_i = 1/s_i$, and $w_i = 1$.

From a fit over one or more spin periods one obtains the time shift of the signal $t = \psi/\omega$ with respect to a reference time best determined by identifying the crossing of smoothed raw data with the mean value. Alternatively, one may apply a two-step process: First find a raw value of this crossing, and then perform a linear fit to the raw data before and after the crossing (where a straight line is a good approximation to a sine curve) and determine the exact time from the crossing between the mean and the linear fit. A second output of these fits to the spin reference is the spin period which is necessary for the fits of the signals in the first place. In theory, the wave data themselves can be used to determine the frequency in a similar manner as from the magnetometer data. If the polarisation rotates fast, the effective period of the wave data will deviate from the spin period by a non-negligible amount and such a fit of the wave data may lead to better results. Typically, however, taking the nominal spin period for the wave data fits introduces only a negligible error. Moreover, the error introduced by determining the period from noisy wave data is more severe than the theoretical advantage of using the correct Faraday period. The intermediate result of all these calculations is the time difference between a sine function fitted to the wave data and the time reference. Figure 2.39 shows the Faraday rotation of the 2.200 MHz signal of flight 41.094 as the phase between the on board gyro and the receiver output. The asymmetry of the receiver output (*cf.* Fig. 2.37, bottom panel, or Fig. 2.38) introduces a phase offset leading to two distinct groups of phase values shown in Figure 2.39 (upper panel). We therefore generally fit the sine wave over a whole spin period (= two Faraday periods), but advance by half a spin period (2nd panel). The ultimate procedure to obtain phase and amplitude information from distorted and noisy receiver outputs is to assimilate the signal by fitting a sine wave over a spin period and to repeat the procedure by advancing by one telemetry sample (with today's telemetry typically 1 or 2 ms). Except for Faraday rotation at the lowest altitudes and frequencies above the local electron gyro frequency f_c (*cf.* Eq. 2.13), the inputs to the magneto-ionic theory have to increase with time (altitude) for the electron density calculations to converge. Consequently, one has to apply some smoothing to the raw data (phase, absorption) before entering them into the calculation. The high resolution curve (bottom panel of Fig. 2.39) shows such reversals of the phase curve. On occasions one can identify the source of the excursions, *e.g.* a particular noise spike which distorts the data for the duration of one spin period.

According to Eq. (2.25) the signal strength of the ordinary wave can be obtained by taking the mean of the reconstructed maximum $\bar{A} = A_o + A_x$ and minimum $\underline{A} = A_o - A_x$. When using this data set for the derivation of electron density the free space attenuation has to be subtracted in order to obtain the absorption of the ordinary wave (AOW).

Smoothing of the raw data can be performed strictly numerically or by drawing an "educated guess" by hand through the data points. Numerical smoothing works well when the scatter is reasonably small and when there are no obvious artefacts such as missed phase discontinuities by 180°. For hand-drawn lines into phase data it is helpful to duplicate the data plot with an offset of 180° to easier identify the new phase range. The input into the magneto-ionic calculation are snapshots of these fitted curves. Often steps of 1s flight time are appropriate, but longer steps may be necessary for heavily scattering data or shorter steps for rapidly rotating polarisation. Under all circumstances using too short step must be avoided, so that the deltas of rotation and absorption reverse sign, as the magneto-ionic formulae can't converge in case of a wrong sign, and moreover produce to misleading results for the following interval where the delta has become artificially high. For the calculation of electron density, we

have to convert the time differences to the relevant quantities dF (change of Faraday rotation per unit length of the propagation path, see Eq. 2.12) and dA (change of absorption per unit path length, see Eq. 2.17). This conversion assumes a purely horizontally stratified ionosphere, using the geometry of the parabolic trajectory of the receiving payload and the position of the ground transmitter.

For each data set and for each time interval (equivalent to a slab in altitude) one applies an iterative procedure: a start value of electron density is assumed and the quantity dF or dA is calculated according to Eq. (2.12) or (2.17), respectively, and compared with the data. The start value of electron density is not critical. The electron density in each height increment is then varied according to standard iteration techniques until agreement with the measurement is achieved. In Figure 2.41 the results of the many raw data sets of flight 41.094 are depicted. Note that more weight is give to Faraday rotation data than those obtained from absorption for the establishment of the line Best Guess. The results of the on-board Langmuir probe (see later Ch. 3.1) roughly follows Best Guess, its modulation is due to the rocket's spin.

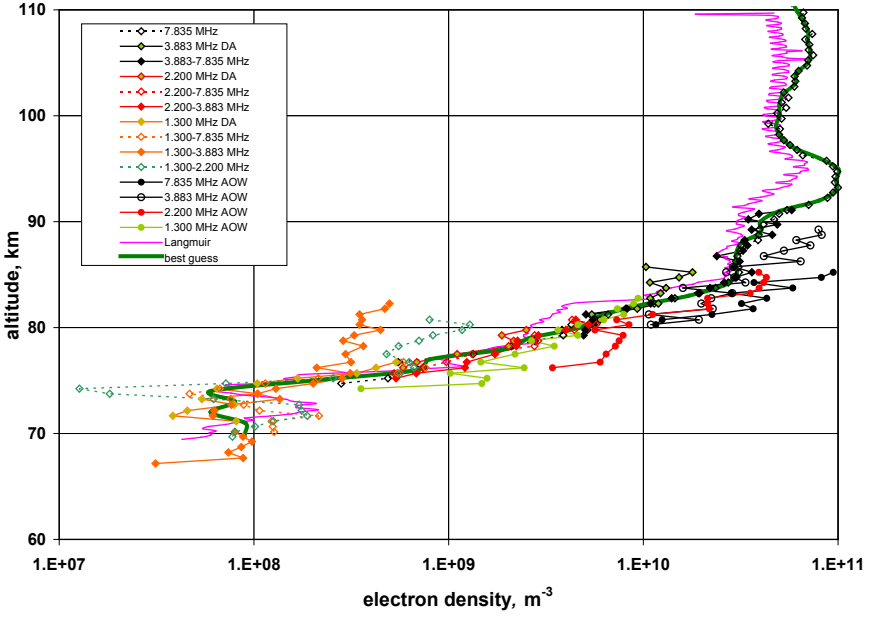


Figure 2.41 Electron densities from the various data sources of flight 41.094. The line "Best Guess" is used to simulate the raw data (cf. Figs. 2.39 and 2.40).

The refractive index is - apart from the magnetic field - a strong function of electron density, whereas the inverse holds for collision frequency. Provided both Faraday rotation and (differential) absorption are available with sufficient accuracy, one can - in principle - derive both electron density N_e and collision frequency ν from the data of only one sounding frequency. Although collision frequency only weakly influences the real part of the refractive index (Faraday rotation), the phase can often be determined more accurately than amplitude (*i.e.* Faraday rotation better than absorption). An iterative procedure to simultaneously derive electron density and collision frequency from Faraday rotation of two different frequencies was developed by Bauer (1972). Because of the weak dependence of Faraday rotation on ν the results scatter widely and this procedure was not pursued further. When the derivation of

collision frequency *v* is aimed for, we simply vary it until the electron densities of the different data sets (mainly Faraday rotation, *vs*. absorption) by and large agree.

Also, the signal received at the payload is the sum of the ordinary and extraordinary wave. For non-vertical transmission into the ionosphere the ray(s) will not be exactly be line-of-sight and ray-tracing should be applied. Since the two wave components experience different refractive indices, the ray paths of the two modes will also be very slightly different and have to be calculated separately. This scenario was taken into account by Bauer (1973), but because of the only very marginal effect, this sophistication was not pursued further.

Summary

- Always more than one sounding frequency should be used, preferably two should still be available at apogee; the results due to the different data sets should essentially agree.
- A frequency below the local gyro frequency will extend the measuring range by a few kilometres to lower altitudes. At night the gain is not so great since measurable electron densities only occur higher up where the impact of the collision frequency has largely vanished.
- Near the (geomagnetic) equator one can rely on differential absorption (with all its shortcomings), or locate the ground transmitter(s) behind the trajectory; this yields Faraday rotation, whereas absorption measurements only provide data from a limited altitude range, irrespective of the sounding frequency.

Pitfalls

- Close to total reflection of a sounding frequency the receiver will measure a standing wave pattern (beat between upgoing and reflected wave). Depending on the actual wavelength, the rocket velocity and the spin rate, this pattern may be akin to the modulation due to the rocket's spin.
- It is always prudent to simulate the input data (rotation, absorption) using the final electron density profile. The simulated values should by and large agree with the raw data ("self fulfilling prophecy"). If not, these particular data should not be used and the procedure be repeated.
- Large electron densities can de-tune the input circuits of the receivers and emulated absorption (for *differential* absorption this does not create a problem).

Outlook

As explained earlier, further reducing the receivers' bandwidth to improve the signal-to-noise ratio is not expedient because of the expected Doppler shift. The height resolution is tied to the rocket's spin rate and can therefore not easily be improved. However, the polarisation can artificially be rotated to obtain more measurement points (*e.g.* Gilchrist and Smith, 1979). Unless the rocket is very accurately stabilised, one - presumably - will in reality still have to form a running mean over a spin period of the many phase values within it. The choice of the highest sounding frequency should not only be determined by the expected absorption, but also the expected Faraday rotation must not exceed 180° within a spin period (undersampling of the phase curve!).

The logistics of integrating the Faraday instrument may be relaxed by using a high impedance, wide-band pre-amplifier located near the antenna which provides low impedance outputs (50 Ω). The tuning (peaking) of the individual receivers in such a scenario can be done in the lab because the actual cable length and the type of antenna does not influence the resonance of the tuned circuit.

Another possibility for improvement is to revive the Seddon experiment of the late 40's (Jackson and Seddon, 1958), however now transmitting the coherent frequencies from the ground to the payload. Contrary to the early experiments one can radiate not circularly polarised waves, but simply use linear ground antennas. One can then observe the "usual" Faraday rotation (with the height resolution determined by the rocket spin), and in addition compare the phase of pairs of (coherent) sounding frequencies. This increases the complexity of the flight hardware, but yields two sets of data. Due to the rotation of the antenna the phase of the received signal will alternate by 180° twice in a spin period, on top of the phase variation due to the electron content between ground and antenna.

At equatorial latitudes the "usual" Faraday rotation arrangement can not be employed, because there is (almost) no component of the Earth's magnetic field in the propagation direction. The Seddon experiment mentioned above does not rely on a magnetic field. Alternatively, one can position the ground transmitters not - as usual - close to the launch site, but rather located *e.g.* 50 or 100 km behind the anticipated rocket trajectory. The ray(s) to the payload will thus have a component of the Earth's magnetic field in the propagation direction. This was for the first time successfully applied in a flight from Kwajalein in 2004 (Friedrich *et al*., 2006). The results from measurements close to the altitude of total reflection must be taken with caution since there the propagation may no longer be strictly line-of-sight. Also, the measurements no longer represent a vertical profile, but have additional susceptibility to horizontal variations of the ionosphere.

At any altitude there will be electron density results from more than one data set. Consequently, one usually draws an intuitive Best Guess profile through the redundant results. Input values simulated using this Best Guess profile should at least not contradict the raw data. An attempt to objectify the establishment of the Best Guess a procedure was developed by which the electron density profile was assimilated by a (logarithmic) polynomial function (Kropf, 1978). Since electron density profiles can have almost any shape, approximation by analytical functions is problematic. Probes, on the other hand, have a good height resolution (one "good" value per spin period or better), but are uncertain in their absolute values. One usually normalises profiles obtained from probes by a constant factor relative to the wave propagation results. A realistic possibility to objectively obtain Best Guess profiles is to use a height dependent probe normalisation factor of low mathematical order and vary the coefficients for best agreement of the simulated wave propagation raw data.

3. Probes

The primary information obtained from the method described in the previous chapter is electron content, and only its derivative is electron density. This method works fine as long as one can assume that the underneath ionosphere is stable for the duration of the measurement. Alternatively, one can measure plasma parameters *in situ* which has its pros and cons which will be listed in the summary to this chapter.

3.1. Langmuir Probe

According to the basic derivation by Smith (1969), the electron current density i to a cylindrical probe, biased with a voltage V *vs*. the ambient plasma is:

$$i = i_0 \sqrt{1 + \frac{eV}{kT}} \tag{3.1}$$

where $i_0 = \frac{N_e e v_{th}}{4}$ is the current density to a stationary probe, and $v_{th} = \sqrt{\frac{8kT}{\pi m_e}}$ is the thermal velocity of the electrons. For $eV / kT >> 1$, the current I collected by a cylindrical probe with length l and diameter d becomes independent of temperature:

$$I = N_e l d \sqrt{\frac{e^3 V \pi}{2 m_e}} \tag{3.2}$$

A thin cylindrical probe is thus largely unaffected by the actual electron temperature, but is sensitive to changes of V due to changes of the payload potential. According to the above relation the collected current depends predictably on the probe potential *vs*. that of the ambient

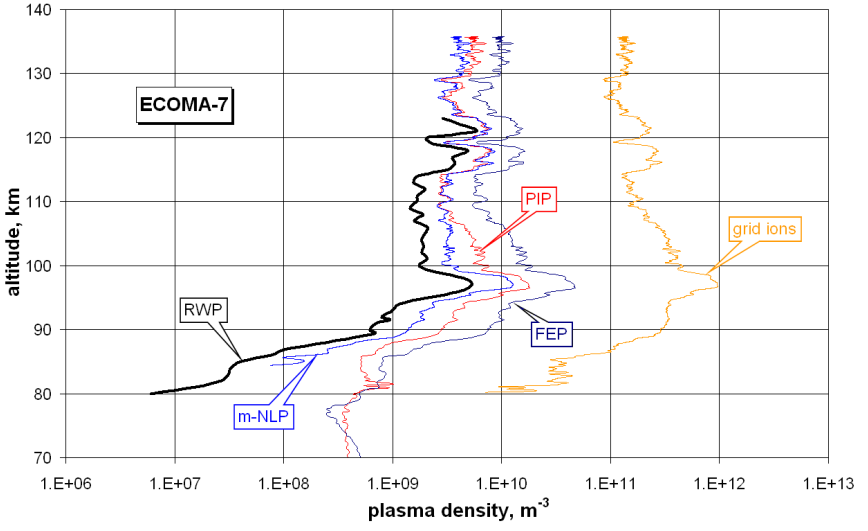


Figure 3.1 Test of various plasma probes. Note that the profile by the multi-needle (Langmuir) probe (m-NLP) comes closest to that of the wave propagation experiment (RWP); Friedrich et al. (2013a).

plasma. Hence, by using two or more probes with different biases the uncertainty of the payload potential can be eliminated. This idea was pursued by Bekkeng *et al*. (2010) and tested against

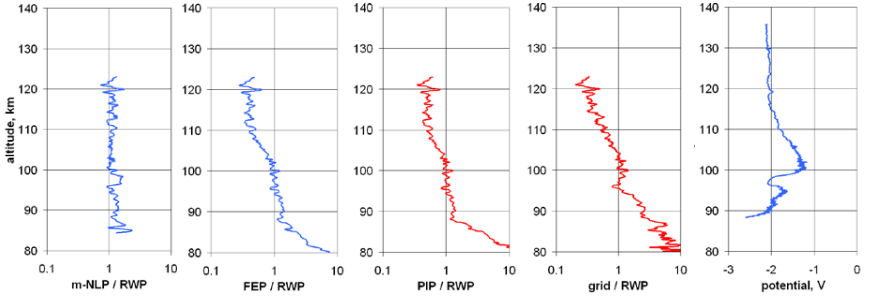


Figure 3.2 Residual deviation from the RWP profile as a function of altitude of the various probes together with the measured payload potential (flight ECOMA-7, Friedrich et al., 2013a). Note that the normalisation factor of the m-NLP profile is almost constant vs. altitude.

absolute electron densities derived by the wave propagation experiment aboard the same rockets (Friedrich *et al*., 2013a). By putting the probes on long enough booms one expects to sample the plasma undisturbed by the shock cone of the fast moving payload. In Figure 3.1 the results of one of the seven investigated cases is shown. Importantly, the multi-needle probe not only yields electron densities closest to the "real" values (Best Guess), but also there is almost no height variation of the normalisation factor (Fig. 3.2).

3.2. Ion Probe

For charge neutrality the number density of negatively charged particles (electrons and negative ions, on occasions also negatively charged dust) must be balanced by positive ions; if negative ions can be ignored (or their number densities are expected to be negligibly low, *cf*. Ch. 5.4), the number density of positive ion density can also be used as a proxy for electrons. The lightest ions in the lower ionosphere are NO^+ and O_2^+ with masses 30 and 32 amu, respectively, and typical temperatures range from 100 to 300 K. For thermal equilibrium - which we assume as one of the characteristic properties of the lower ionosphere - the resulting thermal velocity $v_{th} = \sqrt{8kT/\pi m}$ ranges from 250 to 460 m s^{-1}; these values have to be compared to typical rocket velocities in the mesosphere of 1 to 2 km s^{-1}. A gridded sphere at plasma potential with a negatively biased collector inside will draw a current predominantly determined by the rocket velocity times the sphere's cross-section and the grid's transparency; Figure 3.3 depicts such a probe. The thermal velocity of the ions contributes only marginally to the positive current collected by such an arrangement. Figure 3.4 shows the current ratio (over the one due to the thermal velocity alone) as a function of the probe (rocket) velocity over thermal velocity. One can see that for velocities only twice the thermal velocity the collected current is practically proportional to the rocket velocity. For the computation of ion density from the collected current, v_{th} is calculated to the best knowledge of ion mass m and temperature T. The region where v_{th} may vary largely is below typically 80 km, *i.e.* where the ions may be light ones (NO^+, O_2^+) or many times heavier (water clusters). Fortunately, usual sounding rockets pass this region in excess of 1 km s^{-1}, *i.e.* much larger than v_{th} and a linear relation between (the known) rocket velocity and the collected current holds, and an exact knowledge of v_{th} is not critical (Fig. 3.5).

With few exceptions, the preferred location of most diagnostic instruments is in the central ram position. The ion probe displayed in Figure 3.6 occupied that position which

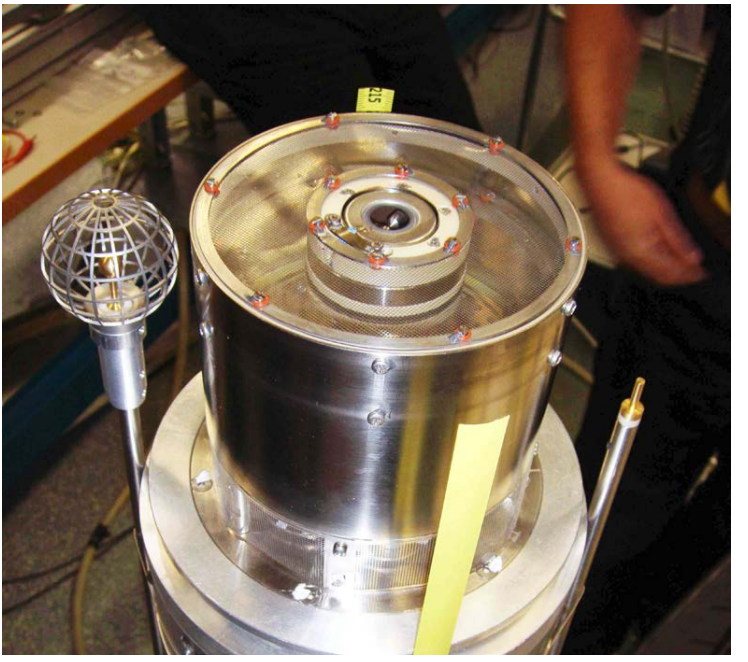


Figure 3.3 Gridded sphere at payload potential with a negatively biased collector for use as an ion probe (payload ECOMA flown between 2006 and 2010 from Andøya). The central instrument is the particle detector with a shielded grid at a negative potential (which therefore also collects positive ions).

resulted in a smooth ion current, virtually without spin modulation. However, without attitude control on the downleg a probe in that position is in the wake and the data taken on the downleg are not comparable to the upleg measurements. More often the probe will be on a deployable boom (*cf.* Fig. 3.7) extending to the side of the payload. The current measured by such an arrangement will generally be modulated by the payload's spin due to the probe's motion in and

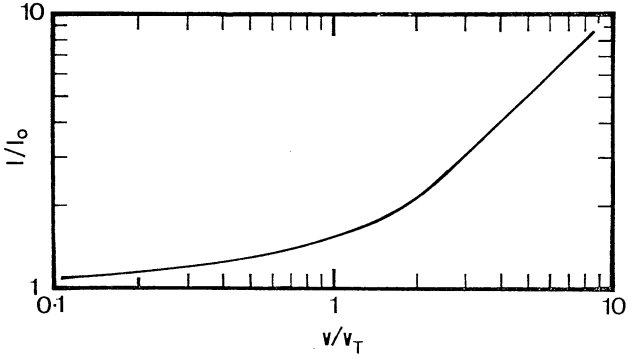


Figure 3.4 Ratio of the ion current to a stationary probe (caused by the thermal velocity v_{th}) to the current actually observed when the probe (rocket) moves with a velocity v (Folkestad, 1970).

out of the shock (Mach-) cone. The usual approach is to consider the highest value in a spin period to represent the undisturbed ion density. While this method ignores the possibility that the probe never reaches the undisturbed environment such that the highest values stem from the enhanced density inside the shock, this approach usually provides reasonable results. In Figure 3.8 the ion current of the whole trajectory (up- and downleg) is shown. The depth of the spin modulation is smallest in the initial part of the flight, *i.e.* when the velocity vector is nearly aligned with the rocket (spin) axis and the probe is always outside the shock cone (or at the same location in the cone). Near apogee the angle of attack is from the side (the payload moves horizontally) and the wake (shadow) within a spin period is most pronounced. Using only the current maxima in each spin period the height resolution is reduced to one value per spin period, but - as Figure 3.9 shows - the downleg data agree excellently with the ones obtained on upleg.

The basic assumption is that the gridded sphere simply sweeps through the plasma and thus collects a current determined by the probe's cross section and the rocket velocity. The bias applied to the collector only serves the purpose of rejecting the unwanted species. However, a more detailed analysis reveals that not all heavy ions (clusters) may be collected for a given bias. Figure 3.10 a trajectory of an ion is depicted for a realistic probe geometry (25/5 mm diameter)

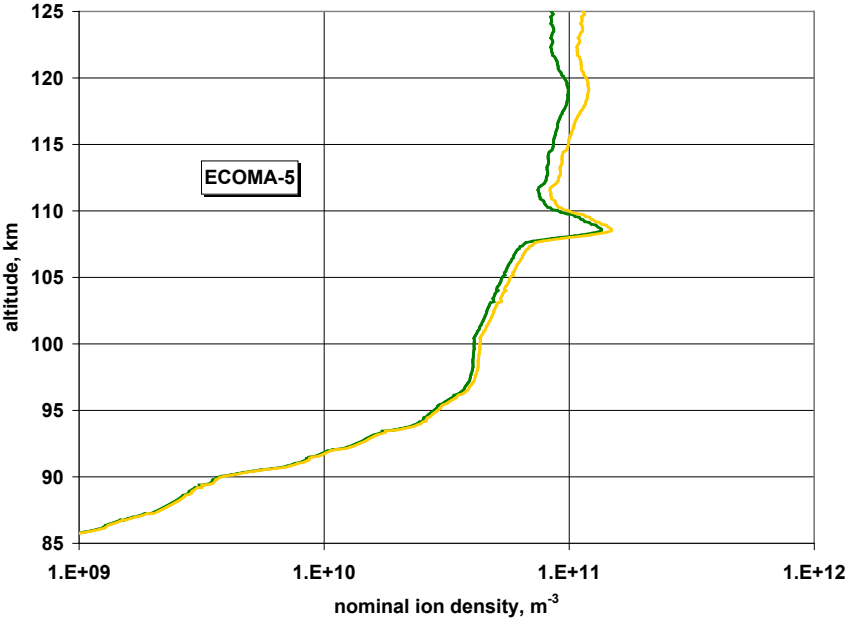


Figure 3.5 Nominal ion densities obtained by assuming two extremes of ion masses (30 and 70 amu). Cleary the actual ion mass only matters near apogee when the rocket velocity is slow, but above ca. 90 km ion masses heavier than 32 amu do not exist.

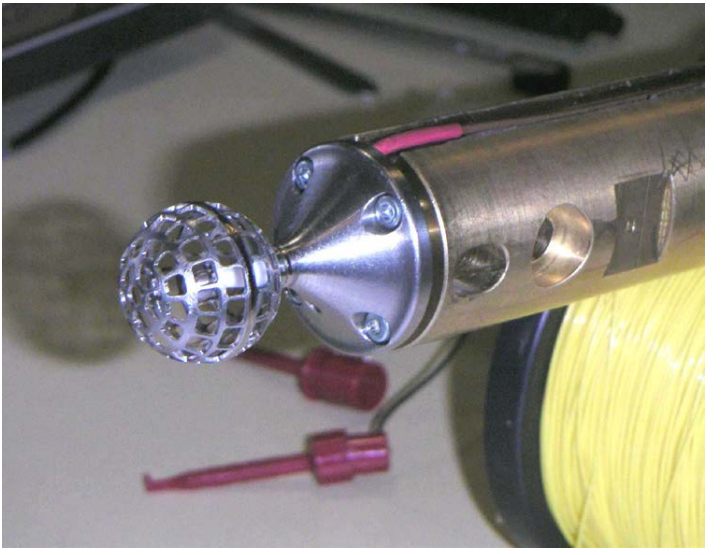


Figure 3.6 Centrally mounted ion probe (MiniMIDAS payload, launched from Andøya with a Viper-III motor, January 18th, 2005).

Figure 3.7 Ion probe mounted on a telescopic boom shown partially deployed (payloads 31.102 to 31.105, launched from Alcântara, Brazil, August 1994).

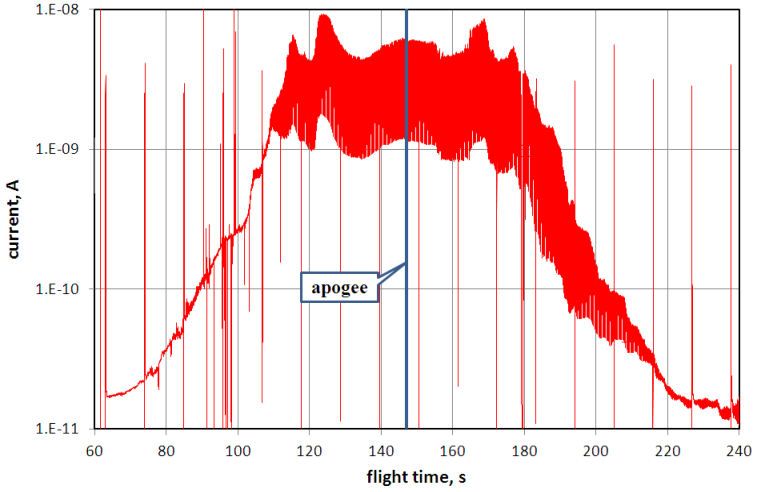


Figure 3.8 Quick look of the current measured by an ion probe on a sideways deployed boom. Note that the depth of the spin modulation is smallest in the initial part of the flight when the rocket axis is aligned with the velocity vector (payload ECOMA-0, October 28th, 2004, Esrange, Sweden).

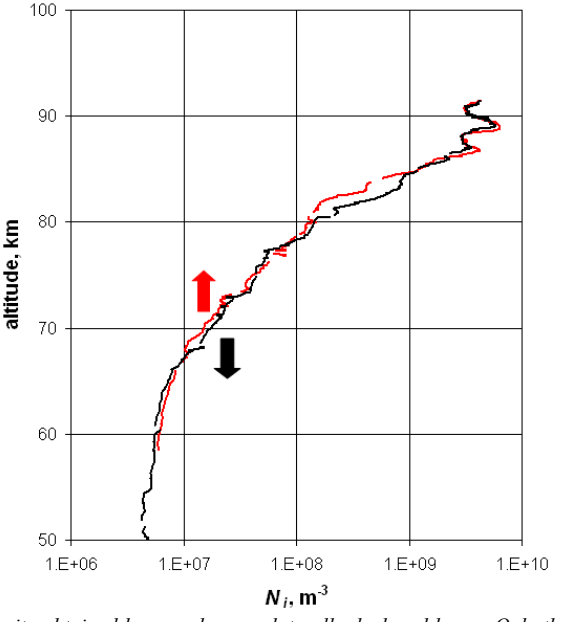


Figure 3.9 Ion density obtained by a probe on a laterally deployed boom. Only the maxima in each spin period are used. Note the excellent agreement between up- and downleg (red and black, respectively), payload ECOMA-0, October 28th, 2004, Esrange, Sweden; cf. the previous figure.

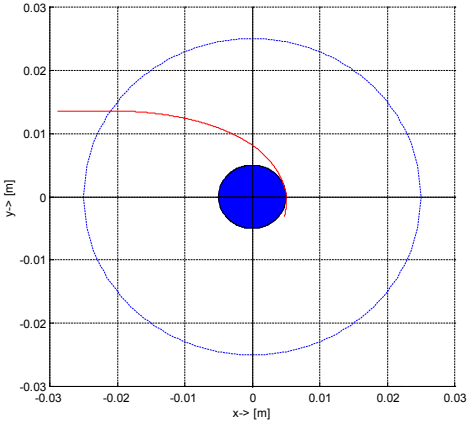


Figure 3.10 Trajectory of an ion "barely" being collected (after Pürstl, 2000).

The effective probe cross section is a function of the bias voltage and varies from that of the outer (gridded) sphere when all positive species are collected, to that of the inner collector for small biases when only positive species directly hitting the collector are measured.

More general, the mass discrimination is shown for a probe with 25 mm outside diameter, 5 mm diameter of the collector sphere, a velocity of 1 km s^{-1} and various biases between -2.5 and -80 V (Fig. 3.11). Clearly there exist maximum masses for full collection tied to the bias employed. Beyond these limits the efficiency decreases asymptotically to the one defined by the diameter of the collector sphere. The mass limit up to which all ions are collected is displayed in Figure 3.12 as a function of bias voltage. Conclusions to be drawn from this mass discrimination issue are: A large bias will collect positive ions of all conceivable masses, but the larger electric field on the collector surface may lead to secondary electrons by impacting ions. The electric field on the inner sphere can be reduced by making it larger; on the other hand, this creates aerodynamic problems (stagnation). The preferred geometry therefore consists of a collector sphere almost as large as the outer gridded sphere, but also gridded to allow free flow of the relatively dense background atmosphere. For such a geometry the actual

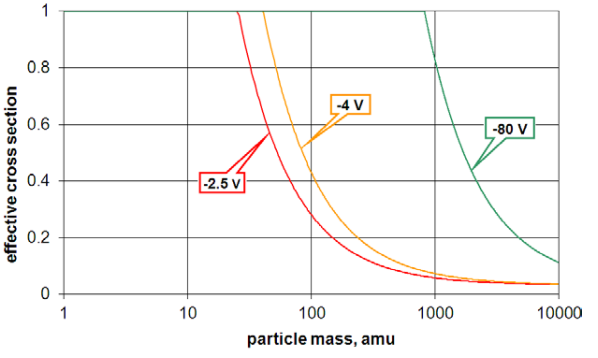


Figure 3.11 Mass discrimination of a gridded spherical ion probe (25/5 mm diameter) at 1 km s^{-1} and different biases (after Pürstl, 2000).

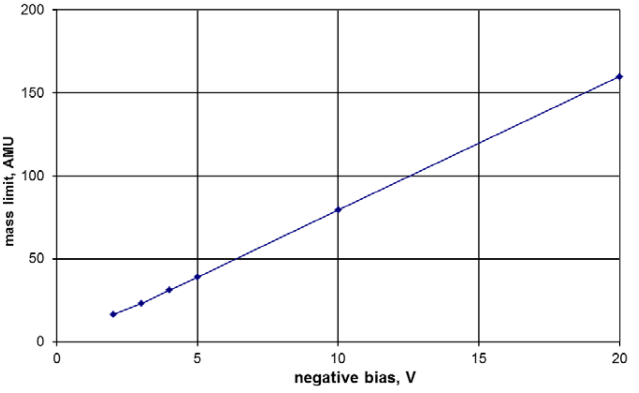


Figure 3.12 Mass limit for complete mass collection as a function of collector bias and the geometry given in Figure 3.11 (after Pürstl, 2000).

bias is not very critical, in the worst case (heavy ions) the effective diameter is that of the inner sphere which is only a little smaller than the outer one. The series of MIDAS payloads used an arrangement with 120 and 100 mm diameter probe (*cf.* Fig. 2.26) which yielded nominal ion densities much closer (to within 10%) to absolute ones than any other gridded probe (*i.e.* calibrated against the wave propagation experiment), whereas for other probes the normalisation factor can be as large as 20. Apparently the large diameter also helps to make these probes less sensitive to payload charging.

A convincing case demonstrating the discrimination of heavy masses is provided by the data of rocket ECOMA-7 (Fig. 3.13). The main instrument of the ECOMA series of sounding rockets was a Faraday cup aimed at measuring charged aerosols (meteoric dust). The negatively charged dust appeared between 82 and 94 km. The number density of negative charges measured by a spherical gridded probe showed little more density that of electrons, whereas the

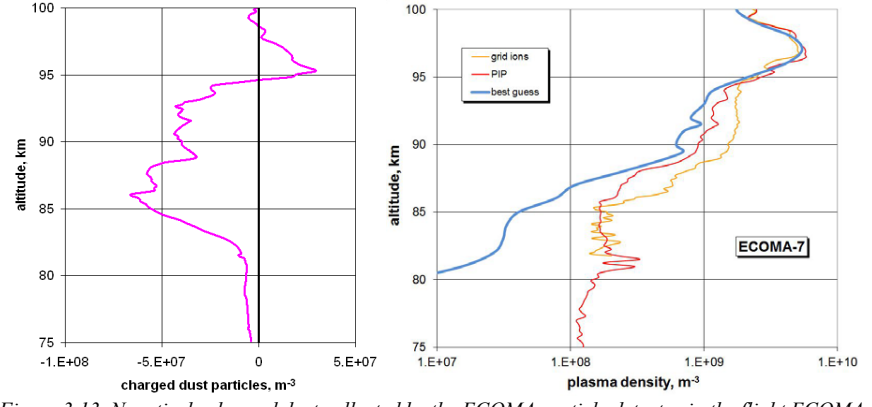


Figure 3.13 Negatively charged dust collected by the ECOMA particle detector in the flight ECOMA-7 (left panel). Electron and ion densities measured in the same flight are displayed in the right panel. Note that in the height region where the ECOMA instrument detects negative dust, the shielding grid of that instrument measures distinctly more positive charges than the gridded-sphere ion probe (Friedrich et al., 2012).

negatively biased, exposed shielding grid in front of the particle detector (Fig. 3.3) measured a current distinctly larger in the height region where negative charged dust was detected. This suggest that the exposed grid collected essentially all positive charges irrespective of their mass. This is supported by the fact that above 94 and below 82 km the densities obtained by the spherical gridded probe and the exposed grid converge (Friedrich *et al*., 2012).

Except for a few degrees, in flight the axis of the sounding rockets generally maintains the elevation it had when on the launcher. Therefore, on upleg the air flow is roughly aligned with the rocket axis, whereas on downleg the angle of attack is close to 180°. The CHAMPS payloads (41.093 and 41.094) launched in 2011 from Andøya were exceptional cases because their rocket axes were actively controlled to be always aligned (to better than 2.3°) with the anticipated velocity vector. The payloads carried two PIPs on sideways deployed booms opposite to each other. For such a scenario one should not expect any spin modulation of the measured ion current. Rather surprisingly, however, the currents of the two probes were slightly modulated (<10%) with the spin frequency in antiphase to each other. We can rule out an instrumental effect for the observed modulation (depth, and phase relative to the on-board gyro) for the following reasons:

(a) The phase *vs*. altitude is symmetric to apogee. *i.e.* the pattern repeats on downleg,

(b) The observed phase of the two identical payloads, but flown two days apart, is completely different (Fig. 3.14).

The best explanation is that the payload's attitude control aligns the rocket axis along the vector of the velocity relative to the ground, but with no relation to the ambient air. We therefore explain the residual spin modulation by horizontal winds which in the mesosphere can reach 100 m s^{-1} or more. Since the collected current is proportional to the relative velocity of the probe to the ambient plasma, the extra velocity component introduced by horizontal winds changes the collection efficiency of the probes (probe spinning against or with the horizontal wind; Friedrich *et al*., 2013b).

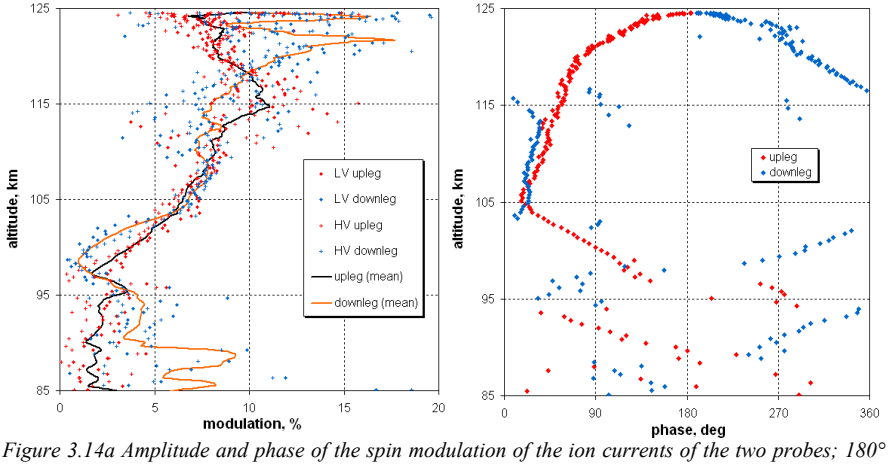


Figure 3.14a Amplitude and phase of the spin modulation of the ion currents of the two probes; 180° are added to the phase of the opposite probe (night-flight 41.094).

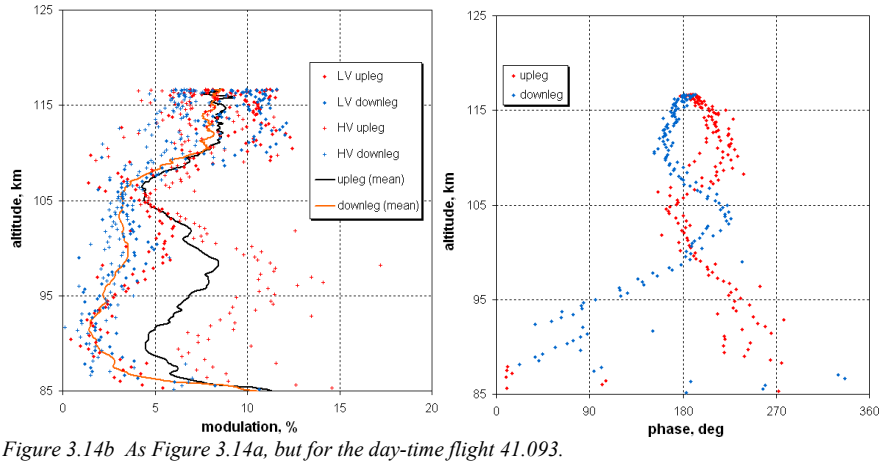


Figure 3.14b As Figure 3.14a, but for the day-time flight 41.093.

Figure 3.15 shows the schematic of the circuit that has been used since 1994. Because of the large required dynamic range (plus the uncertainty of the normalisation), a logarithmic amplifier is used. Although according to the data sheet the amplifier IC is only specified down to 1 nA, extending the range to 10 pA does not seem to deteriorate the linearity or stability. The negative bias for collecting positive ions is achieved by putting the amplifier's GND (pin 1) to

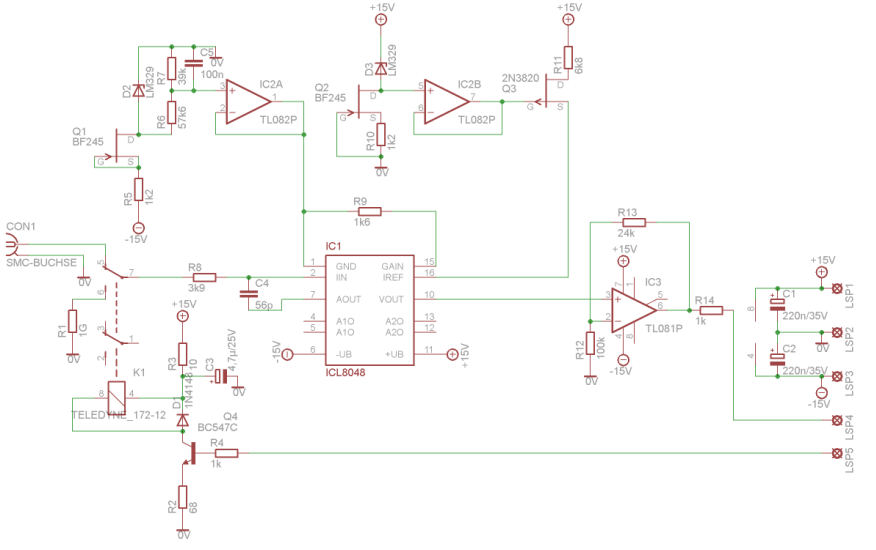


Figure 3.15 Schematic for a probe to measure positive ions.

a -2.5 V source. To be able to check the stability in flight, at specified times the input is connected to an internal calibration resistor (1 GΩ) providing a 2.5 nA calibration current.

3.3 Capacitance probe

Electrostatic probes, such as Langmuir types, always pose the risk of collecting too much electron current that can not be compensated by an equally large return current to the payload provided by positive ions from the ambient plasma; this leads to negative charging of the payload. Another method to measure electron densities *in-situ* makes use of the fact that plasma surrounding a conductor changes the thickness of the sheath and thereby its free-space capacitance. The easiest method is to include a conductive sphere as part of an LC oscillator. The amplitude (AC only!) can be kept very low (*e.g.*, 0.5 V_{pp}) and therefore does not interfere with other diagnostic instruments, let alone impact on the payload potential. The ambient electron density will increase the capacitance and - in consequence - lower the frequency. Such "fixed"-frequency (*i.e.* non sweeping) capacitance probes have been flown successfully many times in the past (Jacobsen, 1972). In ten flights since 1991 we have used the outer grid of the ion probe as the fixed frequency capacitance probe (connected by 100 kΩ to the payload for proper operation of the ion probe). This dual use of one probe to measure two parameters has not led to any noticeable mutually detrimental effect.

The threshold of the measurements depends to some extent on the background collision frequency, itself proportional to the atmospheric pressure (Jacobsen, 1972). Figure 3.16 shows the pressure at which useful measurements began from ten flights as a function of probe frequency (the correlation coefficient for the same analysis with altitude is a little lower). Despite the poor correlation it appears that lower frequencies will be useful at lower altitude (larger collision frequency or pressure).

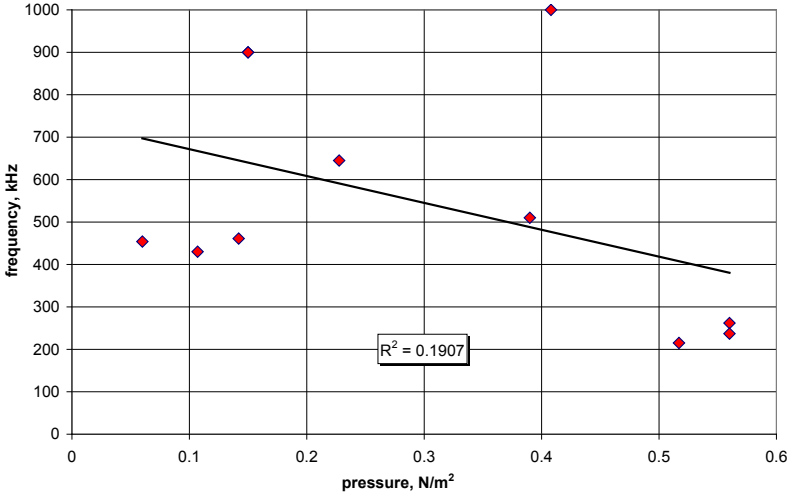


Figure 3.16 Pressure at the beginning of useful measurements with capacitance probes of different frequencies.

We use a modified Clap oscillator in which the sphere is part of the frequency-determining LC-circuit. The changes in capacitance and frequency are minute (<1%); we measure these small changes by dividing the frequency to a frequency slightly above the sampling rate of the telemetry. Using 100 MHz or more we then measure the period of the scaled probe frequency. An up-to-date version of a combined ion-capacitance probe using 250

MHz to determine the period of the scaled oscillator frequency was developed by Hackl (2014) in his Master Thesis. This instrument was developed for the flights of the rocket payloads MaxiDusty flown in Summer 2016 from Andøya. Two probes are used, one with a low bias (for the ion collection) and a low frequency, the other with a somewhat higher bias and a higher frequency of the LC oscillator. The experience gained in these flights will help to decide on the design of future probes.

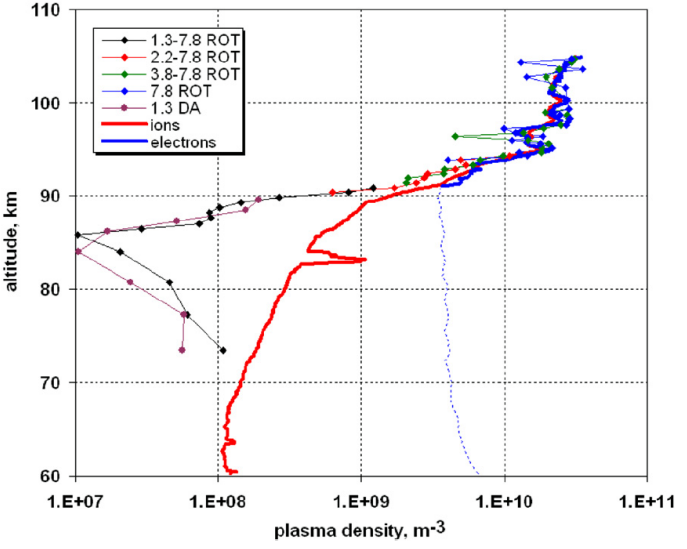


Figure 3.17 Plasma densities obtained from wave propagation data, and due to the combined ion and capacitance probe aboard the payload Decimals-B (August 1st, 1993, Esrange). Note that useable measurements by the capacitance probe only commence above 91 km.

3.4. Probes, Pros and Cons

The current measured by a positively biased cylindrical probe in a plasma is to a good approximation proportional to the square root of the bias voltage relative to the ambient plasma (Smith, 1969). The sensitivity of DC (Langmuir-) probes to payload charging can be

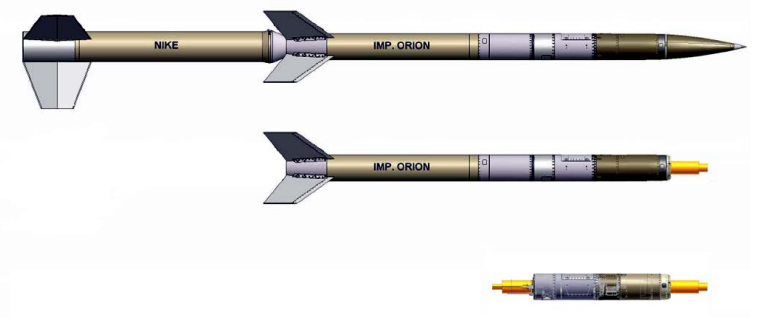


Figure 3.18 Configuration of the ECOMA payloads. Top: at launch, centre: at the beginning of the measurements (ca. 60 km), and (bottom) after payload separation (>70 km).

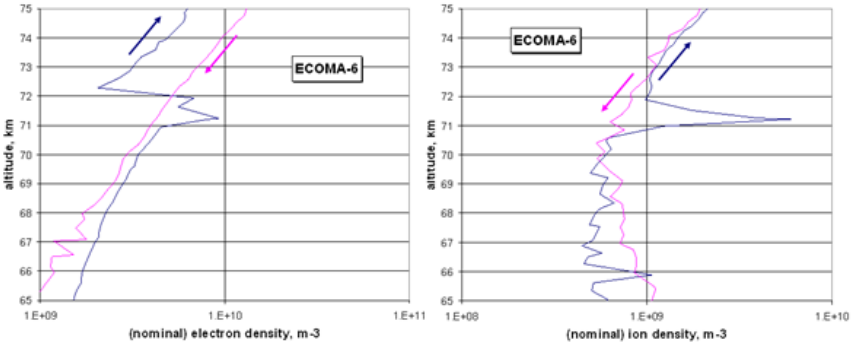


Figure 3.19 Currents of the electron and ion DC probes (left and right, respectively) during payload separation on upleg. Note that no discontinuity occurs on downleg.

demonstrated by the data of the flight of the payload ECOMA-6 (Fig. 3.18). During the first few seconds of measurements the second stage (the motor) was still attached to the payload. At payload separation the current to the Langmuir probe dropped significantly, whereas the ion probe's data are essentially continuous (apart from a transient excursion). The downleg data from the same altitude shows no discontinuity which supports the hypothesis that the reduced surface (reduced return ion current) of the payload is responsible for a negative charging which clearly affects the electron probe more than the ion probe (Fig. 3.19).

instrument	***pro's***	***con's***
wave propagation	- absolute values - not affected by aerodynamic effects or payload charging - only minor to negligible uncertainties due to unknown propagation effects (*e.g.* non line-of-sight propagation)	- poor height resolution - requires temporal stability of N_e during measurement time - therefore, not too good beyond the *E*-region peak (in a valley) - no horizontal stratification of N_e allowed
positive ion probe	- not very sensitive to payload charging - one can infer N_e from N^+ - does not generate a stray field	- slow when currents are low - probe must reach outside shock cone - discriminates heavy masses
Langmuir probe	- collects larger (= easy to measure) current	- sensitive to payload charging - requires large surface for return current - generates an electric field - conversion to electron density also depends on electron temperature and payload potential
capacitance probe	- very fast with high resolution - does not affect the payload potential - produces only a small stray field - can be combined with an ion probe	- readings also depend on electron temperature and payload potential - only operates properly at low collision frequency (> 80 to 90 km)

4. Ionospheric Models

The optical thickness (*i.e.* proportional to pressure) determines the flux of extra-terrestrial, ionising radiation reaching the height under investigation, whereas electron-ion recombination is tied to neutral number density and temperature. Except for sophisticated theoretical ionospheric models which also compute the neutral atmospheric bulk parameters (*e.g.,* Solomon *et al*., 1982, or Kazil *et al*., 2003), for all others - both theoretical and empirical - knowledge of the neutral atmosphere constitutes an important prerequisite. The internationally recognised atmospheric model is CIRA, however MSIS-90-E has recently been extended from the thermosphere to the ground and thus for all practical purposes replaces CIRA. This new NRLMSIS-00 (Picone *et al*., 2002) is available in analytical form and includes diurnal variations; the magnitude of tides in the mesosphere is subject of debate and probably largely an extension from higher altitudes. On the other hand, CIRA is available in tabulated form and essentially represents a diurnal average.

For the European auroral and polar latitudes (*i.e.* geographic >70° N), it was found that both these standard models are inadequate, notably that their mesopause temperatures are not low enough to enable the formation of noctilucent clouds. From the Andøya rocket range a number of falling spheres were launched to derive temperatures, and from the co-located lidar installation ALOMAR temperatures predominantly from winter are available. The climatologies obtained from these measurements are as yet restricted in their seasonal and altitude coverage. Since we seek an atmospheric background which constitutes an improvement over CIRA or MSIS we have to cover all seasons. For this purpose, we use the presently published data (Lübken and von Zahn, 1991; Lübken, 1999) and form the ratio to the (established, but inadequate) CIRA model; perturbations due to tides or gravity waves are conveniently ignored. The deviations (factors) of the various atmospheric parameters (temperature, density) are then plotted *vs*. season for each kilometre (for an example of one height see Fig. 4.1); although in the case of temperatures one could use differences to define the departure of the measurements from the model, for simplicity we use factors as in the cases of density and pressure where it is a necessity. These deviations are assimilated by low-order spline functions which wrap around at the ends of the year and in turn are smoothed with altitude by forming a running mean (25, 50, 25%) with the corresponding results from the layers above

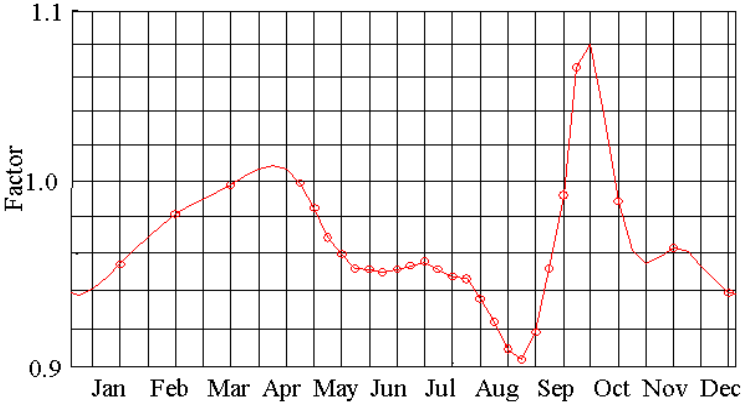


Figure 4.1 Factor between locally measured temperatures and MSIS vs. season at a height where the difference is most pronounced (89 km, 70°N). In the further processing smooth lines obtained with the spline function are used (Friedrich et al., 2004a).

and below. The temperature deviations are forced to approach the original reference (CIRA) above the height of the local measurements, *i.e.* by linear interpolation of the correction factor between the measured value at 90 km and the assumed unity at 105 km. Figure 4.2 shows a contour plot of the correction factor for temperature. The resulting temperature field of this cautious approach, together with MSIS, is depicted in Figure 4.3. It reveals that temperatures at the mesopause are lower and that transitions from and to summer occur more rapidly than in MSIS. Lübken and von Zahn (1991) give a table of derived mass densities from spring to

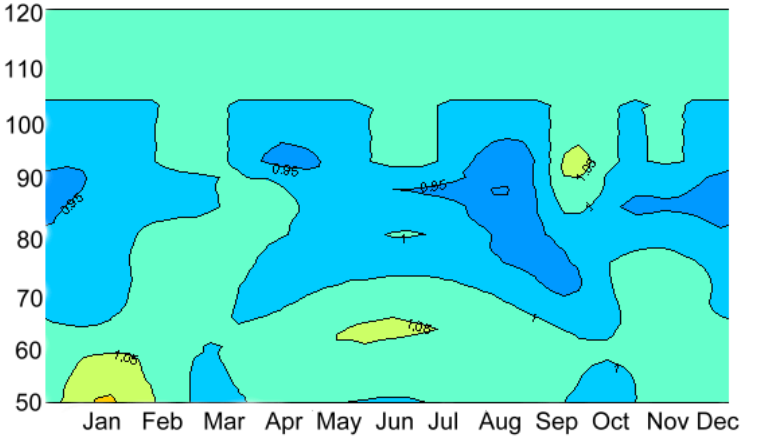


Figure 4.2 Lines of constant correction factors for temperatures at 70°N. Because of the lack of data, the factor is forced to approach unity at 105 km (Friedrich et al., 2004a).

autumn, whereas in MSIS pressures are given. We first convert the former using the ideal gas law and the mean molecular mass of CIRA-72 and then form the ratio to the latter. Returning to the reference density above the height of the measurements would in a physical interpretation mean unrealistic, including negative, temperatures. We therefore maintain the factors established at 90 km into the thermosphere. Figure 4.4 shows the pressures thus obtained together with the variation according to MSIS. It may be added here that using MSIS as the basis essentially yields the same results. The validity of the pressure surfaces obtained by this method at altitudes above 100 km is certainly limited by the lack of local data. In the present

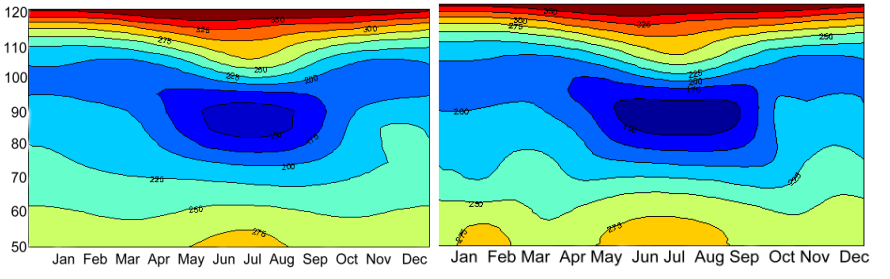


Figure 4.3 Temperature field in K at 70°N according to MSIS (left) and applying the correction field of Figure 4.2 (right). Note that (a) the summer mesopause is colder, and (b) that the seasonal transitions occur more rapidly (Friedrich et al., 2004a).

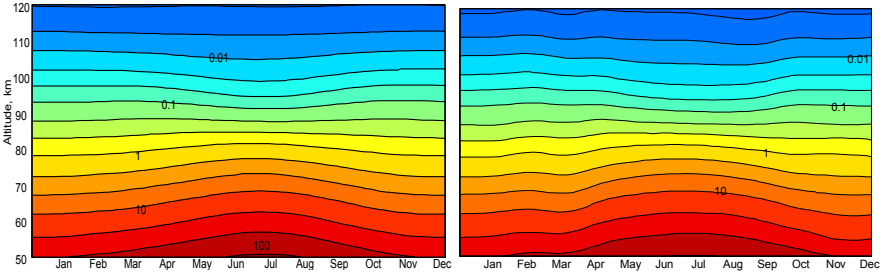


Figure 4.4 Lines of constant pressure in Pa for 70°N according to MSIS (left) and corrected by local measurements (right). The seasonal variation in the thermosphere is essentially an extension of the one at 90 km (Friedrich et al., 2004a).

context, pressure surfaces are used to arrange electron densities, which generally show only small gradients above the *E*-region peak. The derived pressures are most relevant for the analysis of electron densities in the *D*-region, *i.e*. below 90 km where our pressure model is supported by a sufficient amount of measurements.

For number and mass density, we employ the equation of state and the temperatures depicted in Figure 4.3. The extrapolations (>90 km) generally agree with the averaged densities given by Rapp *et al*. (2001) for altitudes up to 110 km above Andøya. Since our emphasis is at lower altitudes (<100 km), for simplicity we use a constant atmosphere at 160 km independent of season and solar activity (800 K, 3.27×10^{-4} Pa). The neutral atmosphere presented here is by no means final and awaits more data from the on-going endeavour for satisfactory coverage; nonetheless, the present cautious approach of establishing an atmosphere defined as a departure from an accepted standard yields reasonable values, yet distinctly different from CIRA or MSIS.

4.1 Theoretical Models

There are various levels of sophistication in modelling the lower ionosphere:

(*a*) One assumes the neutral atmosphere including minor constituents as given. The six-ion model by Mitra and Rowe (1972) is an early example of this approach. A little more recent is the 17 species version by Torkar and Friedrich (1983a). At the time of establishing that model, it was concluded that an exact knowledge of the inputs such as the minor constituents and the solar X-ray fluxes was probably more important than a more complex chemistry.

(*b*) As above, but using a time-dependent calculation. The model by Ogawa and Shimazaki (1975) only shows appreciable differences close to sunrise or sunset.

(*c*) The model by Solomon *et al*. (1982) not only considers the ion chemistry, but also that of the neutrals, as well as transport and tides, and uses a time-dependent calculation. Probably the most recent models of this degree of sophistication are UBAIM (Kazil *et al*., 2003), or SIC in its latest version (Enell *et al*., 2005).

More recently the potential importance of meteoric dust for the loss of free electrons by attachment to these particles is being considered. As the evaporated dust sediments it coagulates and forms a layer whose altitude is predominantly tied to the background neutral pressure. This effect is proposed to explain the ubiquitous electron density ledge, both during the day, but most pronounced at night (Williams *et al*., 2016). Also, recently the absorption cross section of the Earth's major constituents for X-rays has come under scrutiny. A more rigorous treatment

of the cross sections may (will) explain the electron density deficit when modelling the *E*-region (Evans *et al*., 2015).

4.2. Empirical Models

The eventual result of empirical modelling is convergence with theoretical models. The best procedure to create an empirical model is to bin the data for all situations. In reality, however, one is faced with the situation that one has to interpolate - or even extrapolate - for geophysical conditions insufficiently covered by measurements. Hence, also for empirical modelling, one needs some basic qualitative understanding of the expected dependencies on the various geophysical parameters of the behaviour of plasma densities; the empirical model will then provide the quantitative results.

Solar zenith angle χ: The integrated number density of molecules between the altitude under investigation and the Sun varies (roughly) as $1/\cos\chi$ (due to the path length through the overhead atmosphere). Hence also the intensity of an ionising flux of extra-terrestrial origin - assuming absorption to be proportional to the oblique columnar density - will vary in that way. In addition to this gradual variation one can expect an abrupt change in the effective recombination rate at $\chi \geq 98°$, *i.e*. after sunset in the mesosphere (*cf*. Ch. 4.1), and only a very weak χ-dependence during the night (scattered light).

Solar activity: Solar activity expressed as Sun spot number R_z or as solar radio flux $F_{10.7}$ is associated with enhanced fluxes of UV and particularly X-rays (electron densities appear to correlate marginally, but consistently better with $F_{10.7}$ than R_z at all altitudes considered here; Erlacher, 1996). One can therefore expect a good correlation with electron density in the *E*-region where X-rays provide most of the ionisation, and a weaker correlation in the *D*-region. The predominant ionising flux active in the *D*-region is Lyman-α, which only varies by about a factor of two over a solar cycle. However, in addition the concentration of nitric oxide (NO), which is ionised by Lyman-α, is itself tied to solar activity. At very low altitudes where galactic cosmic rays (GCR) provide the production of free electrons, an inverse dependence is expected because GCR are shielded by the more intense solar wind at times of high solar activity. The thermosphere is heated by the absorption of solar X-rays, hence its temperature and density varies with solar activity with a significant time constant (several days). According to theoretical models, this solar activity influence should be detectable down to about 100 km. Since sunspots are cooler areas on the solar surface, the emitted solar wind will have a softer spectrum. Hence even under conditions of darkness, but dominated by charged particles originating from the Sun, the ionosphere varies with solar activity.

Geographic latitude and season: The neutral atmosphere's bulk parameters (pressure, temperature, density) vary with latitude and season as *e.g.,* summarised by the empirical model CIRA. In addition, many minor constituents such as water vapour and nitric oxide have a significant latitudinal and seasonal variation. Notably NO is lost by photo-dissociation and is therefore more abundant in the dark season ("normal winter anomaly"; Schwentek, 1971). An attempt to lump the seasonal and latitudinal dependence for empirical modelling by using "daily integrated insolation" was made by Friedrich and Torkar (1998b); the daily integrated insolation can, however, not account for a seasonally and latitudinally varying temperature, which - depending on the altitude - may also have a significant impact on the recombination chemistry.

Geomagnetic latitude: In the high latitudes (particularly in the auroral zone) ionisation by charged particles, guided by the Earth's magnetic field, generally dominate the ionisation. But also at other latitudes, at least at higher altitudes and notably at night, charged particles, trapped in the geomagnetic field, contribute to the ionisation. Generally, these fluxes increase with geomagnetic latitude and occur more frequently at solar maximum. Galactic cosmic ray fluxes

which dominate the ionisation at the lowest altitudes, also correlate positively with geomagnetic latitude.

UT dependence: Fluxes of trapped charged particles also depend on local geomagnetic time, *i.e.* counted from the meridian through the geomagnetic North Pole. A dependence on geomagnetic time, independent of solar zenith angle, can therefore be expected whenever charged particles dominate the ionisation.

Geomagnetic indices (*Kp*, *ap*): From the readings of a number of selected non-auroral geomagnetic stations a planetary index *Kp* characterising the degree of geomagnetic disturbance in three-hour time slots is formed. Disturbances of the geomagnetic field are due to currents in the ionospheric *E*-region, hence indices such as *Kp* (quasi-logarithmic) or *ap* (linear) will correlate better with the *E*- than the *D*-region. Because of the longer lifetime of free electrons in the *E*-region, *Kp* (or *ap*) from a few hours earlier may actually better correlate with electron densities than instantaneous values.

Hysteresis about noon: The lifetime of ions in the lower ionosphere is short enough not to expect a hysteresis (time lag of electron density *vs*. ionising flux) except near fast sunrise or sunset. However, some of the important, long-lived, neutral trace constituents undergo diurnal variations, and also the diurnal tides of the neutral temperature are not necessarily symmetric about noon (or midnight).

4.2.1. Non-Auroral (FIRI)

This section is the paper by Friedrich and Torkar (2001); it is - expectedly - in places out-of-date, but an updated version is currently being developed.

Introduction

The communications aspect which was the original driver behind the development of ionospheric models has largely disappeared. Only the *F*-region which contributes predominantly to the total electron content (TEC) is still of interest because the delay by TEC constitutes the single largest uncertainty in satellite-borne navigation systems such as GPS. The lower ionosphere, notably the *D*-region, leads to radio wave absorption. For the high frequencies used today, the resulting signal loss of only fractions of a decibel is negligible. It is therefore understandable that the *D*-region has not received much attention in recent years and, if at all included in models, is treated as an appendix of the higher ionosphere. The undoubtedly most ambitious empirical ionospheric modelling effort is the international reference ionosphere (IRI; Bilitza *et al.*, 2011), which covers all altitudes from the *D*-region to the beginning of the plasmasphere. The data entered into IRI over many years stem from various sources such as *in situ* satellite measurements, ionosondes, topside sounders, incoherent scatter, or partial reflection. Owing to the relatively large neutral background density in the mesosphere, all of the above methods are prone to uncertainties in their absolute values, although relative variations (fine structure) or seasonal trends can be deduced from homogeneous data sets. The least questionable results concerning absolute electron densities in the *D*-region (mesosphere) are obtained by partial *in situ* radio wave propagation methods in which an HF signal is transmitted from the ground to the flying rocket payload. The variations of the received HF are due to the total electron content between transmitter (ground) and receiver (rocket payload), and differentiation with respect to altitude yields electron densities. The most common of these methods makes use of the fact that the plane of the polarisation rotates as a function of TEC; this effect is called Faraday rotation. The name FIRI, coined for the present model, simply implies that we attempt to provide a contribution to IRI using electron density data based on Faraday rotation experiments.

The data we use for the establishment of FIRI are restricted to the results from rocket–borne soundings with a radio wave propagation experiment. In the earliest such measurements (late 1940's) the transmitters were on board the rocket payload; now the transmitters are on the ground, and the received signal is telemetered back to the ground. The Earth's magnetic field makes the plasma dually refractive: that is, there exist two modes of propagation. One can, for example, intentionally transmit the two modes alternatingly and measure their amplitudes at the rocket (differential absorption). Alternatively, one can simply radiate using a linearly polarised aerial because a linearly polarised wave can be decomposed into two circularly polarised waves rotating in opposite directions. If the geomagnetic field has a component in the propagation direction, the two partial waves experience different refractive indices, and the composition at the receiver (the flying rocket payload) will maximise at an angle with respect to the original linear polarisation (Faraday rotation). Similarly, the imaginary parts of the refractive indices are generally different (differential absorption); the received polarisation pattern is therefore in general elliptical. The instrumental requirements of this experiment are rather simple; however, a sound magneto-ionic theory is fundamental to derive electron densities for a given geometry, signal frequency, geomagnetic field, and collision frequency. The generally accepted magneto-ionic theory is the one formulated by Sen and Wyller (1960) in which collision frequency is proportional to pressure. A more rigorous treatment of the collision frequency is provided by a further extension of the Sen and Wyller formulation by Friedrich *et al.* (1991); since the results are only marginally different, this new theory is only employed in a very few cases. There is ample literature describing the experiment (*e.g.*, Mechtly *et al*., 1967; Bennett *et al*., 1972); the question of the optimal choice of the sounding frequencies was addressed by Jacobsen and Friedrich (1979). An underlying assumption, which is generally fulfilled, is that the ionosphere is stable during the sounding (about 2 min) and only vertically structured. These restrictions can be greatly relaxed if one measures in a region of steadily increasing electron densities (such as the *D*-region up to the *E*-region peak). Although the authors have extensive experience with this kind of experiment, the bulk of the data were either taken from the literature or made available by the original experimenters by providing unpublished results. To date, a grand total of 274 rocket flights (now 366) have been collected, of which some 50 (now 84) are due to the authors' instruments and about another 20 (now 21) were processed from the raw data by the authors. The majority of the flights were by the University of Illinois at Urbana-Champaign, followed by the Radio and Space Research Station Slough (later the Rutherford-Appleton Laboratory), and various soundings by groups in Norway, Denmark, Sweden, Russia, and India. For the present purpose of establishing a model for the non-auroral, "normal" ionosphere, 118 of the 274 profiles can be used (see Appendix A). Some of the rocket payloads also carried electron density probes, which are uncertain in their absolute values, but can be normalised by the wave propagation data in the height region where there is overlap. Many of the profiles in the present data bank extend to much lower or higher altitudes than covered by the propagation experiment; these extra height ranges are due to the probes assuming that the normalisation is applicable above and below the region of the normalisation. Although some groups have used height-dependent normalisation factors (Mechtly *et al*., 1967; Aikin and Blumle, 1968), no obvious indication of a height dependence was found in a comparison of a number of different probes aboard four identical rocket payloads flown during both day and night (Friedrich *et al*., 1997). The reliability of *in situ* (probe) measurements calibrated by a wave propagation experiment is considered by, for example, Thrane (1974) or Mechtly (1974) to be the greatest in the lower ionosphere; other sources of electron densities such as incoherent scatter radars (Arecibo, EISCAT) not only have a poorer height resolution, but have density thresholds which make them primarily attractive for *E*-region studies.

Other Models and Early Versions of FIRI

As indicated above, the *D*-region is covered by the general empirical ionospheric model IRI. Among the dedicated models, the attempt by McNamara (1979) should be mentioned. The underlying assumption in both models is that the shape of the electron density profiles can be approximated by an analytical function. This condition is only fulfilled if the ionosphere behaves according to the classical Chapman theory of the formation of ionospheric layers. In reality, in the *D*-region the ionisable constituent is nitric oxide, and the constituent, which absorbs the ionising wavelength (Lyman α) is O_2. Whereas the mixing ratio of O_2 is almost constant, that of NO is not only a function of altitude, but is also variable with season, latitude, and solar cycle. Furthermore, the electron densities as a function of solar zenith angle will not follow a Chapman behaviour because the effective electron loss rate is not a constant, but exhibits distinctly different values for day and night. Nonetheless, electron density profiles of the *D*-region display particular shapes, and modellers are tempted to look for characteristic parameters to describe the profiles. One such characteristic is the daytime ledge around 84 km, and in IRI, individual profiles are normalised at that height, and shape functions are defined for the regions above and below.

The model FIRI to be described here is the result of several precursor attempts, all of which have three features in common:

(1) Only "appropriate" data that is, all based on rocket-borne wave propagation experiments are used,

(2) all altitudes are treated independently from each other, and

(3) in general, analytical descriptions are established for neutral density levels (or pressure surfaces) rather than for geometric altitudes.

Hence, we do not attempt to analytically describe electron density profiles, but rather expect that the variation at any altitude (or pressure surface) is an analytically describable function of season, solar activity, and latitude; the analytical description of the solar zenith angle dependence is the most problematic, but this problem is essentially bypassed in the most advanced version of FIRI. The earliest attempt (retrospectively termed FIRI–01) used 72 profiles and was defined as a function of the solar zenith angle χ only (Friedrich and Torkar, 1992). The solar activity dependence was eliminated by normalising the raw data to a mean sunspot number of 60. For this purpose, the ratio between the electron density according to IRI for the conditions of the particular rocket flight and at $R_z = 60$ was formed and applied to the rocket data as a correction factor. For the solar zenith angle variation, Chapman behaviour was assumed (Smith and Smith, 1972) with a somewhat unrealistic extension into full darkness. The influence of season and latitude was hoped to be largely accounted for by processing the data not at constant altitudes, but at constant neutral density levels according to the atmospheric model CIRA–86 (Rees *et al*., 1990). The resulting model profiles appear reasonable for daytime; the densities for night-time conditions are far too large in the *D*-region, obviously due to the unrealistic linearization and extension of the zenith angle variation by a Chapman function.

The next step was more ambitious. It was attempted to fit a total of 99 profiles by a single mathematical function usable at all altitude levels. The function itself is the product of assumed linear dependencies on latitude, solar activity, and the Chapman function, whereas for the seasonal dependence a sinusoidal variation symmetric about solstice is used. The multiplication of these individual dependencies yields 16 coefficients, of which four lead to a seasonal variation at the equator and are therefore forced to zero. The remaining 12 coefficients at each altitude level are still too many to be determined from the limited amount of data. Since not all of the combined terms are expected to contribute significantly to the modelled electron

densities, tests were made with a large number of electron density profiles from a theoretical model, evenly distributed for the various geophysical conditions (solar zenith angle, latitude, solar activity, and season). These artificial electron densities were assimilated by the full function (12 coefficients), and the terms with the smallest contributions were excluded from further considerations. The measured data were then approximated by those eight terms, which were the most significant ones in the analysis of the theoretical profiles (FIRI-02; Friedrich and Torkar, 1998a).

In version FIRI-03 the expected effect of season and latitude are lumped into a parameter "daily integrated insolation", the integral of $\cos\chi$ from sunrise to sunset. Multiplication of the individual functions (solar activity, zenith angle, and insolation) yields only eight coefficients. A search for less important ones is therefore not necessary if one maintains that eight coefficients can reasonably be determined from the data (Friedrich and Torkar, 1998b). Version FIRI-04 was never published: it is simply FIRI-03 with an additional linear variation with geomagnetic latitude. Here again less important combined terms need to be identified and eliminated so as not to exceed a total of eight. Table 4.1 gives a summary of the development of FIRI; FIRI-05 is the version described here.

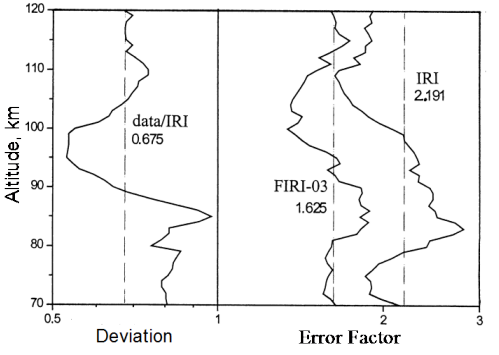


Figure 4.5 Ratio between the rocket data and the corresponding values from IRI and an early version of FIRI as a function of altitude. Note that IRI is on average too large by about 50% and also that the scatter (ability to model the rocket data) is significantly larger that in FIRI-03 (Friedrich and Torkar, 2001a).

The Semi-Empirical Approach

Before embarking on further improvements and refinements in detail, the quality of IRI in the *D*-region was tested with the present set of "good" electron density profiles. For this purpose, at every kilometre of height, the ratio between each available electron density value of the data set and the corresponding value from IRI was established. The mean deviation at each kilometre is depicted in the left panel of Figure 4.5 and shows that the data are on average 0.675 of what IRI predicts, or in other words, IRI is about 50% larger than the rocket data used here. The error factor shown in the right panel is obtained by forming the averaged deviation, or its inverse, such that the factor is always >1. Already the results of an early version of FIRI appear to be better than IRI; the deviation of FIRI from the data it is built from necessarily yields unity. Much refinement effort has been applied to the early FIRI models, however, without greatly improving the error curve displayed in the right panel. More complex functions, which could potentially better reproduce the input data, led to unrealistic excursions for conditions no covered by measurements (extrapolations). Also the linearization of the solar zenith angle dependence by using either $\cos\chi$ or a Chapman function $\mathrm{Ch}\chi$ is not satisfactory

or would require an unrealistically complex mathematical order. Figure 4.6 shows electron densities for various altitudes from a theoretical model (Torkar and Friedrich, 1983a, see below) as a function of solar zenith angle and of Chχ; clearly, even when plotted against ln(Chχ), the relation is generally not linear, not even for zenith angles below 85° where the simple cosχ relation should hold and no Chapman function needs to be involved.

In order to altogether bypass the problem of searching for an analytical approximation of the zenith angle dependence, particularly avoiding higher mathematical orders, we now take the approach that we use the empirical data to provide a correction for a reasonably realistic, but still simple, theoretical mode. For this purpose, we update somewhat the ion-chemical model available to the authors (Torkar and Friedrich, 1983a). In brief, this theoretical model computes 17 charged species and is a steady state calculation in which the neutral atmosphere and ionising fluxes are inputs. The conclusion at the time it was published was that reasonable inputs (ionising fluxes, trace constituents) were more important for correctly reproducing

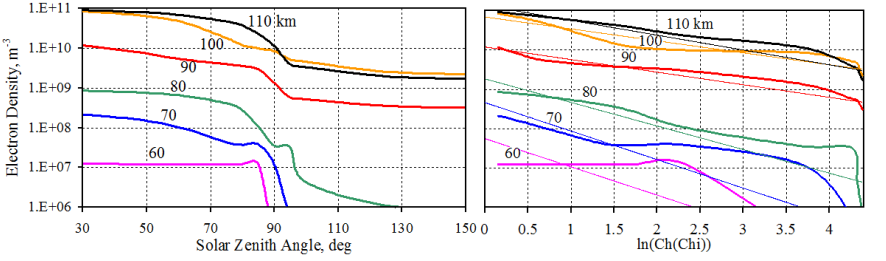


Figure 4.6 Theoretical electron densities at mid-latitude, low solar activity, and equinox at various altitudes (left) as a function of solar zenith angle, and (right) as a function of ln(Ch χ). Note that in the right panel the least mean square lines (thin lines) are at best representative at 110 and 90 km, but are unacceptable approximations elsewhere, even for zenith angles smaller than 85° (or lnCh (χ) = 2.2; Friedrich and Torkar, 2001a).

electron densities than a more sophisticated chemistry; this conclusion is certainly still valid. For the major constituents we now use CIRA-86, of the minor constituents nitric oxide is the most crucial. We use the (idealised) results of the theoretical calculations for quiet conditions (low solar activity, *Ap* = 4) by Rusch *et al.* (1981), vary the (logarithm of the) densities sinusoidally with season (symmetric about solstice), and introduce an enhancement factor of up to 6.3 for the thermospheric NO due to solar activity according to Golshan and Sechrist (1975). A constant mixing ratio is assumed below the mesopause. The other minor species with a low ionisation threshold is $O_2(^1\Delta_g)$ for which we employ the results of the satellite SME by Thomas *et al*. (1984). The only other crucial trace gas is atomic oxygen, which impedes the formation of cluster ions and provides an effective reverse reaction to the attachment of electrons to neutrals; both of these processes have a profound influence on the effective electron recombination rate. Reliable [O] measurements are at least as rare as reliable *D*-region electron densities and the compilation by Gumbel (1997) is probably the most complete. In order to make these [O] data more generally applicable, a model was established using the algorithm described above as FIRI-03 (four coefficients only, that is, the parameters zenith angle and insolation) defined for neutral density levels (Friedrich *et al*., 1999). Apart from that, a temperature–dependent cross section of O_2 for Lyman α according to Carver *et al*. (1977) is now incorporated in the theoretical ionospheric model.

Version	*Main Features*	*Reference*
FIRI-01	Defined for constant neutral density levels, medium solar activity, only solar zenith angle as variable; 72 profiles	Friedrich and Torkar (1992)
FIRI-02	Season, latitude, solar activity and zenith angle considered; 99 profiles	Friedrich and Torkar (1998a)
FIRI-03	Season and latitude replaced by "daily integrated insolation"	Friedrich and Torkar (1998b)
FIRI-04	As above, but with geomagnetic latitude	unpublished
FIRI-05	Any of the above functions used to empirically correct a theoretical model; 118 profiles	this paper; very preliminary results in Friedrich *et al.* (2001)

Table 4.1 Evolution of the FIRI models.

For each of the usable 118 measured electron density profiles a corresponding one is calculated from the theoretical model, and the factor between the two is established at each kilometre (or pressure surface or neutral density level, respectively). These factors are then mathematically assimilated by functions such as the ones described as FIRI-02 to FIRI-04. In the case of FIRI-02 the individual relations for the various parameters are

$$\log N_e = [a + b\ln(Ch\chi)][c + d\lambda][e + f\sin\tfrac{\pi(80+doy)}{182.75}][g + hF_{10.7}] \tag{4.1}$$

where λ is the geographic latitude and $F_{10.7}$ the (observed) solar flux. Expanding Eq. (4.1) by multiplying the individual dependencies, one obtains the following terms for the factor:

$$\begin{aligned} log\, N_e = {} & A + B\, ln(Ch\chi) + C\lambda + D\, sin\tfrac{\pi(80+doy)}{182.75} + EF_{10.7} + F\lambda\, ln(Ch\chi) + \\ & + G\ln(Ch\chi)\sin\tfrac{\pi(80+doy)}{182.75} + H\lambda\sin\tfrac{\pi(80+doy)}{182.75} + I\lambda F_{10.7} + J\ln(Ch\chi)\lambda F_{10.7} \\ & + K\ln(Ch\chi)\lambda\sin\tfrac{\pi(80+doy)}{182.75} + L\ln(Ch\chi)F_{10.7} + M\sin\tfrac{\pi(80+doy)}{182.75}F_{10.7} + \\ & + N\, ln(Ch\chi)\, sin\tfrac{\pi(80+doy)}{182.75}F_{10.7} + O\lambda\, sin\tfrac{\pi(80+doy)}{182.75}F_{10.7} + PF_{10.7}\lambda\, ln(Ch\chi)\, sin\tfrac{\pi(80+doy)}{182.75} \\ & + PF_{10.7}\lambda\ln(Ch\chi)\sin\tfrac{\pi(80+doy)}{182.75}F_{10.7} \end{aligned} \tag{4.2}$$

The terms *D*, *G*, *M*, and *N* have to be set to zero in order to suppress seasonal variations at the equator. The analogous relations for FIRI–04 are:

$$log N_e = [a + b\ln Ch\chi)][c + d\Sigma][e + fF_{10.7}][g + h\lambda_m] \tag{4.3}$$

Here Σ is the daily integrated insolation, and λ_m is the geomagnetic latitude. Multiplication of Eq. (4.3) yields

$$\log N_e = A + B\ln(Ch\chi) + C\Sigma + D\Sigma\ln(Ch\chi) + EF_{10.7} + F\ln(Ch\chi)F_{10.7}$$
$$+GF_{10.7}\Sigma + H\Sigma F_{10.7}\ln(Ch\chi) + I\lambda_m + J\lambda_m\ln(Ch\chi) + K\lambda_m\Sigma + L\lambda_m\Sigma\ln(Ch\chi)$$
$$+M\lambda_m F_{10.7} + N\lambda_m\ln(Ch\chi)F_{10.7} + O\lambda_m F_{10.7}\Sigma + P\lambda_m\Sigma F_{10.7}\ln(Ch\chi) \quad (4.4)$$

A semi-automatic method was devised to find those parameters (*A* to *P*) in the above equations, which led to the mathematical formulation that yields the best agreement between input electron densities and the corresponding model values. For this purpose, the averaged

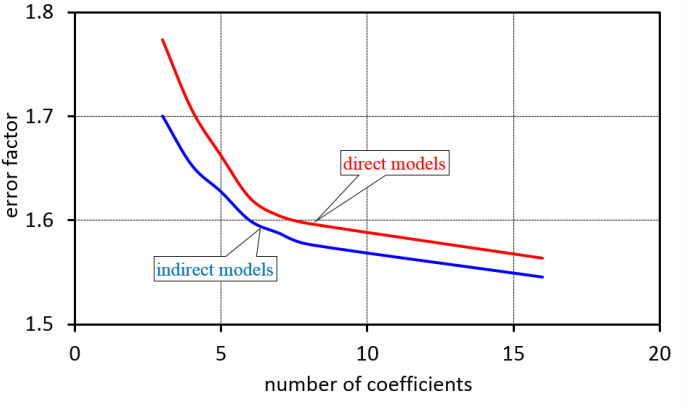


Figure 4.7 Error factors vs. mathematical order of the fitting function in the height region 70 to 120 km. For each number of coefficients, the best solution (e.g., calculated at constant altitudes or constant pressure surfaces) is shown for both the direct models and the model where the data provide the correction (red line: direct, blue line indirect model; Friedrich and Torkar, 2001a).

error over a specified height region is computed for models with different coefficients and orders (number of coefficients). Since we aim at providing an improvement of IRI for the *D*-region, we do not attempt to compete in the *E*-region, which is probably well covered by IRI. We therefore concentrate on the lower altitudes, beginning with 70 km, since below that height, too few data are available, and we limit the height region where we aim to find the optimal function to below 120 km. We simply calculate all possible combinations of coefficients in Eqs. (4.2) or (4.4), respectively, and store the resulting averaged error factor of the specified height region. Since the number of possibilities is very large (and the computing time long), we postulate that the first three terms (*A*, *B*, and *C* in Eq. (4.2) or (4.4)) must always exist. Figure 4.7 shows averaged error factors thus obtained for the best of the direct models (FIRI-02 and FIRI-04) and the present (indirect) model where the data are used to correct theoretical start values (FIRI-05), as a function of the number of coefficients used. The figure clearly reveals that for the same complexity (*i.e.* number of coefficients to be determined) the indirect model always produces the smaller averaged error factor. Although more than 16 coefficients do not occur in the present algorithms, it appears likely that the averaged error factor will never reach unity for higher orders, but rather a value >1. Apart from the trivial possibility of having erroneous measurements in the data, this probably reflects the fact that parameters beyond the ones considered here (*i.e.*, zenith angle, season, latitude, and solar activity) may be needed to fully describe all situations. With the limited number of profiles an inclusion of - for example - solar flares, stratospheric warmings, vertical transport, or an asymmetry between pre-noon and post-noon is not realistic. The choice of the function best describing either the electron densities directly, or the correction of the theoretical model, is not final and is necessarily somewhat arbitrary since it is based on the best agreement in a likewise arbitrarily chosen altitude range.

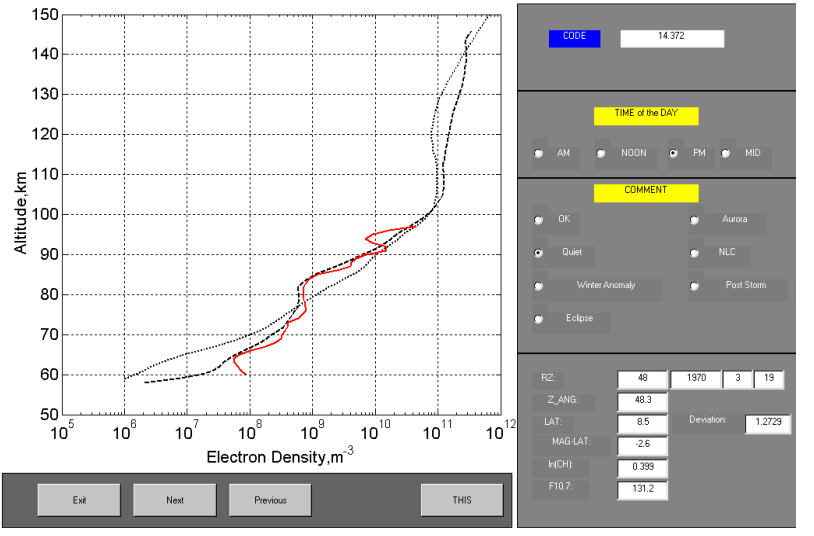


Figure 4.8 Screen shot of the user interface for checking the individual rocket data showing an electron density profile of a daytime flight launched from Thumba (India) at medium solar activity in the afternoon (χ = 48°, red line). The dotted line is the start value (theoretical model for the specific conditions), and the dashed line is the corrected theoretical profile (FIRI-05, five coefficients for pressure surfaces, algorithm of Eq. 4.5). Note that the data are on average only 27% above FIRI-05, whereas the disagreement with the theoretical model is substantial (Friedrich and Torkar, 2001a).

At any rate, in the present version in which only a correction is calculated from the data (FIRI-05), the mathematical formulation does not necessarily need to reflect physical processes and therefore allows a greater freedom in the choice of the function. The FIRI-05 version with a correction according to FIRI-02 for pressure surfaces and five coefficients appears to be a good compromise between stability of the model (no unrealistic excursions) and the ability to reproduce the input values (averaged error factor of 1.6308; the even smaller value shown in Figure 4.5 is due to the fact that there at night a lower limit of 90 km is used in order to make a fair comparison with IRI, which has that height as the lower limit at night).

In this particular case the coefficients A, B, C, H, and L in Eq. (4.2) are calculated (for 12 logarithmically spaced pressure surfaces per decade) such that the resulting formulation for the correction F_{corr} of the theoretical model is

$$\log F_{corr} = A + B\ln(Ch\chi) + C\lambda + H\lambda\sin\frac{\pi(80+day)}{182.75} + L\ln(Ch\chi)F_{10.7} \qquad (4.5)$$

The only smoothing in the resulting profiles is a running mean of each of the coefficients (50% of the layer concerned, and 25% of the layers above and below, respectively). In order to check each version of the model, all input profiles can be plotted together with the corresponding theoretical values (start values) and the corrected theoretical values (*i.e.*, FIRI–05). Figures 4.8 to 10 show comparisons between rocket data, the start values, and the corrected theoretical values (using Eq. (4.2)) for daytime, twilight, and night, respectively. One can see that generally the theoretical start values are corrected in the right direction; the very steep ledge at night can, however, not be reproduced with the present algorithm.

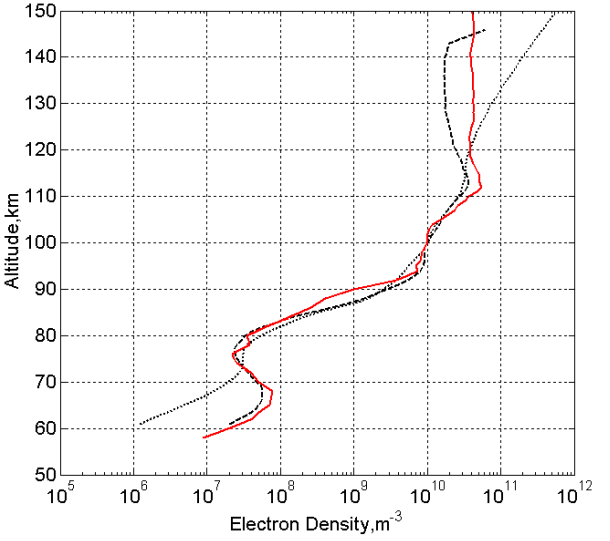


Figure 4.9 Case of a rocket flight from Wallops Island (VA) just after sunrise and at low solar activity (red line). Note the bulge below 70 km, a feature predominantly seen in data from that range and the Urbana group. The start values (dotted line) are successfully corrected accordingly to account for this feature. Values near 150 km vary in altitude when applied to other seasons or latitudes because the altitude of the pressure surface varies. In the tabulated data provided to IRI the height region is restricted to the range 60 to 140 km (Friedrich and Torkar, 2001a).

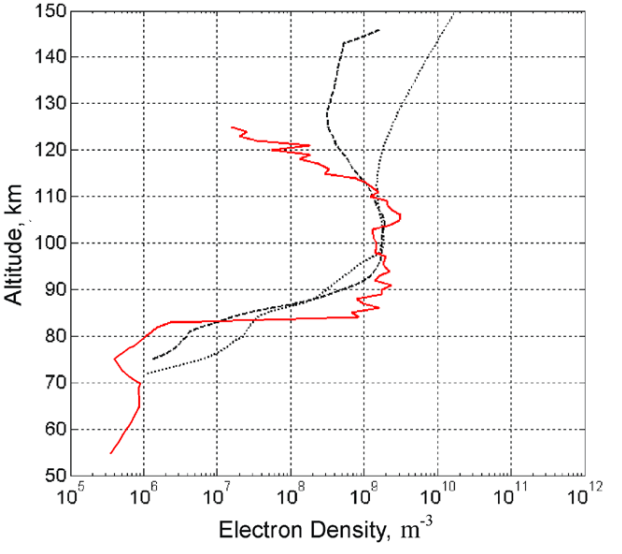


Figure 4.10 Night-time profile from Alcântara (Brazil). Note that the measured ledge is much steeper than the start value, but also than the FIRI–05 result (Friedrich and Torkar, 2001a).

Interpretation of the Results

The factor described by Eq. (4.2) is applied to correct a net of theoretical electron density tables from which the user can interpolate for the desired conditions. The calculations are for latitude steps of 15° from the equator to 60°, for (the middle of) each month, three solar activity levels ($F_{10.7}$ = 75, 130, and 200 sfu (1 sfu [solar flux unit] = 10^4 Jy, 1 Jy = 10^{-26} W m^{-2} Hz^{-1})), and solar zenith angles 0°, 30°, 45°, 60°, 75°, 80°, 85°, 90°, 95°, 100° and 130°.

Figure 4.11 shows electron densities at various heights as a function of solar zenith angle (15° latitude, January, low solar activity). The solid lines are the values from the present model, and the dashed ones represent the start value from the ion-chemical calculation. For the cases displayed, one can see that at 100 km the theoretical results are very realistic and only at night the data lead to slightly lower densities. At 90 km the data (*i.e.*, FIRI) suggest larger values at night, which in a physical interpretation could mean that fluxes of scattered light are larger than assumed in the theoretical computation and values during the day smaller by about the same factor. The discrepancy at 80 km is substantial, but can be brought into agreement by a reduction of the NO density by about a factor of almost 10. At night the number densities of free electrons collapse much more rapidly than predicted by our theoretical calculation. At 70 km, [NO] would have to be increased, and again the densities disappear much faster after sunset. For the huge resulting empirical correction toward larger values for the 60 km electron densities, there are at least two explanations: (1) All data at that altitude are based on *in situ* probe measurements normalised at higher altitudes (a height dependence of the normalisation factor can not be ruled out, but can of course not be ascertained retrospectively), and (2) the ion chemistry incorporated in the theoretical start values is perhaps too simple and conceivably leads to unrealistic electron densities. Figure 4.12 shows electron density profiles for the same conditions. During the day, there is a distinct “knee” in the profiles, plausibly explicable by a corresponding structure in the NO profile. The densities between 60 and 65 km are almost

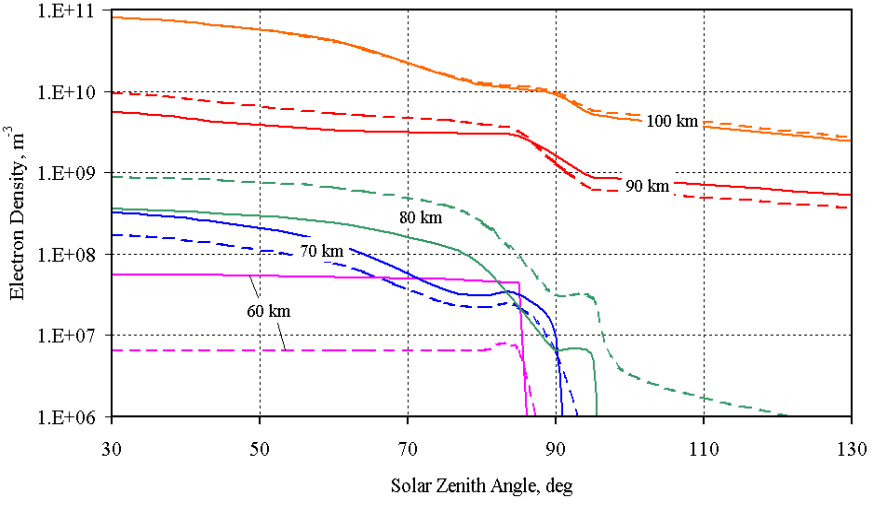


Figure 4.11 Variation of electron densities as a function of solar zenith angle at various heights (low solar activity, 15° latitude, January). The dashed lines are the start values according to the theoretical (ion–chemical) model, and the solid lines contain the empirical correction according to Eq. (4.5) (Friedrich and Torkar, 2001a).

independent of the zenith angle, again realistic since galactic cosmic rays provide the main source of ionisation. As the Sun sets, the region between 70 and 85 km becomes depleted, leaving a bulge below 70 km until the density of atomic oxygen collapses and electrons are lost by attachment to neutrals; an additional loss process is due to the formation of larger cluster ions with their higher recombination rates. After complete sundown, there is a pronounced steep ledge at 83 km, explicable by a corresponding ledge in the density of atomic oxygen. The scenario emerging from this model qualitatively agrees with theoretical calculations (*e.g.*, Ogawa and Shimazaki, 1975).

The solar activity dependence is shown in Figure 4.13. In the *E*-region, enhanced X-ray fluxes are responsible for the larger electron densities; in the *D*-region, larger concentrations of NO are the main cause of the enhancement. As expected from theory, there is indeed an inverse relation between solar activity and electron density at very low altitudes due to the shielding effect for cosmic rays by the solar wind. Figure 4.14 depicts cross sections at solstice from -60° S (Winter hemisphere) to +60° N for various constant solar zenith angles and high solar activity. Whereas at 100 km there is virtually no variation, in the *D*-region there is a significant increase toward the Winter hemisphere because [NO] is larger in the dark hemisphere. The small hemispheric variation at 70 km is probably due to a variation of the neutral temperature which at that height exerts a pronounced influence on the formation of water clusters and hence the recombination rate. Figure 4.15 shows the seasonal variation of electron densities for medium solar activity and 45° latitude. Since the correction function (Eq. 4.5) is symmetric about solstice, the times of the maxima/minima different from solstice, notably at 70 and 80 km, must be due to the theoretical electron densities through the

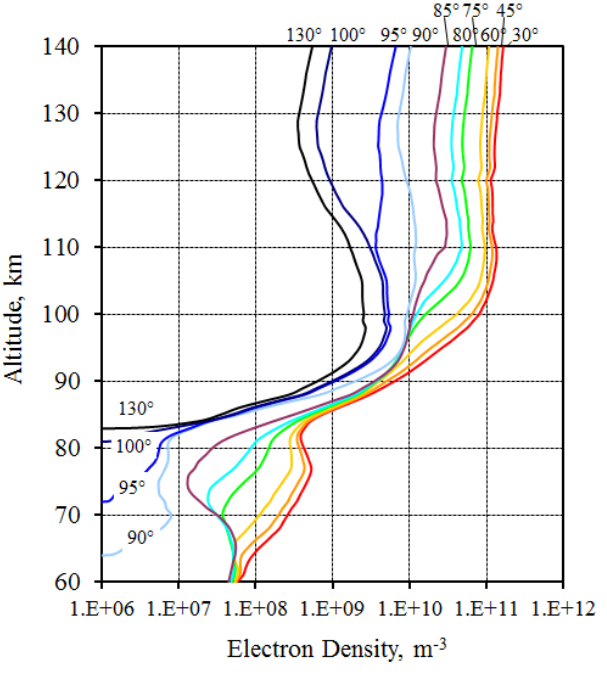


Figure 4.12 Electron density profiles of the conditions of Figure 4.11 for zenith angles between 30° and 130° (Friedrich and Torkar, 2001a).

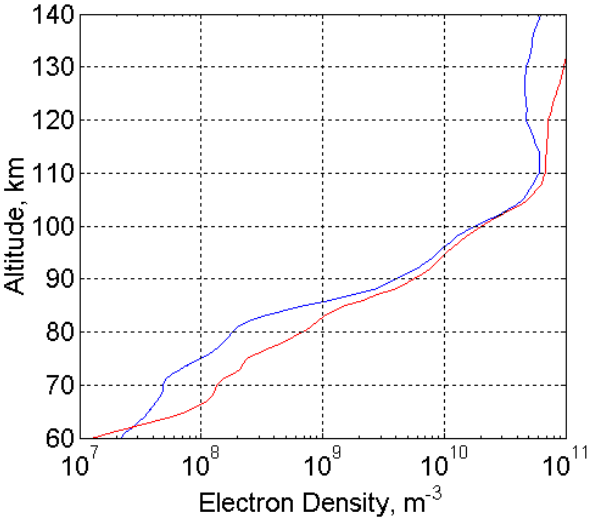


Figure 4.13 Electron densities for low and high solar activity (35° latitude, spring equinox, 75° solar zenith angle, $F_{10.7}$ = 75 and 200 Sfu, respectively). Note the reversal below 62 km (1 Sfu = 10^{-26} W m^{-2} Hz^{-1}; Friedrich and Torkar, 2001a).

underlying neutral atmosphere (CIRA-86). At 80 and 90 km, again the larger densities are seen in Winter, whereas at 70 km the summer densities are larger, probably due to the higher temperatures, which impede the formation of water cluster ions. In spite of the relatively small number of reliably determined electron densities, it has been possible to establish a realistic empirical model by using theoretical values as a basis. The model is available in the form of tables for inclusion in IRI. It is valid for latitudes below 60° both geographic and geomagnetic; hence, the limit of applicability in North America is at lower latitudes than *e.g.* in Europe or Siberia. Since the rocket flights are not evenly distributed over zenith angle, solar activity, *etc.*, the model is, for example, somewhat biased toward low solar activity. One consequence is that the night-time *E*-region emerges with a slightly inverse relation to solar activity, which contradicts the concept that the ionisation is maintained by scattered solar UV, whereas according to observations by the Arecibo incoherent scatter radar the density of the night-time *E*-region peak is independent of solar activity; Titheridge (2001) explains this by starlight providing most of the ionisation. In the small number on flights in full darkness, one profile (flight 14.178) could be identified which produces this inverse behaviour. That particular equatorial flight, which also behaved unexpectedly in other respects, showed an *E*-region peak a factor of 5 above a comparable measurement a few days later at the same solar activity (Aikin and Blumle, 1968), but also higher than all other measurements, including those taken at high solar activity. Since the ionograms also showed such a ratio between the two launch days, we have no justification to exclude that flight and must therefore consider this "wrong" solar activity behaviour an inevitable shortcoming due to the limited number of experimental data.

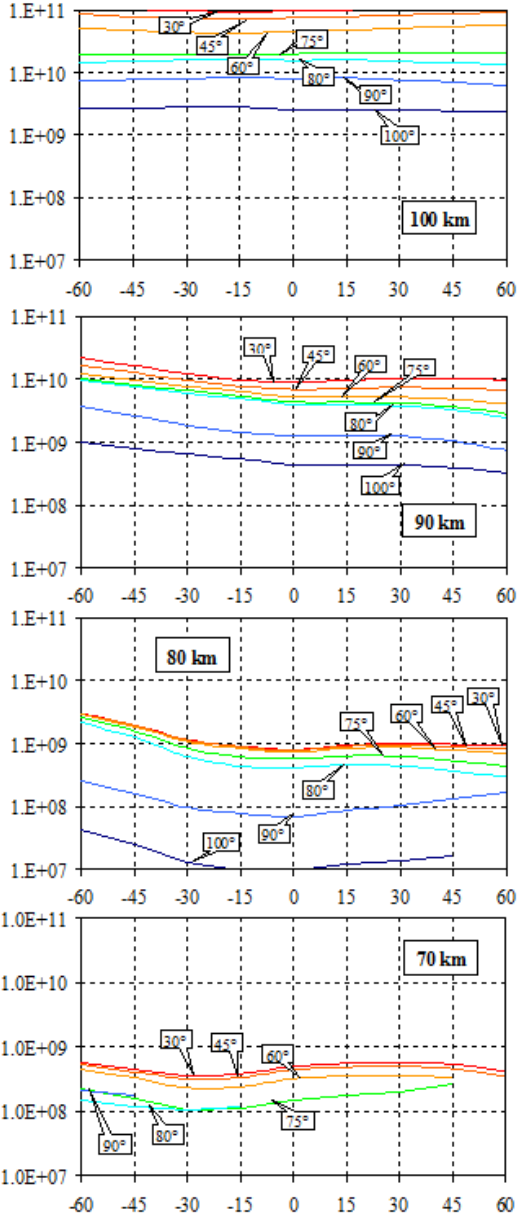


Figure 4.14 Variation of electron densities at solar zenith angles between 30° and 100° and different altitudes at high solar activity and solstice from the winter to the summer hemisphere (left to right). Curves are broken off where zenith angles do not exist or where the model is not valid due to lack of data. Note the increase in electron density toward the winter hemisphere in the D-region, expected due to enhanced NO densities, whereas the E-region shows no obvious hemispheric variation (Friedrich and Torkar, 2001a).

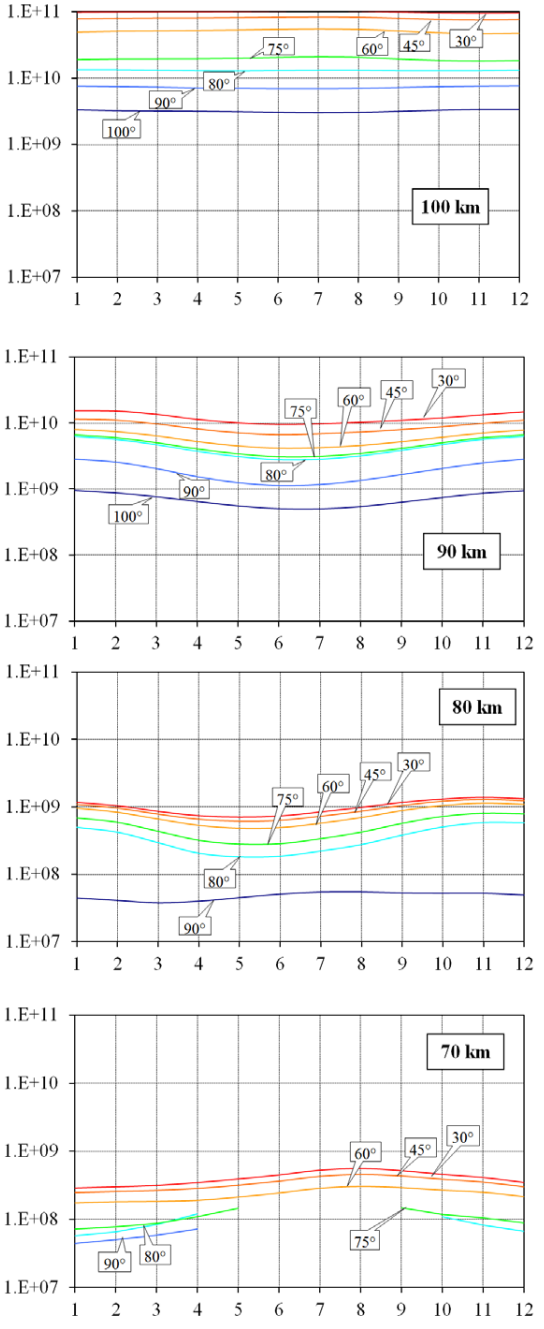


Figure 4.15 Seasonal variation of electron densities at various solar zenith angles between 30° and 100° and four altitudes for medium solar activity and 45° latitude (Friedrich and Torkar, 2001a).

Conclusions

Certain variations in the lower ionosphere are not accounted for by the describing inputs. One such variation surely is the asymmetry between pre-noon and post-noon times. The ion chemistry in the mesosphere has time constants which can only explain a very small hysteresis at sunrise/sunset (*cf.* Ogawa and Shimazaki, 1975); however, radio wave absorption measurements by different groups have clearly revealed substantial differences (Laštovička, 1977, or Torkar *et al*., 1978), whereas Thrane (1969) only found small pre-noon - post-noon differences in *D*-region electron densities from ground-based measurements made on the island of Crete. If one establishes FIRI from pre-noon and post-noon data separately, the afternoon results are indeed consistently larger. The most likely explanation is that [NO] has a hysteresis, an unexpected feature found in the data from the limb-scanning measurements of the Halogen Occultation Experiment (HALOE) aboard the satellite UARS (Friedrich *et al*., 1998). A theoretical explanation for this phenomenon has been put forward by Marsh and Russell (2000). The not uncommon, but unpredictable winter enhancements (winter anomaly) are intentionally not described by the present model. A hemispheric difference can be expected due to the Earth's eccentricity around the Sun and has indeed been seen in *E*-region ionosonde data. Since the data were gathered in a period covering half a century, it may also be tempting to look for long-term trends as was, for example, done by Danilov (1998).

Improvements of FIRI in its latest version (FIRI–05) can be expected from more data, but also from better start values, that is, a better theoretical model and atmospheric inputs consistent with the electron density data (notably NO). An inclusion of the prenoon-postnoon asymmetry needs to be addressed, preferably without unduly increasing the mathematical complexity. Also error bars or ranges of probability should be established and added to the tabulated profiles.

Appendix A

Of the grand total of 274 *D*-region electron density profiles which were measured by sounding rockets using a radio wave propagation method, 106 were launched from sites within or near the auroral zone where generally unpredictable ionisation by energetic particles dominates the ion pair production (geomagnetic latitudes >61°). Of the flights from South Uist (Scotland), three were described as auroral and seven as poststorm, the latter implying enhanced [NO] as a consequence of a geomagnetic disturbance (Dickinson and Bennett, 1978). Twenty-six flights were launched under conditions of identified winter anomaly events (*i.e.* enhanced [NO] by transport phenomena; *cf*. Offermann, 1979), and 15 flights were dedicated to studies of solar eclipses; the remaining total of "normal" profiles is therefore 118 (in cases of sporadic *E*-layers only the 1 or 2 km affected were left out of the analysis). Figure 4.16 depicts all 118 profiles used in the present analysis. Some features of the model profiles in Figure 4.12 are also evident here. One has to bear in mind that the widely scattering data in that mass plot are from all seasons, latitudes, and solar activities, whereas the idealised variation in Figure 4.12 is only valid for one particular season, latitude, and solar activity. A noteworthy feature is the steep ledge of the night profiles which seems to always have the same gradient, but at different heights. Hence, the modelled ledge is somewhere in the middle of the individual ledges and not as steep. The rocket flights are fairly evenly distributed over the whole year with a minimum of five flights in December to a maximum of 20 in March (for the 21 flights from the Southern Hemisphere 6 months were added to the launch dates). Thirty-three profiles are from the equatorial region (±15°), 59 from moderate latitudes (15° to 45°), and 26 from the high mid-latitudes (45° to 60°). Forty-nine flights took place in the morning, 42 in the afternoon, 19 at noon (±1 hour from local noon), and 7 at midnight (±1 hour from local midnight). The solar zenith angle distribution shows a slight dominance of the data near

60° and at twilight. A simple divider for the data is the zenith angle of 98°, that is, where the chemistry governing the recombination changes from day to night behaviour (Stauning, 1996); with this classification there are 102 daytime- as opposed to only 16 night-time profiles. Fifty-two profiles were measured during low solar activity ($F_{10.7}$ < 100 sfu), 37 during medium activity (100 to 150 sfu), and 29 during high solar activity (>150 sfu). The number of electron density values available at each kilometre of height is shown in Figure 4.17. Finally, Figure 4.18 shows the error between data and the corresponding model values of the version described here; obviously, the region around 85 km is the most problematic to reproduce, probably due to the variability of the shape of the NO profile (daytime) and the variability of the ledge of atomic oxygen (night-time). The *E*-layer appears to be the easiest to model; the larger error factors below 65 and above 120 km are explainable since those regions were not included in the search for the optimal function.

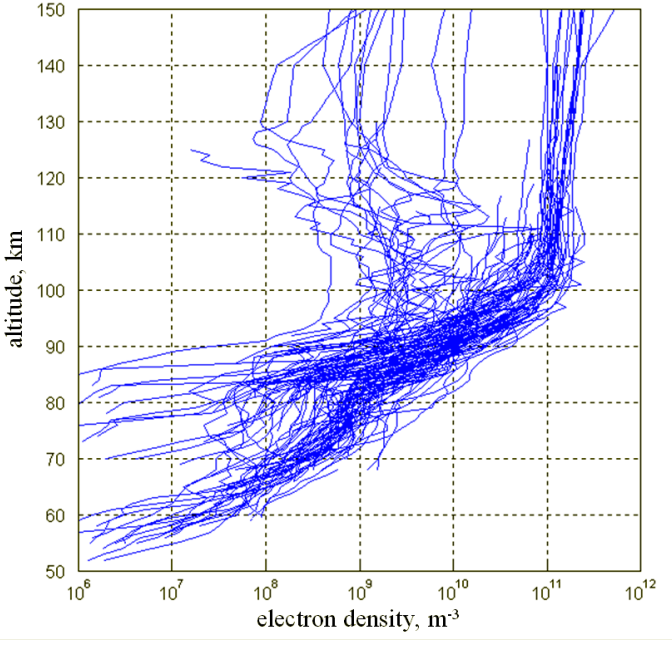


Figure 4.16 Electron density profiles use for the establishment of FIRI (Friedrich and Torkar, 2001a).

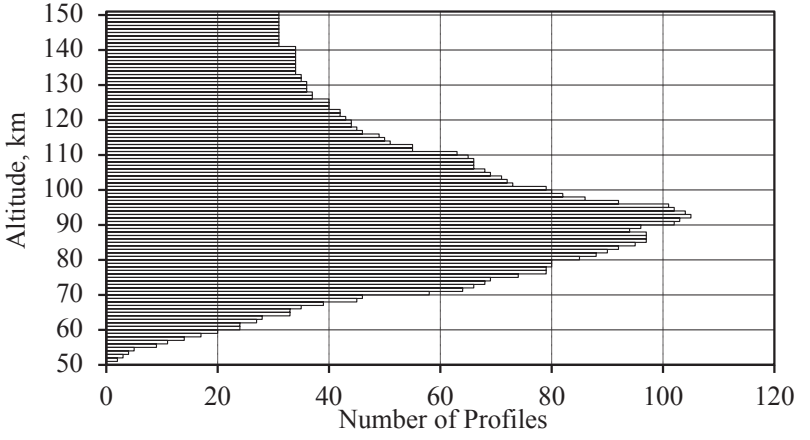


Figure 4.17 Number of electron density values per kilometre of altitude (Friedrich and Torkar, 2001a).

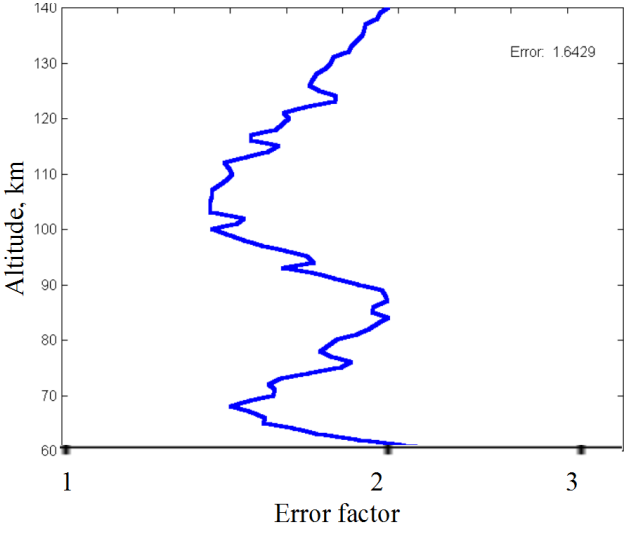
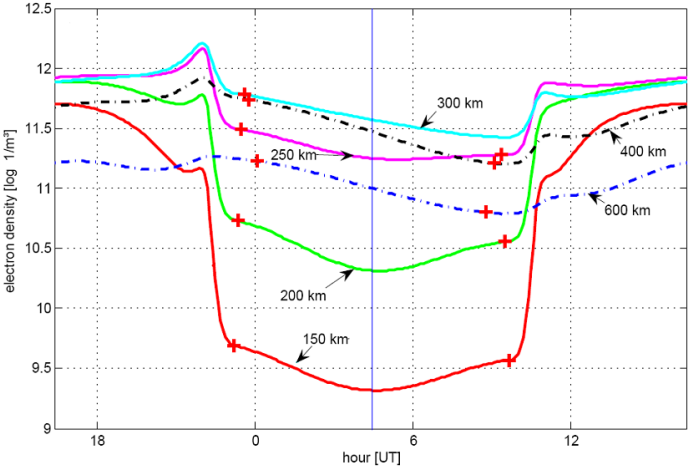


Figure 4.18 Error factor between data and the corresponding model electron density according to the version FIRI-05-02p-5 as described here as a function of altitude (version 05, correction function 02, for constant pressure surfaces, 5 coefficients; Friedrich and Torkar, 2001a).

Outlook

- The theoretical model underlying FIRI is a very simple one; there are better ones "on the market" one could use, but it must *not* be time-dependent as FIRI is a steady-state model.
- Notably the introduction of an electron scavenging meteoric dust layer should reproduce the ubiquituous ledge near the mesopause more relistically.
- The correction function was obtained by minimising the absolute factor to the theoretical model; minimising for a least-mean-square (RMS) is the more reasonable approach.
- within limits the nitric oxide (NO) used in the underlying theoretical model can be modified thus reducing the magnitude of the factor of the correction function.
- a large number of probe measurements are available which reach down to 50 km; with due caution they can be used to extend FIRI to lower altitudes.
- in the present FIRI version at night the electron densities vary a little with solar zenith due to the scattered solar Lyman-α in the underlying theoretical model. A first (time-dependent, empirical) Neural Network model using the Arecibo data suggests the night densities follow two functions (*a*) a variation with solar zenith angle, and (*b*) a simple exponential decay after sunset (Fankhauser, 2010); this behaviour should somehow be accounted for.



Diurnal variation of electron densities according to the Arecibo IS radar (Fankhauser, 2010). Note that at lower altitudes they are largely symmetric to local midnight (blue vertical line), higher up they simply decay exponentially after sunset (red +).

4.2.2. Residual Variations

(This section is largely the paper by Friedrich and Torkar (2001b); again, an updated version is envisaged once the new version of FIRI is available)

An empirical model of the lower ionosphere was established using data from half a century of rocket soundings. The database contains 118 altitude profiles from non-auroral latitudes, and some 100 profiles from the auroral zone. Attempts are made to find long-term trends of electron densities, hemispheric asymmetries, differences between AM and PM, as well as the typical behaviour of winter profiles. The mean behaviour of the empirical model is used as a reference.

Introduction

Among the characteristic features of the lower ionosphere one may mention (*a*) the short lifetime of the free electrons and (*b*) the difficulty of measuring absolute plasma densities. The former makes it customary to ignore transport phenomena and to assume steady-state (which implies no difference between sunrise and sunset for the same solar zenith angle), the latter restricts measurements reliable in their absolute values to rocket-borne wave propagation data. Outside the auroral zone no more than 118 such electron density profiles are available, an extra dozen or so was measured under irregular condition such as winter anomaly, post (geomagnetic) storm or solar eclipse. Based on these 118 profiles an empirical model was built which defines electron densities as a function of latitude, season, solar zenith angle and solar activity (Friedrich and Torkar, 2001a). Figure 4.19 shows a mass plot of the usable profiles (left panel), whereas in the right panel the variation with solar zenith angle of the model is depicted for conditions best covered by the data (low solar activity and low latitude). Although the full height range from 60 to 140 km is displayed, it should be borne in mind that the function describing the electron densities in this model was optimised to best reproduce the input data

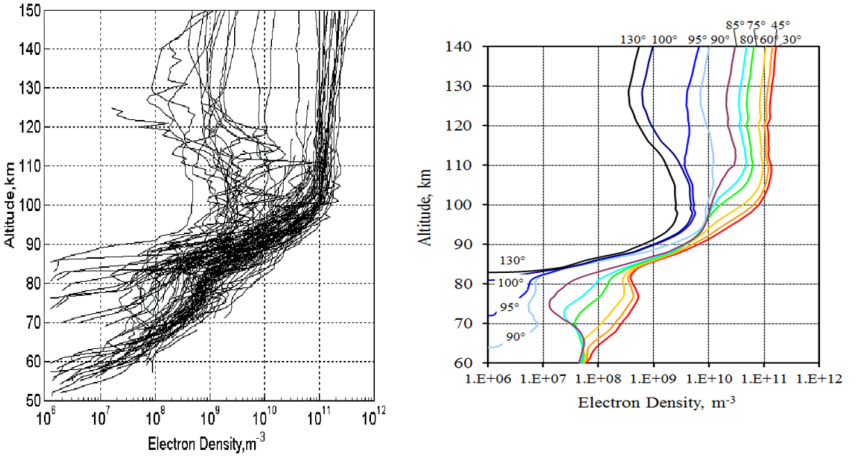


Figure 4.19 Summary of all available electron density profiles (left panel). The right panel depicts the dependence on solar zenith angle for median conditions according to the model built with the data of the left panel (low latitude, winter, low solar activity; Friedrich and Torkar, 2001b).

between 70 and 120 km. We will use this empirical model as a reference in the search for less conventional dependencies.

Long-Term Trends

The well-recorded increase by some 30% in CO_2 over the past 200 years is today generally attributed to anthropogenic causes. These undisputed data stimulated the search for trends in atmospheric parameters of potential interest to the wider public, notably temperature and ozone. Although temperature is the oldest recorded atmospheric parameter, it is very difficult to unambiguously extract genuine trends from errors or biases in the observations; nonetheless, about 0.5° appears to be the mean global rise in temperature during the 20th century. The concentration of ozone and other trace constituents has only become available on a meaningful global basis since the advent of remote-sensing satellites; the relatively short time covered by such measurements is only barely enough to convincingly separate solar cycle effects from possible trends.

The "trace gas" whose concentration was monitored the longest by means of ground-based observations are the free electrons. The longest time series of such measurements by ionosondes is the one from the Radio Research Station at Slough since 1931 to 1995, when it was moved to nearby Chilton, where an ionosonde is in operation to this day. In the absence of a "valley" ionosondes (theoretically) yield complete (bottomside) electron density profiles; however only the critical frequencies of the *E*, *F1* an *F2* layers can unambiguously be determined. The frequency fo*F2* and its (virtual) height hm*F2*, which can most easily be measured, show a large dependence on solar activity. Traditionally measurements are made at noon, hence the data also display a seasonal variation due to the change of the smallest (*i.e.*, noon time) zenith angle as a function of season. Also the scaling of the data prior to the availability of fully automatic digital data logging was inevitably subjective.

Ulich and Turunen (1997) used a unique set of ionosonde data covering 39 years which had been scaled and pre-processed by the same person (*i.e.*, arguably with the same bias). By forming the difference between a carefully chosen empirical relation defined as a function of solar activity and hm*F2* and the instantaneous height of the *F2* peak they derive a residual variation. This residual Δhm*F2* is furthermore smoothed by a running mean over a solar cycle

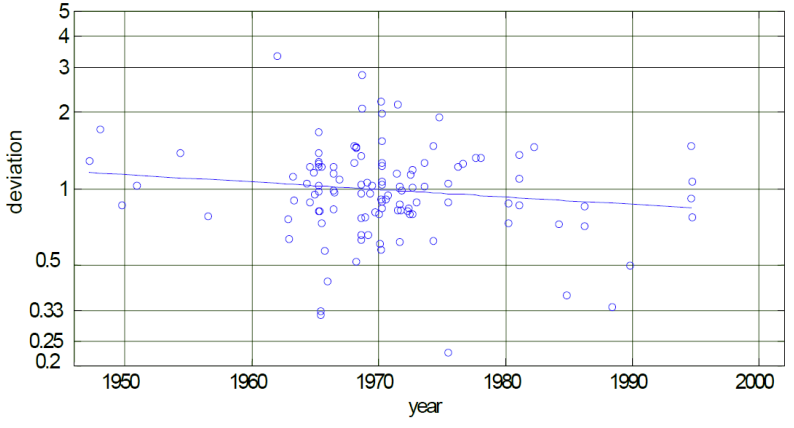


Figure 4.20 Departures of the individual measurements from the corresponding model values as a function of time. Note that data from all zenith angles, solar activities, altitudes and seasons are used (Friedrich and Torkar, 2001b).

and reveals a (seasonally averaged) lowering of 0.39 km a^{-1}. Since other possibilities such as the measured change of the local geomagnetic field could quantitatively be ruled out, their conclusion is a cooling of the upper atmosphere. More recently this trend analysis was repeated and extended by Roininen *et al*. (2015; Fig. 4.22) covering more than 50 years of observations. Jarvis *et al*. (1998) repeated this kind of analysis with data from the southern high latitudes and basically arrived at the same result. A cooling of somewhat lower atmospheric regions was also suggested from *in-situ* temperature measurements (Schmidlin, 1996) and again indirectly by the occurrence frequency of noctilucent clouds (Gadsden, 1990), or the reflection height of LF waves (Taubenheim *et al*., 1990). At auroral latitudes a cooling of the *F*-region has recently been derived from ion temperature data of 33 years of EISCAT data (Ogawa *et al*., 2014). The derived value of about 1 K a year could explain the observed lowering of the ionosphere above Sodankylä (Ulich and Turunen, 1997), and the shrinking of thermosphere of 1.7 % per decade seen in global satellite drag data from 400 km (*e.g.* Marcos *et al*., 2005).

Danilov (1997) analysed the ratio of the partial ion densities NO^+/O_2^+ at 120 km. Within the large scatter nonetheless a decrease during the day, but no apparent variation of the corresponding analysis at night at 150 km, was found. The period of his analysis covers three solar cycles and a distortion of the data due to enhanced neutral NO leading to more NO^+ during solar maximum cannot be ruled out. In a later paper, Danilov (2000) derived long-term trends in electron densities by forming the ratio between the measurement and a model; the model itself is merely described as "multi-dimensional". For the altitudes listed (75 to 90 km) he finds - depending on the selected data sub-set - an annual increase between 3 and 14%. We will essentially also pursue this approach, *i.e.* to analyse trends in the ratio of measurement to corresponding model value.

The simplest (and least revealing) presentation is shown in Figure 4.20. Each point is the deviation (factor) of a particular electron density profile from the corresponding model value averaged over the height region of the measurement. Clearly the heydays of rocket-borne measurements in the lower ionosphere were in the 1960's and 70's, but the slope of the regression line is primarily determined by the earliest such measurements in the late 1940's and our own four from Brazil in 1994. The negative trend of 0.7% p.a. is much smaller (and in the opposite direction) than those deduced by Danilov (2000), but in view of the large scatter in the data certainly not significant. The formation of the ionosphere is radically different during the day and at night; apart from the obvious absence of direct ionising solar radiation, the chemistry

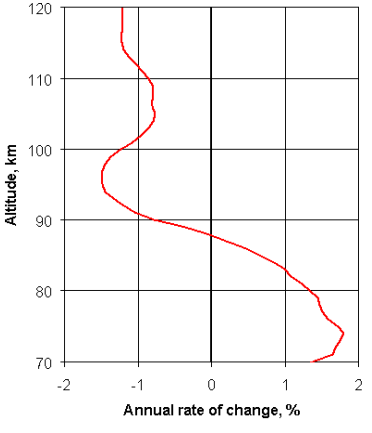


Figure 4.21 Annual rate of change in per cent of electron densities as a function of altitude (after Friedrich and Torkar, 2001b).

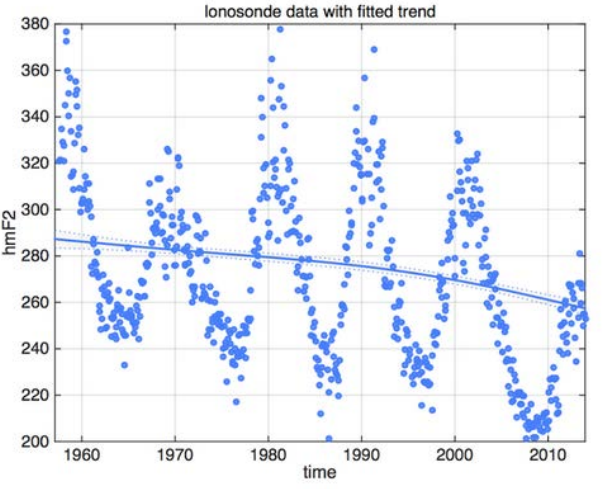


Figure 4.22 Trend of the hmF2 altitude (in kilometres) observed at Sodankylä, 67°N. The large sinusoidal variations are related to the solar cycle (Roinien et al., 2015).

governing the loss of free electrons is different for solar zenith angles beyond 98°. Let us therefore concentrate on the daytime profiles primarily because we have significantly more cases (102 *vs*. only 16). In the height range below, say 80 km both the production and the loss are different from what one expects above say 90 km. In order to generalise the apparent behaviour we form the slopes for each kilometre (Fig. 4.21) which reveals a region between 70 and 85 km with a pronounced increase, whereas the trend above is reversed. In the cluster region, the behaviour is qualitatively similar to the trend deduced by Danilov (2000), *i.e*. an increase, whereas in the upper *D*- and *E*-region the trend is reversed as *e.g.,* seen in fo*F2* data quoted by Danilov (1998). If we are desperate to propose a physical explanation, we may invoke a warming below the cluster ledge which leads to fewer clusters and hence a reduced effective recombination rate, and a cooling in the thermosphere.

Another long-term behaviour one may expect is QBO, the Quasi-Biennial Oscillation; we therefore simply try to find a spectral component with a 24 months period in the data. Only after carefully restricting the geophysical conditions (years 1962 to 1983, daytime, 83 to 85 km) one can indeed find a variation with an amplitude of ±8%.

Sun-Earth Eccentricity

The Earth's orbit around the Sun is slightly eccentric and varies by about ±1.6% with the perihelion in early January. This contributes to the fact that the seasons are less extreme in the northern hemisphere where winter is partly compensated by larger insolation. The ionisation rate is proportional to the flux and hence to the square of the Sun-Earth distance. Since, however, for conditions of steady state the resulting electron densities are proportional to the square root of the ion production rate, to a first approximation the electron densities in this scenario should vary linearly with the Sun-Earth distance. We again apply the usual procedure and form the ratio of the measured electron densities and the corresponding model values (daytime data only). This factor as a function of the day of the year is shown in Figure 4.23 employing data from all altitudes and zenith angles. Since the data cover regions with very diverse physical

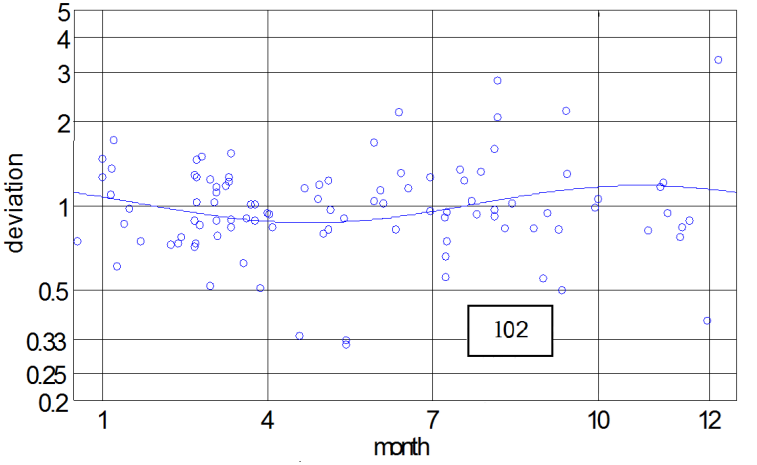


Figure 4.23 Departures of the daytime data from the mean (model) as a function of the day of the year; the number of cases is indicated in the box. Note that in November / December the data appear to be slightly above average (102 = number of cases; Friedrich and Torkar, 2001b).

processes, we will analyse each kilometre of height separately and for daytime only. Figure 4.24 shows the amplitude of the annual variation (left panel and the time of the maximum (right panel) as a function of altitude. Below 100 km the time of the maximum roughly agrees with the perihelion, at 100 km there is no variation, and in the *E*-region, the annual (residual) variation maximises in April; because of the small amplitude, the latter results are certainly not significant. If we only had data from the northern hemisphere the Sun-Earth eccentricity would be included in the seasonal variation. Since the data that entered the model also includes 21 measurements from the southern hemisphere, the present result my indeed be due to the eccentricity.

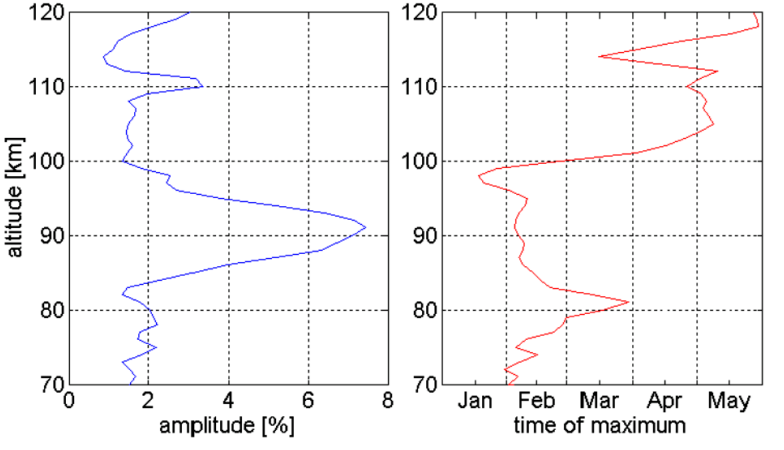


Figure 4.24 Amplitude and phase of the residual annual variation as a function of altitude (Friedrich and Torkar, 2001b).

AM/PM Asymmetry

The lifetime of ions in the lower ionosphere is very short (minutes), hence transport of ions is usually not considered in theoretical models. The time-dependent calculations (*e.g.*, Ogawa and Shimazaki, 1975) only show an appreciable hysteresis at 80 km and below of the order of about a factor of two for local times from 06 to 08 *vs*. 16 to 18 hours in the afternoon. At 90 km and above a hysteresis is only discernible for solar zenith angles a few degrees around twilight. We apply the usual procedure and plot the deviations of the measured densities *vs*. the model as a function of solar zenith angles from 180° to 0° (AM) and from 0° to 180° (PM). Figure 4.25 shows these departures averaged over the altitude region covered by each rocket flight. Fitting separate regression lines to the AM and PM data, respectively, would lead to discontinuities at noon and mid-night. We bypass that problem and assimilate the deviation by a sine wave.

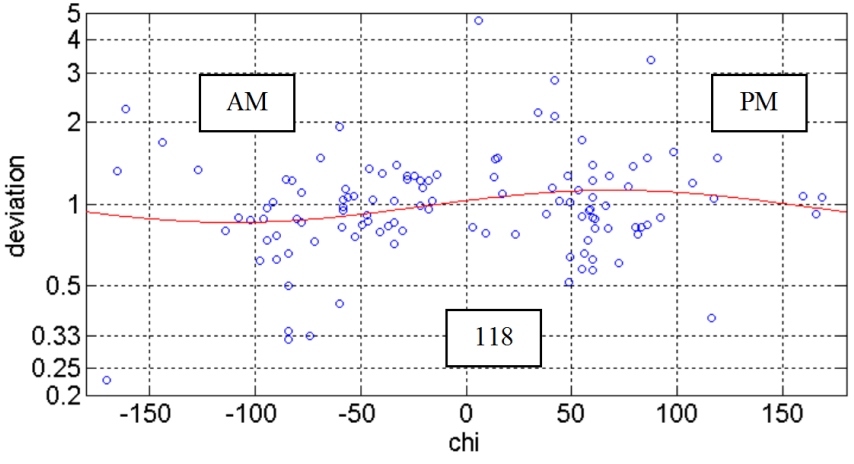


Figure 4.25 Departure of individual electron densities (all altitudes) from the mean (model) as a function of the solar zenith angle. Note that AM values are generally lower than those measured in the afternoon (Friedrich and Torkar, 2001b).

The fitted curve indicates somewhat larger densities in the afternoon. Because of the different physics prevailing above and below the cluster ledge, we repeat the analysis above 90 and below 80 km, respectively (Fig. 4.26). The minute hysteresis in the upper part (region dominated by molecular ions and X-rays) is certainly insignificant, whereas in the lower *D*-region where the ion production hinges on the concentration of NO, the hysteresis is substantial. Asymmetries in radio wave absorption have long been known (*e.g.*, Laštovička, 1977) and persist the whole day and not only around twilight (Torkar *et al*., 1978); this rules out "sluggishness" of the ion chemistry as the cause. The [NO] data from the limb scanning instrument HALOE aboard the satellite UARS indeed showed significant asymmetries particularly pronounced at the equator where the Sun rises fast (Friedrich *et al*., 1998). A theoretical explanation for this diurnal variation by tides was subsequently proposed by Marsh and Russell (2000). However, a hysteresis at night may be due to slow decay of the ionosphere, notably at higher altitudes. In fact, Knight (1972) presented night-time *E*-region electron densities simply with time after sunset as the main parameter. In a more recent study Han and Cummer (2010) presented variations of the night-time electron density ledge based on sferics observations. In most, but not all nights, the altitude of the this ledge drops during the night in the range from

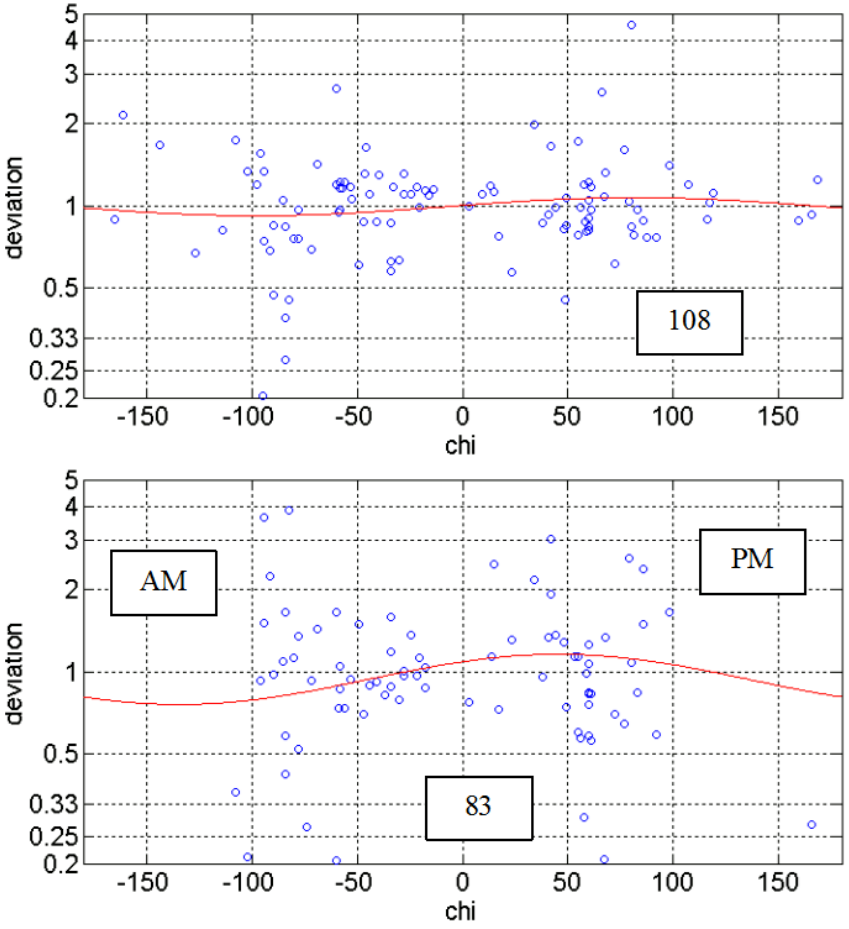


Figure 4.26 Departures as a function of solar zenith angle in the height region above the cluster ledge (90 to 120 km, top) and below (70 to 80 km, bottom panel; Friedrich and Torkar, 2001b).

about 88 to 82 km. These observations are from mid-latitudes and restricted to summer months. A similar lowering of the onset of appreciable atomic oxygen can explain the data, but night-time O-data of a similar coverage are not available to confirm this suggestion.

Hemispheric Asymmetry

According to a recent series of measurements (Lübken *et al*., 1999) the behaviour of the neutral mesosphere at high southern latitudes seems (or rather seemed at the time) to be in-distinguishable from that at northern high latitudes. Intuitively a hemispherical asymmetry may, however, be due to the Sun-Earth eccentricity *e.g.,* leading to a larger seasonal variation in the southern hemisphere. Ideally one should compare two independent models each built from southern and northern hemisphere data, respectively. Because of the paucity of data from the southern hemisphere this test can not be performed with any degree of statistical significance.

Conclusions

The relatively small number of "good" electron density measurements in the lower ionosphere makes the search for residual dependencies subject to a careful choice of the data subset and the anticipated variation. A long-term trend with increasing densities in the cluster region seems real and the decreasing trend in the *E*-region is supported by similar behaviour of other datasets. Signatures of a QBO cannot unambiguously be ascertained, whereas the Earth's orbit eccentricity does seem to be reflected in the data. The clearest empirical behaviour which modellers generally do not consider is the asymmetry between AM and PM. The most convincing explanation is *via* a corresponding asymmetry in global [NO], which the experimenters initially dismissed as an unresolved problem in the data retrieval of the remote-sensing instrument.

The present exercise yields results which are qualitatively acceptable, but quantitatively vague due to the paucity of data. Nevertheless, the analyses might inspire theoreticians to refine their simulations and experimentalists to look more closely into their data.

4.3. High Latitudes (IMAZ)

The ionosphere at high latitudes (*i.e.* of the auroral zone and the polar cap) is more often "disturbed" than not, *i.e.* it does not behave predictably according to the usual parameters such as solar zenith angle, solar activity and season. In order to describe the high latitude ionosphere we need one or more parameters characterising the degree of disturbance in addition to the parameters that suffice to describe the non-auroral ionosphere. For the same reasons as outlined in the description of the FIRI model (Friedrich and Torkar, 2001a) we predominantly rely on electron densities from rocket borne wave propagation experiments; because every geophysical situation is not only characterised by zenith angle, *etc.*, but disturbances of varying degrees can distort the prediction, and one needs considerably more data than for describing the non-auroral situation. These extra data are available from the incoherent scatter radar EISCAT located near Tromsø, Norway, right in the centre of the auroral zone.

Data sources, pre-processing and quality assessment

The validity and quality of any empirical model not only hinges on a physically reasonable mathematical description, but crucially also depend on good input data. A section dealing with the rather elaborate pre-processing and screening of the input data is therefore expedient. To some extent this pre-processing allows the use of lower mathematical orders of the model proper, thus conceivably making it more robust.

4.3.1 Auroral Latitudes

Electron densities

Rocket borne wave propagation data

There are 133 electron density profiles (now 155) available from rocket borne wave propagation experiments of geomagnetic latitudes ranging from 61.03° (Red Lake) to 68.87° (Ft. Churchill; both sites in Canada) but the bulk of the data originates from flights from Esrange, Sweden, and Andøya, Norway; these data are taken at their face value. In some cases the rocket payloads also carried *in situ* electron probes; after establishing a normalisation factor at heights where there is overlap with the wave propagation data, we trust the normalised probe profile above and below the region covered by the propagation experiment. Because the normalisation factor is potentially height dependent (because of aerodynamic effects or varying payload potential) we only extend the "properly" established electron density profiles (*i.e.* due

to the wave propagation data) by the electron probe data by not more than 5 km. The threshold of the wave propagation experiment depends both on the sounding frequency employed and on the collision frequency (background neutral pressure; altitude). Approximate limits for the usual sounding frequencies are 10^7 m^{-3} at 80 km and 10^9 m^{-3} at 60 km (Jacobsen and Friedrich, 1979), whereas normalised Langmuir probe data have yielded credible electron densities as low as 10^6 m^{-3} (Friedrich *et al.*, 1997).

Ion probe data

Measuring the number density of positive ions with probes is in some way easier than measuring electrons, despite the lower currents collected by these probes. The main reason is that - because of the lower mobility of ions - the payload potential is less critical for the collected current (*cf. e.g.*, Folkestad, 1970). Above about 80 to 90 km - notably during the day - negative ions can not exist such that a (normalised) density profile for positive ions can be used as a proxy for electrons. In the presence of appreciable neutral density (<80 km) negative ions are formed by attachment of electrons to neutrals. In full darkness and absence of appreciable concentrations of atomic oxygen (< 80 to 90 km) only the ratio between electron attachment rate β to the ion-ion recombination rate α_i and the neutral background (number) density M determines the ratio between electrons N_e and positive ions N^+ (see Section 5.4, and Friedrich and Torkar, 1995b). After some manipulation of Eq. (5.11) one obtains electron density as a function of positive ion density:

$$N_e = \frac{N^+}{1 + \frac{\beta}{\alpha_i}\frac{M^2}{N^+}} \tag{4.6}$$

Figure 4.27 shows the electron density measured by the sounding rocket F69 together with the positive ion density of that same flight, normalised to the electron density above 90 km. The ratio β/α_i was varied for best agreement between the measured electron density between 80

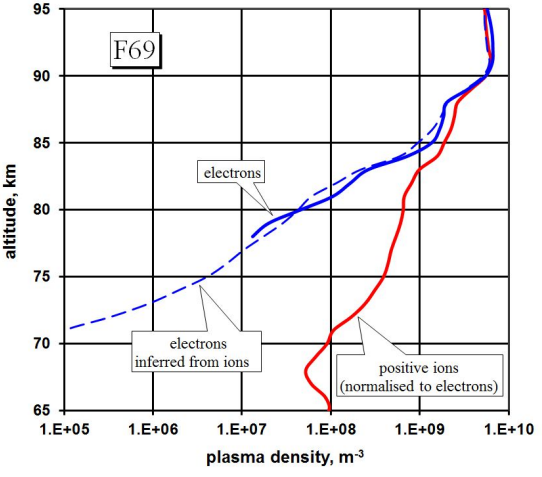


Figure 4.27 Plasma densities measured by the sounding rocket F69. The "inferred electrons" are deduced from the (normalised) ion density assuming ion-ion recombination as the only loss process for negative ions (see text).

and 90 km and the electron density inferred from the ion density using the above Eq. (4.6). We have 15 night-time rocket flights with ion probe data which reach to lower altitudes than the wave propagation measurements. The best value of the ratio β/α_i has to be found by trial-and-error for each rocket flight individually. Unless some of the data should be based on unreliable measurements, the conclusion is that the ratio β/α_i below, say 80 km does not vary with altitude, whereas between 80 and 90 km this ratio can vary which is attributed to additional the electron attachment to meteoric particles (*cf.* Friedrich *et al*., 2011). The procedure is applied to these 15 rockets to somewhat extend the electron density profiles downward, but again by not more than 5 km, and to resulting electron densities not lower than 10^6 m^{-3}.

Incoherent scatter radar (EISCAT)

The incoherent scatter radar EISCAT (= European Incoherent SCATter) is located at Ramfjordmoen (near Tromsø, 69.59°N) and actually consists of two co-located radars, VHF operating on 244 MHz and UHF at 931 MHz. The beam of the former can only be tilted in geomagnetic NS direction, whereas the latter is fully steerable. All EISCAT electron density profiles were interactively inspected, erroneous profiles completely rejected, and unreasonable values notably at the bottom of the profiles, as well as obvious sporadic *E*-layers, or PMSE signatures, eliminated from the rest of the profile. For UHF the generally accepted threshold density is 2×10^9 m^{-3} and much of the available data therefore originate from the *E*- and *F*-regions, whereas the VHF threshold of 3×10^8 m^{-3} makes it better suited for *D*- and *E*-region studies. Incoherent scatter radar measurements have a "system constant", *i.e*. effectively a calibration factor. The established procedure to obtain this constant is to compare the *F*-region peak electron density with the fo*F2* measured by a co-located ionosonde. Because the electron

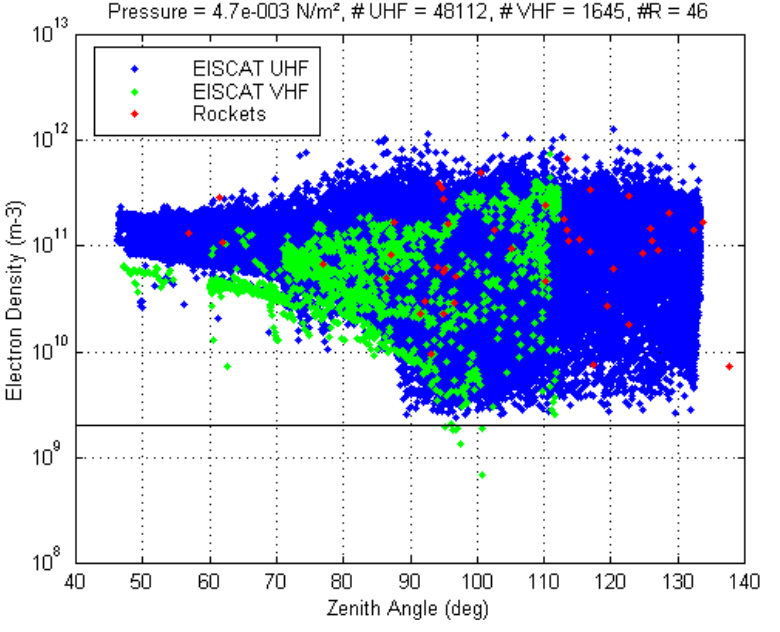


Figure 4.28 Electron densities at a pressure typical for the E-region peak. Note that in this example the EISCAT VHF data are consistently below the corresponding UHF values (Harrich, 2001).

density profiles from the UHF radar always extend beyond the *F*-region peak, its profiles are calibrated more frequently against the local ionosonde, whereas many of the archived VHF data do not extend to the *F*-region and could thus only sporadically be normalised provided a pronounced *E*-layer was seen in an ionogram. Consequently, we consider the UHF data to be better in their absolute values than the VHF data. In order to be able to merge the data from the two radars in a meaningful way we indirectly calibrate the VHF data against the UHF data by simply plotting all available values of the two radars from a pressure surface representative of the *E*-layer (4.7×10^{-3} N m^{-2}; about 100 km) as a function of solar zenith angle (Fig. 4.28). In the example shown in full daytime (solar zenith angle <90°) the VHF results are clearly below the UHF values by about a factor of 1.4. In practice we analyse the VHF data for all the time periods of measurements (each of up to a few days) and plot the values together with *all* available UHF data taken at comparable solar activity. Naturally this procedure can only be applied for daytime conditions, hence not for mid-winter data. Figure 4.29 shows the history of the necessary correction to the VHF data thus obtained between 1990 and 1996; after that time we assume that the VHF radar has become better calibrated, and that by the year 2000 the

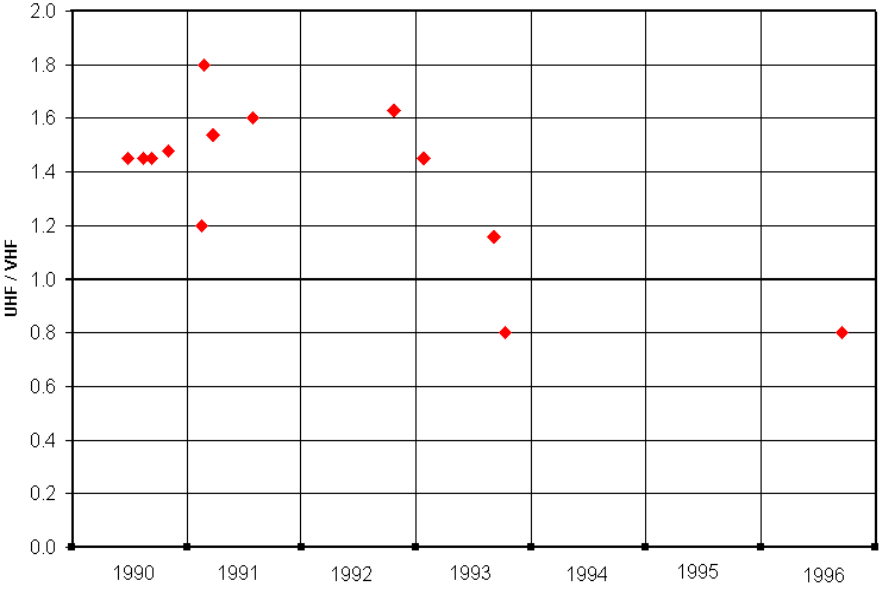


Figure 4.29 Correction to the EISCAT VHF system constant as a function of time.

necessary correction has approached unity. Many of the measurements were taken using oblique beams. These beams cover different location (projections) and the solar zenith angles calculated for the location of EISCAT are different along the beam. Figure 4.30 shows the quartiles of the zenith angle errors as a function of altitude. Since we are here only concerned with altitudes below 150 km the quartiles of this solar zenith angle error only reach 0.5°, an insignificant value which we ignore.

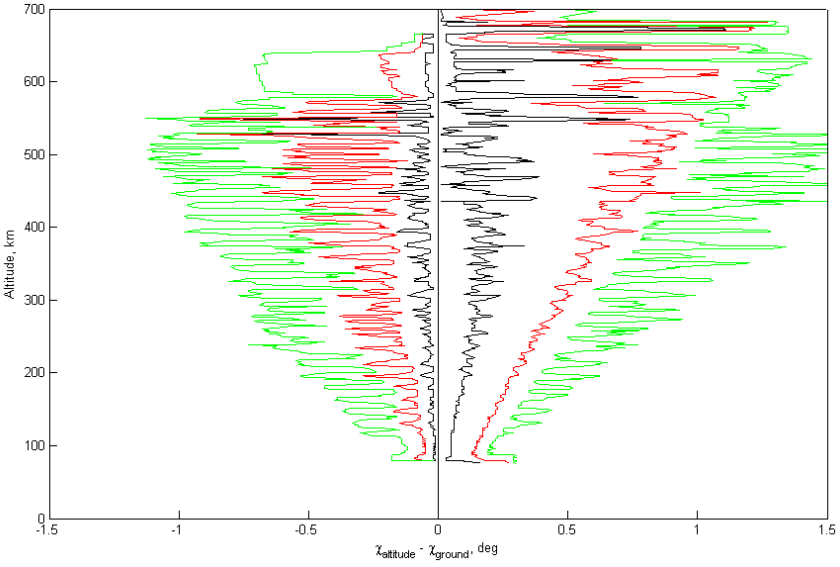


Figure 4.30 Solar zenith angle error as a function of altitude due to the obliqueness of the EISCAT beams (quartiles [black and green] and medians [red]). Note that below 150 km even the upper quartiles of the distribution never exceed 0.5°.

Antarctic probe data

A series of Soviet rocket borne probe measurements from the Antarctic auroral zone (Molodezhnaya) is available (Danilov *et al*., 2003). Although uncalibrated probe measurements are inherently uncertain in their absolute values (factor of 2?), they are numerous (244) and were all launched by the same scientific group using the same type of sounding rocket. These data which are only available to us from three altitudes (75, 80 and 84 km) are used but for a limited purpose.

Disturbance parameters

Riometer

A riometer (= *R*elative *I*onospheric *O*pacity meter) is a ground based instrument which measures excess absorption L_r by the ionosphere relative to a quiet level of natural extra-terrestrial radio sources. The frequencies used are between 20 and 50 MHz, *i.e.* high enough never to be totally reflected, or to deviate appreciably from line-of-sight propagation. Radio wave absorption is to a very good approximation proportional to the product of electron density N_e times electron-neutral collision frequency ν, which in turn is proportional to pressure p. Whereas pressure does not vary much with season or latitude, electron densities vary by orders of magnitude. Nonetheless, the product (specific absorption, dB km^{-1}) varies in magnitude, but the contributing height region is fairly constant and 90 km is a good average altitude (*cf.* Fig. 4.31). In our effort to empirically model the high latitude ionosphere, we use riometer absorption data from three different sources:

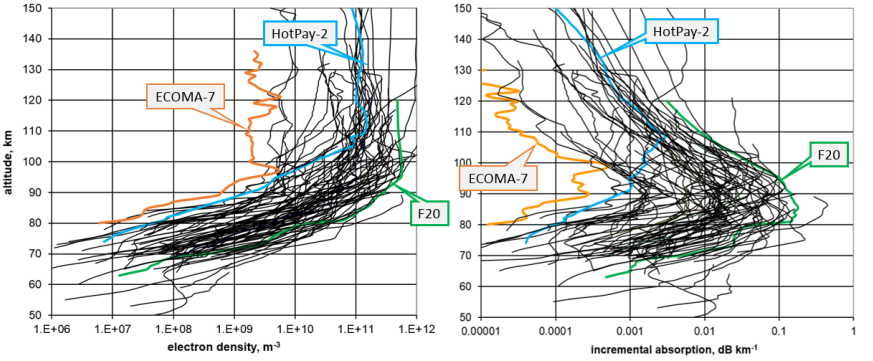


Figure 4.31 Some widely differing high latitude, night-time electron density profiles measured by sounding rockets (left panel) and the resulting specific absorption (@ 27.6 MHz, right panel). Generally larger absorptions tend to peak at lower altitudes.

a) The oldest and simplest instrument used an upward looking three element Yagi. Such an antenna has a typical opening angle of ±30°. An absorbing layer (the *D*-region) with *n* dB. absorption covering the whole field of view will reduce the received signal by *more* than *n* dB.

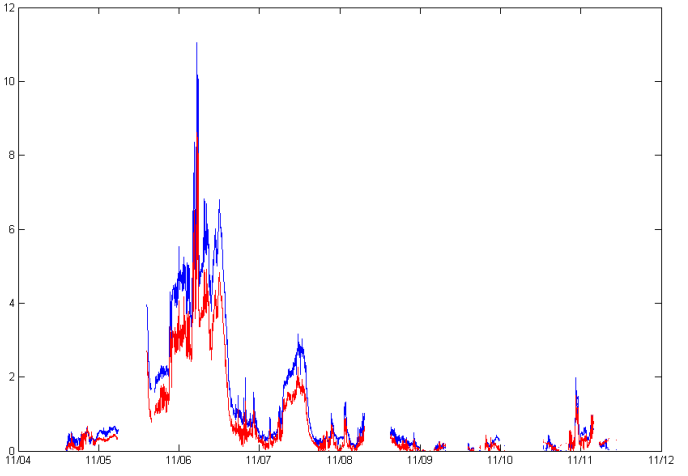


Figure 4.32 Riometer absorption from 8 days in November 2001 measured by the Kilpisjärvi IRIS riometer at 38.2 MHz. Blue values are from the wide beam antenna (beam 50), red ones from the central narrow beam 25.

The reason is that the radio source is not only in the zenith, but mainly originates from sources in the galactic plane (Fig. 4.33). Hence, the received noise power varies with sidereal time in a pattern depending on the geographic latitude of the station. However, as seen from the antenna, there are sources in the zenith and to the side, therefore the noise contributions from off-zenith have a longer path through the ionosphere and consequently will be more absorbed (Fig. 4.34). As a result the effective opening angle of a wide beam riometer becomes smaller for

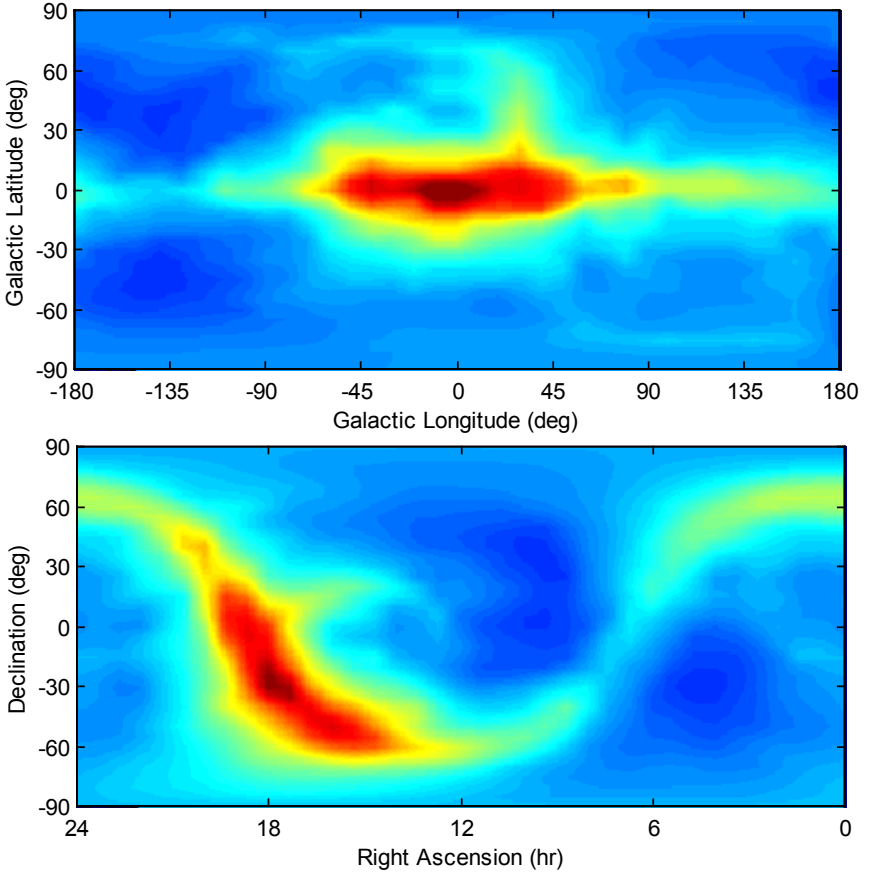


Figure 4.33 Radio noise sources at 30 MHz in the sky shown in two different coordinate systems (Cain, 1978).

large absorptions. A detailed analysis of this effect, calculated for 70°N, was made by Friedrich *et al*. (2002). Using the numerical results of that paper one can convert readings of wide-beam riometers to what one should expect if the source were a point in the zenith, provided one can assume full and uniform coverage of the field of view by an absorbing layer. A test of the collimating effect of the ionosphere is shown in Figure 4.32 where riometer readings (38.2 MHz) from the IRIS installation Kilpisjärvi are shown. "Narrow" refers to the central, upward looking beam with the smallest opening angle, "wide" is a co-located wide-beam riometer. An eight-day period in November 2001 was chosen which included a very strong absorption event. Generally, the wide-beam riometer shows larger absorption, particularly apparent in cases of the large absorption. In Figure 4.35 we plot the values of the wide-beam instrument *vs.* the corresponding ones from the narrow beam (Friedrich *et al.*, 2002). The procedure does indeed bring the wide-beam results closer to the narrow beam values, although the correction would need to be somewhat larger. Possible explanations may be that:

- For the correction in Friedrich *et al.* (2002) the altitude of the absorbing layer was assumed to be at 90 km, whereas particularly strong absorption events are all due to protons (PCA, *cf*. Fig. 4.31); in such cases the peak of the absorption occurs at somewhat lower heights.

- The effective opening angle of the wide-beam riometer may actually be a little wider than the ±30° assumed in the collimation correction; the logarithmic scatter plot (lower panel of Fig. 4.35) reveals that the discrepancy is only of the order of 12%. On the other hand, an even larger opening angle of the wide-beam antenna contradicts the observed maximum-minimum ratio of the quiet day curve (see Figure 5 in Friedrich *et al*., 2002), but of course the actual beam shape may differ from the one assumed and is furthermore most likely not rotationally symmetric.

- Not least, the map of extra-terrestrial radio noise (Cain, 1978) used for the conversion from wide- to narrow-beam absorption is based on measurements at 30 MHz, whereas the riometers used for the present test operate at 38.2 MHz. A more recent sky noise map is the one by Oliveira-Costa *et al*. (2008) which can be interpolated for any frequency between 10 MHz and 100 GHz.

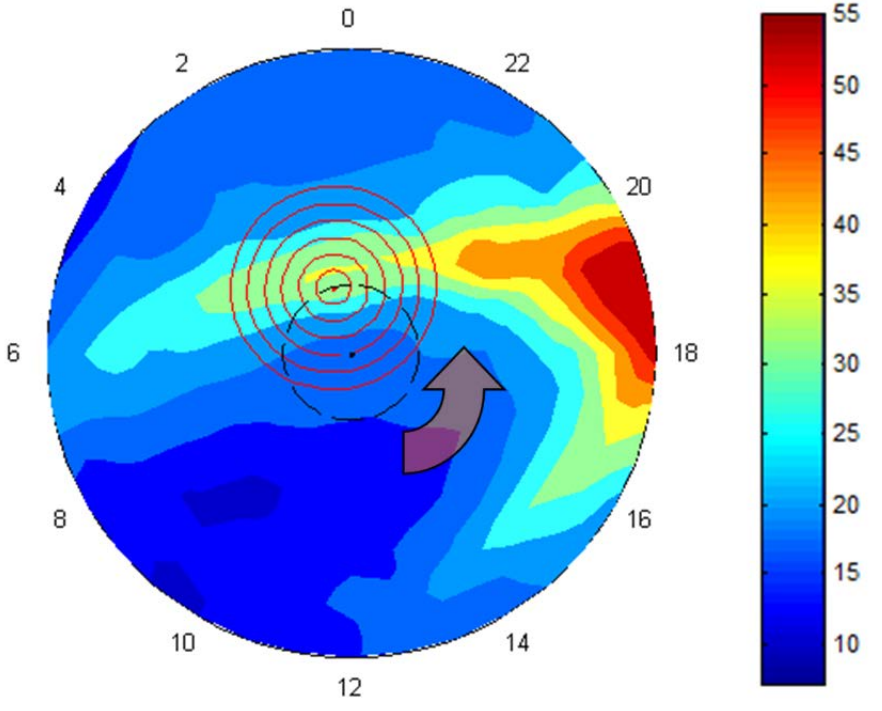


Figure 4.34 Sky noise "seen" by an upward looking antenna located at 70°N during a sidereal day. The latitude 70° is indicated by the dashed circle and the sky temperature at 30 MHz is give by the colour coding (in 1000 K). The red circles indicate the parts of the sky covered by antenna opening angles from ±5 to ±30°.

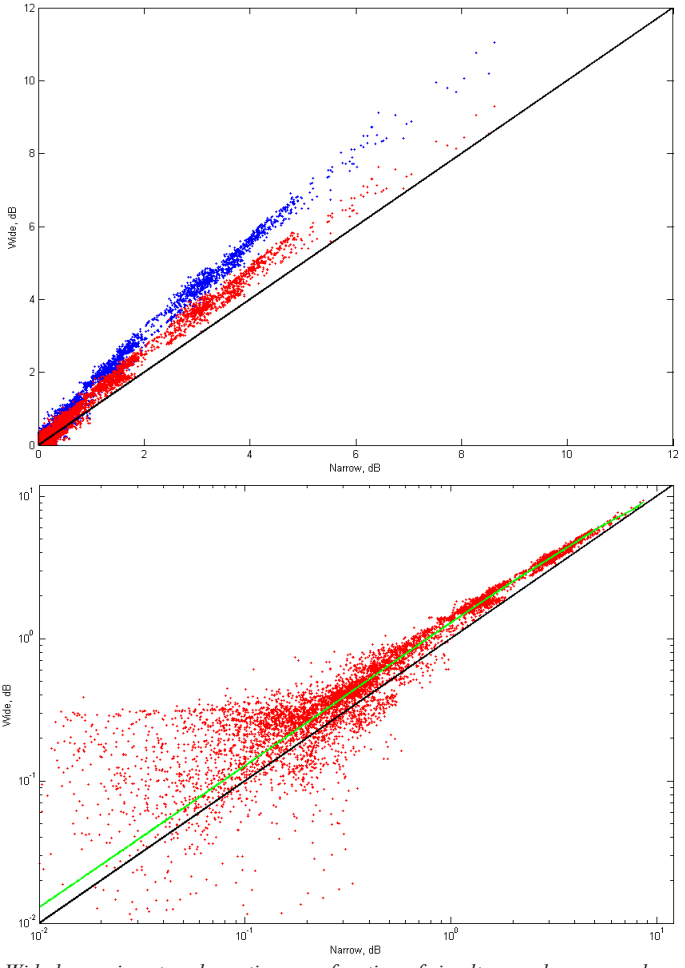


Figure 4.35 Wide beam riometer absorption as a function of simultaneously measured narrow beam absorption; the data are from Figure 4.32. The blue dots are the original wide beam values; the red ones are corrected according to Friedrich et al. (2002). The green line in the lower panel is a 6th order polynomial fitted to the corrected wide beam data and represents the residual error in the conversion from wide to narrow beam (ca. 12%).

Prior to the availability of the imaging riometer IRIS (1995) wide-beam riometers near EISCAT were used. One site was co-located with EISCAT at Ramfjordmoen, the other was Lavangsdalen some 15 km away. Each site had two independent riometers operating on 32 and 40 MHz, respectively. The readings of these four riometers were converted to what a riometer at 27.6 MHz would measure using an inverse square relation of the frequency. This simple conversion holds because all frequencies are considerably higher than the local electron gyro frequency of 1.4 MHz. Other exponents referred to in the literature to relate riometer readings at different frequencies were later dismissed as erroneous interpretation of the data (*e.g.,* Rosenberg *et al*., 1991). Next we form the mean of these four riometer readings, disregard the value that deviates the most from the mean and use the mean of the remaining three riometers.

If only three readings are available, we use the mean of the remaining two, if only one or two values are available their readings were checked - or rejected - if their readings were outside a certain margin compared to the value measured at Andøya some 130 km away. Modern riometers use circularly polarised antennas (crossed Yagis) phased to receive the more sensitive x-mode. According to simulations, the x-mode is typically absorbed about 20% more than the o-mode. Early riometers, which used linearly polarised antennas, received a mean of the two modes. In order to make their data comparable with more recent recordings, we add 10% assuming that this emulates receiving the x-mode.

b) The aurora is a very dynamic process and normally does *not* cover the whole sky. In order to be able to observe the associated ionospheric absorption, imaging riometers are the preferred configuration today. Pixels of the absorption images are generated by narrow beams, which in turn are achieved by appropriate phasing of a large antenna field. The installation IRIS in Kilpisjärvi, Finland, consists of an array of 8 by 8 antennas with which 49 narrow beams are formed. The -3 dB opening angles somewhat depend on the look direction and vary from ±6.3° for the upward looking central beam to about ±7° for oblique beams. The collimation effect of such narrow beams is small and can safely be ignored. The data available on the web are already converted to vertical incidence by the cosine of each beam's zenith angle. The majority of the EISCAT measurements is with beams that are not vertical. In order to achieve a meaningful relation between electron density and riometer absorption we calculate the geographic longitude and latitude where the EISCAT beam intersects the 90 km layer (*i.e.* the layer of maximum absorption, *cf.* Fig. 4.31). We then interpolate the IRIS riometer value for the location of the intersection's footprint. If they fall outside the area covered by IRIS we flag the EISCAT data as lacking associated riometer data. Figure 4.36 shows some ground tracks of actual EISCAT beams intersecting the 90 km layer.

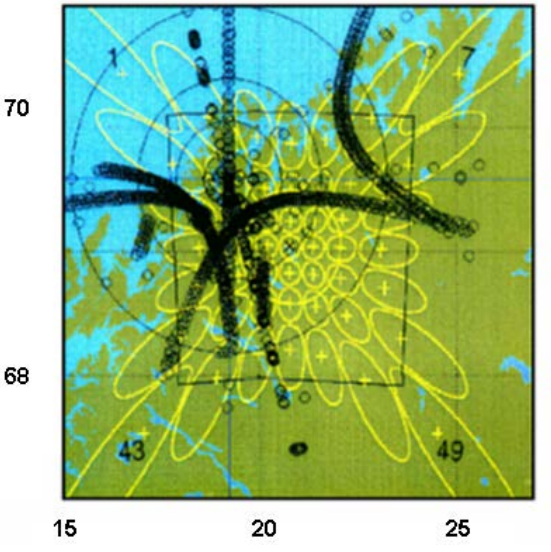


Figure 4.36 Pixels of the imaging riometer IRIS at Kilpisjärvi. The elliptically shaped pixels (1 to 49) represent conic sections of the respective beam opening angles with the 90 km layer. Also indicated are ground tracks of piercing points of some oblique EISCAT beams with that layer. If these intersects fall outside the indicated square, the electron density data are flagged as lacking coincident riometer data.

c) If an electron density profile with sufficient height coverage is available, one can simulate the absorption of a hypothetical source in the zenith. Ideally, the difference between measured riometer absorption (corrected for vertical incidence) and the simulated value should only be very small and only due to the quiet (*i.e.* solar zenith angle controlled) ionosphere. A "sufficient height region" in this context is coverage from below 70 km to beyond 110 km in order to obtain at least 90% of the total absorption. Since even carefully maintained riometers have a typical uncertainty of ±0.1 dB, riometer absorption computed from electron density profiles will more accurately represent barely measurable riometer absorption than data obtained from the ground-based instrument. If a measured rocket profile has the required height coverage we use the simulated values L_i and subtract the calculated quiet absorption L_q (see later) to obtain L_r. Notably for small values of absorption we consider this procedure to yield data closer to "ideal" than can be obtained from an actual riometer. For EISCAT electron densities we rely on corresponding absorption measured by riometers and do not simulate integral absorption. The reason is that only rarely do the IS profiles have the required height coverage (reaching to 70 km or below). However, when they do reach to low enough altitudes the associated measured riometer absorption is always fairly large that even a substantial uncertainty (*e.g.*, 0.2 dB) introduces only a small relative error and we assume that on average they represent correct L_r (*i.e.* as often too large as too small).

As the input to the ionospheric model we use absorption values as they would be measured at 27.6 MHz using the x-mode. The choice of this frequency may appear arbitrary, but it is the frequency used in the early days of riometry. Except for input or output visible to the user, the model internally operates with integral absorption L_i, *i.e.* riometer absorption $L_r + L_q$. L_q is due to the True Quiet ionosphere (see later) and is added to the measured riometer absorption, whereas it is contained in the computed absorption from "full" electron density profiles.

Geomagnetic Index

The accepted explanation for short-term variations in the geomagnetic field is that currents in the *E*-region produce a magnetic field which is superimposed on the Earth's intrinsic field. Such currents require an electric field and enhanced conductivity, *i.e.* enhanced electron densities. Because of the short lifetime of free electrons in the *D*-region, geomagnetic field perturbations must be due to enhanced electron densities in higher layers, *i.e.* the *E*- or *F*-region.

The geomagnetic index *Kp* is a semi-logarithmic index describing departures from a mean diurnal variation and is established from the readings of a number of sub-auroral geomagnetic observatories for 3-hour intervals. The ionosphere at high latitudes varies on much shorter time-scales, and is of course more closely related to magnetic disturbances within the auroral zone. The geophysical observatory Tromsø establishes its own geomagnetic three-hour index K_T. For the purpose of modelling, we use K_T because the bulk of the electron density data originate from EISCAT, which is located only a few kilometres from Tromsø. For the few data from locations outside Scandinavia we convert *Kp* to K_T using an empirically established relation between the two indices (Fig. 4.37). Although K_T still is an average disturbance index over too long a period (3 hours), it does - not surprisingly - correlate better with electron densities from EISCAT (mainland) than *Kp*. In order to make the model compatible with IRI we also produce it in a version where the geomagnetic disturbance is represented by *ap*, essentially *Kp* on a linear scale.

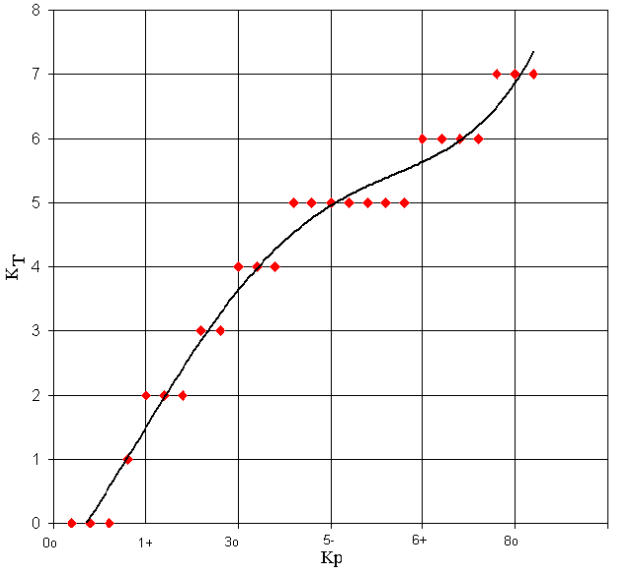


Figure 4.37 Local magnetic index K_T established at Tromsø vs. the planetary geomagnetic index Kp.

Threshold problem

Electron densities

For signal frequencies much larger than the electron gyro frequency, according to the "classical" magneto-ionic theory (Appleton-Hartree; *cf.* Eq. 2.15) radio wave absorption per unit length k_L is:

$$k_L \propto N_e \nu \tag{4.7}$$

Since we use riometer absorption as the prime parameters to quantify disturbance (enhanced electron densities) we can expect N_e to be directly proportional to absorption L. Riometer absorption L_r by definition can be as low as 0 dB, but the ionosphere would still absorb. For processing we therefore always add L_Q of the quiet ionosphere (see later) and work with the integral absorption L_i. As a test we plot electron density values from all three sources (rockets, EISCAT VHF and EISCAT UHF) at a particular pressure level *vs*. integral absorption prevailing at the time of the measurements (Fig. 4.38).

As expected higher L_i is associated with larger N_e and the electron density scatter becomes smaller for larger absorption. This is simply due to the dual-logarithmic presentation where small absolute uncertainties or variations are smaller relative to larger values. The two density thresholds of the EISCAT data are clearly visible, below these only a few rocket data exist. Intuitively, one would assimilate this behaviour by a simple function and we show a linear fit on the logarithmic scale and an overlapping binning procedure. For small absorptions, the binning results in particular, lead to electron densities just above the EISCAT threshold. This effect is simply due to the fact that in the distribution of electron densities associated with small absorption values is incomplete, *i.e.* we lack the values that to fall below the respective thresholds. In order to avoid results thus biased to high values in the distribution we introduce

an absorption-dependent threshold. By simple "eyeballing" we draw envelopes to the electron density distribution and where the lower envelope intersects the respective EISCAT thresholds we define the absorption-dependent threshold for that particular pressure level (indicated by crosses). This rather subjective method was applied at all 59 pressure surfaces for UHF and VHF, and day and night (Fig. 4.39). For further processing we apply an electron density threshold as a function of integral absorption and background pressure as indicated in that figure by the two RMS-lines. This rather drastic, but necessary procedure reduces the number of usable data points by about 30%.

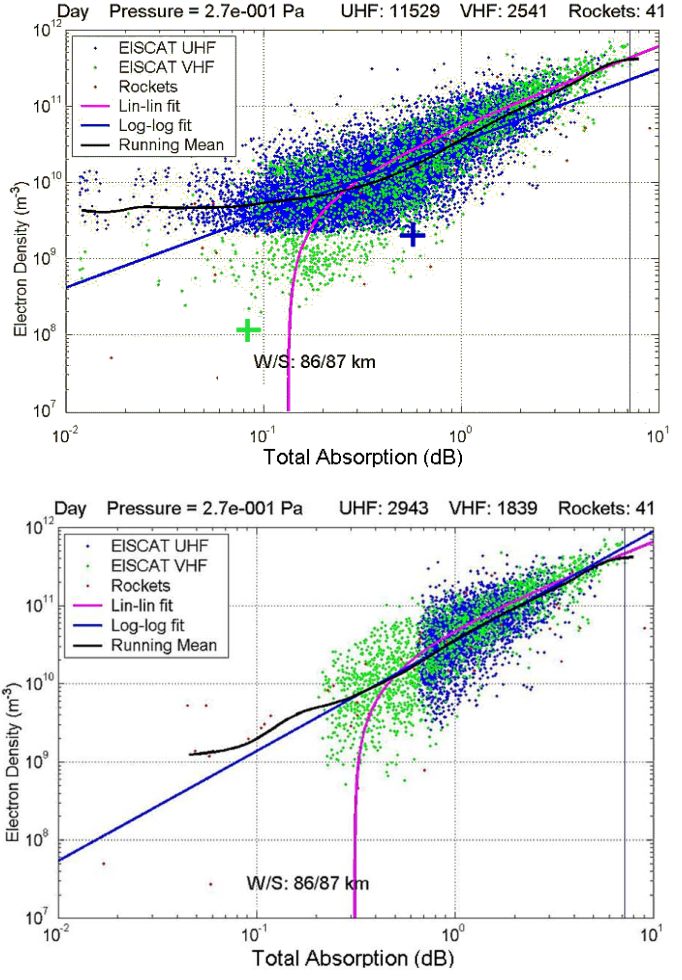


Figure 4.38 Electron densities vs. integral absorption at a pressure typical for 86 to 87 km (winter, summer). Note that the linear fit would lead to negative electron densities for low absorptions, and that binning would always yield a value above the threshold. The crosses indicate the absorption limits beyond which we assume that the full scatter in the data is above the respective threshold. In the lower panel the data are restricted at the indicated thresholds. Here all trends (binning, log-log and linear fit) reach much lower values for small absorptions (Steiner, 2003).

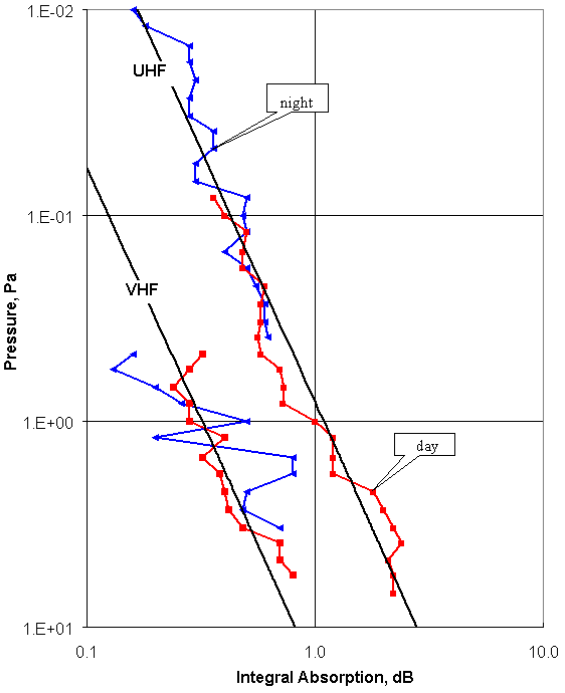


*Figure 4.39 Absorption threshold beyond which electron densities are not expected to be biased by the EISCAT threshold. Integral absorption is plotted vs. pressure for EISCAT UHF and VHF (red: **day**, blue: **night**). The individual values were established by hand for each pressure surface, the lines used in further processing are linear RMS fits with the same slope, but displaced by a factor of 3.415 (Steiner, 2003).*

Riometer

Riometer absorption is the signal loss relative to the quiet diurnal variation (QDC) of the received extra-terrestrial radio noise. Notably small values of riometer absorption crucially depend on a well established QDC. Even the best maintained riometers are uncertain - largely because of the QDC - by about 0.1 dB. Riometer absorption smaller than this uncertainty can therefore lead to negative values which are always flagged as missing measurements. Values close to, or smaller than the uncertainty are most likely biased because - as with the EISCAT electron densities - we miss part of the distribution of values. Perhaps somewhat arbitrarily we consider riometer readings below 0.07 dB as not necessarily wrong, but as not representative, and therefore dismiss them.

Quiet Ionosphere

The ionosphere at high latitudes is much more often "disturbed" than not. By "undisturbed", we imply that it behaves as outside the auroral zone, *i.e.* predictably controlled by solar zenith angle, solar activity and season. Common parameters to identify disturbances are the global geomagnetic index *Kp*, locally measured geomagnetic disturbances or riometer absorption. However, experience has shown that conditions perceived to be quiet based on these data can still vary considerably, *i.e.* some "disturbances" are far below the threshold of ground based instruments.

As discussed in Section 3.4 the benchmark electron density data in the lower ionosphere originate from rocket borne radio wave propagation experiments. The number of such measurements is however, very limited and amounts to no more than 133 in the auroral zone. Electron densities are also available as one of the parameters of the European Incoherent Scatter facility EISCAT. The height resolution of EISCAT is generally poorer (at best 1 km) and the threshold electron density is also significantly larger than that of rocket borne measurements. The sheer volume of the data (60,000 profiles) is impressive and we assume that - although the data scatter more than in the case of *in-situ* measurements - on average they have no systematic bias. EISCAT uses two frequencies (and antennas), namely UHF (931 MHz) and EISCAT VHF (244 MHz). Whereas the UHF radar is better and more often calibrated against a local ionosonde, VHF has a lower density threshold. Figure 4.40 shows electron densities at a pressure surface corresponding to about 110 km (depending somewhat on season) from EISCAT UHF, VHF and sounding rockets plotted against solar zenith angle χ. The figure clearly shows that there is a pronounced lower electron density limit which at night must be somewhere below the EISCAT threshold (2×10^9 m^{-3} for UHF). The variability of the data is large, particularly at night. Although rocket borne measurements have a much lower threshold density, there simply is not enough data to establish an envelope for night conditions. We define

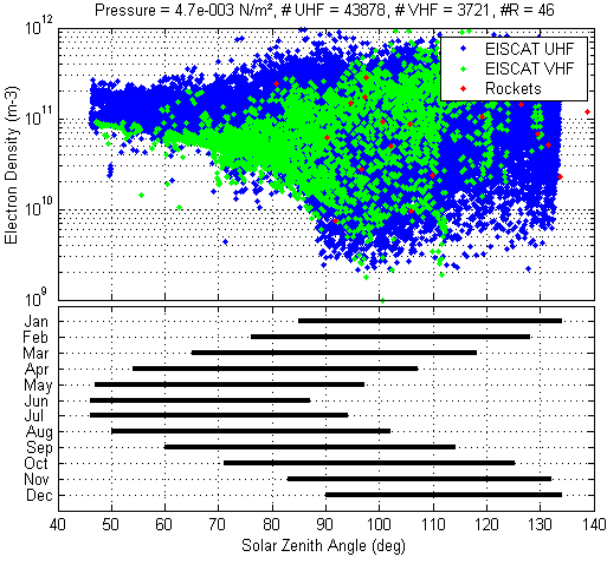


Figure 4.40 Electron densities from various sources near 110 km vs. solar zenith angle. The zenith angle ranges at EISCAT of the different months are also indicated (Friedrich et al., 2004a).

electron densities as "true quiet" by the envelope of the data in that figure. This envelope appears to follow a solar controlled variation such that one can expect to be able to linearize the variation by plotting the electron densities N_e *vs*. log cosχ as predicted by a pure Chapman formation of the form $N_e = N_o(\cos\chi)^n$ (N_o = subsolar electron density). The four panels in Figure 4.41 show the procedure we apply for the purpose of obtaining the envelope: (*a*) We exclude data which are *a-priori* known to be disturbed (riometer >0.05 dB, *Kp* >3$^+$, for profiles without associated riometer absorption data we limit *Kp* to ≤0$^+$), (*b*) the data are plotted *vs*. the logarithm of a 1/Chapman function (to allow the extension of the analysis to about 90°), (*c*) the curve is tilted, and (*d*) a histogram is formed. The data are restricted to ±1.6 σ and the remainder is approximated by a Gaussian distribution. The tilt is iteratively varied for narrowest distribution of the data in panel (*d*). The tangent at the inflection point of the Gaussian distribution is formed and its intersection yields the subsolar electron density N_o, *i.e*. the density for $\chi = 0°$, whereas from the tilt one can derive the exponent *n*. Under quiet conditions, EISCAT does not yield densities below about 88 km. For the *D*-region proper, a set of 244 rocket borne probe measurements exists from the former Soviet Antarctic station Molodezhnaya at the altitudes 75, 80 and 84 km (Danilov *et al*., 2003). For want of better data we still use these measurements because (*a*) the neutral atmosphere background in the conjugate hemisphere seemed indistinguishable from that in the north as evidenced from a series of falling sphere measurements (Lübken *et al*., 1999), although later analysis points to some systematic differences. The most evident difference is near the mesopause (*i.e.* warmer in the southern summer; Lübken and Berger, 2007), (*b*) the data stem from a series of coherent measurements using the same instrument and the same sounding rocket and we expect the data to represent absolute values to within better than a factor of two, and (*c*) Molodezhnaya is at a geomagnetic

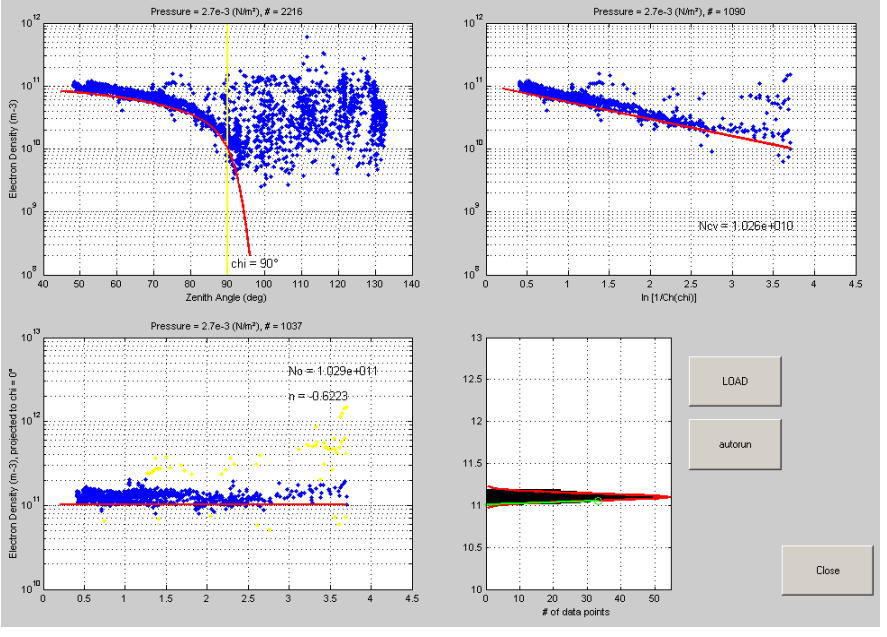


Figure 4.41 User interface of the procedure applied to obtain the envelope (True Quiet, TQ) electron densities (Friedrich et al., 2004a).

latitude comparable to both the Scandinavian rocket ranges Andøya and Esrange as well as EISCAT Tromsø. At any rate, we expect that using these data is preferable to an extrapolation from higher altitudes, lower latitudes or from disturbed to quiet conditions. The analysis demonstrated in Figure 4.41 is carried out for 12 logarithmically spaced pressure levels per decade (*i.e.* about every kilometre of height) and yields subsolar electron density N_o and the exponent n. This procedure is repeated for overlapping solar activity levels (interpolated for the actual time of the day; 20 sfu wide, in 5 sfu increments) and thus yields a solar activity dependence of both N_o and n at each pressure level. The only smoothing with altitude is a weighted running mean (25, 50, 25%) of the coefficients in adjacent layers. The solar activity dependence is assimilated linearly for the exponent and for the logarithm of the subsolar electron density. Figure 4.42 shows N_o and n for two solar activity levels (80 and 180 sfu) as a function of background pressure. Further details involved in the procedure are given by Harrich (2001).

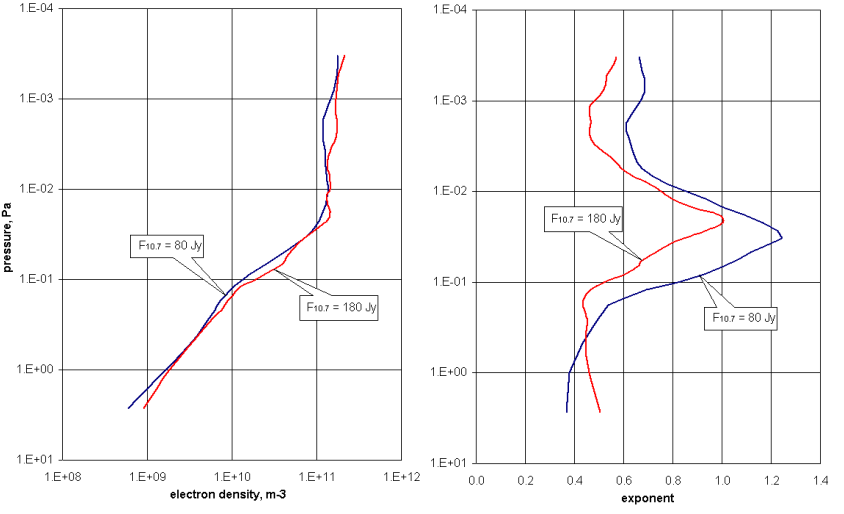


Figure 4.42 Subsolar electron density N_o and the exponent n for two solar activity levels as a function of background pressure.

Figure 4.43 shows electron densities obtained with this method for medium solar activity and different zenith angles, including a very low night-time profile as a proxy for True Quiet (rocket NLTE-2 at $\chi = 117°$). Since the data analysis is carried out for pressure surfaces, one needs to specify the day of the year; in this figure, summer solstice is chosen. Results for other seasons can readily be obtained by using pressure/height relations according to the atmosphere in Figure 4.4. The effect of solar activity is depicted in Figure 4.44 for a zenith angle of 80°; the figure also includes the corresponding results of the non-auroral model FIRI at 60°N (Friedrich and Torkar, 2001a) and clearly shows that also the quiet electron densities in the auroral zone are larger than outside the zone. The real difference is likely even larger since FIRI is essentially a mean of all "normal" values, whereas True Quiet is the low-density envelope of all values expected to be quiet. In an aeronomical interpretation, this suggests that NO densities in the auroral zone (70°) are larger by almost a factor of ten compared to 60° latitude. The seasonal variation for three solar zenith angles at three selected altitudes is shown in Figure 4.45. The displayed variation is only due to the variation of the pressure surfaces neglecting conceivable changes introduced by the seasonally varying neutral atmospheric composition and temperature

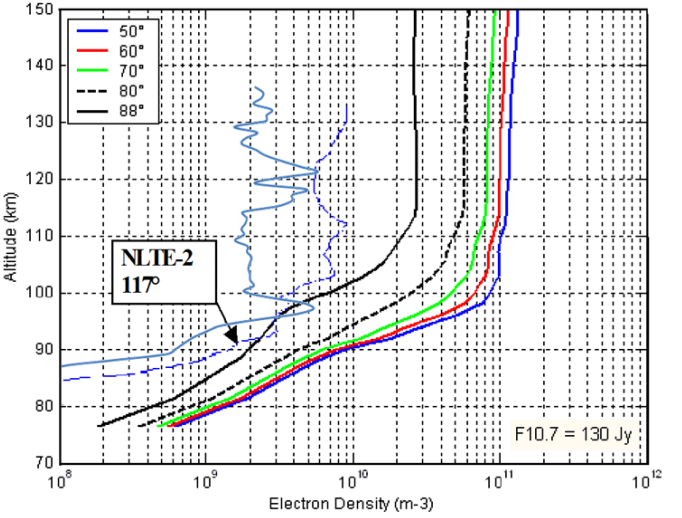


Figure 4.43 True Quiet electron densities for medium solar activity as a function of solar zenith angle. The profile of NLTE-2 is also included (dashed line), but in 2010 ECOMA-7 measured electron densities even lower by about a factor of 2 (full blue line). The height scale applies to summer solstice (Friedrich et al., 2004a).

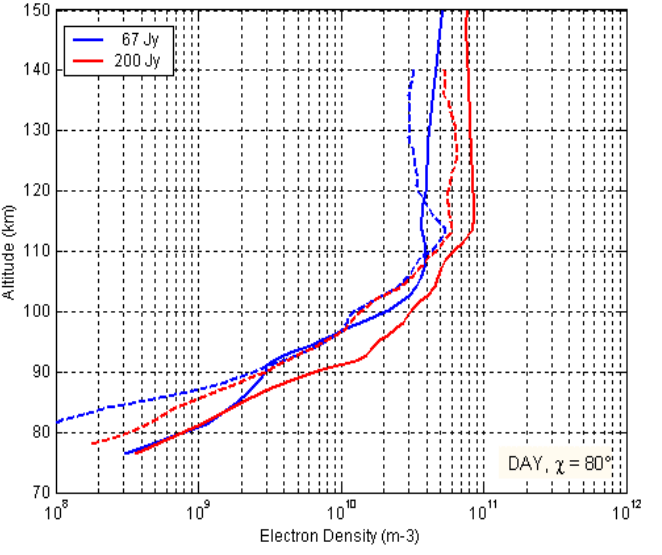


Figure 4.44 Quiet electron densities at a solar zenith angle of 80° for low and high solar activity (mid-May). The dashed lines are corresponding results from the empirical, non-auroral model FIRI at 60° latitude (Friedrich et al., 2004a).

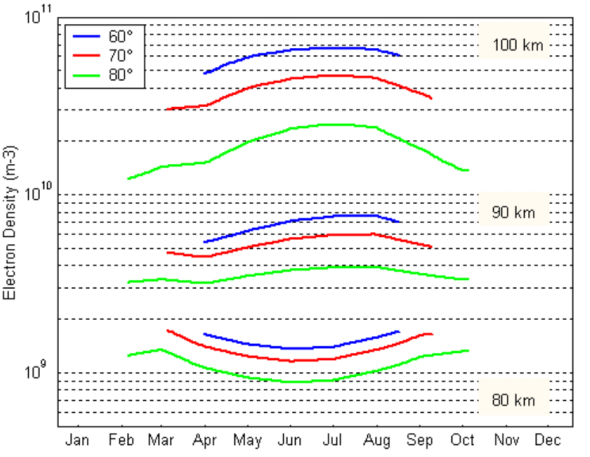


Figure 4.45 Seasonal variation of True Quiet electron densities at three altitudes and for three solar zenith angles (Friedrich et al., 2004a).

The expected True Quiet electron densities for night-time conditions are all below the EISCAT threshold. The rocket flight NLTE-2 from Esrange in March 1998 was the then lowest electron density ever measured by a sounding rocket in the auroral zone (NB: since then ECOMA-7 in 2010 yielded electron densities even lower by about a factor of 2). We therefore use the profile of this flight for the *D*-region, extend it in the *E*-region by EISCAT-VHF data and extrapolate up to 150 km with electron densities from non-auroral night-time measurements in mind. Below about 70 km, we use values somewhat lower than those of NLTE-2 with calibrated Langmuir probe measurements from Alcântara, Brazil, in mind (flights 31.103 and 31.105 of August 1994). The resulting profile (Fig. 4.46) is definitively provisional and may eventually turn out to be even lower.

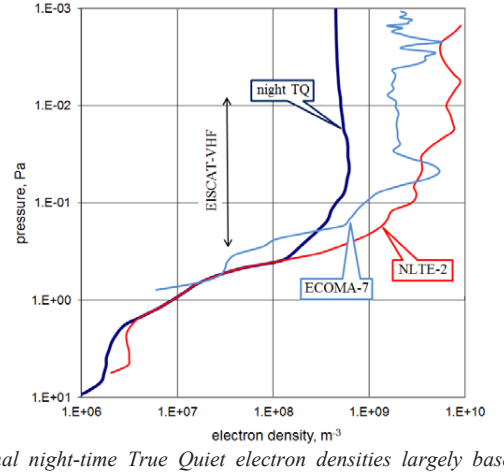


Figure 4.46 Provisional night-time True Quiet electron densities largely based on rocket flight NLTE-2 and EISCAT-VHF. The extrapolations above and below (arrows) are "inspired" by non-auroral electron densities.

Disturbed Auroral Latitudes

The high latitude ionosphere is more often "disturbed" than quiet; in this context "disturbed" means not behaving predictably as a function of solar zenith angle, solar activity and season. On the ground these disturbances are chiefly quantifiable by (*a*) riometer absorption and by (*b*) deviations of the Earth's magnetic field. There is ample literature on the principles behind the riometer (*e.g.,* Hargreaves, 1969), a treatise of the wide-beam riometer is given by Friedrich *et al*. (2002); here we will only repeat the most important features of this instrument.

A riometer (= *R*elative *I*onospheric *O*pacity meter) is a ground based instrument which measures excess absorption L_r by the ionosphere relative to a quiet level of natural extra-terrestrial radio sources. The frequencies used are between 20 and 50 MHz, *i.e.* high enough never to be totally reflected, or to deviate appreciably from line-of-sight propagation. Simulation of absorption (L_i) using measured electron densities or from an empirical model ideally yields a little more than the absorption of the geophysical condition it is meant to represent. The extra absorption is due to the quiet (solar controlled) ionosphere. We use the rest absorption L_q obtained by simulation using the True Quiet electron densities (*cf.* the previous section; Harrich *et al*., 2003); the results of other conceivable procedures to derive L_q scatter far too much to obtain reasonable results.

$$L_i = L_r + L_q \text{ (integral absorption = riometer absorption + quiet absorption)} \qquad (4.8)$$

Riometers use different frequencies, lower ones are more sensitive, but can conceivably experience total reflection at least for contribution from large zenith angles (Schwentek and Gruschwitz, 1970). Today 38.2 MHz is frequently used, on the one hand because this frequency is internationally reserved for radio astronomy, and on the other, because man-made noise (*e.g.* harmonics from switching) is less severe.

There have been several precursor attempts to model the high-latitude ionosphere using riometer absorption as the disturbance parameter. Jespersen *et al.* (1968) simply averaged the earliest rocket borne *D*-region electron density profiles according to high and low absorption. Miyzaki *et al*. (1981) related the logarithm of electron density to riometer absorption at various altitude levels; this assumed relation has no physical justification, rather a *linear* relation can be expected for electron density *vs.* absorption. Friedrich and Torkar (1983a) used the same idea, but related electron density to integral absorption, *i.e.* riometer absorption plus an *a-priori* unknown rest absorption due to the quiet ionosphere. In a later paper, the authors essentially repeated the exercise, but used neutral density rather than altitude (Friedrich and Torkar, 1995a); this way it was hoped to largely account for seasonal variations. The vast amount of EISCAT (UHF) data tempted Kirkwood and Collis (1991) to simply bin the data according to solar zenith angle, riometer absorption and *Kp*; unfortunately, the EISCAT threshold distorted the outcome at low altitudes and low riometer absorptions and the resulting electron densities at low altitudes were always slightly above the EISCAT threshold (*cf*. Fig. 4.38). The first attempt to pool the large number of EISCAT data with the few rocket borne measurements with their much lower threshold was published by Friedrich and Kirkwood (2000). Independently, Hargreaves and Friedrich (2003) attempted to model the lower ionosphere based on EISCAT and riometer only.

Harrich (2001) used EISCAT data from both UHF and VHF, together with - presumably - all electron density profiles measured at the time by rocket borne wave propagation instruments. He assumed a linear relation between integral absorption and electron density for constant pressure levels. Because of the anticipated significant difference between day and night-time recombination rates, the data were split for zenith angles larger and smaller than 98°. Steiner (2003) developed this model further by introducing bins of variable width at each

pressure level, but still retaining separate models for day and night. In addition, for the first time, restricted data at small absorptions were used (*cf.* Fig. 4.39).

Recently, the topical technique of using Neural Networks (NN) has been employed to develop an empirical model for the high latitude lower ionosphere. An NN is a computer algorithm that is trained to "learn" the relationship between a set of given input parameters and the corresponding output. The method is extremely successful in providing prediction tools for functions that are non-linear. The relationship between ionospheric behaviour and the geophysical parameters that govern this behaviour is highly non-linear and therefore lends itself to the use of NNs. The available data (EISCAT plus rocket) for geomagnetic 70°N were used to develop an NN-based model for the high latitude ionosphere. For the purpose of this

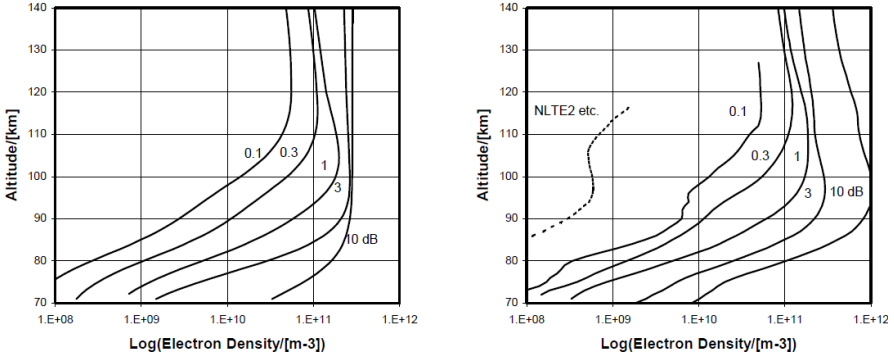


Figure 4.47 Night-time electron densities for various values of riometer absorption in late April. Left panel: NN approach, right panel: binning of data at constant pressure surfaces (MEDAL model, Steiner, 2003; for details see McKinnell and Friedrich, 2003).

development a feed forward back-propagation NN algorithm was employed, with an architecture that contained three layers (input, hidden and output). More information on NNs and the different algorithms can be found in Haykin (1994). In an initial attempt, this model was developed separately for day- and night-time, with the boundary between them at a solar zenith angle of 98° (as for the analytical models). The criteria that were used for determining a good NN model included the root mean square (RMS) error between the measured and predicted electron densities, the ability of the predicted profiles to reproduce the total absorption they are meant to represent, the shape and location of predicted profiles and the variation of the predicted electron densities with input variables. In addition, an NN can only predict reasonable results when interrogated for input data that fall within the input space. Therefore, it is important to understand and know the limits of the input space. To judge the ability of NN to predict an

	RMS error in logN_e *vs.* respective model	RMS error in logN_e *vs.* combined model	degradation *vs.* dedicated model, %
all data	0.2414	0.2414	-
night data	0.2894	0.3005	2.6
day data	0.1787	0.2075	6.8

Table 4.2 RMS errors of the combined and the dedicated models, and the trade-off of the combined model.

average profile given a set of input parameters, comparisons were made with the current available analytical models that were discussed in the previous section. Figure 4.47 shows the predicted profiles for various riometer absorptions as determined by the NN approach and the MEDAL model (Steiner, 2003). These results are from the dedicated night-time model, which intentionally uses the same set of input parameters as the MEDAL model, and are discussed in full by McKinnell and Friedrich (2003). Below 3 dB riometer absorption the results of the two approaches are fairly similar strengthening the confidence that the approach is reasonable. The corresponding results from the daytime model are shown in Figure 4.48 where profiles have been predicted for a mid-April day at a solar zenith angle of 80° and various levels of riometer absorption. Both models established for the comparison with the analytical precursor versions show a convergence of the profiles in the *E*-region. This is to be expected since only riometer absorption and no geomagnetic index was used to characterise the ionosphere. Details of the development of the dedicated daytime model can be found in McKinnell *et al.* (2004).

The new approach not only makes use of the technique of Neural Networks (NN), but also crucially differs from the earlier approaches in that:

- day and night values are treated jointly, and
- for each disturbed electron density value, the corresponding True Quiet (TQ) value (*i.e*. the same geophysical conditions, but zero riometer absorption) is used in training the NN.

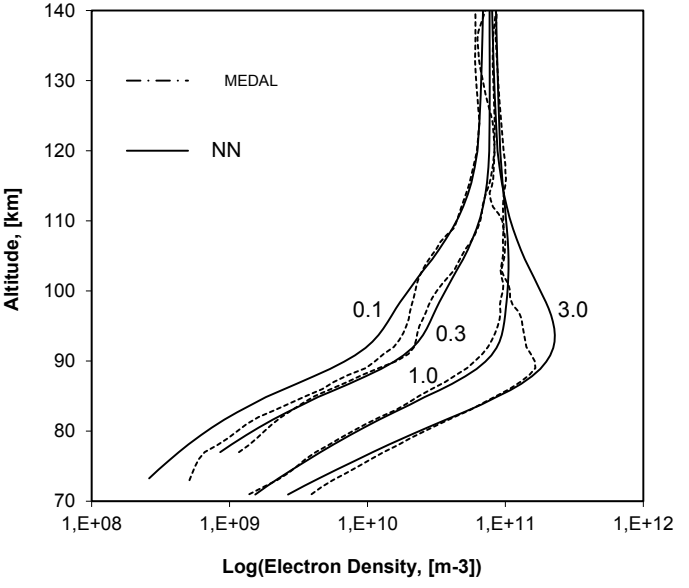


Figure 4.48 Comparison modelling with a conventional approach (MEDAL model, Steiner, 2003) and a neural network with the same inputs. Results are shown for mid-April, a solar zenith angle of 80° and different levels of riometer absorption (McKinnell et al., 2004).

As a result, the NN can be interrogated for any riometer absorption values - including zero dB - which then yields the TQ values; a separate model for True Quiet is thus not required, although TQ was obtained by a procedure separate from the NN. Much of the data at high latitudes is taken under twilight conditions. This is therefore the most common situation to be modelled, but is at the edge of the input space of dedicated models for day and night. A model built using all data has obvious advantages (*a*) there is no need to devise a smooth transition from one model to the other, (*b*) more data are involved in the establishment of the NN, and (*c*) there is no need to define a - perhaps controversial - zenith angle limit to split the dataset. The penalty one has to expect is a poorer prediction ability expressed as a larger RMS error between the measured and predicted output values. The RMS errors listed in Table 4.2 are calculated for the disturbed electron density values only, although the NN was trained for minimum RMS errors against both the disturbed and the corresponding True Quiet values. The table also shows the RMS errors of the two dedicated models and the RMS-"penalty" of the universal model. Clearly, the degradation of a few per cent is acceptable in the light of the advantages gained.

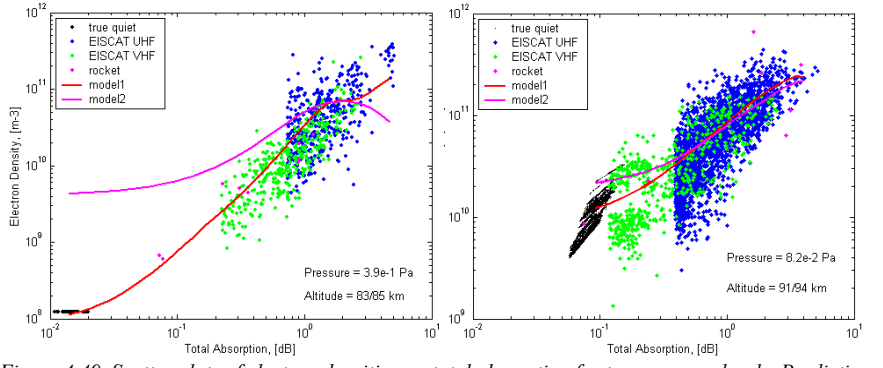


Figure 4.49 Scatter plots of electron densities vs. total absorption for two pressure levels. Predictions have been determined for median winter conditions (left) and for median summer conditions (right). The red line indicates variation according to our optimum model, which included the True Quiet values (black dots) in the input set, and the pink line shows the prediction for the same input parameters, but determined from only the disturbed (actually measured) values.

The reason for including a True Quiet electron density value for each disturbed (measured) value in training the Neural Network can best be demonstrated by Figure 4.49. Scatterplots at two altitudes (84 and 93 km) of electron densities *vs*. total absorption L_i are shown. Clearly, the model that was trained without the True Quiet values extrapolates to unreasonably large values for quiet conditions, whereas in the region of the disturbed values (mostly originating from EISCAT) both curves appear to approximate the data equally well. In Figure 4.50 the ability of the Neural Network, trained with the inclusion of True Quiet values, to reproduce the True Quiet electron densities is demonstrated. The upper panel shows the True Quiet profiles for selected solar zenith angles for mid-April, whereas the lower panel displays these profiles obtained from the Neural Network for the same set of zenith angles. The agreement is good enough to eliminate the necessity of using separate models for True Quiet and disturbed electron densities.

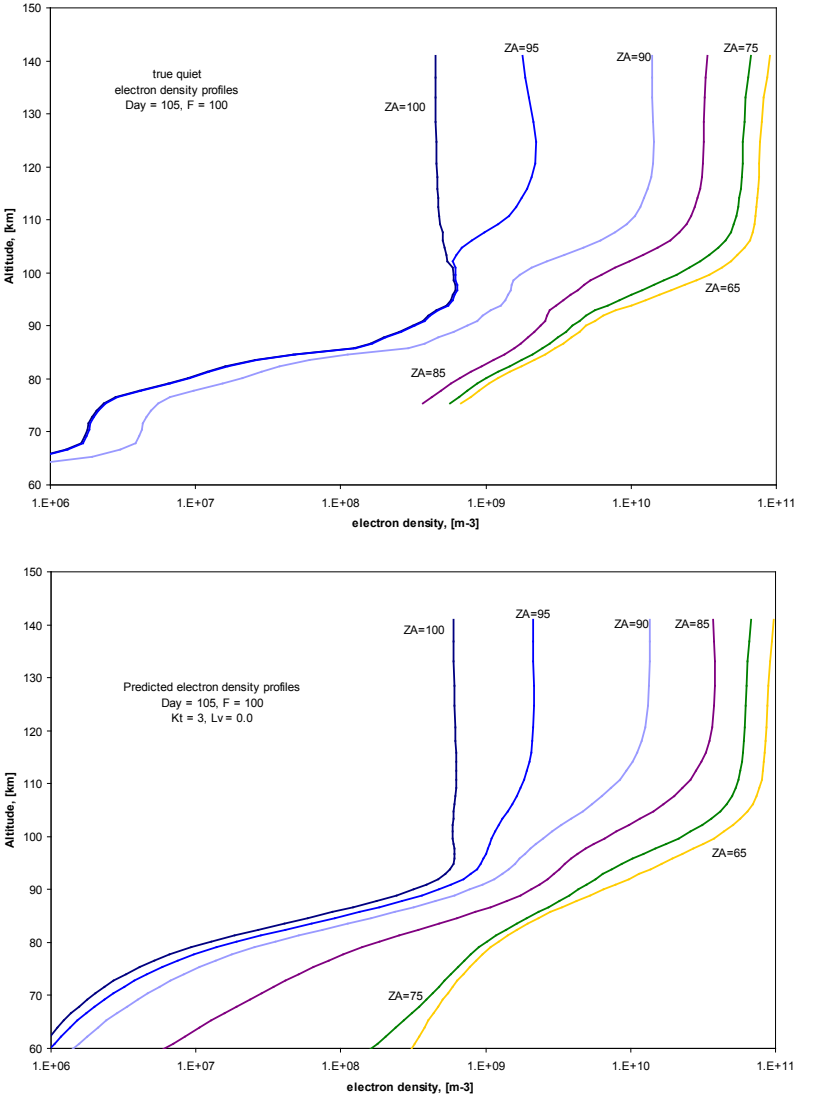


Figure 4.50 True Quiet electron densities at various solar zenith angles for mid-April and median conditions otherwise. Top panel: values used in training the Neural Network, bottom panel: corresponding values obtained from the model.

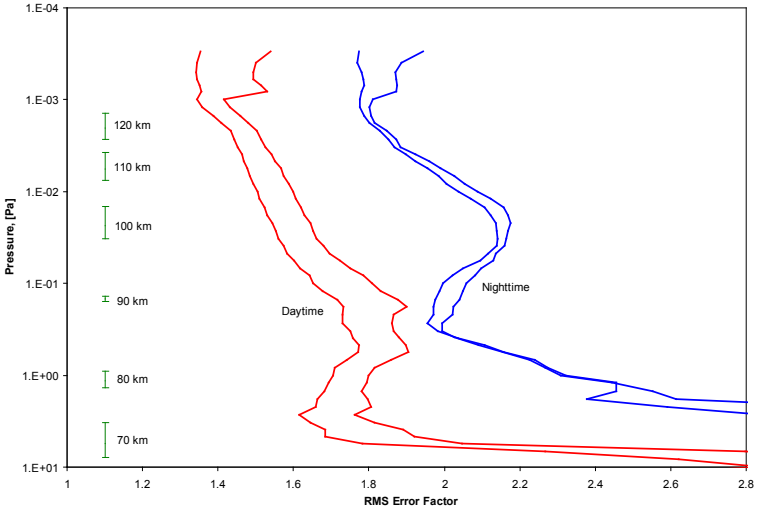


Figure 4.51 RMS error factors as a function of background pressure. Red curves are for daytime, blue ones for night. The somewhat larger errors are for the universal model, the lower ones from the respective dedicated models. The seasonally dependent range of pressures for altitudes from 70 to 120 km is given for orientation.

Figure 4.51 show these RMS errors as a function of background pressure; approximate altitudes ranges are given for guidance. Two curves each are depicted for day and night conditions. The lower values are from the respective dedicated models, the somewhat larger values were obtained by comparison of the results of the combined model with the day and night data, respectively. Clearly, the RMS errors at all pressure levels (altitudes) are slightly larger, but extensive tests have revealed that the advantages definitively outweigh the slightly poorer prediction ability. The model version which appears to be a good compromise between stability (no unrealistic excursions) and the ability to reproduce the input values, has an input layer for five input parameters, one hidden layer with 40 nodes, and one output layer for $\log N_e$. More nodes and a second hidden layer have been tried and could indeed produce lower integrated RMS errors; they, however, often produced very unrealistic values. Although the search for the optimum network architecture is not completed, we do not expect to obtain significantly different results from other model structures. In Figure 4.52, the structure of a Neural Network with input, output and hidden layers and some hidden nodes is shown.

Typical Variations of Electron Densities

The present ***I***onospheric ***M***odel of the ***A***uroral ***Z***one (IMAZ), *i.e.* the one applicable for both day and night, uses the following input parameters to predict electron densities:

(*a*) solar activity in $F_{10.7}$ (observed),

(*b*) solar zenith angle χ,

(*c*) geomagnetic disturbance parameter K_T derived at Tromsø,

(*d*) geomagnetic time, and

(*e*) riometer absorption.

The electron density N_e is obtained as a function of atmospheric pressure, which relates to altitude according to a local model, which is a function of season only (*cf.* Friedrich *et al.*, 2004a). In contrast to the precursor models in which fitting or binning procedures were carried out for each pressure level, we here use all input values (including pressure) jointly and no interpolation (or smoothing) between adjacent pressure levels is required. Solar zenith angle and solar activity implicitly enter the calculations twice: The quiet rest absorption L_q, needed to obtain integral absorption L_i from the riometer absorption, is a function of $F_{10.7}$ and χ, and the Neural Network itself is trained with $F_{10.7}$ and χ. Using the Chapman function of the solar zenith angle yield marginally smaller RMS errors than using χ alone. We here employ the Chapman function as described by Titheridge (1988). The riometer values are all converted to 27.6 MHz, vertical incidence (infinitely narrow beam), and x-mode. In order to highlight the dependencies of N_e on any of the parameters (*a*) to (*e*), we vary one parameter at a time and fix all others at their median values. These medians and their upper and lower quartiles in our dataset are given in Table 4.3.

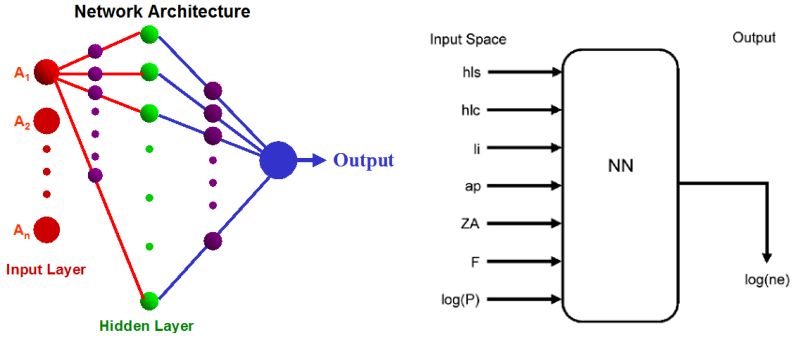


Figure 4.52 A generalized representation of a 3-layer neural network architecture, with n input nodes and m hidden nodes. This example shows only 1 node in the output layer as all of the nets we trained only had 1 output. The number of nodes, n, in the input layer is determined by the size of the input space and the number of hidden nodes, m, is determined as part of the training process. The size of the purple dots symbolises the weights. Right panel: input of the IMAZ NN (McKinnell and Friedrich, 2007).

In Figure 4.53 electron densities are shown for various riometer absorption values from 0 dB (True Quiet) to 2 dB at local noon ($\chi = 60°$). We chose a day in mid-April because at this time of the year both full night (*i.e.* $\chi > 98°$) and a wide range of daytime zenith angles occurs; otherwise the median conditions of Table 4.3 apply. Figure 4.54 shows the same conditions as before, but for local midnight ($\chi = 100°$). There are two noteworthy differences between day and night (*a*) at small riometer values (<0.1 dB) the densities at night are significantly lower, but (*b*) also for larger absorptions where particle ionisation by far dominates over all solar controlled processes, the shape of the profiles is distinctly different (*cf.* the 2 dB curves). The most plausible explanation of the different shape is that during the night the effective recombination rate below typically 80 km is larger due to cluster ions and because of the possible attachment of electrons to neutrals forming negative ions (*cf.* Ch. 5.1).

The relevance of the geomagnetic disturbance as a descriptor of the ionosphere is tested in Figures 4.55 and 4.56 for day and night conditions, respectively. Daytime data vary by not more than 20% with K_T within the quartiles, at night the dependence is somewhat larger; this may, however, be an artefact due to the poorer data coverage.

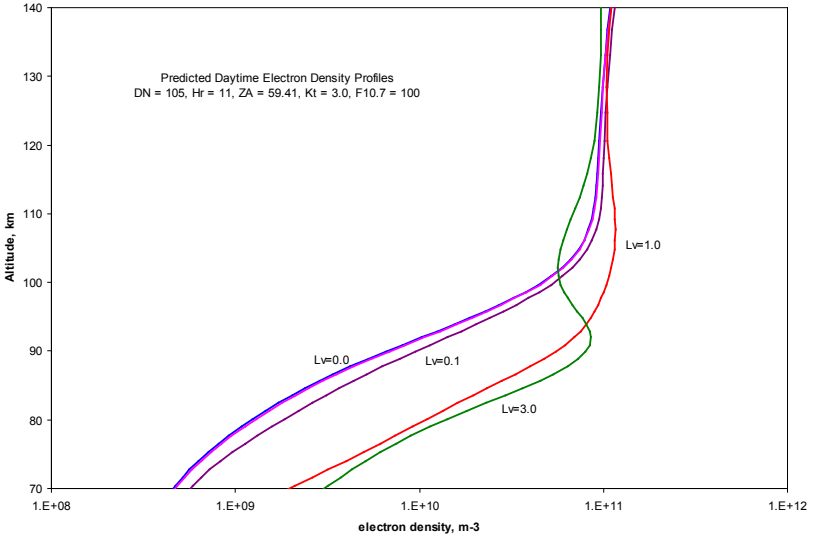


Figure 4.53 Electron density profiles for mid-April noon (χ = 60°) and various degrees of riometer absorption; otherwise median conditions apply. The profiles for 0 and 0.01 dB are virtually indistinguishable.

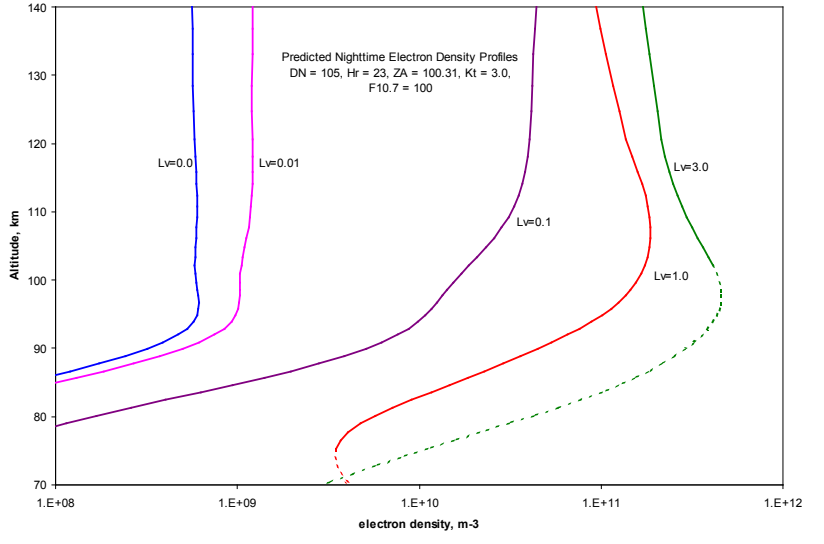


Figure 4.54 As Figure 4.53, but for χ = 100° (local midnight).

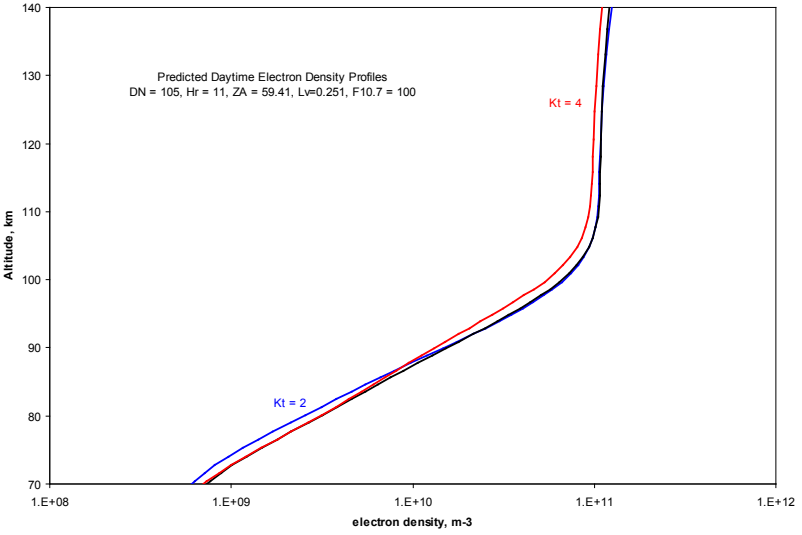


Figure 4.55 Dependence of daytime electron densities (χ = 60°) on the local magnetic disturbance parameter K_T in mid-April; all other parameters are at their median values. Cases of lower, median, and upper quartiles are shown in blue, black and red, respectively (K_T = 2, 3 and 4).

Solar activity is related to the flux of UV and particularly of X-rays. A good correlation at non-auroral latitudes is known and well understood. At auroral latitudes, most of the time, irregular fluxes of charged particles dominate the ion production and one should not expect a dependence except for True Quiet conditions, *i.e.* when the ionosphere behaves as outside the

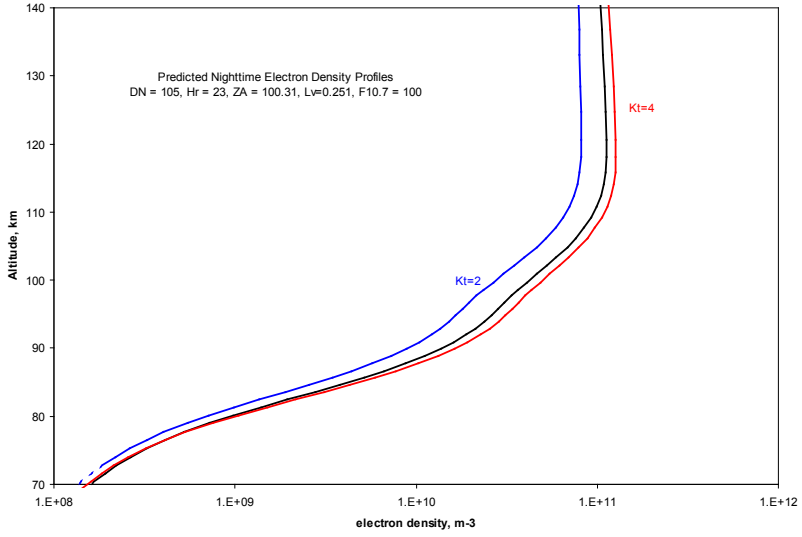


Figure 4.56 As Figure 4.55, but for night-time (χ = 100°).

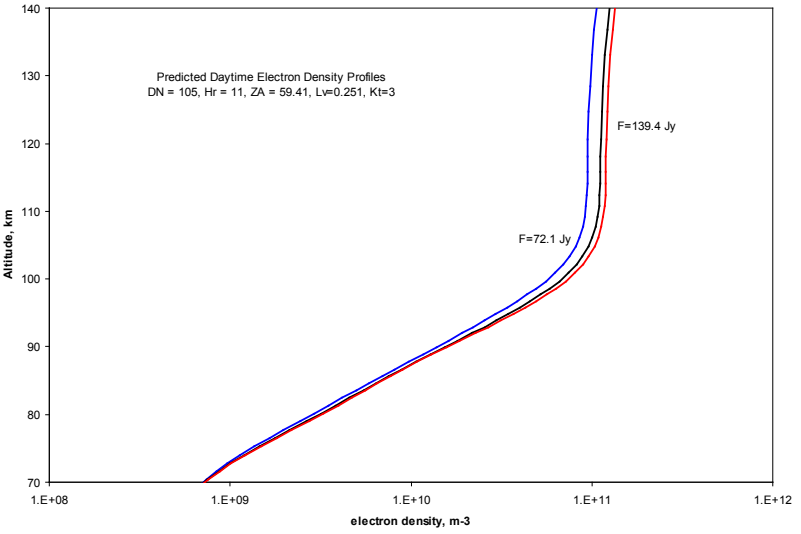


Figure 4.57 Solar activity dependence for noon (χ = 60°) in mid-April and median conditions otherwise. Profiles are shown for median, lower and upper quartile of the solar activity.

auroral regions. In Figure 4.57 the dependence on solar activity expressed by $F_{10.7}$ is again shown for mid-April, noon and median conditions otherwise. $F_{10.7}$ is at the median (100 sfu) and the lower and upper quartiles (90 and 130 sfu). No clear trend is discernible in the daytime results; one mechanism by which a dependence can be explained is *via* the rest absorption L_q. The corresponding test for midnight (Fig. 4.58) shows a clear trend of enhanced N_e at higher

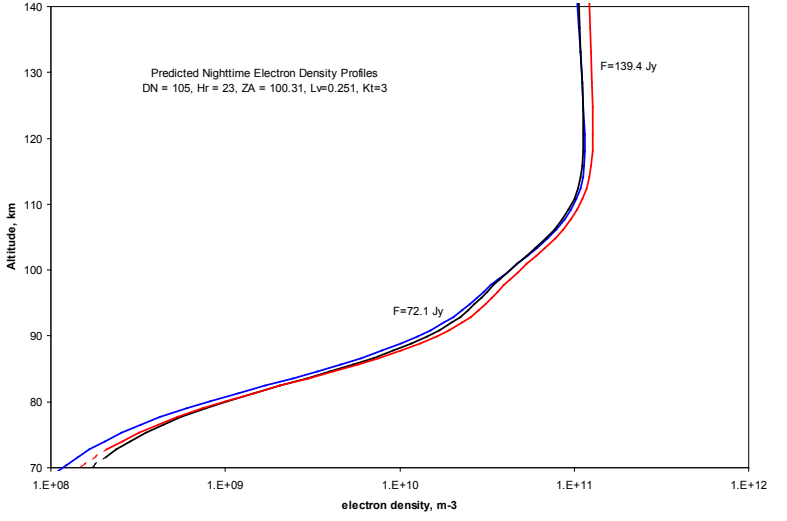


Figure 4.58 Solar activity dependence for midnight (χ = 100°) in mid-April and median conditions.

altitudes and towards lower values in the *D*-region. Since the fluxes of UV and X-rays hardly contribute to the ionisation, notably not at night, a softening of the particle spectra of solar origin at higher sun spot numbers may provide the explanation.

For quiet conditions, a dependence of N_e on solar zenith angle is to be expected, notably of course for the case of True Quiet. For completeness, we also test the χ-dependence for median conditions, *i.e.* at 0.25 dB absorption when ionisation by charged particles definitively dominates. At high latitudes the diurnal variation of the solar zenith angle is small; in order to cover full daytime ($\chi = 60°$) to full night-time ($\chi = 100°$) conditions, we interrogate the model for different seasons. The smallest χ occurs at local noon (11:25 UT for the bulk of our data), whereas the largest zenith angle occurs at local midnight. The IMAZ model in its present form uses geomagnetic time as an input, hence a comparison of noon data (for small χ) with midnight

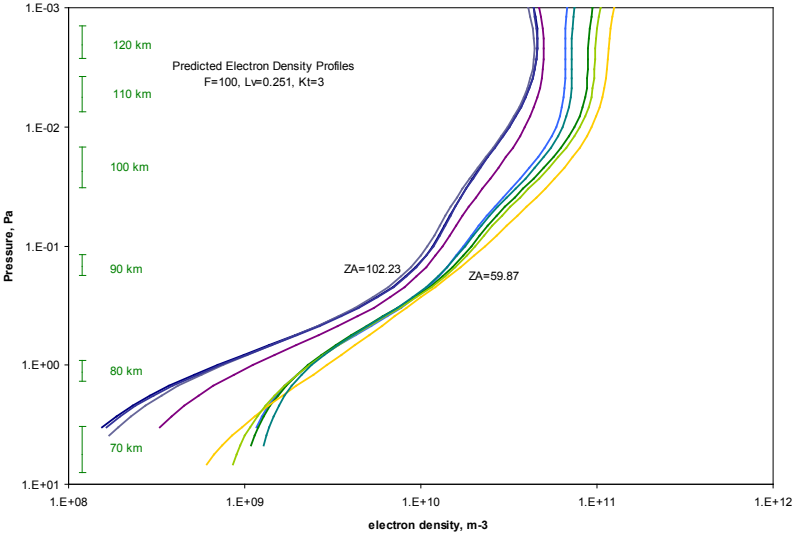


Figure 4.59 Dependence on solar zenith angle for median conditions. Lines between the two extreme solar zenith angles are in 5° increments.

data (for large χ) is potentially distorted by different geomagnetic times. We aim at bypassing this problem by forming the mean of the results at 07 and 15 UT (*i.e.*, symmetric to local noon). Figure 4.59 shows the variation with zenith angle thus obtained. Clearly, at night the *D*-region is depleted more than the *E*-region; again this can be attributed to the larger recombination rate in the nocturnal lower ionosphere.

The Figures 4.55 to 4.59 were calculated for the median riometer absorption of 0.29 dB, although the electron density data that were fed to the NN included the same number of profiles, but for 0 dB of riometer absorption (TQ profiles). Hence on the one hand the median riometer absorption of the data entered was only half of 0.29 dB, and on the other hand the 0.29 dB that prevailed at the measurements represents reasonably disturbed conditions (*e.g.* in darkness associated with visible aurora). Therefore the variation of the other inputs to IMAZ (K_T, $F_{10.7}$ and χ) only had a minor effect on the prediction.

The diurnal variation at different altitudes and different levels of riometer absorption is depicted in Figure 4.60. For quiet conditions (≤0.01 dB) clearly the sunrise/sunset variation is

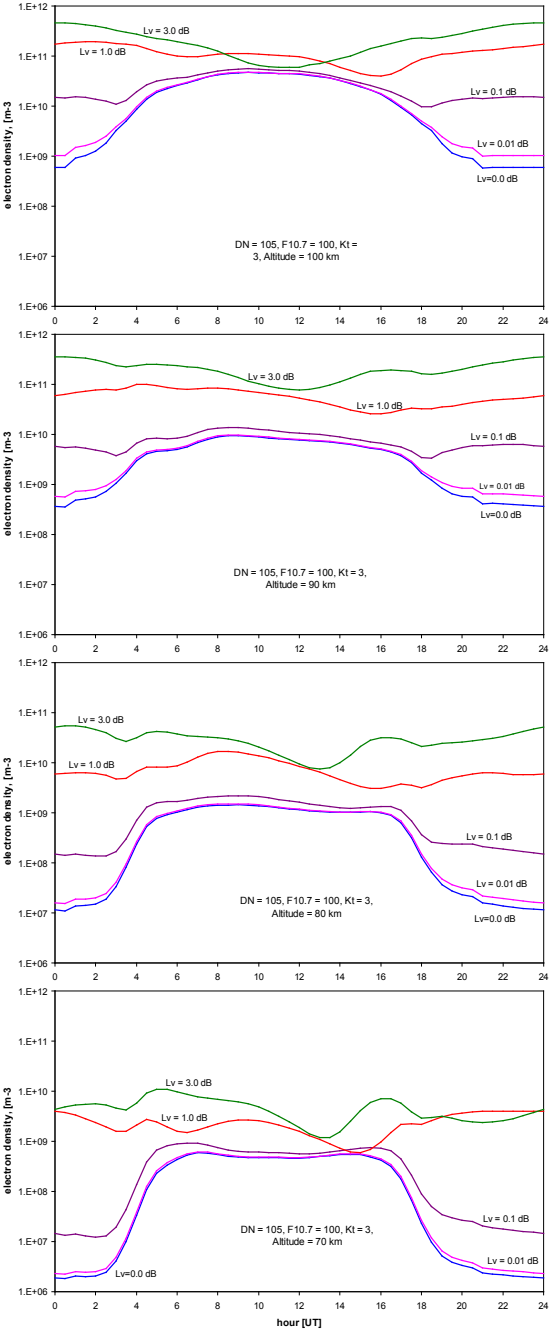


Figure 4.60 Diurnal variation of electron densities for different levels of riometer absorption in mid-April at four heights (median conditions otherwise).

more abrupt at 70 km and very gradual at 100 km, but also the ratio between day and night is smaller at higher altitudes. With increasing altitude and absorption, this ratio becomes smaller and for disturbed conditions (≥1 dB) the diurnal variation does not seem to be related at all to solar zenith angle, but rather to geomagnetic time only. One can also see that a very marginal riometer absorption (0.1 dB) makes an order of magnitude difference in darkness, whereas during the day an electron density increase associated with this riometer absorption is negligible.

Data Coverage and Prediction Quality

In empirical models, data are either binned for all possible combinations of geophysical conditions, or the data are assimilated by functions obtained by procedures, which seek to minimise the RMS error between each data point and its corresponding predicted value. The present Neural Network seeks the minimum RMS deviation of all data entered, *i.e*. both the disturbed values and their corresponding True Quiet counterparts. However, for the evaluation of the quality of the present model, only its ability to predict disturbed values is considered. In Figure 4.61 the RMS errors are displayed as a function of pressure (altitude) for all data, as well as separately for the day and night data. The model predicts better for daytime conditions and at higher altitudes. The averaged RMS error factors are 1.509, 1.947 and 1.744 for day, night, and all conditions, respectively. The poorer prediction ability at night may be due to the fact that the model is based on more than twice the number of day than night data (20362 *vs*. 9588 cases) and may have "learned" a behaviour from the daytime data which is not ideal at night (although the errors of the dedicated night model are only marginally smaller, *cf.* Fig. 4.51).

Our most important input parameter is riometer absorption; at the same time this parameter is the only one among the five we use (the others being K_T, χ, UT, $F_{10.7}$) which can be checked for consistency. Ideally, the simulated riometer absorption $L_{r\text{-}sim}$ calculated by using a profile which represents a particular riometer absorption must equal L_r. The test of this consistency is shown in Figure 4.62. Because of the large range of riometer absorption in our data we chose a dual logarithmic scale and plot integral absorptions L_i (rather than riometer absorption L_r) such that even for $L_r = 0$ dB, L_i has a finite value. Generally, the agreement seems to be better for larger absorption; however, this simply means that we have an absolute error in L_i which naturally appears smaller for large values. This absolute uncertainty in the riometer values (measured and predicted) can also very likely be due to the uncertainty in the

	lower quartile	***median***	***upper quartile***
solar activity, $F_{10.7}$, Sfu	72	109	139
solar zenith angle, χ, deg	74	89	104
K_T	2	3	4
Kp	1^+	2^+	3^+
ap	7	15	27
geomagnetic time	6.43	11.78	17.88
UT	6.58	12.20	18.08
riometer absorption, dB (@27.6 MHz)	0.16	0.29	0.56

Table 4.3 Medians and the upper and lower quartiles of the input parameters in the present data set. The median and quartiles given for the time (both UT and geomagnetic) are irrelevant: the measurements are essentially evenly distributed over the day.

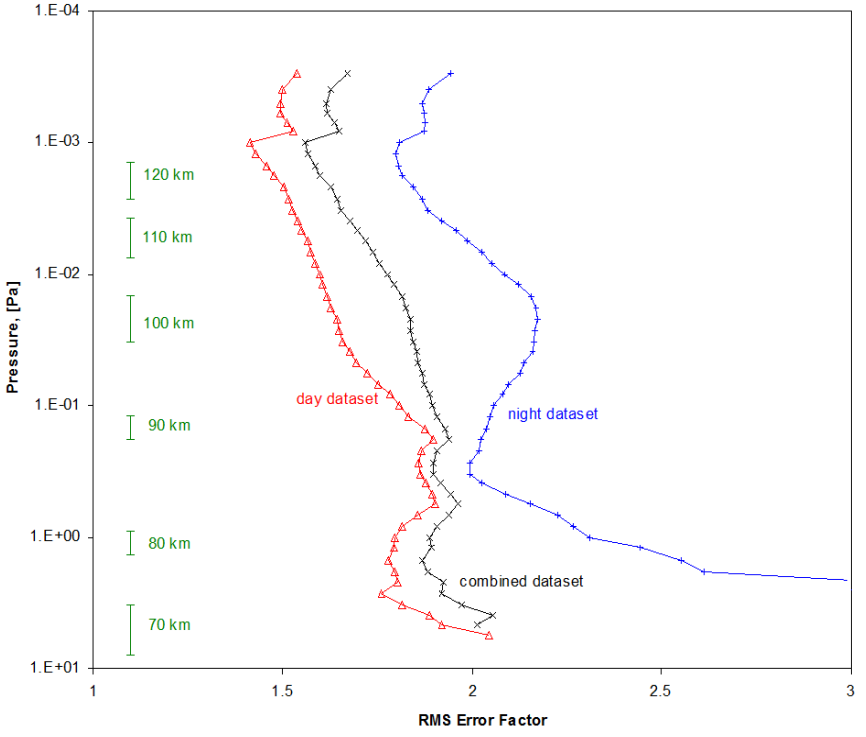


Figure 4.61 RMS error in the prediction as a function of background pressure. Daytime and higher altitudes are obviously easier to predict.

determination of the absorption measurement, or more likely still, in the uncertainty of establishing the quiet day curve. For the well maintained IRIS riometer an uncertainty of 0.05 dB (at 38.2 MHz) is quoted, which corresponds to about 0.1 dB at our frequency (27.6 MHz). Hence predicting electron densities for situations with riometer absorption below 0.1 dB is largely academic since such small values cannot be obtained with any degree of reliability from even the best riometers; nonetheless, IMAZ predicts realistic profiles between True Quiet and 0.1 dB essentially by interpolation. In addition, one has to bear in mind that riometer absorption is calculated for each case (or geophysical condition) in our data set regardless of whether or not the profile for that condition actually had full height coverage. Notably for small absorption values and at night profiles only start at 80 or even 90 km, whereas that height region contributes predominantly to the absorption. For such cases, we use predicted values for the simulation of riometer absorption for which we do not have corresponding measurements.

Estimating the uncertainty in the predictions

Neural Networks (NNs) have also been used to estimate the uncertainty on a prediction. This uncertainty can be determined by finding a statistical measure of the differences between the predicted and measured values. For each input vector (local magnetic time, absorption, *etc.*) of the original dataset used to train, the NN that became our IMAZ model (referred to as NN1 of that model), the difference between the predicted and measured output value ($\log N_e$) is

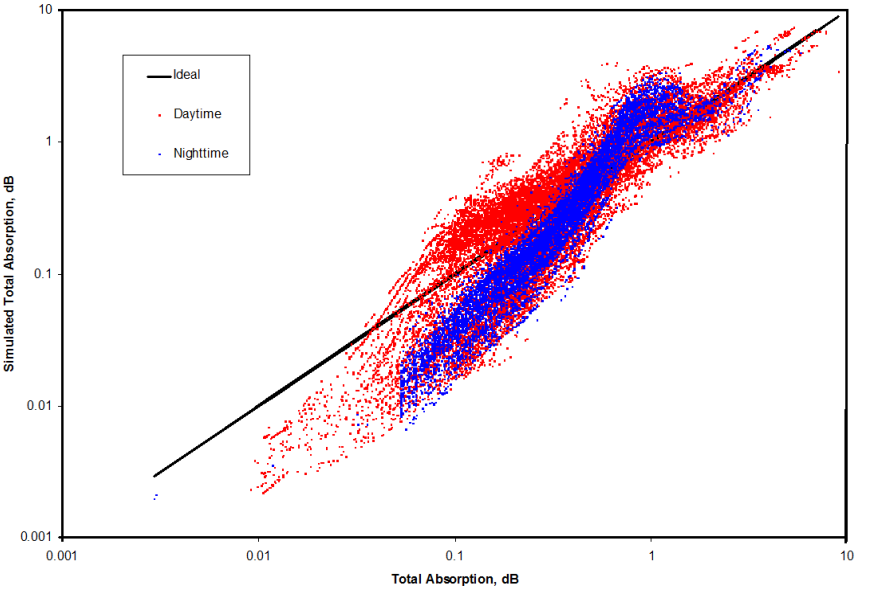


Figure 4.62 Simulated integral absorption vs. measured (vertical) absorption as an input. Red dots are for daytime conditions (χ <98°), blue ones for night.

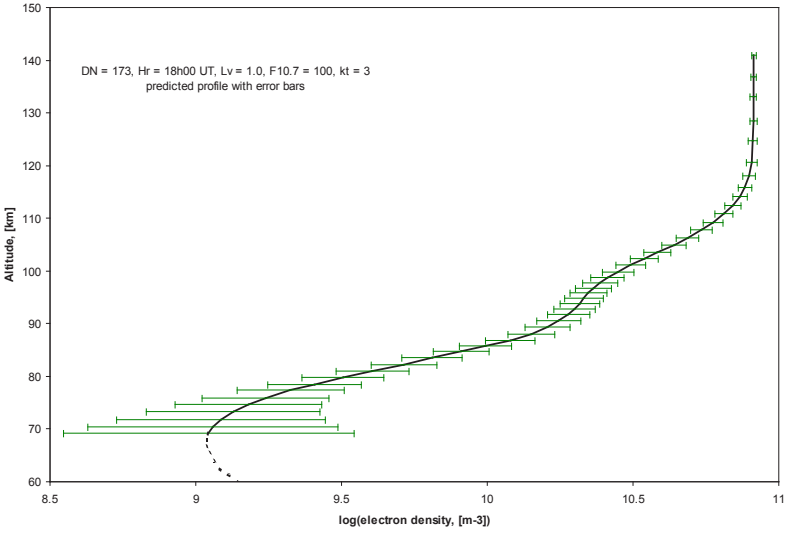


Figure 4.63 Electron density profile predicted for Summer solstice at 18:00 UT (χ = 75°), 1 dB absorption and median conditions otherwise. Note that the error bars are much larger in the D- than in the E-region.

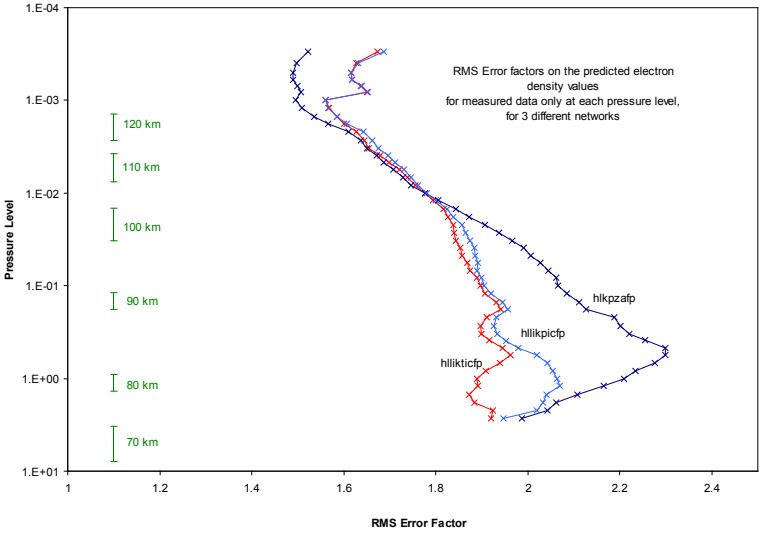


Figure 4.64 RMS prediction error as a function of background pressure, the corresponding altitude ranges are given for orientation. The full version of IMAZ as described here is shown in red, replacing the locally established magnetic index K_T by the global Kp in blue, and using Kp as the only disturbance parameter (no riometer absorption) is depicted in black. Note that in the E-region K is apparently quite sufficient to predict the electron densities, whereas the prediction error increases significantly in the D-region if no absorption data are available.

evaluated and squared. A second NN (referred to as NN2 of IMAZ) is then trained with the same input data as NN1 but with the squared differences as the output. Since it is the nature of NNs to find the mean predicted value, the square root of the output of NN2 is a root mean squared (RMS) difference. This difference represents a measure of the variation that can be expected between any predicted and measured value. More details of this technique and its application to the ionospheric parameter, *fo*F2, can be found in Poole and McKinnell (2000). The output from NN2 of the IMAZ model provides an estimated uncertainty on the logarithm of the electron density, which is dependent on all parameters within the input space, including altitude (in the form of the pressure surface input). This provides the opportunity to include error bars on the predicted electron density values. These error bars represent the maximum possible statistical variation of the average predicted electron density for a given input set. A cut-off value for the estimated uncertainty can then be established to indicate the limitations of the prediction. For example, predictions that have uncertainties that are more than this cut-off value are flagged as unreliable. Figure 4.63 shows an electron density profile with the errors thus calculated. The mean RMS error for all conditions was found to be 0.2414 in logN_e (or a factor of 1.74, *cf.* Table 4.2); unless otherwise indicated we only display the results if the RMS error is less than twice the mean value, *i.e.* <0.4828 in logN_e (or below a factor of 3).

The overwhelming amount of data upon which the IMAZ model was built originated from the EISCAT installation near Tromsø, hence using the locally determined magnetic index K_T is appropriate. In order to make the model usable for users at other location, but still in the auroral zone, another Neural Network was trained with the commonly available magnetic index *Kp*. For users who have no access to riometer measurements, a third version with *Kp* as the only the disturbance parameter was established. The resulting over-all degradation in the RMS error

factors is 1.759 and 1.776 compared to 1.743 for the full IMAZ model as described here. These factors appear to only marginally deteriorate the results, however they are determined as the average over all input values, most of which originate in the *E*-region. Plotting the error *vs*. background pressure (altitude) is more revealing (Fig. 4.64) and clearly shows that the disturbed *D*-region cannot be described satisfactorily by a geomagnetic index only. On the other hand, the fact that the models with *K* only (*i.e.* without riometer absorption) yield somewhat better results at the highest altitudes (>120 km) suggests that the Neural Network has "learned" an absorption-dependent behaviour at lower altitudes which is irrelevant at higher altitudes. Rather surprisingly the locally determined K_T also helps to somewhat better describe the *D*-region than *Kp*; this is of course not related to currents in the *D*-region, but rather points to a statistical coincidence of hard particles (which ionise the *D*-region) with softer ones which provide the additional electron densities in the *E*-region and in consequence currents and (local) magnetic disturbances.

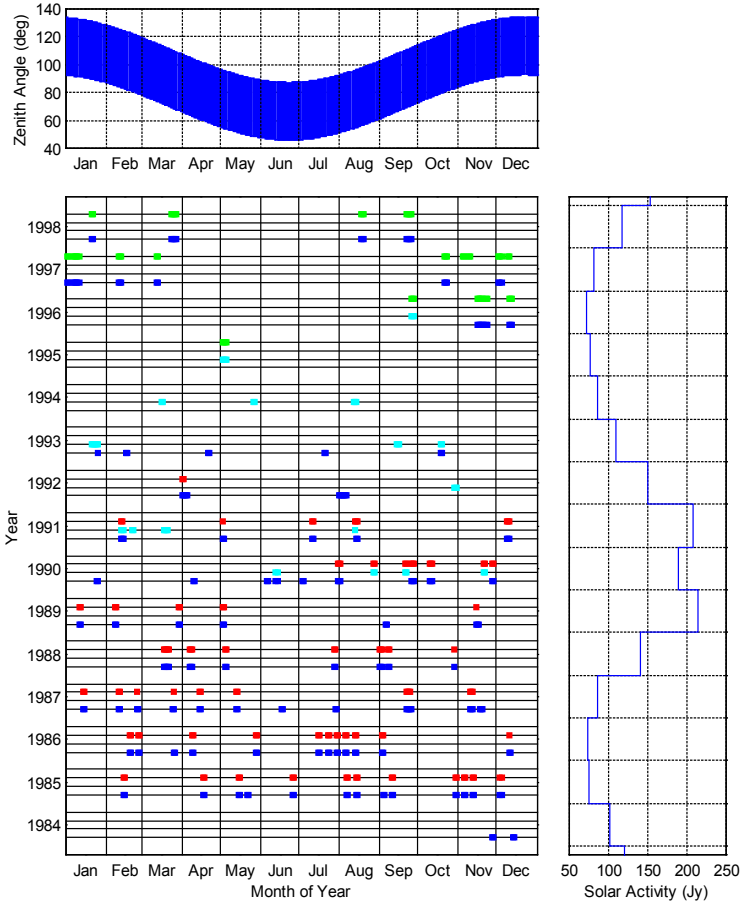


Figure 4.65 Coverage of EISCAT data used in the IMAZ model. Dark blue = UHF, light blue = VHF, red = wide beam riometer, green = IRIS. The zenith angle range as a function of season is given in the top panel, the monthly median solar activity is displayed in the right panel (Harrich, 2001).

Statistics

The data used in the present version of IMAZ are the ones pre-processed and screened by Harrich (2001) for his thesis and later also used by Steiner (2003). Therefore, data from EISCAT from the beginning of its operations in 1984 to only 1998 inclusive are used, *i.e.* the first solar cycle (Fig. 4.65). We have electron density data from 126 sounding rockets with wave propagation experiments (Faraday rotation, differential absorption), namely (arranged in descending geomagnetic latitude): 13 from Ft. Churchill (Canada), 67 from Andøya (Norway), 45 from Esrange (Sweden), and one from Red Lake (Canada). Of these 121 can be used for the IMAZ model, for the remaining five no riometer values are available or could not be calculated (insufficient height coverage). For the determination of the True Quiet electron densities values at 75, 80 and 84 km are available from 244 rocket borne probe measurements taken at the former Soviet Antarctic station Molodeznaya. Riometer values prior to IRIS (1995) were from wide beam instruments located at the rocket ranges or in the vicinity of EISCAT, respectively. Until

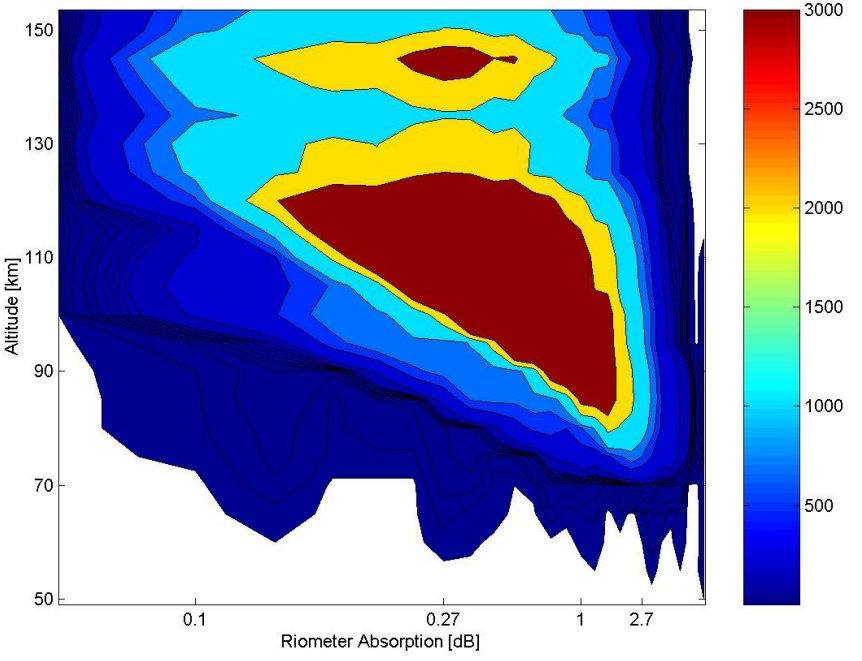


Figure 4.66 Number of usable electron density values at each kilometre as a function of riometer absorption. As a crude guide >500 values should be available to ensure representative model results. The lower limit of the number of values is due to the restriction of the data for small absorptions (cf. Figures 4.38 and 4.39).

1992 four wide beam riometers were operated near EISCAT, namely in Ramfjordmoen and Lavangsdalen, each operating on 32 and 40 MHz. The mean of all readings - converted to 27.6 MHz - was formed and the data with the largest deviation from the mean eliminated. The mean of the remaining three data is then used. If only three values are available, the mean of the "best" two is used; should only one riometer have provided data, the readings of the riometer located at the Andøya rocket range (at 100 km distance) was used to judge its credibility (Harrich,

2001). In our data set, about half the riometer values are from wide-beam instruments and the other half from IRIS.

The grand total of the EISCAT measurements is about 65,000 profiles (55,000 UHF, 10,000 VHF), whereof only the ones with concurrent riometer readings are used for IMAZ (*i.e.* 40,000), whereas the other half (measurements without riometer values) was only used for the determination of the True Quiet electron densities. All profiles are interpolated to 59 common pressure levels (12 per decade from 8.2×10^{-4} to 5.6×10^{1} Pa); due to the restrictions applied because of the scatter at small absorption values (*cf.* Figures 4.38 and 4.39) not the whole profiles are used and a grand total of 694,849 data points are entered to train the IMAZ Neural Network. Figure 4.66 shows the number of available data as a function of altitude and riometer absorption. The lower edge of the well-covered region is due to the restriction of the profiles at small values (*cf.* Fig. 4.39). Roughly, 10% of the data are used to determine True Quiet, and about 40% to establish disturbed profiles. In Figure 4.67, the electron density contributions from the various sources (rockets, EISCAT VHF and EISCAT UHF) are given as function of altitude. Below 66 km, rockets soundings are the dominating source of data. Most EISCAT VHF data in our dataset are restricted by the operating mode to 115 km; this abrupt disappearance of the potentially better data in the lower ionosphere probably causes the discontinuity of the RMS error at that height.

IMAZ is published by McKinnell and Friedrich (2007) and is available online from the IRI server (spdf.gsfc.nasa.gov/pub/models/iri/iri2007/IMAZ).

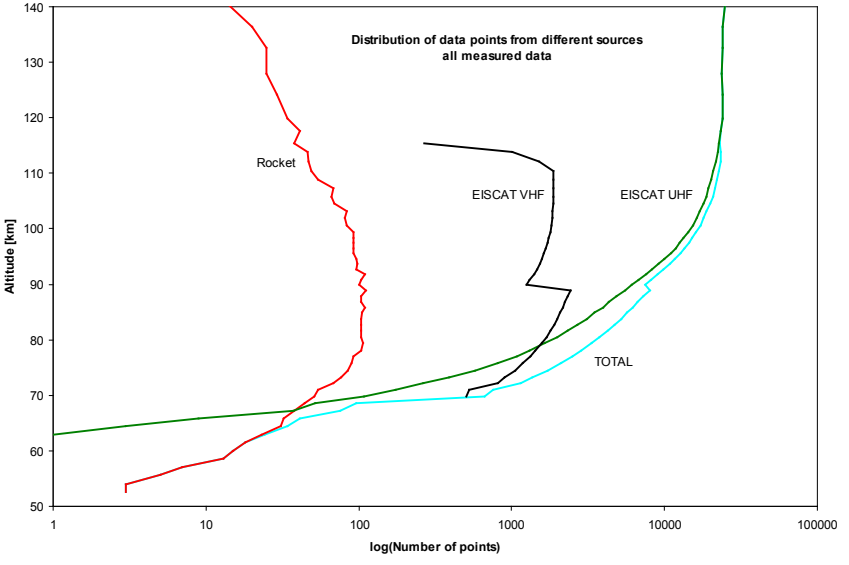
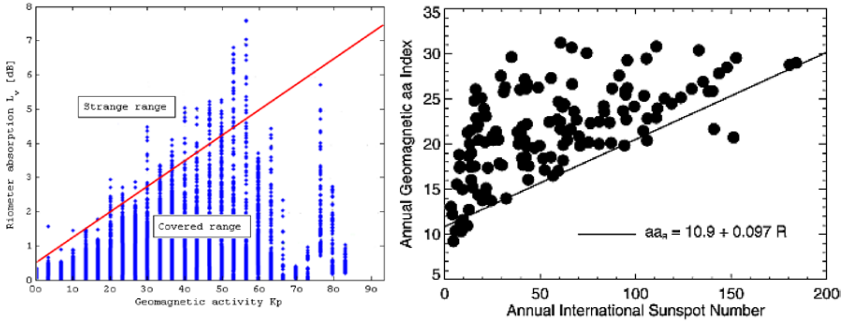


Figure 4.67 Number of electron density values from various sources. Below 65 km all data are due to sounding rockets, however for low disturbances the dominance of rocket data reaches to much greater heights.

Outlook

- The True Quiet electron densities should be re-calculated, notably at night where the flight of ECOMA-7 provides the lowest densities ever measured at auroral latitudes.
- more data, both from EISCAT and rockets should be included
- The MF radar at Saura (near the Andøya rocket range) provides *D*-region electron density data if only for specific conditions (too much disturbance absorbs the signal, whereas when there is little electron density there is interference from far away stations)
- ultimately one could include data from other auroral incoherent scatter radars, such as Søndre Strømfjord, Poker Flat (PFISR), but also from the polar cap station EISCAT Svalbard and conceivably also the sub-auroral station Millstone Hill. This requires the introduction of geomagnetic latitude as an additional variable.
- The IMAZ version *without* riometer as an input that is currently available uses the same data as for the "full" IMAZ. There are, however, much more data that could be integrated in such a version of IMAZ (*i.e.* without concurrent riometer data).
- With practicable definitions for the limits of the input space one can extrapolate the neural network for the conditions of the parameter that exceeds that limit. This was successfully exercised for large riometer absorption in an Honours Project by G. Landauer (Friedrich and Landauer, 2011).
- Careful statistics can reveal unrealistic combinations of input parameters. By not allowing such unrealistic input combinations the model will produce more of the credible results. For riometer absorption and *Kp* such an analysis has been carried out (Friedrich and Landauer, 2011), but conceivably can also be performed for *Kp* and $F_{10.7}$; a correlation between the annual magnetic index and the annual Sun spot number has already been exercised.



Riometer absorption vs. K_p (left panel), annual geomagnetic index aa vs. sunspot number (right panel). Note that large absorption never occurs without large K_p, whereas large K_p can occur without simultaneous riometer absorption. Similarly, high sunspot numbers are always associated with enhanced (annual) geomagnetic activity.

4.3.2. Polar Cap

The behaviour of the ionosphere inside the auroral oval (the polar cap) is in many ways similar to that of the auroral zone. Notably the ionisation is more often than not dominated by energetic charged particles, although their behaviour is somewhat different. Not withstanding these differences, one needs the same parameters to characterise the ionosphere as for the auroral zone, *i.e.* solar zenith angle, solar activity, season as well as the "disturbance parameters" riometer absorption and *Kp* primarily describing the *D*- and *E*-regions,

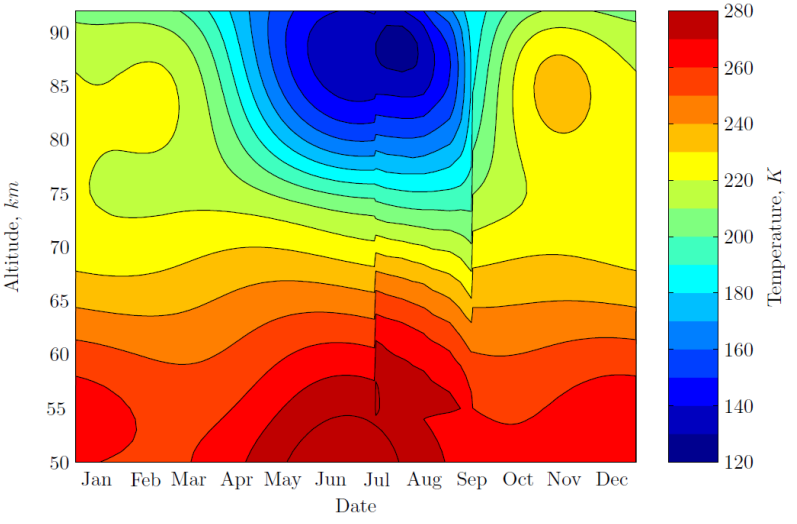


Figure 4.68 MSIS temperatures for 78°N with the local summer measurements inserted (Egger, 2004).

respectively. Available electron densities from the polar cap are primarily due to the incoherent scatter radar EISCAT located near Longyearbyen (geomagnetic latitude 75.2°). This installation on the island of Svalbard/Spitsbergen (Norway) is in near-continuous operation since 1996. Due to the electron density threshold measurements only begin slightly below 100 km. For lower altitudes probe data of rocket soundings are available, performed from Heiss Island (Russia), further north, but geomagnetically at a comparable latitude (72.4°; Fig. 4.69). The (371!) soundings were performed between 1979 and 1994 and cover 51 to 90 km. Since at the present time no coincident reliable data from co-located riometers are available, only two modelling attempts were performed, namely (a) the True Quiet, and (b) a time-dependent model of the whole ionosphere, but lacking the riometer input which therefore makes it not representative in the *D*-region.

The True Quiet derived by Egger (2004) is obtained by similar procedures as described in Ch 4.2.3.1 beginning with the establishment of the neutral background atmosphere. Again the MSIS atmospheric model provides the backbone, but for summer a total of 30 local measurements are used (Lübken and Müllemann, 2003) which - again - show lower mesopause temperatures than MSIS. For other seasons a smooth transition from the local summer

measurements of temperature, pressure and density was made by Egger (2004) to be able to cover all seasons (Fig. 4.68 shows the case of temperatures).

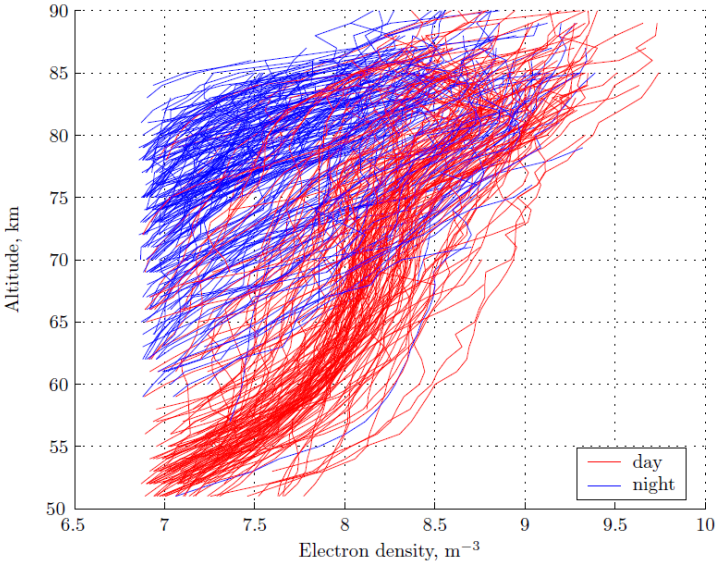


Figure 4.69 Electron densities from Heiss Island measured by rocket borne Langmuir probes (Egger, 2004).

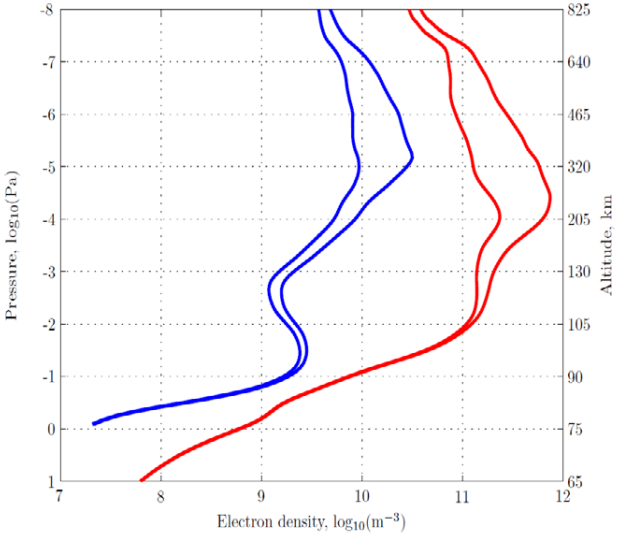


Figure 4.70 True Quiet electron densities for the polar cap. The blues lines represents night time (solar zenith angles >98°), the red ones subsolar densities. Two solar activities are considered (75 and 200 Jy; Egger, 2004).

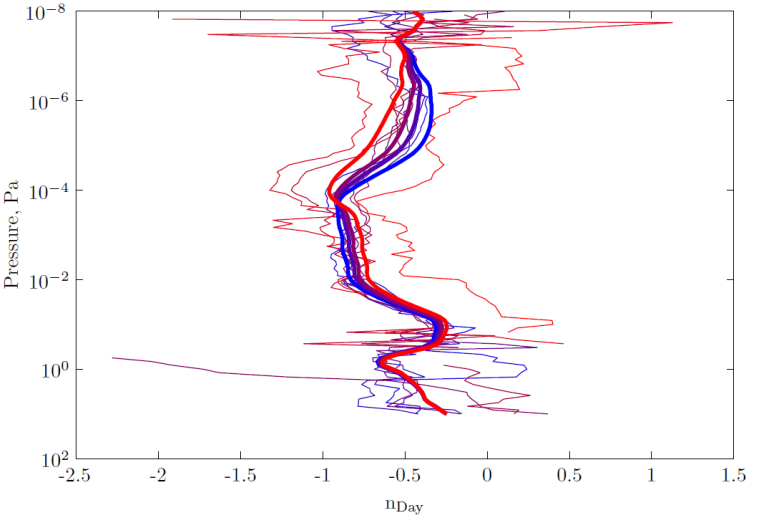


Figure 4.71 Exponent of the function $N_e = N_o(\cos\chi)^n$ to compute daytime True Quiet electron densities. Blue and red lines are for 75 and 200 Jy, respectively (Egger, 2004).

The resulting True Quiet values using both the Heiss Island rocket data and EISCAT Longyearbyen from 1996 to 2004 are shown in Figure 4.70. Daytime values can be computed using the subsolar electron density, an extrapolation to a fictitious overhead Sun ($\chi = 0°$). With the value shown in that figure one can readily compute N_e for any χ within the range of solar

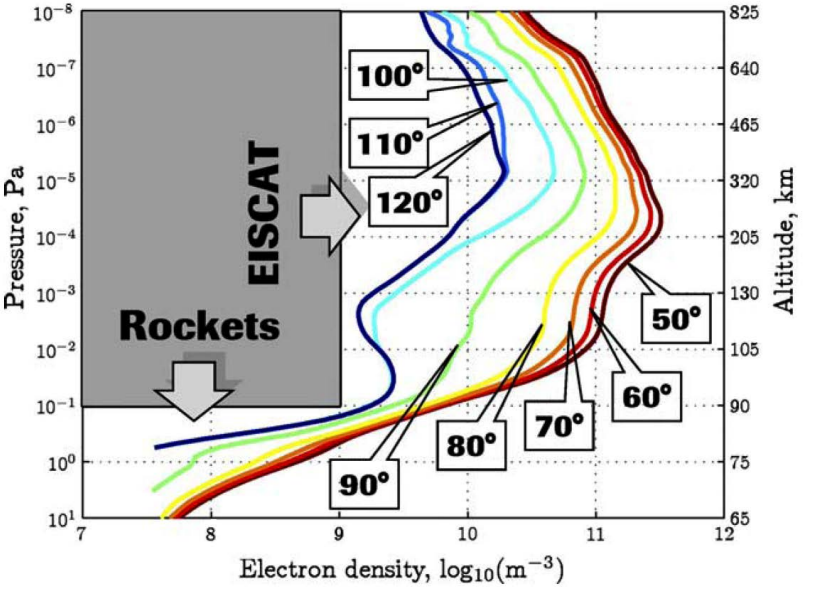


Figure 4.72 True Quiet polar cap electron densities as a function of pressure for the median activity of the data used (170 Jy; Egger, 2004).

zenith angles occurring at a latitude of 75° (*i.e.* χ = 55 to 125°) by multiplication with $(\cos\chi)^n$, using the exponent n. Theoretically this exponent should have a value of 0.5 which is roughly confirmed by the data (Fig. 4.71); representative True Quiet electron density profiles are given in Figure 4.72. A first result of the time-dependent, neural network model by Fankhauser (2010) is shown in Figure 4.73 where the electron density decay is displayed in hours after sunset.

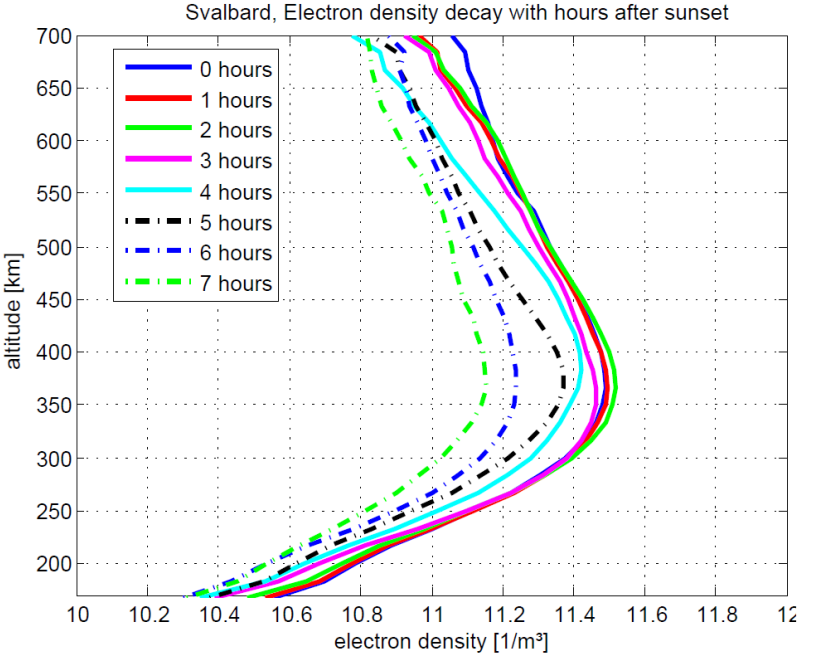


Figure 4.73 Electron densities above Longyearbyen as a function of time after sunset (Fankhauser, 2010).

5. Models of Other Bulk Parameters

In order to understand the behaviour of the density of free electrons, it is advantageous to have some insight into the behaviour of the other charged species, the positive ions, but also their nature of the positive ion is of importance. At low altitudes the negative species can also be in the form of ions. For purposes of wave propagation not only the electron density, but also the collision frequency matters, notably absorption only occurs in the presence of collisions.

5.1. Effective Recombination Rate

The number density of free electrons in the atmosphere is the result of a balance between ionisation and electron loss. Whereas the former is reasonably predictable at non-auroral latitudes, at high latitudes - where ionisation by energetic particles dominates - one often aims to infer the ionising particle spectra from the resulting electron densities. For this procedure, as well as for the verification of ion-chemical models, a sound knowledge of the recombination rate is crucial, notably in the *D*-region where variations of the recombination rate by orders of magnitude must be expected. Empirically one obtains this rate by forming the ratio of the ionisation rate to the square of the electron density. Therefore, both parameters have to be known with good accuracy. The number of cases is thus essentially limited to those reliable electron densities for which in addition all important ionisation processes are known. From the 49 such cases we attempt to deduce representative recombination rate profiles for day and night conditions

Introduction and Data

The upper atmosphere is constantly ionised by fluxes of various types. This ionisation rate q (ion-electron pairs per m^3 and s) is balanced by the reverse process in which the ionisation products (electrons, N_e and positive ions N^+, both in m^{-3}) recombine (rate α_D, $m^3\ s^{-1}$). For charge balance ($N_e = N^+$), and in the absence of negative ions, the effective electron loss rate is:

$$\Psi = \frac{q}{N_e N^+} = \frac{q}{N_e^{\,2}} \tag{5.1}$$

From the point of view of the electrons they can also be lost through attachment to neutrals (including aerosols) thus forming negative ions; eventually these negative ions will recombine with positive ions (rate α_i, $m^3\ s^{-1}$). If one expresses the number density of negative ions as the fraction $\lambda = N^-/N_e$, after a little manipulation one obtains the effective electron loss rate Ψ as:

$$\Psi = (1+\lambda)(\alpha_D + \lambda\alpha_i) = \frac{q}{N_e^{\,2}} \tag{5.2}$$

This effective recombination rate can thus be determined when both the electron density and the ionisation rate is known. For the conditions of the lower ionosphere (*D*- and *E*-regions) steady-state can safely be assumed, *i.e.* changes in the ionisation are slow compared to the lifetime of free electrons. We restrict our analysis to such electron density profiles, which were obtained from rocket borne radio wave propagation experiments because uncertainties in N_e are more critical than in q since electron density enters the formula for the effective recombination rate quadratically, whereas ionisation only enters linearly. Of the total of 49 cases available for the present analysis, 29 are based on the authors' own flights and most of them were used in an earlier paper (Torkar and Friedrich, 1988). In these cases, it was ensured that more than 50% of the ionisation rate was known from measurements (particle fluxes, nitric oxide) and only less than 50% was based on standard values from the literature (X-rays, cosmic rays). The other 20

cases were taken from the literature at face value. The enlarged database and the improved model of the atmospheric background is the motivation for the present re-analysis.

The knowledge of the recombination rate basically serves two purposes: (*a*) modellers can assess the quality of their theoretical computations, and (*b*) one can infer the ionising particle spectra from electron densities. In particular, there is a need for the latter application since many groups try to establish charged particle spectra from continuous measurements of

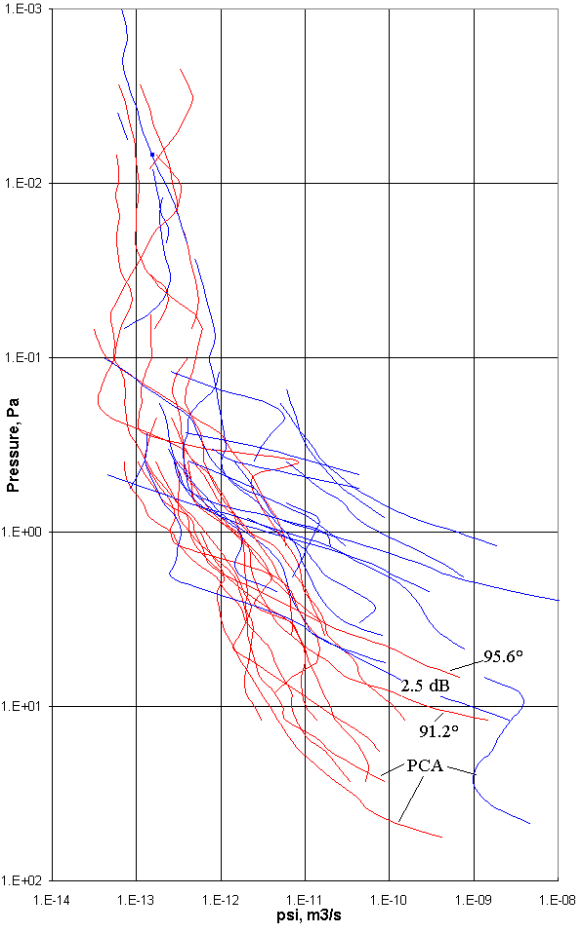


Figure 5.1 Empirical electron loss rates as a function of background pressure. Day and night conditions are indicated in red and blue, respectively; some special cases are also indicated (see text; Friedrich et al., 2004b).

electron densities (*e.g.*, EISCAT; Osepian *et al*., 2001), or indeed to infer particle spectra from riometer data (Collis *et al*., 1997; Kirkwood and Osepian, 2001). A sound knowledge of the electron loss rate thus constitutes an important link between magnetospheric and atmospheric physics.

Data Preparation

Figure 5.1 shows the quantity q/N_e^2 as a function of background pressure. One can clearly see that below 0.1 Pa (above typically 90 km) the values converge, whereas below that height (above that pressure) they increase and show a much larger variability. Because electron loss has primarily chemical causes, we surmise that it should behave comparably at constant densities rather than constant altitudes. We employ an atmospheric model based on measurements above the Andøya rocket range (*cf.* Ch. 3 and Friedrich *et al*., 2004a) for data taken at that range and from Esrange, for other latitudes we use CIRA-1986 (Rees *et al*., 1990). The ψ values originally tabulated for each kilometre are interpolated for 12 pressure levels per decade (using neutral number density might be more appropriate, but we use pressure for compatibility with other ionospheric models). Figure 5.2 depicts a very simple scheme, yet adequate to qualitatively explain the loss of free electrons in the mesosphere. The primarily produced ions NO^+ and O_2^+ can not only recombine with electrons, but also may form cluster ions with recombination rates orders of magnitude larger. Possible paths for the formation/destruction of negative ions are indicated. From that simple scheme the following qualitative behaviour of ψ emerges:

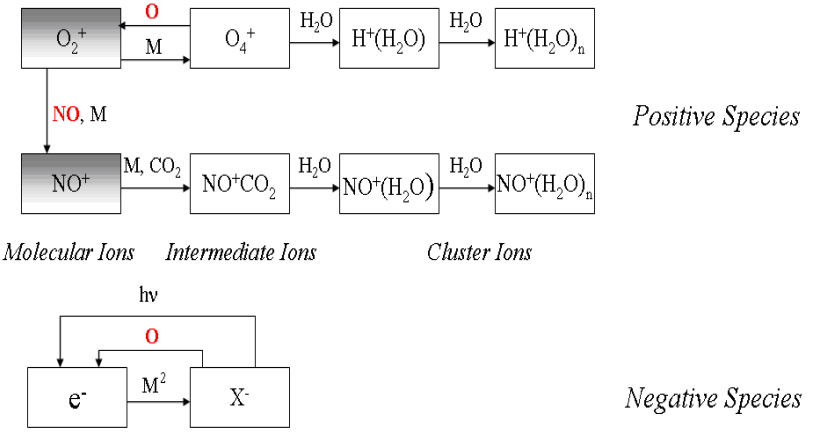


Figure 5.2 Simplified scheme for the loss of free electrons. For graphical reasons recombination paths from electrons or negative ions to each of the positive species are not shown (Friedrich et al., 2004b).

1) For the formation of clusters and negative ions a background neutral number density M is required, whereas
2) the presence of atomic oxygen inhibits the formation of clusters and negative ions, and
3) negative ions are photo-dissociated by visible sunlight.

By using pressure rather than altitude we feel confident to analyse all data together, despite the significant scatter of the data. In Figure 5.3 a rare example of an ion mass spectrometer measurement and a full profile of effective recombination rate is shown. Clearly ψ is drastically different in the regions where cluster or molecular ions dominate (above and below 77.3 km in this case).

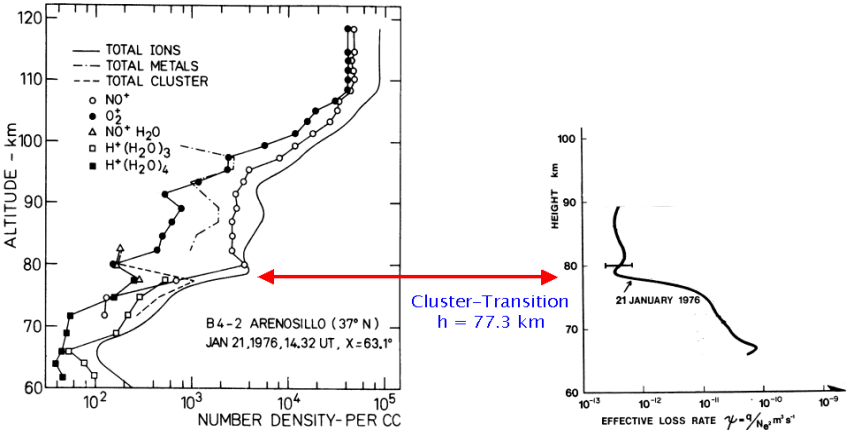


Figure 5.3 Partial ion density measured by a rocket borne mass spectrometer (left panel, Arnold and Krankowsky, 1979) and the effective electron recombination rate derived from data of instruments aboard the same sounding rocket (right panel; Thrane et al., 1979). Note that below the transition height (i.e. where cluster ions dominate) the recombination rate increases drastically.

Search for Dependencies

The density of atomic oxygen changes drastically below a ledge typically located at about 83 km from immeasurably low values at night to about 10^{16} m^{-3} during the day. According to theoretical calculations (*e.g.*, Bowman and Thomas, 1973; Isaksen, 1973) this transition occurs at a zenith angle of 98°. For the photo-detachment of electrons from negative ions, sunlight in the visible spectrum is responsible. Hence, we can expect that process to be effective as soon

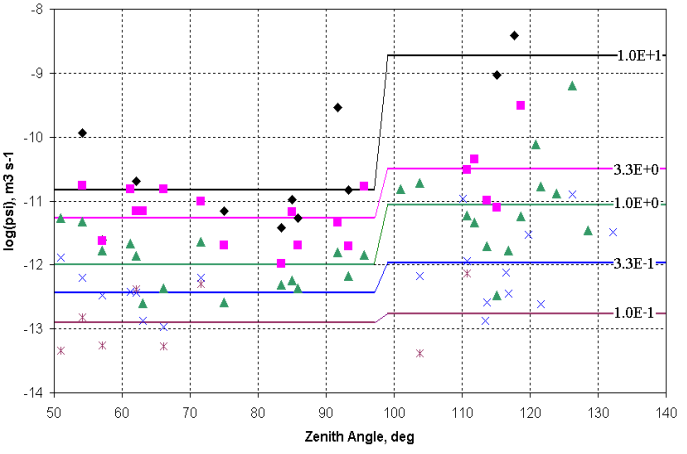


Figure 5.4 Effective electron recombination rates between 0.1 and 10 Pa at selected pressure levels vs. solar zenith angle. The lines are the means over all values below and above 98°, respectively (Friedrich et al., 2004b).

as the mesosphere is illuminated from below (zenith angle 98° at 70 km and above). For the loss of free electrons both processes (detachment of negative ions and fewer clusters) have the same effect, namely to reduce the effective electron loss rate during the day. The limiting solar zenith angle of 98° was also found empirically by Stauning (1996) who studied a stable auroral event, which showed a marked decrease of riometer absorption beyond a zenith angle of 98°. In Figure 5.1, we therefore highlight day- and night-time values as defined here (red and blue, respectively). The figure generally shows the expected behaviour, *i.e.* lower recombination rates during the day. At night there seems to be a slope with a constant gradient ($\sim p^{-6}$), however, at different pressure levels (altitudes) in each case. Two twilight profiles indicated by their zenith angles (91.7° and 95.6°) appear to have slopes characteristic for night conditions. The night profile "2.5 dB" shows very low, almost daytime, values and is a case of a large auroral ionisation. The very lowest profiles are all cases of Polar Cap Absorption (PCA) events. In Figure 5.4 the recombination rates at five pressure levels are plotted *vs.* solar zenith angle. Because of the paucity of data, the theoretically expected drastic change at 98° is not covered by the data. In the figure we therefore indicate the means of the day and night values. Clearly, the day-night difference is larger at low altitudes (higher pressures), *i.e.* in the *D*- rather than the *E*-region (or in the region dominated by cluster, rather than by molecular ions).

All chemical rates (recombination, attachment, charge exchange, clustering) have negative temperature dependencies, hence higher temperatures should lead to a lowering of the effective recombination rate. This can be expected primarily at those altitudes (pressures) where clusters dominate whose formation rates depend more strongly on temperature ($T^{-4.7}$; Reid,

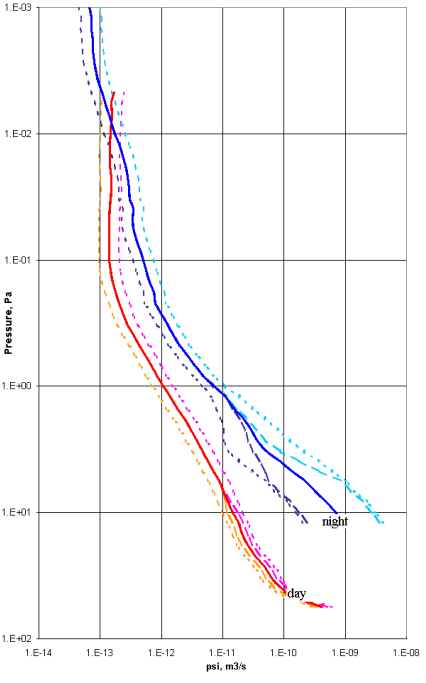


Figure 5.5 Mean effective recombination rates for day and night (full red and blue lines, respectively). Estimated variations due to temperatures higher/lower by 10 K are indicated by dotted lines and due to ionisation rates higher/lower by a factor 30 by dashed lines (Friedrich et al., 2004b).

1977) than the actual recombination rate (typically $T^{-0.5}$). Because of the very limited number of cases it is not expedient to derive an empirical temperature dependence.

In the scheme in Figure 5.2 the primarily produced ions are the molecular types NO^+ and O_2^+ with low recombination rates, whereas the rapidly recombining clusters are formed by charge exchange in the presence of a background neutral number density M. Hence, we expect that for large production rates q, the balance will be shifted away from clusters to molecular ions because the molecular ions are produced faster than they can convert to clusters. Because of the larger fraction of molecular ions, the net result must be a lowering of the effective recombination rate. Again, the limited number of data neither confirms nor contradicts the predicted dependence on q (Friedrich and Torkar, 1995b).

Provisional Model

Ideally, the effective electron loss rate should be described by a five-dimensional function obtained by an RMS fit to the data (*i.e.* a function of pressure [number density], temperature, ionisation, and zenith angle).

Figure 5.5 displays the mean ψ-profiles as a function of pressure for both day and night using the theoretical model by Torkar and Friedrich (1983a). Also depicted in this figure are the estimated deviations from this mean for temperature variation of ±10 K from the mean temperature T_o and factors from the mean ionisation q_o of 30 higher or lower.

Discussion and Outlook

The number of reliable profiles of effective electron loss rates from the lower ionosphere is very limited. The daytime data scatter much less than those of night-time and an average over all data from the auroral zone to the equator is quite feasible. In the height region where negative ions and clusters can be expected, the night profiles show steep gradients in the effective recombination rate at various altitudes. The inclusion of dependencies on temperature and ionisation rate only partly reproduces this variability.

Because of the very small number of data, another approach is probably more promising. One could conceivably calculate ψ from a theoretical ionospheric model for all conditions prevailing during the measurements and form the factor to the actually measured data. Rather than describing ψ by an analytical function, one can more easily assimilate the factor between theory and measurement by an analytical function and with it adjust the results of the theoretical model. Such an approach has been successfully applied by Friedrich and Torkar (2001b) in the semi-empirical model FIRI of the non-auroral ionosphere and has, *e.g.* made it possible to circumvent the problem of linearizing the zenith angle variation between day and night.

5.2. Collision Frequency

Collisions of free electrons with neutrals, ions or other electrons are important for various macroscopic processes. At higher altitudes (*E*-region) they determine the thermal and electrical conductivity, in the *D*-region they lead to absorption of radio waves. Conversely, the collision frequency can be deduced from radio wave propagation data with an appropriate theory. We use this method to derive collision frequencies. The collision cross section σ for electrons in N_2, the dominant atmospheric constituent up to at least 100 km, was measured in the laboratory by Phelps and Pack (1959) to be proportional to electron velocity v. Sen and Wyller (1960) applied this finding to all the air's constituents (*i.e.* $\sigma \sim v$) and furthermore assumed a Maxwellian velocity distribution. In general, one can express:

$$\nu_m = n v \sigma \tag{5.3}$$

$$n = \frac{p}{kT}$$

$$v = \sqrt{\frac{2kT_e}{m_e}}$$

Strictly speaking, the velocity v in Eq. (5.3) is the relative one between neutrals and electrons. If thermal equilibrium between the two prevails ($T = T_e$) - because of the larger mass of the neutrals - v in the above equations is for all practical purposes simply the thermal velocity of the electrons, whereas neutrals and ions can be considered to be stationary. Since σ was found to be ~ v, ν_m can be expressed as

$$\nu_m = n v \sigma = C v^2 \frac{p}{kT} = Kp \tag{5.4}$$

Based on a collection of more recent laboratory measurements of collision cross sections of various atmospheric gases (Aggarwal and Setty, 1980) one can expect the proportionality factor K between pressure and collision frequency to range from 7.10×10^5 to 6.47×10^5 $m^2\ s^{-1}\ N^{-1}$ for temperatures between 150 and 300 K. The collision frequency is one of the parameters, which determine the refractive index of the ionosphere to radio waves (the others being signal frequency, electron density, the magnetic field strength, and its angle to the propagation direction; see Section 2). Hence ν can be obtained as a by-product of wave propagation

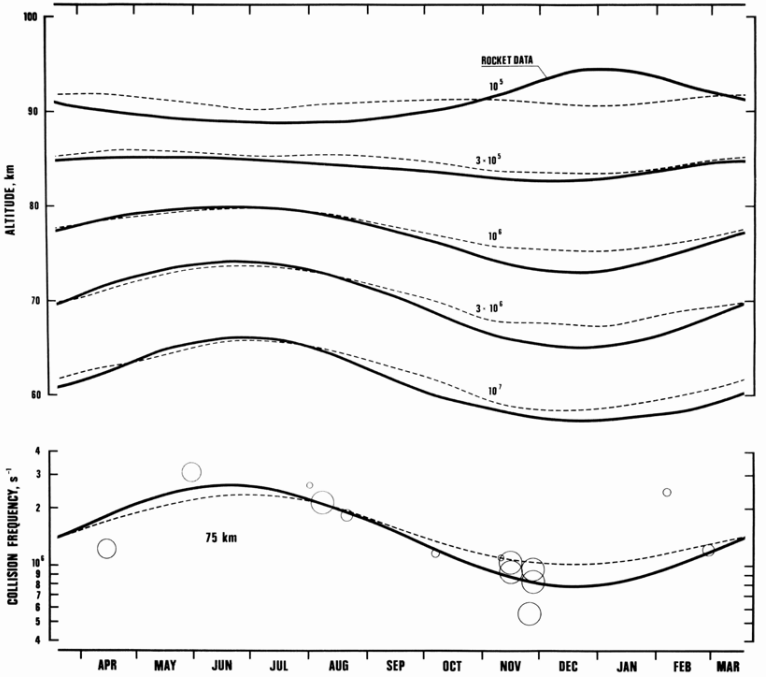


Figure 5.6 Seasonal variation of collision frequencies at selected heights in the auroral zone (Friedrich and Torkar, 1983b). Solid lines are fitted results of rocket borne wave propagation measurements, the dashed lines are proportional to a pressure model. In the bottom panel individual values at 75 km are displayed.

experiments that primarily aim at yielding electron densities. Thrane and Piggott (1966) found from early such rocket measurements from Andøya that collision frequencies were larger in summer than in winter, which they rightly attributed to the seasonally varying pressure. This finding was generalised (Friedrich and Torkar, 1983b) by assuming a sinusoidal seasonal variation of ν at each kilometre of height. Using 19 rocket soundings from the two Scandinavian ranges (Andøya and Esrange) the obtained variation was very similar to the corresponding variation of p from the (then) recent atmospheric Air Force Geophysical model (Cole and Kantor, 1978) and the laboratory K inserting an average temperature of 200 K (Fig. 5.6). To make the finding generally applicable - also for other latitudes - the averaged value of the ratio between collision frequency and the pressure model used was established; the result was

$$K = 6.41(\pm 0.11)\times 10^5 \text{ m}^2 \text{ s}^{-1} \text{ N}^{-1}. \qquad (5.5)$$

Values of K reported in the literature range from 5.8×10^5 m^2 s^{-1} N^{-1} (Smith *et al.*, 1978) to 9.9×10^5 (Bennett *et al.*, 1972), or even twice these values (Vuthaluru *et al*., 2002). However, the two extreme values refer to pressures obtained by the somewhat controversial method of measuring the extinction of solar Lyman-α by the air's major constituent O_2 and we therefore probably do not need to aim at reconciling these results with our own data. Mechtly (1974) used the atmospheric model CIRA-65 and rocket data from Wallops Island; his K appeared to increase with altitude. No explanation for the height dependence of K was put forward. In a similar analysis with ionospheric data from the high latitudes, but using the AFGL model, again K was found to increase with altitude, and was interpreted as an indication of an inadequacy of the pressure model, at least for the European sector at auroral latitudes (Friedrich and Torkar, 1983b; Friedrich, 1985). An attempt to provide a correction to the Cole and Kantor model, based on derived collision frequencies, was undertaken (Torkar and Friedrich, 1983b) and yielded colder mesospheric temperatures in Summer (Fig. 5.7). This interpretation of the feature was much later confirmed by dedicated measurements by falling spheres above Andøya (Lübken and von Zahn, 1991; *cf.* Section 3). The method of deriving pressure, and in consequence temperature, from collision frequency data was not pursued further, although much more collision frequency data have since become available.

The studies concerning momentum transfer collision frequency, the quantity that enters the description of the ionosphere's refractive index, we consider closed. The results one can reasonably obtain from the rocket borne wave propagation experiments are less certain than the pressures from modern atmospheric models. We therefore use pressures (from a measurement

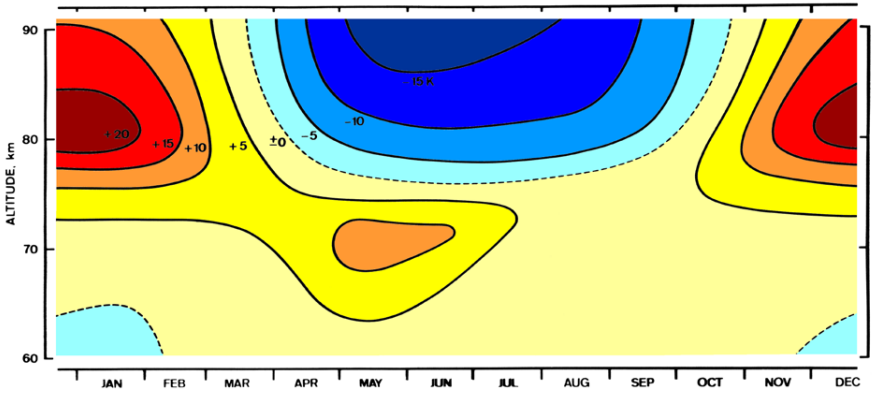


Figure 5.7 Temperature differences to the AFGL atmospheric model based on collision frequency measurements at high latitudes (Torkar and Friedrich, 1983b).

close in time and space, or a suitable model) times $K = 6.41 \times 10^5$ m^2 s^{-1} N^{-1} when using the magneto-ionic theory according to Sen and Wyller. According to the new laboratory results the collision cross section is a linear function of thermal velocity, but has also has a term independent of temperature. K is then not a constant and this finding required the revision of the Sen and Wyller equations, which is briefly described earlier (*cf.* Chapter 2).

5.3. Cluster Transition

The lower *D*-region is chemically complex; whereas towards the *E*-region the dominant ions are NO^+ and O_2^+ whose chemical behaviour is rather well known, below a marked transition height massive ions of the type $H^+(H_2O)_n$ and $NO^+(H_2O)_n$ dominate. Because these ions recombine with electrons orders of magnitude faster than moleculars, a knowledge of the location of the transition is essential in order to understand electron densities N_e resulting from a known ion production rate q. From the very simplified scheme shown in Figure 5.2 it follows that (*a*) cluster ions require a sufficient background density M, (*b*) they cannot form in the presence of atomic oxygen O, and (*c*) they are efficiently broken up by light in the visible range by $h\nu$. Because of the larger recombination rate, the effective electron loss rate q/N_e^2 increases markedly below the transition height. Figure 5.3 shows a rare case where most ionisation processes were known, a mass spectrometer provided the distribution between molecular and cluster ions, and electron densities were established with good accuracy by a wave propagation experiment.

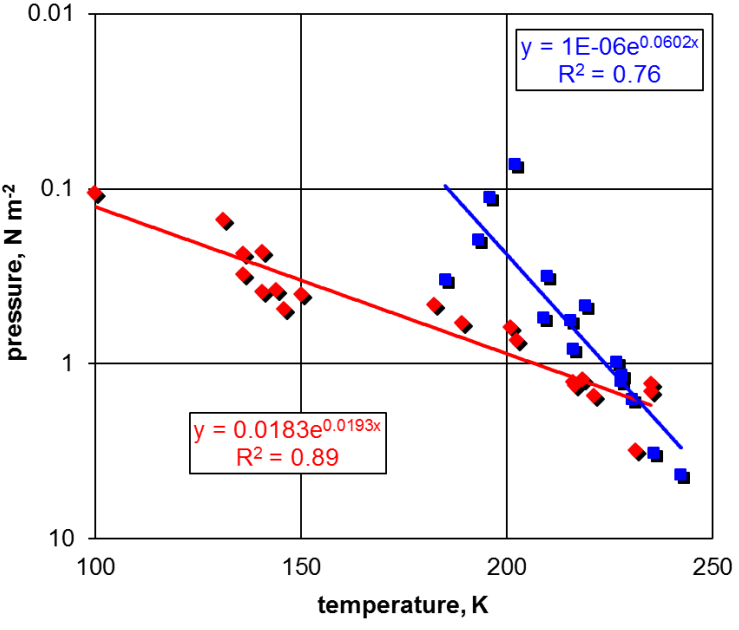


Figure 5.8 Pressure at the transition from cluster to molecular ions as a function of temperature. Note that for very low temperatures the daytime transition (red symbols) occurs at about the same height as under night conditions (blue). Pressures of 10, 1 and 0.1 Pa are typically at altitudes 65, 80 and 93 km, respectively (Friedrich and Torkar, 1988, updated).

The measurement of the distribution between heavy and light positively charged ions requires a cryogenically cooled mass spectrometer in order to enable measurements beginning well below the transition height. World-wide (probably) only three research institutions have ever successfully built such instruments, namely NASA's Goddard Space Flight Center, the Max-Planck Institute for Nuclear Physics, Heidelberg, and the University of Bern. The most recent such flight took place from Esrange, Sweden, in 1993 by the University of Bern and one thus has to work with a very limited data set of only 35 cases. On theoretical grounds one expects a marked difference between day and night, not only due to photo-detachment of the precursor ions leading to water clusters, but also because the density of mesospheric atomic oxygen changes from immeasurably low values at night to about 10^{16} m^{-3} when sunlit. We therefore split the available data according to zenith angle below and above 98° (*i.e*. Sun at the horizon for an observer at ≥70 km). All reactions leading from the initially produced molecular ions to clusters have a strong temperature dependence, *i.e.* the reactions occur faster at higher temperatures. In Figure 5.8, the transitions (equal number of clusters and moleculars) are plotted *vs*. background pressure and temperature at the transition height (For compatibility with other ionospheric models, we chose pressure to replace altitude, although on theoretical grounds we expect number density to be important for the transition). It appears that at night the temperature dependence is stronger than during the day, but the reduced temperature range of the night measurements may somewhat bias the outcome. If the data were restricted to the range common to both data sets (*e.g*., 180 to 230 K and 0.3 to 3 Pa), the resulting gradients were considerably closer. An extension of the fitted lines to smaller pressures (greater heights) is not advisable, because above about 83 km atomic oxygen appears and significantly inhibits the formation of cluster ions.

In cases of very large ionisation rates such as in a PCA event, we expect the primary (molecular) ions to be formed faster than they can be transformed to clusters. Hence, one would expect the transition height to be lower for large ionisation (*e.g.,* Brown, 1970). Applying a 3D fit (*i.e*., *vs. T* and *q*) to the scarce data only produced inconclusive results (Friedrich and Torkar, 1988).

5.4. Negative Ions

Free electrons produced in an ionisation process can recombine with positive ions that were generated in the process (primary ions), or with positive ions resulting from various processes starting from these primary ions, notably with cluster ions (*cf*. Ch. 4.1). Additionally, however, free electrons can also be lost by attachment to neutrals. Because of the larger mass, the negative ions thus created no longer impact on the propagation of radio waves, but do contribute to the conductivity of the ionosphere. The very simple scheme in Figure 5.2 shows that the attachment requires a substantial neutral background, and that during the day negative ions virtually disappear due to photo-dissociation and much larger densities of atomic oxygen. At any rate, daytime negative ions are only expected - and were indeed only detected - at very low heights where the reliability of the measurements is poor (<60 km).

Negative ions can be detected by appropriate mass spectrometers, whereas simple electrostatic probes cannot distinguish between the negative charges, *i.e*. between electrons and negative ions. For quasi-neutrality of the ionospheric plasma the number density of positive ions N^+ must equal the sum of electron density N_e and that of negative ions N^-. One can therefore derive the number density of negative ions by measuring the number density of positive ions and subtracting that of electrons.

$$N^- = N^+ - N_e \tag{5.6}$$

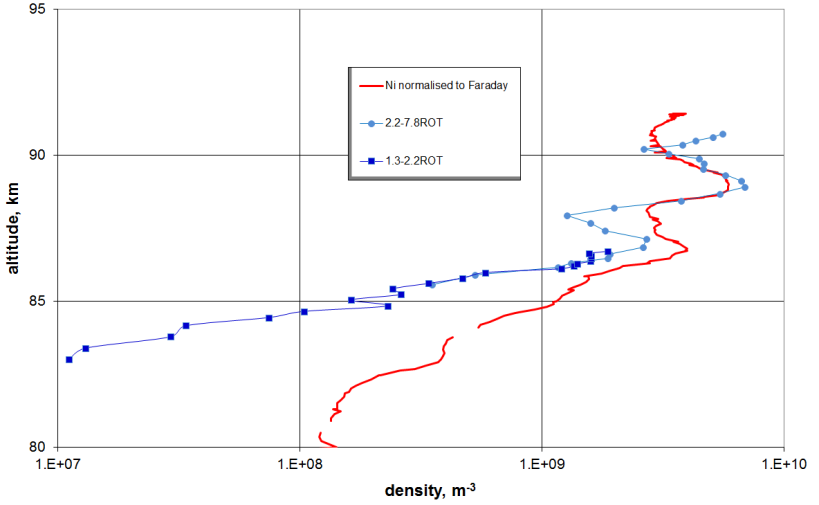


Figure 5.9 Electron and ion densities measured by the rocket ECOMA-0 from Esrange, Sweden, in 2004. The ion densities are normalised near apogee to the electron densities.

Figure 5.9 shows a case of a night-time flight where both electrons and positive ions were derived down to 80 km. The profile of the positive ion density was measured with an electrostatic probe and normalised to the absolute values obtained with the wave propagation experiment. Above about 88 km both profiles show the same relative variations and the probe data were therefore normalised by the wave propagation data with a factor obtained between 88 km and apogee (93 km). The difference between N^+ and N_e below 88 km we interpret as being due to negative ions N^-. Because the density of negative ions is generally measured jointly with that of electrons, it is customary to use the ratio of the two negative species

$$\lambda = \frac{N^-}{N_e} \tag{5.7}$$

For conditions of steady-state (which we always tacitly assume), the formation of negative ions by three-body collision is principally balanced by three loss processes

$$N_e \beta M^2 = N^- (\gamma + \alpha_i N^+ + k[O]) \tag{5.8}$$

β ... attachment rate
M ... neutral number density
α_i ... ion-ion recombination rate
k ... electron detachment by atomic oxygen

For night-time conditions, *i.e.* for which we have sufficient data, [O] and γ are negligible and Eq. (5.8) reduces to

$$N_e \beta M^2 = N^- \alpha_i N^+ \tag{5.9}$$

We now replace N^- and N^+ by N_e and λ (Eqs. 5.6 and 5.7), and one obtains a quadratic equation for λ

$$\lambda^2 + \lambda - \frac{\beta M^2}{\alpha_i N_e} = 0 \tag{5.10}$$

where only one solution has a physical meaning

$$\lambda = \sqrt{\frac{1}{4} + \frac{\beta M^2}{\alpha_i N_e}} - \frac{1}{2} \tag{5.11}$$

In the above equation one can empirically determine β/α_i using measured N_e and λ as well as M from a suitable atmospheric model. The outcome as a function of altitude is shown in Figure 5.10 (Friedrich and Torkar, 1995b); because no systematic variation is evident, at best a mean value of $\beta/\alpha_i = 10^{-25}$ cm^3 can be derived. Numerically λ can thus be described as a function of background neutral number density M and electron density N_e (both in m^{-3})

$$\lambda = \sqrt{0.25 + 10^{-31} \frac{M^2}{N_e}} - 0.5 \tag{5.12}$$

It is of interest to compare the empirically derived β/α_i to laboratory values. Inserting 4×10^{-7} cm^3 s^{-1} for α_i (Peterson *et al*., 1971), one obtains 4×10^{-32} cm^6 s^{-1} for β, which is of the same order of magnitude as laboratory values (Phelps, 1969). Eq. (5.11) is only valid for solar zenith angles >98° and has an upper height limit where one can expect the ledge in the night-time [O] density, *i.e*. at about 83 km.

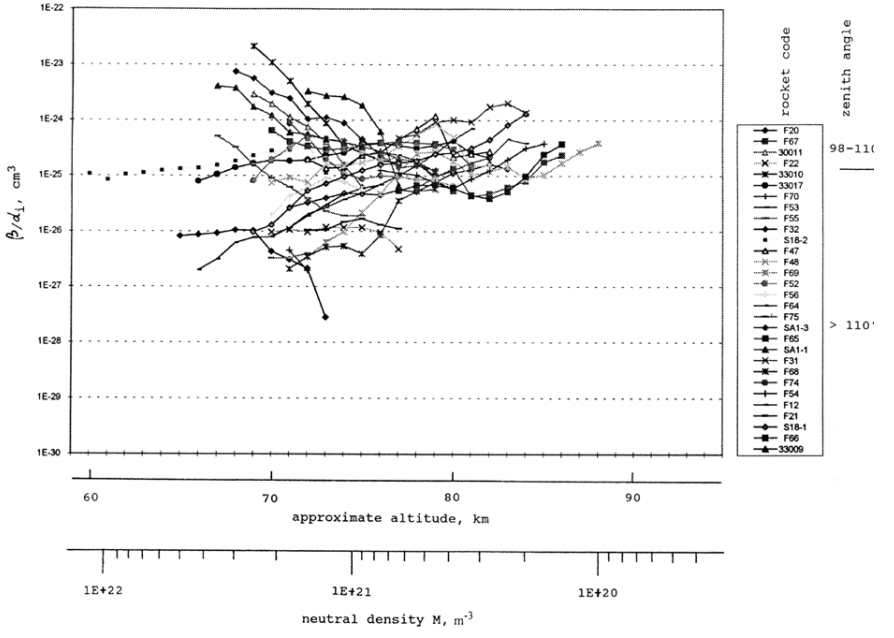


Figure 5.10 Ratio between attachment and ion-ion recombination rate as a function of neutral density obtained from night-time flights. Altitude is for orientation only (Friedrich and Torkar, 1995b).

During the day more loss processes for negative ions are active, which are not as easily describable as the simple ion-ion recombination. Nevertheless, one can try to establish an "effective" β/α_i with the daytime data. If we leave out the twilight values (χ <80°) one obtains a corresponding relation for daytime λ of about two orders of magnitude higher in density (or about 25 km lower in altitude)

$$\lambda = \sqrt{0.25 + 3 \cdot 10^{-35} \frac{M^2}{N_e}} - 0.5 \qquad (5.13)$$

Again, this empirical relation must not be extended beyond, say, 83 km, *i.e.* into a region where the density of atomic oxygen becomes appreciable.

The huge scatter in the β/α_i values (*cf.* Fig. 5.10) triggered the search for a plausible explanation. A topical candidate is meteoric dust which is known to be ubiquitous and prevalent in the mesosphere due to ablating and coagulating meteors. The estimates concerning the meteoric matter hitting the Earth range from 44 tons a day (according to Hunten *et al.*, 1980) and to only 4 tons a day (Asmus *et al.*, 2015). The latter figure turned out to be sufficient to explain the balance of charges measured *in situ* in the mesosphere. If we accept that aerosols (ice or meteoric dust) impact on the ion-electron ratio λ and in consequence the effective recombination rate ψ, we need to invoke a similarly large variability of their concentration in the mesosphere; neither indirect ground based measurements, nor global models of their occurrence rule out such variability. The simple ion-chemical scheme of the lower ionosphere depicted in Figure 5.1 has to be expanded by the inclusion of particles (Fig. 5.11). Note that (a) the loss of free electrons to particles (*P*) is by direct attachment and (b) that - in contrast to the electrons attached to molecules - atomic oxygen does *not* provide a reverse reaction.

The arguably most convincing data demonstrating the relevance of meteoric particles for the ion chemical balance is provided by the data from the rocket flight 41.094 (Andøya, October 11, 2011, at χ = 116°). Figure 5.12 shows the measured ion and electron densities. Also indicated is the inferred electron density derived from the ion density by Eq. (4.6). Clearly between 82 and 92 km there are fewer electrons than expected from the ion density (hatched

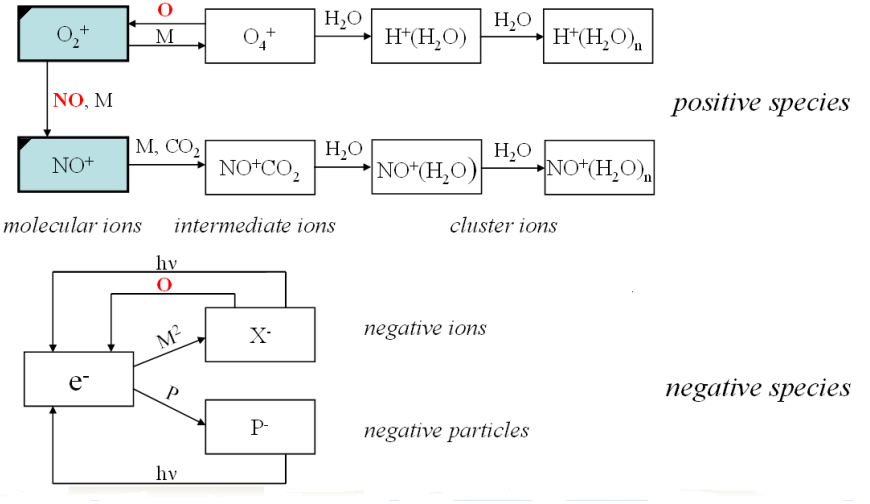


Figure 5.11 The simple ion-chemical scheme of the D-region (cf. Fig. 5.2) expanded by the inclusion of particles (Friedrich et al., 2012).

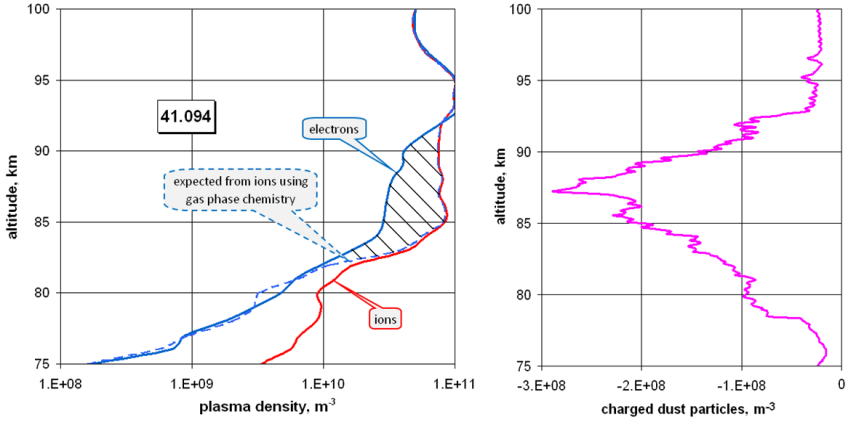


Figure 5.12 Night-time plasma densities in the mesosphere (left panel) and measured heavy charged particles (right panel). Clearly the deficit of free electrons is related to the observed negatively charged heavy particles (after Friedrich et al., 2012).

area). This altitude coincides with the appearance of negatively charged particles in the mass range 500 to 2000 amu measured by a mass spectrometer (Friedrich *et al.,* 2012).

Other additional electron loss mechanisms have been identified earlier, namely the so-called bite-outs. They can unambiguously be tied to visible noctilucent clouds (NLC) in the cold summer mesopause consisting of ice crystals. Since they occur as spectacularly thin layers (1 or 2 km), their relevance for the ion chemistry had been recognised much earlier than the ubiquitous, but much broader regions where meteoric dust impacts.

A recent study of the potential importance of meteoric dust for the ion chemistry in the mesosphere is presented in the paper by Friedrich *et al*. (2011). Data from 29 night-time rocket soundings which provided both ion and electron densities are used to study the variation of the ratio β/α_i. According to the simple theoretical scheme in Figure 5.1 one does not expect an influence of atomic oxygen on the attachment of electrons onto meteoric dust, whereas the formation of negative ions will be drastically reduced at altitudes of appreciable [O]. However, both negative species which scavenge free electrons are affected by light *via* hν. Applying Eq. 4.6, for each rocket sounding, and each kilometre of altitude, the ratio β/α_i was established which produced the best agreement between measured and inferred electron density. The effective attachment rate β will be reduced by the reverse reaction *via* photo-detaching light, hence we also expect β/α_i to vary with both scattered sunlight after sunset and - conceivably – also by moon light. The widely scattering values make a 3-dimensional fit (*i.e. vs.* solar zenith angle and moon light) unrealistic. Instead we first analyse the dependence of β/α_i *vs.* solar zenith angle only using data that were taken in absence of any moonlight (Fig. 5.13).

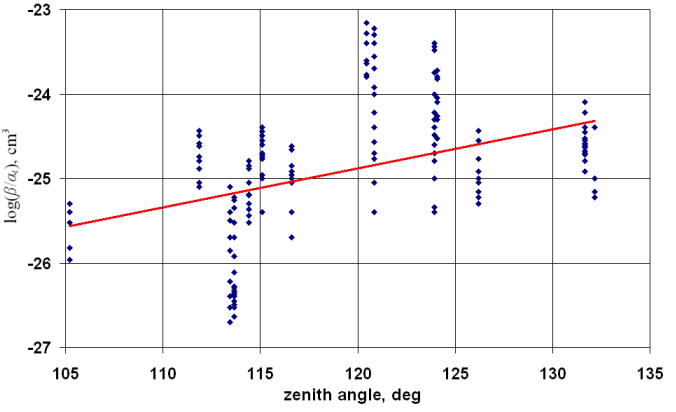


Figure 5.13 Ratio β/α_i as a function of solar zenith angle after sunset (no moon data only). The values at constant zenith angles are from different altitudes of the same rocket flight (Friedrich et al., 2011).

The figure shows an increase of β/α_i with increasing solar zenith angle, *i.e.* most likely an increase of β due to the reduced reverse reaction by photo detachment. The RMS line in the figure follows (in cm^3)

$$\log(\beta/\alpha_i) = 0.046\chi\text{-}36.45 \qquad (5.14)$$

This variation by about a factor of ten can be explained by a similar factor of the decrease of the scattered Lyman-α over the same zenith angle range (*e.g.* Strobel *et al.*, 1974); a similar relative variation of the wavelengths responsible for photo detachment can reasonably be assumed. Next we apply the above relation to normalise all β/α_i values to a mean solar zenith angle of 120°. Having thus - on average - eliminated the solar angle influence, we now plot these values *vs.* moon albedo (Fig. 5.14). Again, the gradient suggests that moonlight may indeed reduce the effective attachment rate β.

The huge scatter one can best explain (a) by the varying size and number density of meteoric dust, and (b) by an unaccounted variation of atomic oxygen. Despite the scatter of the data the results may inspire the inclusion of meteoric dust in theoretical atmospheric models.

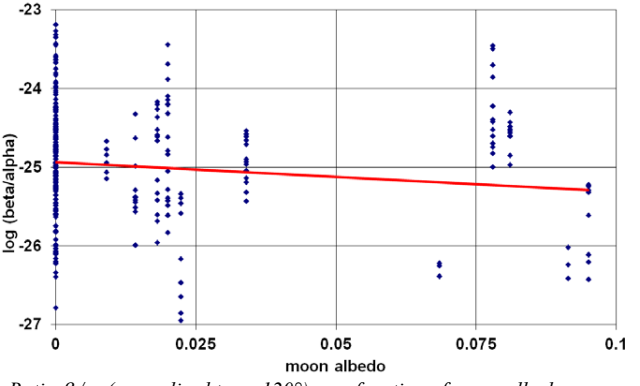


Figure 5.14 Ratio β/α_i (normalised to χ=120°) as a function of moon albedo.

References

Aggarwal, K.M. and C.S.G.K. Setty:
Collision & Transport of Electrons in the Ionosphere,
Indian J. Radio Space Phys. **9**, pp. 105-111, 1980.

Aikin, A.C. and L.J. Blumle:
Rocket Measurements of the *E*-Region Electron Concentration Distribution in the Vicinity of the Geomagnetic Equator,
J. geophys. Res. **73**, pp. 1617-1626, 1968.

Arnold, F. and D. Krankowsky:
Mid-Latitude Lower Ionosphere Structure and Composition Measurements During Winter,
J. atmos. terr. Phys. **41**, pp. 1127-1140, 1979.

Asmus, H., S. Robertson, Shannon Dickson, M. Friedrich, and Linda Megner:
Charge Balance for the Mesosphere with Meteoric Dust Particles,
J. atmos. solar terr. Phys., **127**, pp. 137-149, 2015.

Bauer, A.:
Untersuchungen zur Berechnung von Elektronendichte und Stoßzahl in der Ionosphäre mit Hilfe des Faradayeffektes,
PhD Thesis, Technical University Graz, 1972.

Bauer, A.:
Berücksichtigung der verschiedenen Wellenpfade der ordentlichen und außerordentlichen Komponente bei der Berechnung von Elektronendichte und Stoßzahl mit Hilfe des Faradayexperimentes,
Kleinheubacher Ber. **16**, pp. 353-358, 1973.

Bekkeng, T.A., K.S. Jacobsen, J.K. Bekkeng, A. Pedersen, T. Lindem, J.-P. Lebreton, and J.I. Moen:
Design of a Multi-Needle Langmuir Probe System,
Meas. Sci. Technol. **21**, doi:10.1088/0957-0233/21/8/085903, 2010.

Bennett, F.D.G., J.E. Hall, and P.H.G. Dickinson:
D-Region Electron Densities and Collision Frequencies from Faraday Rotation and Differential Absorption,
J. atmos. terr. Phys. **34**, pp. 1321-1335, 1972.

Bilitza, D., Lee-Anne McKinnell, B. Reinisch and T. Fuller-Rowell:
The International Reference Ionosphere Today and in the Future,
J. Geod. **85**, pp. 909-920, doi 10.1007/s00190-010-0427-x, 2002.

Bowman, M.R. and L. Thomas:
Sunrise Changes in Concentration of Minor Neutral Constituents in the Mesosphere,
J. atmos. terr. Phys. **35**, pp. 347-352, 1973.

Brown, R.R.:
On the Influence of Energetic Electron Precipitation on the Water Cluster Ion Production in the Upper *D*-Region,
J. atmos. terr. Phys. **32**, pp. 1747-1753, 1970.

Cain, H.V.:
A 30 MHz Map of the Whole Sky,
Austr. J. Phys., **31**, pp. 562- 565, 1978.

Carver, J.H., H.P. Gies, T.I. Hobs, B.R. Lewis, and D.G. McCoy:
Temperature Dependence of the Molecular Oxygen Photoabsorption Cross Section Near the H Lyman α Line,
J. geophys. Res. **82**, pp. 1955-1960, 1977.

Cole, E.A. and A.J. Kantor:
Air Force Reference Atmospheres,
AFGL-TR-0051, 1978.

Collis, P.N., J.K. Hargreaves, W.G. Howarth, and G.P. White:
Joint Imaging Riometer - Incoherent Scatter Radar Observations: A Four-Dimensional Perspective on Energetic Particle Input to the Auroral Mesosphere,
Adv. Space Res. **20 (6)**, pp. 1165-1168, 1997.

Danilov, A.D.:
Long-Term Changes of the Mesosphere and Lower Thermosphere and Composition,
Adv. Space Res. **20 (11)**, pp. 2137-2147, 1997.

Danilov, A.D.:
Review of Long-Term Trends in the Upper Mesosphere, Thermosphere and Ionosphere,
Adv. Space Res. **22 (6)**, pp. 907-915, 1998.

Danilov, A.D.:
New Ideas on the *D*-Region Modelling,
Adv. Space Res. **25 (1)**, pp. 5-14, 2000.

Danilov, A.D., Natasha V. Smirnova, T.A. Blix, E.V. Thrane, and Ljudmila B. Vanina:
Some Features of Electron Density Behaviour in the High Latitude *D*-Region Derived from *in-situ* Measurements,
J. atmos. solar terr. Phys. **65**, pp. 417-427, 2003.

Dickinson, P.H.G. and F.D.G. Bennett:
Diurnal Variation in the *D*-Region During a Storm After-Effect,
J. atmos. terr. Phys. **40 (5)**, pp. 549-551, 1978.

Egger, G.:
Empirical Model of the Polar Cap Ionosphere,
MSc Thesis, Graz University of Technology, 2004.

Enell, C.F., A. Kero, E. Turunen, T. Ulich, P.T. Verronen, Annika Seppälä, S. Marple, Farideh Honary, and A. Senior:
Effects of *D*-Region RF Heating Studied with the Sodankylä Ion Chemistry Model,
Ann. Geophys. **23**, pp. 1575-1583, 2005.

Erlacher, R.:
Empirisches Elektronendichtemodell der unteren Ionosphäre mit variabler Ordnung und Anzahl der Parameter,
MSc Thesis, Technical University Graz, 1996.

Evans, J.S., B.M. McLaughlin, and R.R. Meier:
Why do Models Fail to Predict Sufficient *E*-Region Electron Densities?
Proposal 1613970 to the National Science Foundation, 2016.

Fankhauser, M.:
Time-Dependent Models of the Polar Cap and Low Latitude Ionosphere,
PhD Thesis, Graz University of Technology, 2010.

Folkestad, K.:
Ionospheric Studies by *in situ* Measurements in Sounding Rockets,
PhD Thesis, University of Oslo (and Internal NDRE Report **59**), 1970.

France, L.A. and E.R. Williams:
An Improved Faraday Rotation Experiment for Rocket Measurements of the Lower Ionosphere Electron Concentration,
J. atmos. terr. Phys. **39**, pp. 957-964, 1976.

Friedrich, M.:
A Compact Faraday-Slant Range Receiver Unit as Flown on Rocket F27,
Internal Report **INW 7205**, 1972.

Friedrich, M.:
On-Board Phase Measurement and Maximum-Minimum Detection for the Faraday Experiment,
Royal Norwegian Council for Scientific and Industrial Research (NTNF), **SAD-47-T**, pp. 12-19, 1973.

Friedrich, M.:
Raketenmessungen in einem REP Ereignis,
PhD Thesis, Technical University Graz, 1974.

Friedrich, M.:
Recent Considerations Concerning *D*-Region Collision Frequencies,
Adv. Space Res. **5 (7)**, pp. 91-93, 1985.

Friedrich, M.:
The Lower Ionosphere: Abandoned by Communication - To be Rediscovered for Aeronomy,
Radio Sci. Bull. **309**, pp. 38-46, 2004.

Friedrich, M. and T.A. Jacobsen:
On the Accuracy of the Determination of Electron Density and Collision Frequency with the Faraday Experiment,
Kleinheubacher Ber. **18**, pp. 333-343, 1975.

Friedrich, M. and Sheila Kirkwood:
The *D*-Region Background at High Latitudes,
Adv. Space Res. **25 (19)**, pp. 15-23, 2000.

Friedrich, M., and G. Landauer:
An Extension of the Model IMAZ for Large Absorption,
Earth Planets Space **63**, pp. 383-390, 2011.

Friedrich, M. and K.M. Torkar:
Electron Densities in the *D*- and *E*-Regions,
Paper presented at 5th EGS Meeting, Strasbourg, August 1978, also *Internal Report*
Internal Report **INW 7809**, 1978.

Friedrich, M. and K.M. Torkar:
High-Latitude Plasma Densities and Their Relation to Riometer Absorption,
J. atmos. terr. Phys. **45 (2/3)**, pp. 127-135, 1983a.

Friedrich, M. and K.M. Torkar:
Collision Frequencies in the High-Latitude *D*-Region,
J. atmos. terr. Phys. **45 (4)**, pp. 267-271, 1983b.

Friedrich, M. and K.M. Torkar:
Empirical Transition Heights of Cluster Ions,
Adv. Space Res. **8 (4)**, pp. 235-238, 1988.

Friedrich, M. and K.M. Torkar:
An Empirical Model of the Nonauroral *D*-Region,
Radio Sci. **27**, pp. 945-953, 1992.

Friedrich, M. and K.M. Torkar:
Typical Behaviour of the High Latitude Lower Ionosphere,
Adv. Space Res. **16 (1)**, pp. 73-81, 1995a.

Friedrich, M. and K.M. Torkar:
An Attempt to Parameterise Negative Ions in the Ionospheric *D*-Region,
ESA **SP-370**, pp. 257-261, 1995b.

Friedrich, M. and K.M. Torkar:
Equatorial Ionospheric Measurements, Part III, Preliminary Data,
Internal Report **INW 9501**, 1995c.

Friedrich, M. and K.M. Torkar:
Comparison Between an Empirical and a Theoretical Model of the *D*-Region,
Adv. Space Res. **21 (6)**, pp. 895-904, 1998a.

Friedrich, M. and K.M. Torkar:
Empirical *D*-Region Modelling, A Progress Report,
Adv. Space Res. **22 (6)**, pp. 757-766, 1998b.

Friedrich, M. and K.M. Torkar:
FIRI: A Semiempirical Model of the Lower Ionosphere,
J. geophys. Res. **106**, pp. 21409-21418, 2001a.

Friedrich, M. and K.M. Torkar:
Long-Term Trends and Other Residual Features of the Lower Ionosphere,
ESA **SP-471**, pp. 357-362, 2001b.

Friedrich, M., R. Finsterbusch, K.M. Torkar, and P. Spöcker:
A Further Generalisation of the Sen and Wyller Magneto-Ionic Theory,
Adv. Space Res. **11 (10)**, pp. 105-108, 1991.

Friedrich, M., K.M. Torkar, R.A. Goldberg, J.D. Mitchell, C.L. Croskey, and G. Lehmacher:
Comparison of Plasma Probes in the Lower Ionosphere,
ESA **SP-397**, pp. 381-386, 1997.

Friedrich, M., D.E. Siskind, and K.M. Torkar:
HALOE Nitric Oxide Measurements in View of Ionospheric Data,
J. atmos. solar terr. Phys. **60**, pp. 1445-1457, 1998.

Friedrich, M., J. Gumbel, and R. Pilgram:
Atomic Oxygen in the Mesosphere and its Relevance for the Ionosphere,
ESA **SP-437**, pp. 287-290, 1999.

Friedrich, M., R. Pilgram and K.M. Torkar:
A Novel Concept for Empirical *D*-Region Modelling,
Adv. Space Res. **27 (1)**, pp. 5-12, 2001.

Friedrich, M., M. Harrich, K.M. Torkar, and P. Stauning:
Quantitative Measurements with Wide-Beam Riometers,
J. atmos. solar terr. Phys. **64**, pp. 359-365, 2002.

Friedrich, M., M. Harrich, R.J. Steiner, K.M. Torkar, and F.-J. Lübken:
The Quiet Auroral Ionosphere and its Neutral Background,
Adv. Space Res. **33 (6)**, pp. 943-948, 2004a.

Friedrich, M., K.M. Torkar, and R.J. Steiner:
Empirical Recombination Rates in the Lower Ionosphere,
Adv. Space Res. **34 (9)**, pp. 1937-1942, 2004b.

Friedrich, M., K. Torkar, G. Lehmacher, C. Croskey, J. Mitchell, E. Kudeki, and M. Milla:
Rocket and Incoherent Scatter Common-Volume Electron Measurements of the Equatorial Lower Ionosphere,
Geophys. Res. Lett. **33 (8)**, S. L08807 - L08807, 2006.

Friedrich, M., M. Rapp, J.M.C. Plane, and K.M. Torkar:
Bite-Outs and Other Depletions of Mesospheric Electrons,
J. atmos. solar Terr. Phys. **73**, pp. 2201-2211, 2011.

Friedrich, M., M. Rapp, T. Blix, U.-P. Hoppe, K. Torkar, S. Robertson, Shannon Dickson and Kristina Lynch:
Electron Loss and Meteoric Dust in the Mesosphere,
Ann. Geophys. **30**, pp. 1495-1501, 2012.

Friedrich, M., K.M. Torkar, U.-P. Hoppe, T.-A. Bekkeng, A. Barjatya, and M. Rapp:
Multi-Instrument Comparison of *D*-Region Plasma Measurements,
Ann. Geophys. **31**, pp. 135-144, 2013a.

Friedrich, M., K. Torkar, S. Robertson, and Shannon Dickson:
Bonus Results from Ion Probes,
ESA **SP-221**, pp. 75-80, 2013b.

Friedrich, M., K. Torkar, and S. Ulrich:
A Rocket Borne Experiment to Measure Plasma Densities in the *D*-Region,
J. Geophys. **44 (1/2)**, pp. 91-98, 1977.

Gadsden, M.;
A Secular Change in Noctilucent Cloud Occurrence,
J. atmos. terr. Phys. **52 (4)**, pp. 247-251, 1990.

Gilchrist, B.E. and L.G. Smith:
Rocket Radio Measurements of Electron Density in the Nighttime Ionosphere,
University of Illinois Urbana-Champaign, *Aeronomy Report* **85**, 1979.

Goldberg, R.A., G.A. Lehmacher, F.J. Schmidlin, D.C. Fritts, J.D. Mitchell, C.L. Croskey, M. Friedrich, and W.E. Swartz:
Equatorial Dynamics Observed by Rocket, Radar, and Satellite During the CADRE/MALTED Campaign: 1. Programmatic and Small Scale Fluctuations,
J. geophys. Res. **102 (D22)**, pp. 26179-26190, 1997.

Golshan, N. and C.F. Sechrist:
Seasonal and Solar Cycle Variation of the *E*-Region Nitric Oxide,
Radio Sci. **10**, pp. 305-315, 1975.

Gumbel, J.:
Rocket Borne Optical Measurements of Minor Constituents in the Middle Atmosphere,
PhD Thesis, Stockholm University, 1997.

Hackl, H.:
A Combined Ion- and Capacitance Probe for Ionospheric Research,
MSc Thesis, Graz University of Technology, 2014.

Han, F. and S.A. Cummer:
Midlatitude Nighttime *D* Region Ionosphere Variability on Hourly and Monthly Time Scales,
J. geophys. Res. **115 (A09323)**, doi:10.1029/2010JA015437, 2010.

Hargreaves, J.K.:
Auroral Absorption of HF Radio Waves in the Ionosphere: A Review of Results from the First Decade of Riometry,
Proc. IEEE **57 (8)**, pp. 1348-1373, 1969.

Hargreaves, J.K. and M. Friedrich:
The Estimation of *D*-Region Electron Densities from Riometer Data,
Ann. Geophys. **21**, pp. 603-613, 2003.

Harrich, M.:
Empirical Modelling of Electron Densities in the High Latitude Mesosphere,
PhD Thesis, Technical University Graz, 2001.

Harrich, M., M. Friedrich, S.R. Marple, and K.M. Torkar:
The Background Absorption at High Latitudes,
Adv. Radio Sci. **1**, pp. 325-327, 2003.

Haselgrove, Jenifer:
Oblique Ray Paths in the Ionosphere,
Proc. Phys. Soc. B **70**, p. 653, 1957.

Haykin, S.:
Neural Networks: A Comprehensive Foundation,
Macmillan, Indianapolis, IN, 1994.

Heilmann, A.:
Antennen, Erster Teil (1),
Hochschultaschenbuch-Verlag, Mannheim, 1970.

Hunten, D.M., R.P. Turco, and O.B. Toon:
Smoke and Dust Particles of Meteoric Origin in the Mesosphere and Stratosphere,
J. atmos. Sci. **37**, pp. 1342-1357, 1980.

Isaksen, I.S.A.:
Diurnal Variation of Atmospheric Constituents in an Oxygen-Hydrogen-Nitrogen-Carbon Atmospheric Model, and the Role of Minor Neutral Constituents of the Lower Ionosphere,
Geofys. Publ. **30**, pp. 1-63, 1973.

Jackson, J.E. and J.C. Seddon:
Ionospheric Electron Density Measurements with the Navy Aerobee III Rocket,
J. Geophys. Res. **63**, pp. 197-208, 1958.

Jacobsen, T.:
High Latitude Ionospheric Studies by Rocket Measurements with Emphasis on RF-Capacitance Probes,
MSc Thesis, University of Oslo (also Internal NDRE Report **E-2049**), 1972.

Jacobsen, T.A. and M. Friedrich:
Practical Limits to the Use of the Faraday Technique,
Internal Report **INW 7403**, 1974.

Jacobsen, T.A. and M. Friedrich:
Electron Density Measurements in the Lower *D*-Region,
J. atmos. terr. Phys. **41 (12)**, pp. 1195-1200, 1979.

Jarvis, M.J., B. Jenkins, and G.A. Rodgers:
Southern Hemisphere Observations in the *F* Region Altitude and Thermospheric Wind Providing Possible Evidence for Global Cooling,
J. geophys. Res. **103 (A9)**, pp. 20774-20787, 1998.

Jespersen, M., O. Petersen, J. Rybner, B. Bjelland, O. Holt, and B. Landmark:
Electron Density Distribution in Ionosphere and Exosphere,
E. Thrane (ed.), North Holland, Amsterdam, 1964.

Jespersen, M., J.A. Kane, and B. Landmark:
Electron and Positive Ion Density Measurements During Conditions of Auroral Absorption,
J. atmos. terr. Phys. **30**, pp. 1955-1963, 1968.

Johnson, R.C. (ed.):
Antenna Engineering Handbook,
McGraw-Hill, 3rd edition, 1993.

Kazil, J., E. Kopp, S. Chabrillat, and J. Bishop:
The University of Bern Atmospheric Ion Model: Time-Dependent Ion Modelling in the Mesosphere and Lower Thermosphere,
J. geophys. Res. **108 (D14)**, doi:10.1029/2002JDD003024, 2003.

Kirkwood, Sheila and P.N. Collis:
The High Latitude Lower Ionosphere Observed by EISCAT,
Adv. Space Res. **11 (10)**, pp. 109-112, 1991.

Kirkwood, Sheila and Alya Osepian:
Quantitative Description of Electron Precipitation During Auroral Absorption Events in the Morning/Noon Local-Time Sector,
J. atmos. solar terr. Phys. **63**, pp. 1907-1922, 2001.

Knight, P.:
A Classification of Night-Time Electron Density Profiles,
J. atmos. terr. Phys. **34**, pp. 401-410, 1972.

Kropf, W.:
Berechnung von Elektronendichte- und Stoßzahlprofilen aus redundanten Datensätzen mit Hilfe der Methode der kleinsten Quadrate,
MSc Thesis, Technical University Graz, 1978.

Laštovička, J.:
Seasonal Variation in the Asymmetry of Diurnal Variation of Absorption in the Lower Ionosphere,
J. atmos. terr. Phys. **39**, pp. 891-894, 1977.

Lübken, F.-J.:
Thermal Structure of the Arctic Summer Mesosphere,
J. geophys. Res. **104**, pp. 9135-9149, 1999.

Lübken, F.-J. and U. Berger:
Interhemispheric Comparison of Mesospheric Ice Layers from the LIMA Model,
J. atmos. solar terr. Phys. **69**, pp. 2292-2308, 2007.

Lübken, F.-J. and A. Müllemann:
First *in situ* Temperature Measurements in the Summer Mesosphere at Very High Latitudes (78° N),
J. geophys. Res. **108 (D8)**, 8448, doi:10.1029/2002JD002414, 2003.

Lübken, F.-J. and U. von Zahn:
Thermal Structure of the Mesopause Region at Polar Latitudes,
J. geophys. Res. **96**, pp. 20841-20857, 1991.

Lübken, F.-J., M.J. Jarvis, and G. Owen:
First *in situ* Temperature Measurements at the Antarctic Summer Mesopause,
Geophys. Res. Lett. **26**, pp. 3581-3584, 1999.

Marcos, F.A., J.O. Wise, M.J. Kendra, N.J. Grossbard, and B.R. Bowman:
Detection of a Long-Term Decrease in Thermospheric Neutral Density,
Geophys. Res. Lett. **23**, L04103, doi:10.1029/2004GL021269, 2005.

Marsh, D.R. and J.M. Russell III:
A Tidal Explanation for the Sunrise/Sunset Anomaly in HALOE Low-Latitude Nitric Oxide Observations,
Geophys. Res. Lett. **27 (19)**, pp. 3197-3200, 2000.

Mayer, G., R. Hommel, L. Schöberle, and P. Marutschek:
The DEOS Radio Beacon Experiment on Rocket (RABER) of DEOS, First Results,
ESA **SP-437**, pp. 423-428, 1999.

McKinnell, Lee-Anne and M. Friedrich:
Towards Neural Network Based Models for the Ionospheric *D*-Region,
ESA **SP-530**, pp. 369-372, 2003.

McKinnell, Lee-Anne, and M. Friedrich:
A Neural Network Based Ionospheric Model of the Auroral Zone,
*J. atmos. solar-terr. Phys***. 69**, pp. 1459-1470, 2007.

McKinnell, Lee-Anne, M. Friedrich, and R.J. Steiner:
A New Approach to Modelling the Daytime Lower Ionosphere at Auroral Latitudes,
Adv. Space Res. **34**, pp. 1943-1948, 2004.

McNamara, L.F.:
A Statistical Model of the *D*-Region,
Radio Sci. **14**, pp. 1165-1173, 1979.

Mechtly, E.A.:
Accuracy of Rocket Measurements of Lower Ionosphere Electron Density Concentration,
Radio Sci. **9**, pp. 373-378, 1974.

Mechtly, E.A., S.A. Bowhill, L.G. Smith, and H.W. Knoebel:
Lower Ionosphere Electron Concentrations and Collision Frequency from Rocket Measurements of Faraday Rotation, Differential Absorption, and Probe Current,
J. geophys. Res. **72**, pp. 5239-5245, 1967.

Mitra, A.P. and J.N. Rowe:
Ionospheric Effects of Solar Flares - Changes in the *D*-Region Ion Chemistry,
J. atmos. terr. Phys. **34**, pp. 795-806, 1972.

Miyazaki, S., A. Ogawa, and H. Mori:
Some Features of Nighttime *D*- and *E*-Region Electron Density Profiles in the Polar Ionosphere,
Mem. National Inst. Polar Res. **18**, pp. 304-311, 1981.

Nutz, K.-D.:
Digitale Phasenmessung mit automatischer Spitzenwertübertragung für das Faraday-Raketenexperiment,
MSc Thesis, Technical University Graz, 1972.

Offermann, D.:
A Study of *D*-Region Winter Anomaly in Western Europe, 1975/76,
J. atmos. terr. Phys. **41**, pp. 1-13, 1979.

Ogawa, T. and T. Shimazaki:
Diurnal Variation of Odd Nitrogen and Ionic Densities in the Mesosphere and Lower Thermosphere: Simultaneous Solution of Photochemical-Diffusive Equations,
J. geophys. Res. **80**, pp. 3945-3960, 1975.

Ogawa, Y., T. Motoba, S.C. Buchert, I. Häggström, and S. Nozawa:
Upper Atmosphere Cooling Over the Past 33 Years,
Geophys Res. Lett. **41**, pp. 5629-5625, doi:10.1002/2014GL060591, 2014.

Oliveira-Costa, Angélica, M. Tegmark, B.M. Gaensler, J. Jonas, T.L. Landecker, and Patricia Reich:
A Model of Diffuse Galactic Radio Emission from 10 MHz to 100 GHz,
Mon. Not. R. Astron. Soc. **388 (1)**, pp. 1-16, 2008.

Osepian, Alya, Sheila Kirkwood, and Natasha Smirnova:
Variations of Electron Density and Energetic Spectra of the Precipitating Electrons During Auroral Substorms by Incoherent Scatter Data,
Cosmic Res. **39 (3)**, pp. 311-315, 2001.

Peterson, J.R., W.H. Abert, J.T. Moseley, and J.H. Sheridan:
Ion-Ion Mutual Neutralization Cross Section Measured by Superimposed Beam Technique. III $O_2^+ + O_2^-$, $O_2^+ + NO_2^-$, and $NO^+ + NO_2^-$,
Phys. Rev. **3A**, pp. 1651-1657, 1971.

Phelps, A.V.:
Laboratory Studies of Electron Attachment and Detachment Processes of Aeronomic Interest,
Can. J. Chem. **47**, pp. 1783-1793, 1969.

Phelps, A.V. and J.L. Pack:
Electron Collision Frequencies in Nitrogen and in the Lower Ionosphere,
Phys. Rev. Lett. **3**, pp. 340-342, 1959.

Picone, J.M., A.E. Hedin, D.P. Drob, and A.C. Aikin:
NRLMSIS-00 Empirical Model of the Atmosphere: Statistical Comparison and Scientific Issues,
J. geophys. Res. **107 (A12)**, doi:10.1029/2002JA009430, 2002.

Poole A.W.V. and Lee-Anne McKinnell:
On the Predictability of fo*F*2 Using Neural Networks,
Radio Sci. **35 (1)**, pp. 225-234, 2000.

Preiß, D.:
Impedanzbestimmung elektrisch kurzer Antennen und Link Budget,
Project, Graz University of Technology, 2004.

Pürstl, F.:
MARIPROBE-D a 3D-Ion-Spectrometer for *in-situ* Investigation of Cold Ions in the Martian Ionosphere,
PhD Thesis, Graz University of Technology, 2000.

Rapp, M., J. Gumbel, and F.-J. Lübken:
Absolute Density Measurements in the Middle Atmosphere,
Ann. Geophys. **19**, pp. 571-580, 2001.

Rapp, M., Irina Strelnikova, B. Strelnikov, P. Hoffmann, M. Friedrich, J. Gumbel, Linda Megner, U.-P. Hoppe, S. Robertson, S. Knappmiller, Mareile Wolf and D.R. Marsh:
Rocket Borne in situ Measurements of Meteoric Smoke; Charging Properties and Implications for Seasonal Variation,
J. geophys. Res. **115**, doi 10.1029/2009JD01275, 2010.

Rees, D., J.J. Barnett, and Karin Labitzke:
COSPAR International Reference Atmosphere II, Middle Atmosphere Models,
Adv. Space Res. **10 (12)**, 1990.

Reid, G.C.:
The Production of Water-Cluster Positive Ions in the Quiet Daytime *D* Region,
Planet. Space Sci. **25**, pp. 275-290, 1977.

Reisinger, J.:
Phasenmessung mit Hilfe von analogen Verzögerungsleitungen,
MSc Thesis, Technical University Graz, 1977.

Roininen, L., M. Laine and T. Ulich:
Time-Varying Ionosonde Trend: Case Study of Sodankylä hmF2 Data 1957-2014,
J. geophys. Res. (Space), doi:10.1002/2015JA021176, 2015

Rosenberg, T.J., D.L. Detrick, D. Venkatesan, and G. von Biel:
A Comparative Study of Imaging and Broad-Beam Riometer Measurements: The Effect of Spatial Structure on the Frequency Dependence of Auroral Absorption,
J. geophys. Res. **96 (A10)**, pp. 17793-17803, 1991.

Rusch, D.W., R.G. Roble, J.-C. Gérard, and I.A. Stewart:
A Two-Dimensional Model of Odd Nitrogen in the Thermosphere and Mesosphere,
in: *Handbook for MAP* **2**, pp. 442-449, 1981.

Schwentek, H.:
Regular and Irregular Behaviour of the Winter Anomaly in the Ionospheric Absorption,
J. atmos. terr. Phys. **33**, pp. 1647-1650, 1971.

Schwentek, H. and E.H. Gruschwitz:
Measurement of Absorption in the Ionosphere on 27.6 MHz at 52°N by Means of a Riometer and a Corner Reflector Antenna Directed to the Pole Star,
J. atmos. terr. Phys. **32**, pp. 1385-1402, 1970.

Schmidlin, F.J.:
Rocketsonde Temperature and Long-Term Stratospheric Behavior,
Paper C2.1-0002 presented at 31st COSPAR, Birmingham, 1996.

Seddon, J.C.:
Propagation Measurements in the Ionosphere with the Aid of Rockets,
J. geophys. Res. **58**, pp. 323-335, 1953.

Sen, H.K. and A.A. Wyller:
On the Generalization of the Appleton-Hartree Magnetoionic Formulas,
J. geophys. Res. **65 (12)**, pp. 3931-3950, 1960.

Smith, L.G.:
Langmuir Probes in the Ionosphere,
in: *Small Rocket Techniques*, N. Holland, pp. 1-15, 1969.

Smith, L.G.:
Electron Density Measurements in the Middle Atmosphere by Radio Propagation Techniques,
in: *Handbook for MAP* **19**, pp. 173-185, 1986.

Smith III, F.L. and Cody Smith:
Numerical Evaluation of Chapman's Grazing Incidence Integral ch(X,χ),
J. geophys. Res. **77**, pp. 3592-3597, 1972.

Smith, L.G., E.K. Walton, and E.A. Mechtly:
Vertical Incidence Absorption Calculated Using Electron Density Profiles from Rocket Experiments and Comparison with Observations During the Winter Anomaly,
J. atmos. terr. Phys. **40 (10/11)**, pp. 1185-1197, 1978.

Solomon, Susan, P.J. Crutzen, and R.G. Roble:
Photochemical Coupling Between the Thermosphere and the Lower Atmosphere 1. Odd Nitrogen,
J. geophys. Res. **87**, pp. 7221-7227, 1982.

Spöcker, P.:
Die verallgemeinerte Behandlung der Stoßzahl in der magneto-ionischen Theorie der Wellenausbreitung,
MSc Thesis, Technical University Graz, 1984.

Stauning, P.:
High-Latitude *D*- and *E*-Region Investigations Using Imaging Riometer Observations,
J. atmos. terr. Phys. **58 (6)**, pp. 765-783, 1996.

Steiner, R.J.:
Novel Procedures for Modelling the High-Latitude Ionosphere,
PhD Thesis, Graz University of Technology, 2003.

Strobel, D.F., T.R. Young, R.R. Meier, T.P. Coffey, and A.W. Ali:
The Nighttime Ionosphere; *E* Region and Lower *F* Region,
J. geophys. Res .**79 (22)**, pp. 3171-3178, 1974.

Taubenheim, J., G. von Cossart, and G. Entzian:
Evidence of CO_2-Induced Progressive Cooling of the Middle Atmosphere Derived from Radio Observations,
Adv. Space Res. **10 (19)**, pp. 171-174, 1990.

Thomas, R.J., C.A. Barth, D.W. Rusch, and R.W. Sanders:
Solar Mesospheric Explorer Near-Infrared Spectrometer: Measurements of 1.27 μm Radiances and the Inference of Mesospheric Ozone,
J. geophys. Res. **89**, pp. 6569-9580, 1984.

Thrane, E.V.:
Diurnal Variation of Electron Production Rates in the *D*-Region,
J. geophys. Res. **74**, pp. 1311-1316, 1969.

Thrane, E.V.:
Ionospheric Profiles up to 160 km: A Review of Techniques and Profiles,
in: *Methods of Measurements and Results of Lower Ionosphere Structure*, Akademie Verlag, Berlin, 1974.

Thrane, E.V. and W.R. Piggott:
The Collision Frequency in the *D*- and *E*-Regions of the Ionosphere,
J. atmos. terr. Phys. **28**, pp. 721-737, 1966.

Thrane, E.V., B. Grandal, O. Hagen, F. Ugletveit, W. Bangert, M. Friedrich, A. Loidl, H. Schwentek, and K.M. Torkar:
Ion Production and Effective Loss Rate in the Mesosphere and Lower Thermosphere During the Western European Winter Anomaly Campaign 1975/76,
J. atmos. terr. Phys. **41**, pp. 1097-1103, 1979.

Titheridge, J.E.:
An Approximate Function for the Chapman Grazing Incidence Function,
J. atmos. terr. Phys. **50 (8)**, pp. 699-701, 1988.

Titheridge, J.E.:
Production of the Low-Latitude Night *E* Layer,
J. geophys. Res. **106 (A7)**, pp. 12781-12786, 2001.

Torkar, K.:
Auswertung von Raketenrohdaten,
MSc Thesis, Technical University Graz, 1975.

Torkar, K. and M. Friedrich:
A Computer Programme to Evaluate Faraday Rotation and Absorption,
Internal Report **INW 7602**, 1976.

Torkar, K.M. and M. Friedrich:
Refinements in the Derivation of *D*-Region Plasma Densities Based on the Rocket-Borne Faraday Experiment,
in: *Sounding Rocket Program Aeronomy Project: Energy Budget Campaign 1980, Experiment Summary*, D. Offermann and E.V. Thrane (eds.), BMFT-FB-W 81-052, pp. 244-255, 1981.

Torkar, K.M. and M. Friedrich:
Tests of an Ion-Chemical Model of the *D*- and Lower *E*-Region,
J. atmos. terr. Phys. **45**, pp. 369-385, 1983a.

Torkar, K.M. and M. Friedrich:
Suggested Correction to the AFGL Reference Atmosphere for Use in the High-Latitude European Sector Based on Collision Frequency Measurements,
ESA **SP-183**, pp. 185-188, 1983b.

Torkar, K.M. and M. Friedrich:
Empirical Electron Recombination Coefficients in the *D*- and *E*-Region,
J. atmos. terr. Phys. **50 (8)**, pp. 749-761, 1988.

Torkar, K.M., M. Friedrich, W. Wallner, G. Rose, and H.U. Widdel:
Preliminary Results of Absorption Measurements of a Central European A3 Path,
Kleinheubacher Ber. **21**, pp. 155-161, 1978.

Ulich, T. and E. Turunen:
Evidence for Long-Term Cooling of the Upper Atmosphere in Ionosonde Data,
Geophys. Res. Lett. **24 (9)**, pp. 1103-1106, 1997.

Verronen, P.T., M.E. Anderson, D.R. Marsh, T. Kovács, and J.M.C. Plane:
WACCM-D Whole Atmosphere Community Climate Model with *D*-Region Ion Chemistry,
J. Adv. Modelling Earth Syst., doi 10.1002/2015MS000592, 2016.

Vuthaluru, Rupa, R.A. Vincent, R.A. Holdsworth, and I.M. Reid:
Collision Frequencies in the *D*-Region,
J. atmos. solar terr. Phys., **64 (18)**, pp. 2043-2054, 2002.

Williams, E.R., Wu, Yen-Jung, R.-R. Hsu, and M. Friedrich:
Ablation of Sub-Visual Meteors as the Origin for the Upper Boundary of the VLF Global Waveguide
J. geophys. Res., submitted, 2016.

Acknowledgements

The undisputed initial driver of our ionospheric activities was Professor W. Riedler. But soon support and encouragement of one kind or another came notably from E.V. Thrane, T.A. Jacobsen, and of course K.M. Torkar who has been involved in these activities almost since their beginning. The described activities were made possible by several grants provided by the Austrian Research Fund (FWF) between 1970 and 2014 (see list below). Much of the nitty-gritty work involved in the described fields was carried out within Master and PhD theses of students of the institute. The alphabetical list below is inevitably incomplete, but to the best knowledge.

Alfred M. Bauer 1972 (PhD)

Georg Egger 2004 (MSc)

Robert Erlacher 1996 (MSc)

Martin Fankhauser 2010 (PhD)

Herbert Hackl 2014 (MSc)

Markus Harrich 1998 (MSc), and 2001 (PhD)

Wolfgang Kropf 1977 (MSc)

Gerhard Landauer 2010 (Hon. Proj.)

Dieter Nutz 1972 (MSc)

Dieter Preiß 2004 (Bacc)

Franz Pürstl 2000 (PhD)

Jochen Reisinger 1977 (MSc)

Peter Spöcker 1984 (MSc)

René Steiner 2003 (PhD)

Klaus M. Torkar 1975 (MSc), and 1977 (PhD)

FWF Projects that supported participation in ionospheric sounding rockets

1970	P-955 (*Riedler*): Measuring Electron Density Profiles in the Ionosphere Using Sounding Rockets
1974	P-2498 (*Riedler*): Measuring Electron Density and Collision Frequencies in the Lower Ionosphere
1974	P-2671 (*Riedler*): Measuring Parameters of the *D*-region Under Conditions of Winter Anomaly
1978-1983	S-20 Research Phocus (*Riedler*, co-ordinator): Space Research: Investigation and Utilisation of Near-Earth Sspace
1987	P-6328-TEC (*Friedrich*): Trace Gases in the Upper Atmosphere and Their Transport
1990	P-7991-PHY (*Friedrich*): NLC, Plasma Structures and Their Relation to Noctilucent Clouds

1992 P-9315-PHY (*Friedrich*): CADRE, Coupling and Dynamics in Regions Equatorial

1996 P-11827-TEC (*Friedrich*): NLTE, Non-Local Thermal Equilibrium

1998 P-13226-PHY (*Friedrich*): DROPPS, Role of Particles in the Polar Mesopause Under Summer Conditions

2001 P-14922 (*Friedrich*): MIPMAG, Modification of Ionisation Properties in the Middle Atmosphere by Gravity Waves

2002 P-15468 (*Friedrich*): McWAVE, Gravity Wave Forcing of the Mesosphere and Lower Thermosphere: Mountain and Convective Waves Ascending Vertically

2002 M708-N06 (*McKinnell*): A Neural Network Approach to Empirically Model the Lower Ionosphere

2003 P-15985 (*Friedrich*): LEMMA, Layers in the Equatorial Mesosphere, Motion and Aerosol Rocket and Radar Study

2005 P-18560-N10 (*Friedrich*): ECOMA, Existence and Charge State of Meteoric Dust Grains in the Middle Atmosphere

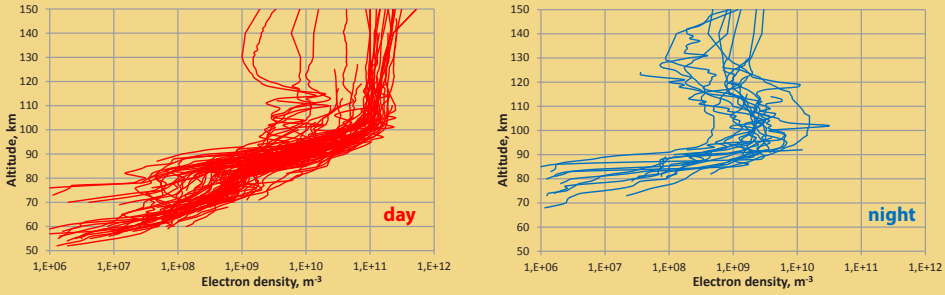
2010 P-22613-N21 (*Friedrich*): PHOCUS, Particles, Hydrogen and Oxygen Chemistry in the Upper Summer Mesosphere

2010 P-23100-N21 (*Friedrich*): Electron Scavenging in the Mesosphere

2014 P-26932 (*Friedrich*): Dust and Oxygen, New Players in the Mesosphere: Assessing the Role of Meteoric Dust and Atomic Oxygen

SUMMARY

The present Handbook of the Lower Ionosphere is not only a review of almost half a century of *in-situ* ionospheric research performed by the Institute of Communications and Wave Propagation (now the Institute of Communication Networks and Satellite Communications), but it summarises the various aspects of ionospheric/atmospheric investigation with an outlook of what could be topics of future research. The described activities are generally restricted to altitudes below, say 150 km, not least because higher altitudes are well covered by decades of ionosonde measurements and more recently by dozens of scientific satellites. To some extent the layers below about 100 km are accessible by world-wide only a few dedicated radar or lidar installations, but for *in-situ* information one has to rely on sporadic rocket soundings. The above Institute has participated with scientific instruments in more than 100 of such soundings from rocket bases ranging from Alcântara (2°S) to Spitsbergen (79°N). The information thus obtained is essential for assessing the performance of theoretical atmospheric models which predict potentially more important parameters, not least global warming and ozone. As an example for this effort the two panels illustrate the huge variability of the number density of free electrons which one ultimately would like to be able to understand and predict.



Electron densities from non-auroral latitudes obtained by rocket soundings

Verlag der Technischen Universität Graz
www.ub.tugraz.at/Verlag

ISBN 978-3-85125-485-3

ISBN 978-3-85125-485-3